



TECHNISCHE UNIVERSITÄT MÜNCHEN

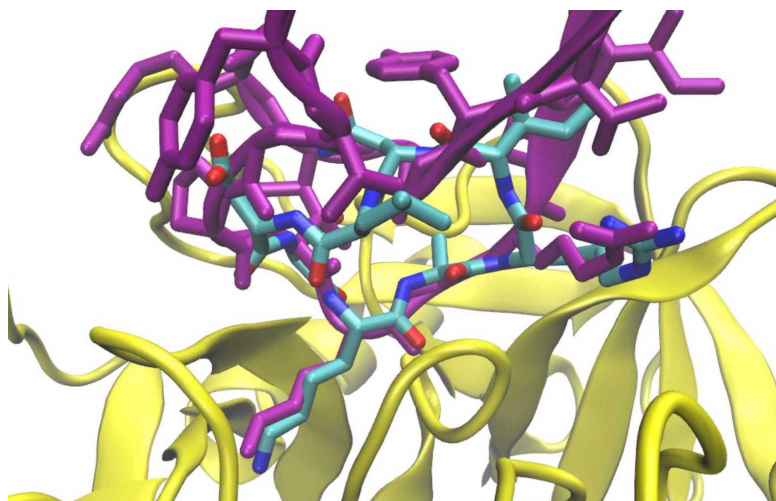
TUM SCHOOL OF NATURAL SCIENCES

Doctoral Thesis

Computational Design of Cyclic Peptides for Protein-Protein Interaction Modulation

Brianda Paola López Santini

Supervisor: Prof. Dr. Martin Zacharias





TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM SCHOOL OF NATURAL SCIENCES

Computational Design of Cyclic Peptides for Protein-Protein Interaction Modulation

Brianda Paola López Santini

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften genehmigten Dissertation

Vorsitz:	Prof. Dr. Matthias Feige
Prüfende der Dissertation:	1. Prof. Dr. Martin Zacharias 2. Prof. Dr. Dmitrij Frischmann

Die Dissertation wurde am 27.06.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 29.07.2024 angenommen.

I confirm that this doctoral thesis is my own work and I have documented all sources and material used.

Munich, 27.06.2024

Brianda Paola López Santini

Abstract

This thesis centers on the computational design of cyclic peptides as modulators for protein-protein and protein-ligand interactions. At the core of this study is the cPEPmatch method, a novel computational approach for identifying cyclic peptide templates that effectively modulate protein interactions. This method is based on structural matching at protein interfaces and is specially optimized to target hot spots within these interfaces, enhancing binding specificity and effectiveness. A comprehensive database of over 400 cyclic peptides from the Protein Data Bank supports this research, providing insights into their structural diversity and laying the foundation for future drug discovery applications.

The research initiates with the validation of the cPEPmatch method against a benchmark of known protein-protein interactions, demonstrating its capability for the automated identification of potential cyclic peptide templates for protein-protein modulation. Two direct applications of the method led to computational predictions which were subsequently validated experimentally by collaborators. Firstly, the design of cyclic peptides that mimic glycosaminoglycan-binding proteins was refined, improving the cPEPmatch algorithm to identify and characterize GAG-protein interaction sites. This process identified promising peptide binders, notably 2LWT version 2, and validated our computational approach. Additionally, the study presents cyclic peptides as effective ICOS/ICOSL inhibitors. The application of cPEPmatch to the ICOS/ICOSL interface resulted in the identification of Match 41, a first-of-its-kind cyclic peptide ICOS/ICOSL inhibitor, highlighting its potential in immunomodulation and cancer therapy.

Lastly, we detailed the capabilities of the cPEPmatch tool and webserver, showcasing its relevance to target a variety of medically relevant protein-protein interactions. The findings of this thesis underscore the versatility and potential of cyclic peptides as therapeutic agents. By exploring their role in modulating protein-protein and protein-ligand interactions, this research opens new avenues in drug discovery and biomedical research, paving the way for the development of innovative therapeutic strategies based on cyclic peptides.

Contents

Abstract	v
1. Introduction	1
2. Cyclic Peptides and Protein-Protein Interactions	3
2.1. Cyclic Peptides	3
2.1.1. Structure and Stability	3
2.1.2. Therapeutic Potential	6
2.2. Protein-Protein Interactions	7
2.3. cPEPmatch Virtual Database	8
3. Theory and Methods	11
3.1. Molecular Dynamics Simulations	11
3.1.1. Force Field in Molecular Dynamics Simulations	12
3.1.2. Integrating the Equations of Motion	14
3.1.3. Simulation Box and Periodic Boundary Conditions	15
3.1.4. Temperature and Pressure Coupling	17
3.2. Free Energy Calculations	18
3.2.1. End-Point Method: MMPB/SA and MMGB/SA	18
3.2.2. Advanced Sampling Method: Hamiltonian Replica Exchange Molecular Dynamics	20
3.3. Cyclic Peptide Preparation	23
3.3.1. Head to Tail Cyclization	23
3.3.2. Parametrization of Non-Standard Amino acids	24
3.3.3. Point Mutations	24
4. cPEPmatch	25
4.1. Introduction	25
4.1.1. Cyclic Peptides as PPI Modulators	25
4.2. Materials and Methods	27
4.2.1. Cyclic Peptide Matching (cPEPmatch) Approach	27
4.2.2. Cyclic Peptide Library and Protein-Protein Complexes	27
4.2.3. Molecular Dynamics Simulation, RS-REMD, and MM/GBSA Evaluation	28
4.2.4. MD	28
4.2.5. MM/GBSA	29
4.2.6. RS-REMD	29
4.3. Results and Discussion	30

4.4. Conclusions	37
5. Cyclic Peptides Mimicking Protein-Protein Hot Spots	39
5.1. Introduction	39
5.2. Materials and Methods	41
5.2.1. Identification of hot loops for cPEPmatch	41
5.3. System preparation	41
5.3.1. Trajectory analysis to score the matches	42
5.4. Results and Discussion	43
5.4.1. Trajectory analysis and binding hot spot identification	43
5.4.2. Application of cPEPmatch to hot loop regions	45
5.4.3. Selected results for targeting the MHC α chain	45
5.5. Conclusions	46
6. Design of Glycosaminoglycan-Binding Cyclic Peptides	49
6.1. Introduction	49
6.2. Materials and Methods	50
6.2.1. Structures	50
6.2.2. Cyclic Peptide Matching Update	51
6.2.3. cPEPmatch Software Availability	52
6.2.4. ATIII Cyclic Peptide Matching Study	52
6.2.5. MD Simulations	53
6.3. Results and Discussion	54
6.3.1. Adapting cPEPmatch to Complex Binding Sites	54
6.3.2. MD and Optimization	54
6.3.3. Specificity of the ATIII-GAG Interactions	56
6.3.4. Experimental Validation	59
6.4. Conclusions	60
7. Design of ICOS-ICOSL Cyclic Peptide Modulators	61
7.1. Introduction	61
7.2. Materials and Methods	63
7.2.1. Structure Model	63
7.2.2. Matching	63
7.2.3. Molecular Dynamics (MD) Simulations	63
7.2.4. Trajectory Analysis	64
7.3. Results and Discussion	65
7.3.1. Identification of ICOS/ICOS-L Hot Spots	65
7.3.2. cPEPmatch Selected Matches	65
7.4. Experimental Validation	70
7.4.1. Conclusions	71
8. cPEPmatch Code and Webserver: Application to Medically Relevant Systems	73
8.1. Development and Features of cPEPmatch	73
8.1.1. Cyclic Peptide Database Creation	74
8.1.2. Characterization of Protein Interfaces	74
8.1.3. Backbone Match Motif	75

8.1.4. Superimpose and Mutate	76
8.1.5. Result Analysis and Output	76
8.1.6. Installation and Setup	77
8.2. cPEPmatch Webserver and Source Code	77
8.2.1. Source Code Structure	77
8.2.2. Webserver for cPEPmatch	79
8.3. cPEPmatch Webserver Tutorial	80
8.3.1. Step 1: Identify Your Target Complex	80
8.3.2. Step 2: Inspect and Prepare the PDB File	81
8.3.3. Step 3: Select cPEPmatch Parameters	81
8.3.4. Step 4: Run cPEPmatch	83
8.3.5. Step 5: Post cPEPmatch Evaluation	84
8.4. Application of cPEPmatch to Medically Relevant Case Studies	84
8.4.1. Materials and Methods	86
8.4.2. Results and Discussion	89
8.5. Conclusions	93
9. Conclusions and Outlook	95
Bibliography	97
A. Supplementary Information for Chapter 2: Cyclic Peptides and Protein-Protein Interactions	113
B. Supplementary Information for Chapter 4: cPEPmatch	127
C. Supplementary Information for Chapter 7: Design of ICOS-ICOSL Cyclic Peptide Modulators	133
D. Supplementary Information for Chapter 8: cPEPmatch Code and Webserver	137
Acknowledgements	147
List of Publications	149

1. Introduction

This thesis explores the potential of cyclic peptides as therapeutics, with a focus on their unique properties and advantages in drug development. The aim is to demonstrate the feasibility of designing cyclic peptides based on existing templates, hypothesizing that these structures can be effectively used to disrupt specific protein-protein and protein-ligand interactions. The creation and utilization of the cPEPmatch algorithm is a cornerstone of this research, offering a novel approach to identifying cyclic peptide candidates.

Chapter Overview:

- **Chapter 2 - Cyclic Peptides:** This chapter highlights the significance of cyclic peptides in pharmaceutical research. It discusses their structural uniqueness and stability, which are crucial for therapeutic applications, and outlines the development of the cPEPmatch database.
- **Chapter 3 - Theory and Methods:** Here, we delve into the theoretical aspects of molecular dynamics simulations and free energy calculations. The chapter serves as a foundational framework for the methodologies applied in subsequent chapters.
- **Chapter 4 - cPEPmatch:** This chapter introduces the cPEPmatch method, a computational approach for designing cyclic peptides to inhibit protein-protein interactions. The chapter demonstrates the effectiveness of this method using a variety of protein-protein complexes.
- **Chapter 5 - Mimicking Protein-Protein Hot Spots:** The focus here is on refining the cPEPmatch method to target hot spots within protein-protein interfaces. The application of this enhanced method is illustrated through its use on an MHC class I protein complex.
- **Chapter 6 - Design of GAG-Binding Cyclic Peptides:** Discussing the design of cyclic peptides targeting glycosaminoglycans, this chapter explores the potential of these peptides in therapeutic and diagnostic applications.

- **Chapter 7 - ICOS-ICOSL Cyclic Peptide Modulators:** This chapter describes the application of the cPEPmatch approach to design cyclic peptides inhibiting the ICOS/ICOSL interaction, with implications for cancer drug discovery.
- **Chapter 8 - Webserver and Medically Relevant Applications:** Development and deployment of the cPEPmatch webserver, along with its application in identifying cyclic peptide matches for medically relevant protein-protein interactions.

Conclusion and Outlook: The thesis concludes with a summary of findings and future perspectives on the cPEPmatch approach and the role of cyclic peptides in therapeutic applications.

2. Cyclic Peptides and Protein-Protein Interactions

Cyclic peptides stand at the forefront of modern pharmaceutical research, epitomizing the confluence of natural biochemistry and cutting-edge drug design. These molecules, distinguished by their cyclic structures, offer a unique combination of specificity, stability, and efficacy, differentiating them from conventional linear peptides and small molecule drugs. This chapter provides a focused overview of cyclic peptides, particularly emphasizing their structural uniqueness and stability, which underpin their substantial therapeutic potential. We will explore the key attributes that position cyclic peptides as valuable entities in drug discovery, highlighting their superior stability and specificity. Additionally, the chapter outlines the development of our cPEPmatch database, detailing the methodology and criteria employed in its construction. This database represents a significant effort to catalog and understand the structural diversity of cyclic peptides, laying the groundwork for their future exploration and application in pharmaceutical research.

2.1. Cyclic Peptides

2.1.1. Structure and Stability

Peptides are chains of amino acids that play vital roles in biological systems. Among peptides, cyclic peptides have attracted significant attention due to their distinct characteristics. Unlike linear peptides which have a beginning and end, cyclic peptides form a continuous loop. This cyclization can occur in various ways (Figure 2.1): by connecting the head to the tail, linking side-chains to each other, bonding side-chains to the backbone, or through the formation of disulfide bridges. Modern methods also include the use of staples or non-standard amino acids to achieve this circular structure [1].

These cyclic structures offer cyclic peptides certain advantages, particularly in the realm of drug design. They tend to be more stable and can bind to their targets with

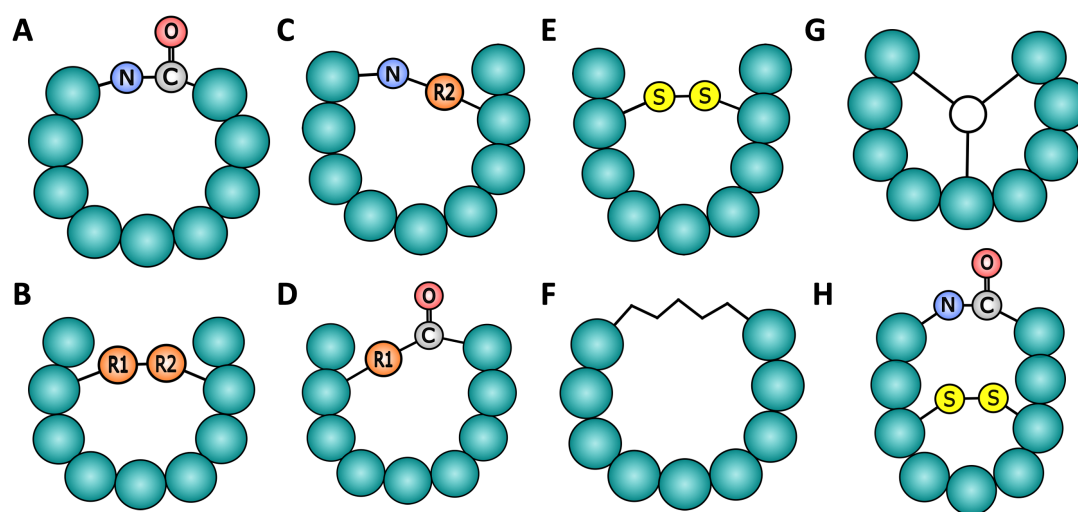


Figure 2.1.: Types of Cyclization in Cyclic Peptides. This figure delineates the diverse cyclization mechanisms employed in cyclic peptides, illustrating the structural versatility and complexity of these biomolecules: (a) Head-to-Tail, where the N-terminus (head) is linked to the C-terminus (tail). (b) Side-Chain to Side-Chain, involving connections between the side chains of amino acids. (c) Head-to-Side Chain, featuring a link between the N-terminus and a side chain. (d) Tail-to-Side Chain, where the C-terminus connects to a side chain. (e) Disulfide Bonded, characterized by disulfide bridges typically formed between cysteine residues. (f) and (g) Staples, denoting one or two holds, respectively, often used to stabilize specific conformations. (h) Combination, such as Disulfide + Head-to-Tail, representing a mix of different cyclization methods to achieve specific structural and functional properties.

higher specificity compared to linear peptides[2, 3, 4, 5, 6]. This stability arises from the cyclic structure, which can be formed using different chemical modifications like disulfide bridging, amide bonding, and side-chain covalent binding. Because of these properties, cyclic peptides are increasingly being considered as promising candidates for new drugs and therapeutics[7, 8, 9]. Each cyclization method imparts unique characteristics to the peptides, influencing their biological activity and therapeutic potential[10].

Cyclic peptides, despite being smaller than full-sized proteins, possess well-defined secondary structures like α -helices, β -sheets, and specific turns [11] (Figure 2.2). Their compact size doesn't compromise their structural integrity. In fact, the defined structure, along with their circular nature, makes them more resilient against enzymes that might otherwise break them down[12, 13]. This blend of smaller size and distinct, robust structures gives cyclic peptides a considerable advantage, especially when it comes to therapeutic applications and drug design[14].

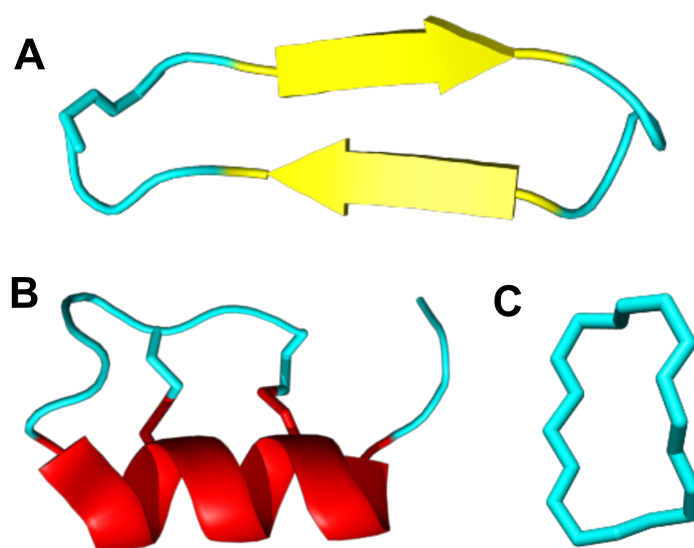


Figure 2.2.: Illustration of Secondary Structures in Cyclic Peptides. This figure showcases three prominent types of secondary structures observed in cyclic peptides: (a) β -hairpin, demonstrating a structure with two strands of amino acids connected by a sharp turn, resembling a hairpin. (b) α -helix, depicting a right-handed helical structure, where each amino acid residue is aligned perpendicular to the helix's axis. (c) Coil, representing a structure that can be repetitive, such as a turn, but does not form a fixed or regular conformation like the α -helix or β -hairpin.

2.1.2. Therapeutic Potential

Small drugs and antibodies, each with their unique advantages and disadvantages, are pivotal in therapeutic applications. Small drugs, often low in molecular weight, offer easy synthesis and modification, with good oral bioavailability and cell membrane permeability [15]. However, they may lack specificity, leading to off-target effects and are limited in disrupting large protein-protein interfaces. Antibodies, in contrast, are highly specific, minimizing off-target effects and are ideal for targeting large protein complexes [5]. Their disadvantages include production complexity, high costs, poor oral bioavailability, and potential immune responses[16]. In drug design, ADMET considerations are critical, and cyclic peptides excel in several aspects. They demonstrate enhanced bioavailability due to their resistance to enzymatic degradation. Their compact size allows for better tissue penetration, and their metabolic stability ensures a longer duration of action[14]. Predictable excretion pathways and reduced toxicity due to high specificity are additional benefits, making them favorable candidates in therapeutics.

Cyclic peptides have carved a niche in various therapeutic areas, evidenced by several notable drug approvals. Telavancin, dalbavancin, and oritavancin, semi-synthetic cyclic lipoglycopeptides, represent a new generation in the treatment of bacterial infections, building on the foundational structure of vancomycin and teicoplanin [17, 18]. Anidulafungin, belonging to the echinocandin class, showcases advancements in antifungal therapy, inhibiting fungal cell wall synthesis with a structurally unique lipophilic side chain [19]. In the oncology domain, lanreotide and pasireotide, analogues of the cyclic peptide hormone somatostatin, have been effective in treating neuroendocrine tumors [20, 21]. Romidepsin, a depsipeptide isolated from *Chromobacterium violaceum*, marks a breakthrough in the treatment of T-cell lymphomas, functioning through the inhibition of histone deacetylase [22]. Linaclotide, derived from enterotoxins, has provided novel relief for patients with chronic constipation and irritable bowel syndrome, illustrating the versatility of cyclic peptides in gastrointestinal disorders [23, 24]. These examples, encompassing a wide spectrum of diseases, underscore the significant role cyclic peptides play in current therapeutic landscapes and highlight the potential for future drug development.

Beyond these, there are numerous cyclic peptides in various stages of clinical trials, addressing a wide range of conditions from cancer to metabolic disorders. The cyclic peptide targeting p53, a key protein in cancer biology, exemplifies this [25]. p53 is often referred to as the "guardian of the genome" due to its role in preventing cancer formation, and cyclic peptides targeting this protein have shown potential in reactivating its function in cancer cells where it is typically mutated or inactivated.

This approach represents a novel strategy in cancer therapy, exploiting the specific interaction capabilities of cyclic peptides to modulate crucial protein functions.

Moreover, the field of cyclic peptides has been enriched by the discovery and development of peptides that inhibit critical protein-protein interactions, a new frontier in therapeutic interventions [7]. These interactions, often challenging to target with traditional small-molecule drugs, are now being effectively modulated by specially designed cyclic peptides. The advancements in synthetic methodologies and computational design tools have further broadened the scope of cyclic peptide applications [26].

2.2. Protein-Protein Interactions

Protein-protein interactions (PPIs) serve as the basis for nearly all biological processes; they play key roles in intercellular communication, cell-to-cell signaling, metabolic and developmental control, and programmed cell death [27]. Abnormal regulation of PPIs results in many human diseases such as cancers, immune disorders, and neurodegenerative diseases. Modulating these aberrant interactions is of clinical relevance; however, targeting PPIs may be challenging because of the intrinsic properties of protein-protein interfaces [28, 29].

PPI interfaces typically involve large, flat binding sites which are devoid of any major binding pocket [30, 31], making them challenging targets for conventional drug modalities because small molecules generally do not bind to large binding sites with high affinity. Larger molecules with more extended binding surfaces, e.g., monoclonal antibodies, cannot easily cross the cell membrane to reach the targets [16]. Progress has, however, been made in targeting protein-protein interactions (PPI) by realizing that not all the surface area of a PPI contributes equally to the strength of the interaction between the protein partners [32, 33, 34]. Their binding is mediated by strong packing or electrostatic interactions where only a few neighboring amino acids are crucial for binding and recognition. These small areas of disproportionately high-affinity binding at the protein-protein interface are referred to as “hot spots” [33, 35]. Nevertheless, targeting specifically protein-protein interaction hot spots still tends to require complex drug molecules [36].

In the realm of cyclic peptide therapeutics, key challenges include enhancing oral availability and improving cell permeability. To address these issues, innovative methods such as the use of cell-penetrating peptides and the stabilization of cyclic peptides with hydrocarbon linkers are being explored to increase oral bioavailability

[37]. For cell permeability, strategies like the development of N-methylated peptides are showing promise [38]. Future research is also looking into the development of chimeric peptide/polyketide macrocycles, exemplified by compounds like tacrolimus, which achieve effective drug delivery through innovative structural designs [5]. The advancement of cyclic peptide therapeutics will also depend on refining synthetic techniques and employing high-throughput screening methods, crucial for identifying and optimizing compounds with optimal therapeutic properties.

2.3. cPEPmatch Virtual Database

The design and development of cyclic peptides have been significantly enhanced through the integration of computational tools. These tools bridge the gap between the vast combinatorial space of potential peptide sequences and the specific needs of effective PPI modulators. Techniques such as virtual screening, pharmacophore matching, and molecular dynamics simulations predict the behavior of cyclic peptides in biological systems, providing insights into their binding modes and efficiencies [39, 40, 41, 16]. In the realm of rational drug design, these techniques elucidate how peptides interact with their targets, and innovative approaches, like matching secondary structures, enable the identification of peptides that effectively mimic complex protein-protein interfaces.

For the purpose of our cyclic peptide computational studies presented in the subsequent chapters, we have meticulously constructed a virtual database of cyclic peptides. This database has been progressively assembled, starting initially with 30 structures in our first study. As of now, it encompasses 432 distinct cyclic peptides. This gradual expansion has been driven by a methodical search in the Protein Data Bank (PDB), a renowned repository for 3D structures of large biological molecules. Our search strategy employed keywords such as "cyclic peptides," "macrocyclic peptides," and "cyclized peptides," enabling us to collate a comprehensive and diverse collection. The accumulated database is a rich source of information on the sizes, secondary structures, and types of cyclization of these peptides. This extensive compilation is invaluable for our ongoing research, offering critical insights into the structure-function relationships of cyclic peptides and aiding significantly in the design of new therapeutic agents.

The characteristic analysis presented in Figure 2.3 underscores the diversity of cyclic peptides within our virtual database. Notably, the histogram in subplot A reveals a prominent frequency of peptides containing 10-20 amino acids, which coincides with the typical length observed in naturally occurring cyclic peptides known for

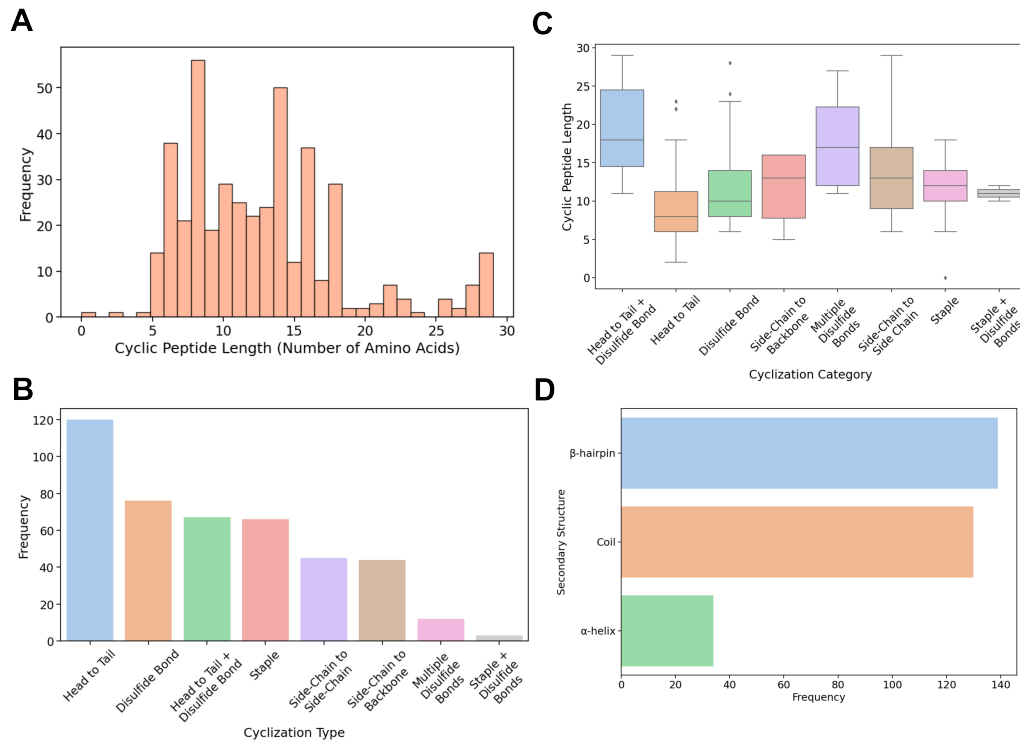


Figure 2.3.: Overview of Cyclic Peptide Characteristics within the cPEPmatch Database. (A) Histogram of cyclic peptide lengths, detailing the frequency distribution of peptides based on the number of amino acids. This plot highlights the common sizes of cyclic peptides encountered. (B) Bar chart illustrating the frequency of different cyclization types, such as head-to-tail and disulfide bonding, demonstrating the diversity of cyclization mechanisms. (C) Box plots that correlate cyclization category with cyclic peptide length, offering insights into the relationship between cyclization type and peptide size. (D) Stacked bar chart presenting the frequency of secondary structure types, such as β -hairpin, α -helix, and coil, indicating the prevalence of each structure type within the dataset.

their bioactive properties. Subplot B's bar chart reflects the predominant cyclization technique to be the head-to-tail method, a classical approach that is often favored due to its simplicity and stability in peptide structures. The box plots in subplot C indicate a broad range of lengths across different cyclization categories, with some methods, like side-chain to side-chain cyclization, associated with shorter peptide lengths, potentially influencing their conformational flexibility and target specificity. Lastly, the stacked bar chart in subplot D highlights a significant representation of β -hairpin structures within our dataset, suggesting a possible correlation with the structural stability required for effective PPI modulation.

The work presented in the following chapters will delve deeply into the realm of cyclic peptide research. Cyclic peptides, as a unique and versatile class of biomolecules, hold significant potential for therapeutic applications. Their distinct structural features, stability, and high specificity not only make them ideal candidates for drug development but also serve as a focal point for our comprehensive studies. The ongoing research and advancements in cyclic peptide synthesis and design, as detailed in our subsequent discussions, are set to further enhance their role in modern medicine. These studies aim to explore and elucidate the promising avenues cyclic peptides offer for treating a wide range of diseases, highlighting their critical importance in the future of medical research and therapy.

3. Theory and Methods

In this study, we utilized advanced computational methodologies, primarily molecular dynamics (MD) simulations and free energy calculations, to investigate diverse biomolecular systems. Understanding the fundamental principles and assumptions underlying these techniques is pivotal, as they aim to accurately model the complex dynamics of biological molecules. This chapter delves into the theoretical underpinnings of MD simulations, outlining the key concepts, algorithms, and computational strategies implemented in our analysis. It serves as a foundational framework for the methodologies applied in the subsequent chapter.

3.1. Molecular Dynamics Simulations

Quantum mechanics provides a robust framework for simulating molecules at the atomic level, but its computational intensity limits its application for large biomolecules. Methods such as Hartree-Fock [42, 43, 44], post-Hartree-Fock [45], and density functional theory (DFT) [46], provide numerical solutions to the time-independent Schrödinger equation, revealing detailed insights into electron distribution changes associated with molecular processes like bond formation or cleavage. However, these methods are computationally demanding, with QM simulations progressing at a slow pace, often used only for single-point calculations or geometry optimization. Even when QM calculations are combined with molecular mechanics (QM/MM)[47], the simulation speed remains constrained, achieving only a few picoseconds per day, even on GPU-accelerated platforms.

In contrast, Molecular Dynamics (MD) simulations, which rely on classical mechanics, offer a more computationally efficient alternative for studying dynamic molecular systems. Instead of the detailed electronic interactions required for bond formation or breaking, this approach reduces the system's complexity, enabling more practical computations for large-scale simulations. The core of MD simulations involves solving Newton's second law of motion for each atom in a system, represented as:

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2}, \quad (3.1)$$

where $\mathbf{F}_i$ denotes the force on atom i , m_i its mass, and $\mathbf{r}_i$ its position vector. This approach allows for the calculation of the dynamics of biomolecules within their physiological environments, such as aqueous solutions or lipid membranes. In MD simulations, biomolecules are placed in a simulation box that replicates their biological environment, and the Newtonian equations of motion are solved iteratively. This process captures the dynamic behavior of particles within the simulation box, providing insights into their motion and interactions over time.

By focusing on the classical treatment of atomic interactions, MD simulations efficiently model the behavior of large biomolecular systems. An integral part of these simulations is the use of force fields, which define the specific interatomic forces and potentials. These force fields, crucial for the accuracy of the simulations, are detailed in the following subsection.

3.1.1. Force Field in Molecular Dynamics Simulations

Force fields are fundamental in molecular dynamics (MD) simulations, converting complex quantum mechanical interactions into classical force potentials that influence atomic behavior in biomolecules. These empirical models, commonly shaped by quantum calculations and validated by experimental data, allow simulations of biomolecules to strike a balance between computational feasibility and physical realism. In MD simulations, atoms interact through pairwise forces, and the total potential energy (U_{total}) of the system is described as the sum of bonded and non-bonded energy terms [48]:

$$U_{total} = U_{bonded} + U_{non-bonded}, \quad (3.2)$$

Though the specific forms of these energy functions vary across different MD software, their general structure and parameters are often interchangeable. For instance, in the AMBER [49] MD framework, the bonded energy (U_{bonded}) is composed of bond stretching, angle bending, and dihedral angle terms (Figure 3.1, left):

$$U_{bonded} = \sum_{r \in bonds} K_r (r - r_{eq})^2 + \sum_{\theta \in angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{\phi \in dihedrals} V_n [1 + \cos(n\phi - \gamma)], \quad (3.3)$$

where K_r and K_θ are the force constants for bonds and angles, r_{eq} and θ_{eq} are their respective equilibrium values, and V_n , γ , and n parameterize the dihedral potential.

Non-bonded interactions, comprising electrostatic and van der Waals forces (Figure 3.1, right), are expressed as:

$$U_{non-bonded} = \sum_{i < j} \frac{q_i q_j}{\epsilon R_{ij}} + \sum_{i < j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right], \quad (3.4)$$

where q_i and q_j are atomic charges, ϵ is the dielectric constant, R_{ij} is the interatomic distance, ϵ_{ij} is the potential well depth, and σ_{ij} defines the van der Waals radius. The electrostatic interactions follow Coulomb's law, and the van der Waals forces are modeled using the Lennard-Jones potential [50].

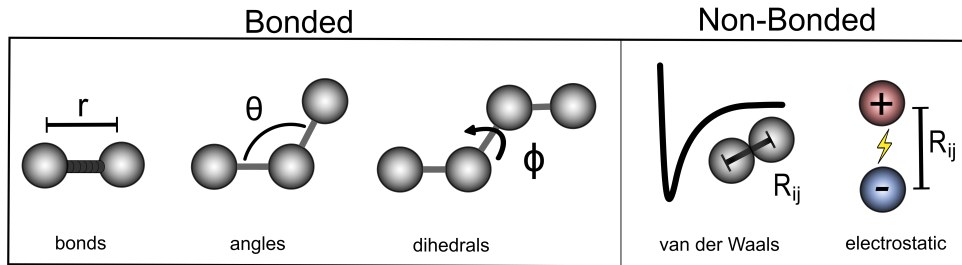


Figure 3.1.: Depiction of force field components in molecular dynamics simulations. Bonded interactions include bonds, angles, and dihedrals. Non-bonded interactions are represented by van der Waals forces and electrostatic interactions.

Atom types in a force field are designated based on their chemical and physical characteristics, with each type assigned a specific set of parameters. For example, an oxygen atom in a carbonyl group is treated differently from one in a hydroxyl group. Various force fields have been developed to accurately represent diverse molecular systems, from aqueous solutions to DNA-protein complexes. To encompass larger time and space scales, polarizable force fields [51] and coarse-grained models [52] are also utilized, capturing effects like polarization and simplifying large biomolecular structures.

3.1.2. Integrating the Equations of Motion

In molecular dynamics simulations, the behavior of a biomolecular system is deciphered by numerically solving Newton's equations of motion for the particles, using methods such as the Verlet algorithm [53]. This finite difference method predicts the future position of a particle based on its current and previous states, a crucial step in mapping the dynamic trajectory of each atom within the simulation.

For a particle i with mass m_i and velocity $\mathbf{v}_i(t)$ under a force $\mathbf{F}_i(t)$, the Verlet algorithm uses the Taylor expansion to approximate its future position at $t + \Delta t$ as:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\mathbf{F}_i}{2m_i}\Delta t^2 + O(\Delta t^3), \quad (3.5)$$

and similarly, its past position at $t - \Delta t$ can be expressed as:

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \mathbf{v}_i(t)\Delta t + \frac{\mathbf{F}_i}{2m_i}\Delta t^2 - O(\Delta t^3). \quad (3.6)$$

By combining these equations, the velocity terms are eliminated, providing the future position of the particle directly from the force acting on it:

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i}\Delta t^2 + O(\Delta t^4). \quad (3.7)$$

The force $\mathbf{F}_i$ on each particle is calculated from the derivative of the potential energy function U of the system:

$$\mathbf{F}_i = -\nabla U(\mathbf{r}_i), \quad (3.8)$$

where $\nabla U(\mathbf{r}_i)$ represents the gradient of the potential energy at the position $\mathbf{r}_i$.

The choice of the time step Δt is critical in MD simulations. It needs to be small enough to accurately capture the motion of the fastest-moving atoms, such as hydrogens, but large enough to allow the simulation to cover biological timescales efficiently. Techniques like the SHAKE [54] algorithm, which constrains the bond lengths to hydrogen atoms, and the Hydrogen mass repartitioning method, which redistributes the mass of heavier atoms to bonded hydrogens, enable larger time steps to be used without losing system stability [55].

3.1.3. Simulation Box and Periodic Boundary Conditions

Biological entities are often defined by the solvent that envelops their constituent molecules. Solvation can be represented in two primary ways: by explicitly modeling water molecules or by employing a continuum model of the solvent environment. The explicit representation enables the precise depiction of hydrogen bonding, which is crucial for the stabilization of molecular structures and interaction facilitation among various molecular sites. However, implicit solvent models do not typically account for such specific solvent-solute interactions.

Several explicit water models are available that vary based on the number of interaction sites, the degree of atomic flexibility, and their capacity to include polarization effects [references]. This study employs the widely used rigid TIP3P water model for its simplicity and efficiency. It features a single interaction site per atom, adequately serving our modeling purposes.

A key aspect of MD simulations is the implementation of periodic boundary conditions (PBC) [56], particularly when using explicit solvents. PBC allows for the simulation of extensive solvent systems without the limitations of a finite space. This is achieved by conceptually replicating the simulation box, often cubic or truncated octahedral in shape, in all three dimensions. As a result, particles seamlessly transition from one box to its replicated neighbor upon reaching a boundary (Figure [creffig:pb](#)c).

The infinite replication of the box in theory would lead to a quadratic increase in computational load due to the non-bonded interactions. To address this, the Coulomb interaction is divided into several components via the Ewald summation. This method splits the electrostatic potential into real space and reciprocal space contributions, along with a self-energy term and a dipolar correction. The Ewald formula is expressed as:

$$U_{electrostatic} = U^{(r)} + U^{(k)} + U^{(self)} + U^{(dipolar)}, \quad (3.9)$$

where $U^{(r)}$ and $U^{(k)}$ respectively represent the real and reciprocal space contributions, $U^{(self)}$ is the self-energy, and $U^{(dipolar)}$ is the polar correction. In practice, a cut-off distance is chosen to truncate the short-range interactions, and a cut-off wave number is selected for the long-range interactions. The particle-mesh Ewald (PME) method [57], utilizing fast Fourier transforms and interpolation techniques, further reduces the computational complexity of these interactions from $O(N^2)$ to $O(N \ln N)$, significantly enhancing efficiency.

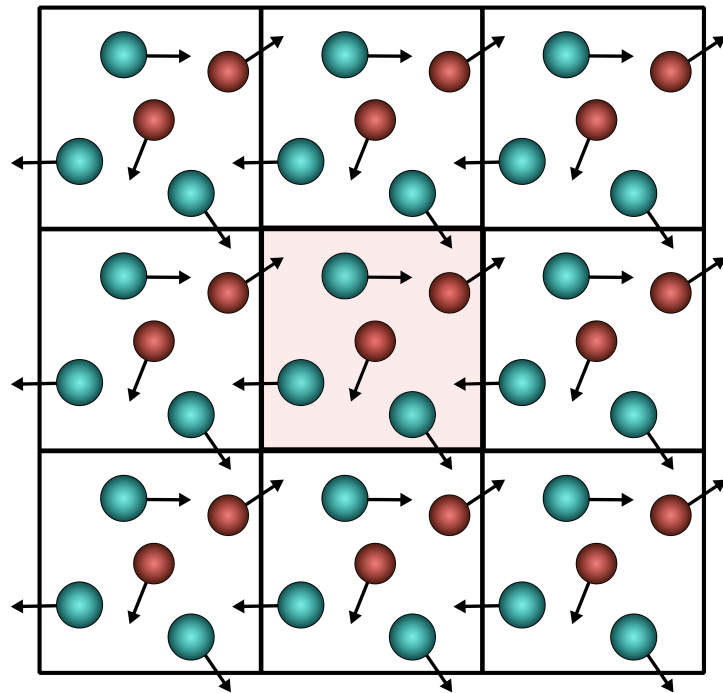


Figure 3.2.: Illustration of Periodic Boundary Conditions. The central box (highlighted in pink) represents the primary simulation area where particles (shown in teal and maroon) move freely. When a particle exits one side of the primary box, it re-enters from the opposite side, as indicated by the arrows, simulating an infinite system within a finite computational domain.

3.1.4. Temperature and Pressure Coupling

To model the physiological condition barostats and thermostats are usually coupled to the systems to control the pressure and the temperature respectively to generate a canonical ensemble (NPT). A common choice of barostat is the Berendsen barostat [58] by first calculating the pressure of the system with a Volume V by

$$P = \frac{1}{3V} \sum_{i,j>i}^N \left(\mathbf{F}_{ij} \cdot |\mathbf{r}_i - \mathbf{r}_j| + \frac{|\mathbf{p}_i|^2}{m_i} \right) \quad (3.10)$$

where m_i and $\mathbf{p}_i$ are the mass and momentum of a particle respectively. With a pressure relaxation constant τ_p and the isothermal compressibility β , the targeted pressure P_0 can be reached by rescaling the box lengths d ($d = V^{1/3}$ for a cubic box) and the center of mass coordinates of each molecule by a scaling factor μ

$$\mu = \frac{d_{\text{new}}}{d_{\text{old}}} = \left[1 - \frac{\beta \Delta t (P_0 - P)}{\tau_p} \right]^{1/3} \quad (3.11)$$

A targeted temperature T_0 of the system can as well be controlled by the Berendsen thermostat [58] in a similar approach by rescaling the velocity

$$\mathbf{v}_{\text{new}} = \mathbf{v}_{\text{old}} \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)} \quad (3.12)$$

where τ_T is the temperature relaxation constant.

Although controlling the pressure and temperature using the Berendsen scheme is relatively simple to implement, such integration will generate deterministic trajectories and is not convenient for the calculation of statistical observables. Instead, a Langevin thermostat is a more robust way to control the temperature through a stochastic Langevin equation of motion [59]

$$m \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i - \gamma \frac{d \mathbf{r}_i(t)}{dt} m_i + \mathbf{R}(t) \quad (3.13)$$

The equation of motion compared to the Newtonian one (equation 3.1) includes two additional terms: the frictional forces mediated by the collision frequency γ and the Brownian random force $\mathbf{R}$. The Brownian random force is a delta-correlated stationary Gaussian process with zero-mean $\langle \delta \mathbf{R}(t) \rangle = 0$ and $\langle \delta \mathbf{R}_i(t) \delta \mathbf{R}_j(t') \rangle = 2\gamma k_B T m \delta(t - t')$

where k_B is the Boltzmann constant.

3.2. Free Energy Calculations

Molecular dynamics (MD) simulation provides a pathway to understanding the dynamics of a given biomolecular system with atomistic precision. Utilizing trajectories generated under physiological conditions, one can extract various dynamic and thermodynamic properties using principles of statistical mechanics. A pivotal thermodynamic property is the free energy difference required to transition a system from one state to another. This includes the solvation free energy, binding free energy, and conformational free energy. Under isothermal-isobaric (NPT) conditions, the free energy of a state is typically expressed in terms of Gibbs free energy:

$$G = H - TS = U + PV - TS \quad (3.22)$$

where G is the Gibbs free energy, H is the enthalpy, U is the internal energy calculable from the force field, and S is the entropy of the system. The free energy difference, often computed using end-point methods, alchemical transformations, or biased MD simulations, is pivotal for our understanding.

3.2.1. End-Point Method: MMPB/SA and MMGB/SA

For binding energy calculations, we consider a ligand-receptor pair. The end-point method reassesses trajectories at the bound and unbound states using molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) and molecular mechanics generalized Born surface area (MM/GBSA) methods [60, 61]. The binding free energy in a solvated environment is expressed as:

$$\Delta G_{\text{bind,solv}} = G_{\text{bound,solv}} - G_{\text{unbound,solv}} = \Delta H_{\text{bind,solv}} - T\Delta S_{\text{bind,solv}} \quad (3.23)$$

Here, $\Delta H_{\text{bind,solv}}$ and $\Delta S_{\text{bind,solv}}$ denote the enthalpy and entropy changes between the solvated unbound and bound states, respectively. To counteract the noise from solvent-solvent interactions, we integrate a mean solvent response based on a continuum model for each conformation in the trajectory.

The binding free energy is decomposed into the ligand-receptor interaction energy

in vacuum ($\Delta G_{\text{bind,vacuum}}$), the solvation free energy changes for the receptor and ligand, and the conformational entropy changes in vacuum. The solvation free energy is further divided into polar and non-polar contributions:

$$\Delta G_{\text{solv}} = G_{\text{solv,npol}} + G_{\text{solv,pol}} \quad (3.24)$$

The non-polar solvation energy, related to the cavity formation in the solvent, is linearly proportional to the solvent-accessible surface area (SASA):

$$\Delta G_{\text{npol}} = \gamma \cdot \text{SASA} \quad (3.25)$$

In the MM/PBSA approach, the polar solvation free energy ($\Delta G_{\text{solv,pol}}$) is determined by solving the Poisson-Boltzmann (PB) equation [42]:

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) - 4\pi\lambda(\mathbf{r}) \sum_i z_i c_i \exp(-z_i \phi(\mathbf{r})/k_B T) \quad (3.26)$$

Where $\epsilon(\mathbf{r})$ is the spatially varying dielectric constant, $\phi(\mathbf{r})$ the electrostatic potential, $\rho(\mathbf{r})$ the charge density of the solute, $\lambda(\mathbf{r})$ a Stern layer masking function, and z_i , c_i are the charge and concentration of ion type i , respectively.

The MM/GBSA method approximates this with a generalized Born (GB) formula [62], simplifying calculations:

$$\Delta G_{\text{pol,GB}} = \sum_{i < j} \frac{q_i q_j f_{\text{GB}}}{\epsilon_{\text{out}} R_{ij}} \quad (3.27)$$

$$f_{\text{GB}} = \sqrt{r_{ij}^2 + R_i R_j e^{-\alpha_{ij}}} \quad (3.28)$$

$$\alpha_{ij} = \frac{r_{ij}}{4R_i R_j} \quad (3.29)$$

Where r_{ij} is the distance between atoms i and j , R_i and R_j are their Born radii, and q_i , q_j are their respective charges.

Developments in MM/PBSA and MM/GBSA methods have led to more efficient approaches, such as the one-average MM/PBSA (1A-MM/PBSA). This method requires fewer simulations and improves precision, focusing on the ensemble average of the bound complex by omitting free ligand and receptor conformations. This is particularly useful in cases where conformational changes upon binding are minimal

or can be reasonably neglected.

$$\Delta G_{\text{bind}} = G_{\text{PL}} - G_{\text{P}} - G_{\text{L}} \quad (3.30)$$

Here, G_{PL} , G_{P} , and G_{L} represent the free energies of the complex, free protein, and free ligand, respectively, calculated using the implicit solvent models. These methods, initially developed for AMBER, are now available in various software packages like GROMACS and NAMD, making them accessible for a wide range of computational studies.

3.2.2. Advanced Sampling Method: Hamiltonian Replica Exchange Molecular Dynamics

Hamiltonian replica exchange molecular dynamics (H-REMD) is an advanced simulation technique that combines molecular dynamics (MD) with Monte Carlo methods to enhance sampling in complex phase spaces [63]. This approach is particularly effective in biological systems characterized by rugged potential energy landscapes with numerous local minima separated by substantial energy barriers. H-REMD enables systems, especially those trapped in potential wells, to overcome such barriers more efficiently than traditional MD simulations. This is achieved by not just drifting over the barriers but by enabling coordinate switching between different system replicas.

In H-REMD, a set of n simulations, or replicas, each with a unique Hamiltonian, is run in parallel. Periodically, these replicas assess the probability of exchanging configurations with their neighbors, using a Metropolis criterion [64]:

$$P_{\text{acc}} = \min \left(1, \frac{e^{-\beta[U_i(r_j) + U_j(r_i)]}}{e^{-\beta[U_i(r_i) + U_j(r_j)]}} \right), \quad (3.14)$$

Here, U_i and U_j represent the potential energies of replicas i and j , respectively, and $\mathbf{r}_i$, $\mathbf{r}_j$ are their corresponding configurations. The exchange is favorable if the sum of potential energies of the swapped configurations is lower than the original, ensuring a more efficient exploration of the energy landscape.

H-REMD can be integrated with free energy perturbation methods to improve the sampling of intermediate states in alchemical transformations [65, 66, 67]. This is valuable for computing free energy differences in processes such as ligand binding or

conformational changes. Regular application of H-REMD, particularly in conjunction with umbrella sampling methods, enhances the convergence and efficiency of simulations, making it a powerful tool in computational studies of complex biological systems [68, 69, 70].

A specialized variant, the Repulsive Scaling (RS-)REMD, is employed for studying biomolecular complexes, as proposed by Siebenmorgen and Zacharias, 2020 [71]. This technique incrementally increases the effective pairwise van der Waals radii across the replica set. This adjustment introduces a physical dissociation in the ligand-receptor interactions while maintaining the integrity of intramolecular and solvent interactions. RS-REMD thus enhances phase space sampling, particularly at higher replicas, by steering clear of local minima on the receptor surface. This strategy effectively drives the system towards the global minimum, providing a more comprehensive understanding of the binding mechanisms and energetics.

The bias implemented in RS-REMD is achieved through adjustments to the Lennard-Jones potential, specifically by reducing the potential well depths (ϵ) and increasing the effective pairwise van der Waals radius. This results in the ligand disassociating in the higher replicas, effectively losing contact with the receptor protein. The reference replica, which represents the unaltered state between the ligand and receptor ($\alpha = 1$), serves as the baseline for the bound state of the complex. As the replica number increases, the system transitions towards the unbound state. A perturbative approach along the biasing (α) coordinate, correlated with increasing replica numbers, allows for the calculation of the binding free energy differentials between the bound and unbound states of the complexes [72].

In the RS-REMD simulations, explicit solvent conditions are used to more accurately represent the biological environment and to account for the flexibility of the interacting molecules [73, 74]. It is important to note that the free energy calculations derived from these simulations provide an approximate binding free energy rather than an absolute value. This approximation arises because only a fraction of the full translational and rotational freedom of the partners in the unbound state is sampled. Despite this, the RS-REMD method, when applied to a benchmark set of complexes, has demonstrated a good correlation with experimental binding data, validating its effectiveness [72]. The use of multiple replicas with increasing repulsive bias allows for efficient sampling of the energy landscape, ensuring that the system transitions from bound to unbound states.

The bias for each replica is determined by re-evaluating its trajectory. Specifically, the trajectory for replica $\alpha + 1$ is analyzed using the LJ parameters for replica α and then for replica $\alpha + 1$. The free energy variation associated with these two sets of

Table 3.1.: Lennard-Jones scaling parameters for the different RS-REMD simulation setups. For the repulsive scaling simulations starting far from the binding geometry and for the refinement of a docking ensemble the 16 replica scheme was used (columns 2 and 3). In the refinement simulations of individual docking poses, the 8 replica setup was used (columns 4 and 5).

Replica Number	16 replicas		8 replicas	
1	0.0	1.0	0.0	1.0
2	0.01	0.99	0.015	0.99
3	0.02	0.98	0.03	0.985
4	0.04	0.97	0.045	0.98
5	0.08	0.96	0.06	0.97
6	0.12	0.94	0.075	0.96
7	0.16	0.92	0.09	0.95
8	0.2	0.9	0.12	0.935
9	0.24	0.88		
10	0.28	0.86		
11	0.32	0.84		
12	0.38	0.82		
13	0.44	0.8		
14	0.5	0.78		
15	0.58	0.76		
16	0.68	0.74		

parameters is quantified using the Bennet-acceptance ratio (BAR) method [75]. This method is extended to encompass not only adjacent biases in the α series $\Delta U_{\alpha,\alpha+1}$ but all possible bias pairings $\Delta U_{i,j}$, where $i, j = 1, \dots, N$, with the corresponding free energy differences being calculated using the Multistate Bennet Acceptance Ratio (MBAR) [76]. For the computation of relative binding free energies, only the latter half of the simulation timeframe is considered, treating the initial half as the equilibration period. Statistical errors are estimated by dividing the simulation into five segments and computing the mean and standard deviation.

This method offers a computationally demanding but valuable free energy-like score that includes contributions from the flexibility of the partners and the surrounding solvent, providing a robust foundation as a complimentary scoring method of predicted binding affinities. In this work, the RS-REMD technique employed a series of 8 or 16 parallel-running replicas with increasing repulsive biasing, according to Table 3.1[71].

3.3. Cyclic Peptide Preparation

3.3.1. Head to Tail Cyclization

Special steps have to be added when dealing with the preparation of head-to-tail cyclized peptides:

1. Modification of the AMBER forcefield `leaprc.ff14B` parameter file to eliminate the mapping of terminal residues to allow the cyclic bond. First, a new copy of the file, which should be found at `$AMBERHOME/dat/leap/cmd/`, must be saved into the current working directory. A new name should be given to the file, e.g. `leaprc.cPep`. Then, the section that contains the residue mapping that defines terminal residues as the N- or C-terminal variants of those residues must be eliminated from the copy. It looks as follows:

```
addPdbResMap {  
    { 0 "HYP" "NHYP" } { 1 "HYP" "CHYP" }  
    { 0 "ALA" "NALA" } { 1 "ALA" "CALA" }  
    { 0 "ARG" "NARG" } { 1 "ARG" "CARG" }  
    { 0 "ASN" "NASN" } { 1 "ASN" "CASN" }  
    { 0 "ASP" "NASP" } { 1 "ASP" "CASP" }  
    { 0 "CYS" "NCYS" } { 1 "CYS" "CCYS" }  
    { 0 "CYX" "NCYX" } { 1 "CYX" "CCYX" }  
    { 0 "GLN" "NGLN" } { 1 "GLN" "CGLN" }  
    { 0 "GLU" "NGLU" } { 1 "GLU" "CGLU" }  
    { 0 "GLY" "NGLY" } { 1 "GLY" "CGLY" }  
    { 0 "HID" "NHID" } { 1 "HID" "CHID" }  
    { 0 "HIE" "NHIE" } { 1 "HIE" "CHIE" }  
    { 0 "HIP" "NHIP" } { 1 "HIP" "CHIP" }  
    { 0 "ILE" "NILE" } { 1 "ILE" "CILE" }  
    { 0 "LEU" "NLEU" } { 1 "LEU" "CLEU" }  
    { 0 "LYS" "NLYS" } { 1 "LYS" "CLYS" }  
    { 0 "MET" "NMET" } { 1 "MET" "CMET" }  
    { 0 "PHE" "NPHE" } { 1 "PHE" "CPHE" }  
    { 0 "PRO" "NPRO" } { 1 "PRO" "CPRO" }  
}
```

2. Manual editing to the input PDB file: Removal of the last OXT atom, and addition of a CONECT bond between the C-terminal carbon and the N-terminal

nitrogen at the end of the file.

3. Standard tleap preparation protocol is run for the edited PDB file, sourcing the modified leaprc.cPep parameter file.

3.3.2. Parametrization of Non-Standard Amino acids

Custom non-standard and/or methylated amino acids were parameterized using Amber's Antechamber toolkit AM1- BCC charges and the gaff2 forcefield [77, 78]. tLeap was then used to build the starting structures

3.3.3. Point Mutations

To ensure our cPEPmatch cyclic peptide binders accurately mimic the protein of interest, we introduce specific mutations into the native template. In the realm of computational simulations, this is a relatively streamlined process. Mutations are made by directly removing or adding atoms within the structure file. For the mutations in our study, we performed this task by deleting all side-chain atoms beyond the β -carbon from the existing amino acid in the file. The atoms necessary for the new residue are then automatically generated using the tleap module from AmberTools20 [49]. This method presupposes that the mutation does not significantly alter the overall folding of the protein, which is an important consideration. Particular attention is required when introducing multiple mutations simultaneously into a protein due to the potential for more significant structural impacts. When a more detailed approach to mutations was required, we employed Modeller software to exchange and refine the side chains.

4. cPEPmatch

This chapter introduces the first proof of principle for the cPEPmatch method. Rational design of specific inhibitors of protein-protein interactions is desirable for drug design to control cellular signal transduction but also for studying protein-protein interaction networks. cPEPmatch is a rapid computational approach to rationally design cyclic peptides that potentially bind at desired regions of the interface of protein-protein complexes. The methodology is based on comparing the protein backbone structure of short peptide segments (epitopes) at the protein-protein interface with a collection of cyclic peptide backbone structures. A cyclic peptide that matches the backbone structure of the segment is used as a template for a binder by adapting the amino acid side chains to the side chains found in the target complex. For a small library of cyclic peptides with known high-resolution structures, we found at least one very well-fitting match for the majority ($\approx 82\%$) of 154 protein-protein complexes for a cyclic peptide template to a protein-protein interface segment. The majority of the constructed protein-cyclic peptide complexes was very stable during Molecular Dynamics simulations and showed an interaction energy score that was typically more favorable compared to interaction scores of typical peptide-protein complexes. Our cPEPmatch approach could be a promising approach for rapid suggestion of cyclic peptide binders that could be tested experimentally and further improved by chemical modification.¹

4.1. Introduction

4.1.1. Cyclic Peptides as PPI Modulators

A promising approach for PPI targeting is the design of molecules that closely mimic epitopes, i.e., binding segments, of a partner protein found as part of the interface. These molecules are typically derived from existing peptides and tend to conserve a protein-like chain, but with its chemical structure modified to adjust the molecular properties to become more drug-like. Following this strategy, many investigators

¹This chapter includes material published under a CC-BY license by Frontiers in Chemistry.

have turned their attention to macrocycles, particularly cyclic peptides, as potential PPI inhibitors [7, 9, 8]. Cyclic peptides are among the most promising PPI modulators owing to their ability to potentially bind to large surfaces with reasonable affinity and specificity and due to the enhanced stability and bioavailability as compared to linear peptides [3, 2, 79]. The advantages of cyclic peptides in terms of conformational rigidity and proteolytic resistance can also be further enhanced by modifying the peptide backbone. Proteases act upon cleavage sites in the peptide sequence, which may be at the N or C-terminus, or sometimes at a particular motif within the peptide. Peptide cyclization often reduces the accessibility to cleavage sites in proteases. There is also evidence that cyclic peptides may have improved membrane permeability, compared with linear peptides, in cases where the cyclic peptide can internally satisfy its hydrogen bonds [14].

Early and recent work has demonstrated the possibility to use cyclic peptides that could bind at protein-protein interface regions and interfere with protein-protein binding [80, 81, 82, 83, 84, 85, 41]. Identification of potential appropriate cyclic peptides can be achieved using techniques used in small-molecule drug design efforts such as virtual screening and pharmacophore matching approaches [40, 16, 41]. In this study, we propose a straightforward automated process for the rapid search and optimization of cyclic peptides as protein-protein interaction inhibitors based on known structures of cyclic peptides. Our main goal is to start with a library of existing cyclic peptides and use them as templates for the construction of molecules that closely mimic the epitopes of a partner protein in PPI complexes. We start our process by characterizing backbone motifs of the backbone of epitopes found on the PPI interfaces and comparing them to backbone motifs in a database of cyclic peptides. If a backbone match is found, the cyclic peptide structure is superimposed, and the corresponding interface amino acids of the cyclic peptides are replaced by those of the binding epitope, to closely mimic the specific interface. The generated cyclo-peptide-protein complexes are refined and evaluated by Molecular Dynamics (MD) simulations and energetic evaluation. In this proof-of-principle study, our automated approach is tested on 154 protein-protein complexes. For the majority of 71%, we identify cyclic peptide motifs that result in stable complexes during MD refinement and closely mimic the natural interface epitope.

4.2. Materials and Methods

4.2.1. Cyclic Peptide Matching (cPEPmatch) Approach

Many protein-protein interactions are at least in part mediated by short linear motifs that can be part of α -helices, β -strands, or loops [11]. The workflow of the cPEPmatch approach proposed in this work is based on the idea that the shape of the backbone structure of the PPI interfaces plays a key role in the binding interaction between these short motifs [86]. Our cyclic peptide matching approach is divided into 4 main steps, as outlined in **Figure 4.1**, including a final refinement and evaluation step using the AmberTools18 package [87]. The tool *backbo* characterizes the backbones of each cyclic peptide in terms of distances between backbone atoms of 4 consecutive residues, creating the motifs that are then stored in a database. A second tool, *int_analysis*, identifies all the neighboring protein residues from the PPI complexes as the interface and, subsequently, characterizes its backbone atom motifs also in sets of 4 consecutive residues and identifies matching backbone motifs in the cyclic peptide data file. For each protein-protein interface motif within up to 7 Å distance from the partner protein was considered. For the matching of backbone atom distances of a motif at the protein-protein interface and a corresponding motif in a cyclic peptide, an average distance deviation of up to 0.5 Å was used as a threshold to accept only sufficiently precise matches (termed F-RMSD). In a third step, a cyclic peptide is superimposed on the matching backbone segment of a partner protein to form an appropriate binding placement. This step returns a coordinate file of the cyclic peptide in complex with the receptor partner protein. The side chains of the 4 consecutive (matching) residues at the interface between cyclic peptide and protein are then replaced by the side chains found in the original protein-receptor complex. Exceptions for residue substitutions are made when the matched residue of the cyclic peptide is one involved in a disulfide bond that is part of the peptide cyclization (not changed). The resulting complex was further refined and evaluated by energy minimization and Molecular Dynamics (MD) simulation (see below).

4.2.2. Cyclic Peptide Library and Protein-Protein Complexes

Different chemical strategies can be used to create cyclized peptides, including head-to-tail, disulfide, other side-chain to side-chain, and side-chain to terminus bonding [40, 88, 89, 90, 91, 34]. Incorporating multiple cyclizations to generate peptides that are bicyclic, tricyclic, etc., can provide additional restraints to rigidify the peptide and provide further complexity of design space [92, 93, 29]. We extracted structures of 30

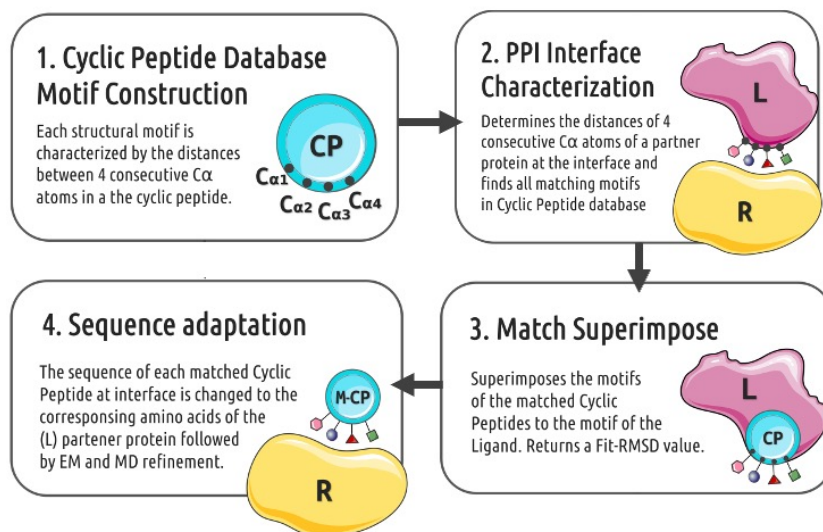


Figure 4.1.: Workflow of the in silico cyclic peptide binder construction. CP indicates cyclic peptide, L and R represent the two partners of a protein-protein interaction pair.

cyclic peptides that vary in cyclization type and sizes as a representative set of cyclic peptides with known well-resolved 3D structure. The cyclic peptides' structure types have been chosen to represent common portions of protein-protein interaction hot-spots such as α -helices, β -strands, and turns. All structures used were downloaded from the Protein Data Bank (PDB), and a complete list is found in Table S1. We applied the method to the analysis of protein-protein complexes chosen from the protein-protein interface (PPI) Benchmark constructed by Tikk and collaborators [94], and the complete list of complexes is given in Table S2.

4.2.3. Molecular Dynamics Simulation, RS-REMD, and MM/GBSA Evaluation

4.2.4. MD

All cyclic peptide-protein complexes were processed for energy minimization and MD simulations using the *tleap* module of Amber18. Protein parameters were retrieved from the ff14SB force field [95]. The complexes were neutralized by the addition of Na⁺ or Cl⁻ ions and solvated in an orthorhombic box with a minimum distance to box-boundaries of 10 Å using explicit TIP3P water molecules [96]. All simulation systems were first energy-minimized with the steepest descent method in 2000 steps by using the Amber18 *Sander* module. All subsequent MD simulations were

performed with the *pmemd.cuda* module. Initially, the systems were heated up to 310 K in three stages (in 100 K steps). Each stage was simulated for 100 ps and included positional restraints on all non-hydrogen atoms with respect to the starting conformation. Subsequently, positional restraints were gradually reduced from 25 kcal·mol⁻¹·Å⁻² to 0.5 kcal·mol⁻¹·Å⁻² in five consecutive simulations of 100 ps at 310 K and at a constant pressure of 1 bar. The equilibrated structures served as input for the production runs for each system, with no restraints. Data gathering simulations were carried out for 5 ns. Coordinates were written out every 500 steps. A time step of 2 fs was used, and all bonds involving hydrogens were constrained to the optimal length using Shake [54]. In order to estimate the stability of a bound cyclic peptide, the L-RMSD (root-mean-square deviation of all non-hydrogen atoms of the cyclic peptide after the best superposition of the receptor binding site on the start structure) was calculated to assess the deviation of the peptide ligand along the 5 ns of trajectory for each match. For the superposition on the receptor structure, the receptor binding site, which includes all atoms within 6 Å of any peptide ligand atom, was considered.

4.2.5. MM/GBSA

MM/GBSA (molecular mechanics Generalized Born surface area) binding free energy calculations were carried out to estimate the strength of cyclopeptide-receptor interactions. Binding free energies were estimated using 250 snapshots retrieved from the last 2.5 ns of MD simulation production using the *MMPBSA.py* module in Amber18. GB calculations were carried out using the modified GB model (*igb=5*) with *mbondi2*, and α , β , and γ values of 1.0, 0.8, and 4.85, respectively. Dielectric constants for the solvent and the solute were set to 80 and 5, respectively.

4.2.6. RS-REMD

We conducted RS-REMD for our highlighted systems. Simulations were conducted in explicit TIP3P water¹² at 300 K, together with neutralizing Na⁺ counterions in water and no additional salt. The prepared system structures were placed within periodic octahedral solvent boxes and a minimum distance between peptide and boundary of 15 Å. The systems were then minimized via Steepest Descent for 10 k steps, heated up step-wise to a target temperature of 300 K within 60 ps and equilibrated within NVT and NPT conditions for 1 ns each. The RS-REMD simulations were executed with 25 ns per replica and exchange attempts every picosecond. These simulations were carried out using 8 and 16 replicas, each exhibiting a progressive increase in the bias applied to the ligand and receptor Lennard-Jones parameters (see Table 3.1).

The trajectories were analyzed using pytraj and the associated relative binding free energies were calculated with MBAR as proposed by Siebenmorgen and Zacharias, 2020 [71].

4.3. Results and Discussion

Identification of cyclic peptides that bind to the interface region of protein-protein complexes may provide a promising route for modulating PPIs [16]. In this work, we present an automated rapid workflow (*cPEPmatch*) that identifies and optimizes cyclic peptides as starting templates to construct modulators of protein-protein interactions. It is based on the matching of small backbone motifs at protein-protein interfaces and corresponding motifs in cyclic peptides (**Figure 4.1**). It has been observed that many protein-protein interaction surfaces are dominated by short segments of peptides [11] and that the constrained backbone often plays a key role in the binding interactions [86].

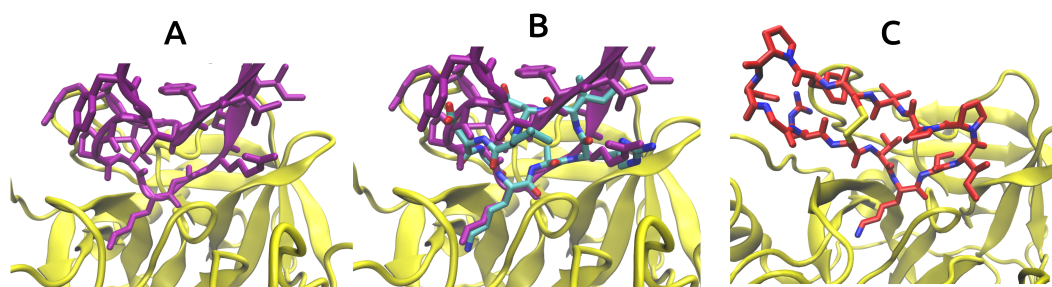


Figure 4.2.: (A) Interface structure of the complex (pdb4dg4) of trypsin inhibitor protein (pink) and trypsin (yellow). (B) same view as in (A) but with a superimposed cyclic peptide (pdb3avb, atom color coded) and adapted interface side chains after 5 ns MD simulation. (C) Complex of sunflower cyclic peptide (red/blue stick model) and trypsin (pdb1sfi).

We first tested the approach on a complex formed by bovine trypsin inhibitor (BPTI) protein and the trypsin protease (pdb4dg4). Our matching approach identified a cyclic peptide backbone that closely resembled the protein backbone of the BPTI protein at the interface (**Figure 4.2 A,B**) with a root-mean-square deviation (RMSD) of the 4 residue backbone segment of only 0.4 Å with respect to the corresponding segment in the BPTI protein. The side chain replacement resulted in an excellent steric fit closely resembling the original interface structure even after the MD simulation refinement in explicit solvent. The predicted structure of the cyclic-peptide trypsin complex is

also in excellent agreement with an experimentally known structure of the sunflower cyclic peptide in complex with trypsin (pdb1sfi, **Figure 4.2 C**), with an RMSD of the interface segment $< 0.5 \text{ \AA}$ (**Table 4.1**). Note that the sunflower peptide is not included in our cyclic peptide database. The evaluation of the cyclic peptide interaction with trypsin resulted in a MMGBSA interaction energy of $-43.8 \text{ kcal}\cdot\text{mol}^{-1}$ (**Table 4.1**). This is less than what is obtained for the sunflower peptide ($-68.8 \text{ kcal}\cdot\text{mol}^{-1}$), but the sunflower peptide is also significantly larger and forms a more extended interface compared to the generated potential small cyclic peptide binder. In addition to the 3avb trypsin inhibitor match, a second putative binder was identified: the 1ebp match, exhibiting a MMGBSA of -38.74 kcal/mol . These findings are particularly significant as the calculated MMGBSA interaction energy for these cases was substantially more favorable than that typically observed for peptide-protein complexes with micromolar binding affinities (**Table 4.1**).

Table 4.1.: Performance of the cyclo-peptide matching approach

MM/GBSA Comparison	
Trypsin - Sunflower Cyclic Peptide Inhibitor (1sfi)	$-68.85 \text{ kcal}\cdot\text{mol}^{-1}$
Trypsin - Matched Cyclic Peptide (3avb)	$-43.79 \text{ kcal}\cdot\text{mol}^{-1}$
Trypsin - Matched Cyclic Peptide (1ebp)	$-38.74 \text{ kcal}\cdot\text{mol}^{-1}$
FBP11 WW Domain (1ywi)	$-30.06 \text{ kcal}\cdot\text{mol}^{-1}$
FE65 WW Domain (2ho2)	$-32.02 \text{ kcal}\cdot\text{mol}^{-1}$
Overall Performance	
Putative match found	82%
Stable binding during MD	86%
Average F-RMSD of best matches	$0.38 \text{ \AA}$
Average L-RMSD of best matches	$2.77 \text{ \AA}$
Average MM/GBSA of best matches	$-43.09 \text{ kcal}\cdot\text{mol}^{-1}$

Further analysis using the Multistate Bennett Acceptance Ratio (MBAR) based on RS-REMD trajectories from 16 replicas substantiated the binding affinity of our identified matches, 3avb and 1ebp. The respective G values, as detailed in Table 4.2, underscore their potential as promising lead molecules for further development. Although their binding affinities did not surpass that of the sunflower cyclic peptide, which remains the most potent binder with a ΔG of -17.958 kcal/mol , the findings for 3avb (δG of -11.345 kcal/mol) and 1ebp (δG of -15.184 kcal/mol) are especially noteworthy. These results affirm the efficacy of our matching approach in pinpointing viable cyclic peptide binders. Additionally, it's worth noting that RS-REMD simulations were also tested with 8 replicas; however, as these matches are all very strong binders, only the 16 replicas were able to sufficiently sample unbound conformations (see Supplementary Figures B.2-B.6). Further studies to assess the sampling perfor-

mance of 8 vs. 16 replicas were conducted for several systems, and the results are presented in Chapter 8.

Next, we applied our *cPEPmatch* approach systematically to 154 protein-protein complexes with known structures to identify putative matches that could potentially bind at the PP interface and interfere with PP interaction. According to our threshold criteria (see Methods), we were able to identify at least one putative cyclic peptide match for 82% of all the cases studied (**Table 4.1**). The putative cyclic peptides matched with the best RMSD and steric fit to each PPI were superimposed, interface residues were substituted, and the interface structure was refined by short MD simulations in explicit solvent. We found that during the 5 ns of MD, 86% of the matched complexes were stable (no dissociation and L-RMSD < 5 Å). Hence, the overall yield of stable predicted complexes of cyclic peptides with protein binding partners was 71% (0.82×0.86). These identified complex structures had an average F-RMSD, L-RMSD, and MMGBSA interaction energy of 0.38 Å, 2.77 Å, and -43.09 kcal mol⁻¹, respectively.

Attempts to replicate this analysis using 8 replicas were less successful in sampling unbound conformations. This limitation suggests that these binders are exceptionally potent, as even with added repulsions post-8 replicas, most sampled states remained bound, as illustrated in Figures S1-12. This further emphasizes the strength of the binding affinity exhibited by 3avb and 1ebp.

Table 4.2.: Binding Free Energy Calculations using MBAR for RS-REMD Trajectories: Results and Uncertainties for 8 Replicas

System	$\Delta G(\text{kcal}\cdot\text{mol}^{-1})$	+/-
Sunflower	-17.96	2.1
3avb	-11.35	0.91
1ebp	-15.18	0.91

Examples of predicted complexes between cyclic peptides that mimic a β -hairpin, a turn, and an α -helical interface segment and partner proteins pdb1cgI, pdb2hle, and pdb2nz8, respectively, are illustrated in **Figure 4.3 A-C**. A more detailed list of the top 30 matches in terms of calculated MMGBSA interaction energy is found in **Table 4.3**. The L-RMSD of all these complexes is in between 1 and 5 Å. The calculated MM/GBSA interaction energies vary significantly from system to system (from -80 to -30 kcal mol⁻¹) which might be due to different sizes but may also be due to the targeted interface region in each complex. However, the MMGBSA interaction energies are, in most cases, more favorable compared to small linear peptides that bind with micromolar binding constants to target domains (**Table 4.1**).

To assess the consistency of our simulations, we expanded our original MD productions by 5 ns and performed a second independent MD simulation of 10 ns for each of the highlighted systems in our study, 1CGI, 2HLE, 2NZ8, and 4DG4. Three MM/GBSA binding free energy calculations were then estimated using 250 snapshots retrieved from 2.5-5.0, 5.0-7.0, and 7.5-10 ns of every simulation (**Figure 4.4 A-D**). Overall, the binding free energy values lie closely to the average for each case, suggesting that the sampled conformations of the complexes can be represented by short simulations.

Furthermore, we evaluated the importance of the generated corresponding mutations of the ligand-fitted cyclic peptides constructed by our cPEPmatch approach by comparing their binding free energies to cyclic peptides mutated with alanines instead of the fitted mutations and to the wild-type cyclic peptides (**Figure 4.5**). The alanine mutant cyclic peptides proved to be considerably weaker binders when the secondary structure is an α -helix or a turn (2HLE, 2NZ8 and 4DG4), but a strong binder in the cyclic peptide with a β -hairpin interface (1CGI). This could be due to the high amount of backbone-backbone interactions that take place in β -hairpin interfaces, where alanines do not get in the way of such binding, however, not having fitted mutations in this case could produce a highly promiscuous binder. The wild-type cyclic peptides for all systems are not as strong binders as the ligand-fitted cyclic peptides but nevertheless showed a strong binding free energy, suggesting that even when the side chains of the ligand are unknown, a backbone fit approach could still work to find suitable modulators of PPIs. Overall, **Figure 4.5** shows that in all cases the binding free energy is stronger with the fitted mutations generated by our cPEPmatch approach.

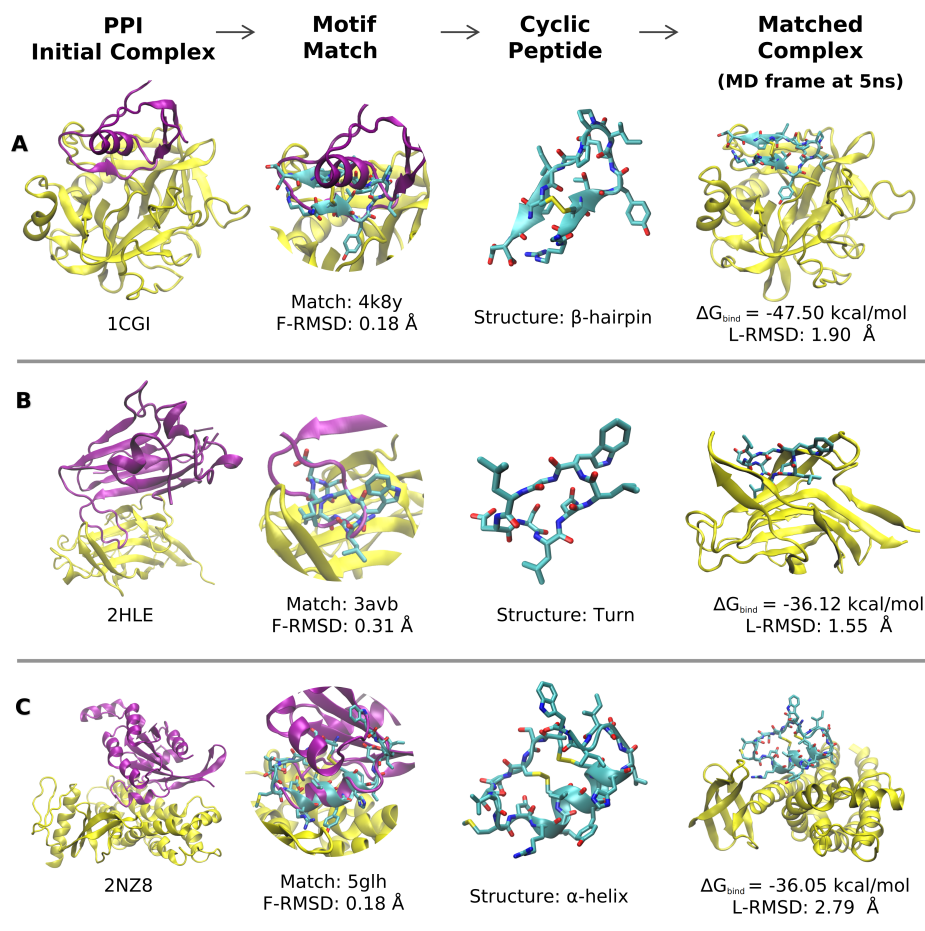


Figure 4.3.: Representative matches and modelled structures of protein-cyclic-peptide complexes. (A) Example of a cyclic peptide with an β -hairpin motif at the interface (pdb-entries of the complex and the cyclic peptide template are indicated; the calculated MMGBSA interaction energy and final deviation from the start structure are also included). (B) Same as in (A) but for a cyclic peptide with a turn motif as binder. (C) same as in (A, B) for a cyclic α -helix binding motif. In each case the target protein-protein complex is shown as cartoon (yellow: receptor; pink: ligand protein). The superimposed matching cyclic peptide is indicated in the second column and the cyclic peptide (with adapted interface sequence) is shown in the third column. The last column represents the final structures of the receptor protein (yellow) in complex with the cyclic peptide after 5 ns MD simulation in explicit solvent.

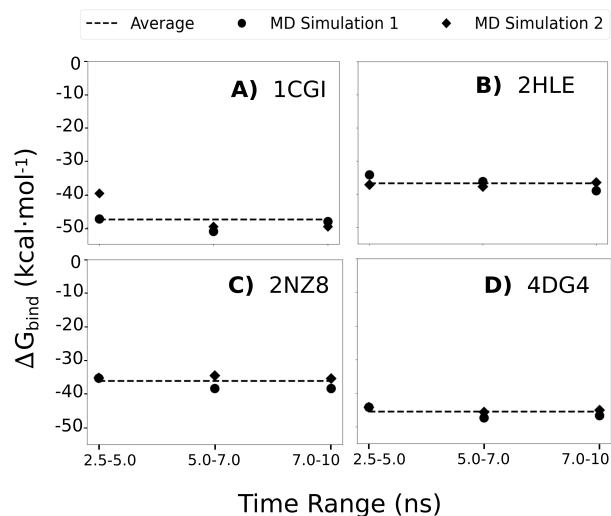


Figure 4.4.: MM/GBSA binding free energy calculations estimated using 250 snapshots retrieved from 2.5-5.0, 5.0-7.0 and 7.5-10 ns of two independent MD simulation productions for each of the studied systems (A) 1CGI, (B) 2HLE, (C) 2NZ8, and (D) 4DG4. The average binding free energy for each case is shown as a dashed line along the graphs.

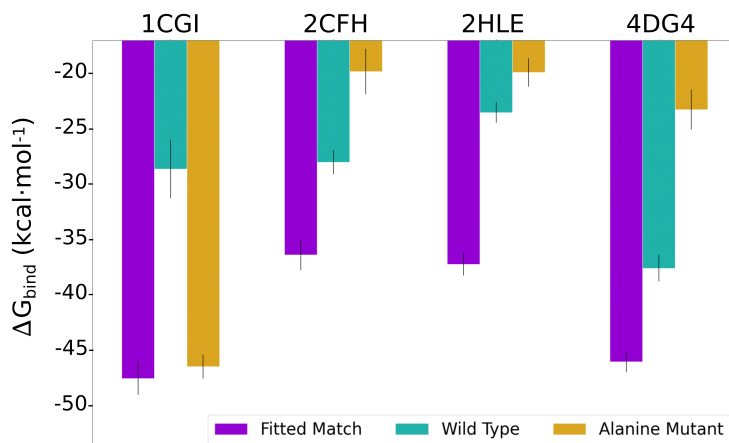


Figure 4.5.: Average of 3 MM/GBSA binding free energy calculations estimated using 250 snapshots retrieved from 2.5-5.0, 5.0-7.0 and 7.5-10 ns of MD simulation productions for the 1CGI, 2CFH, 2HLE and 4DG4 PPI receptors in complex with the Ligand-Fitted Cyclic Peptide Match (Purple Bar), Wild Type Cyclic Peptide (Cyan Bar), and an Alanine Mutant Cyclic Peptide (Orange Bar).

Table 4.3.: Results for the 30 top matches

PPI	Cyclic Peptide	F-RMSD (Å)	MM/GBSA ΔG_{bind} (kcal/mol)	L-RMSD (Å)
1E96	2lwu	1.56	-66.46 ± 0.52	3.53
1ACB	4k8y	0.11	-50.78 ± 0.33	1.17
1AHW	1qx9	2.38	-29.30 ± 0.54	3.42
1AZS	5glh	0.17	-46.10 ± 0.33	3.98
1BKD	5glh	0.07	-82.38 ± 0.46	2.43
1CGI	4k8y	0.18	-47.50 ± 0.34	1.90
1D6R	4k8y	0.09	-53.57 ± 0.60	2.07
1E6E	5glh	0.12	-42.75 ± 0.28	3.95
1EAW	4k8y	0.43	-49.78 ± 0.35	2.23
1FAK	6pio	1.44	-36.27 ± 0.59	2.73
1FSK	3p8f	0.36	-40.38 ± 0.29	2.36
1K74	5glh	0.34	-39.63 ± 0.33	4.77
1LFD	1ebp	0.43	-30.43 ± 0.33	4.00
1NW9	5glh	0.25	-31.99 ± 0.43	4.51
1PXV	4k8y	0.18	-40.25 ± 0.30	3.99
1R0R	4k8y	0.10	-40.74 ± 0.38	1.89
1VFB	5glh	0.18	-34.36 ± 0.24	3.23
2B4J	1npo	0.13	-29.26 ± 0.25	4.16
2C0L	5glh	0.05	-54.09 ± 0.51	1.77
2CFH	5glh	0.18	-50.49 ± 0.41	2.53
2FD6	2lwt	0.75	-37.97 ± 0.47	1.83
2HLE	3avb	0.16	-36.12 ± 0.25	1.55
2NZ8	5glh	0.18	-36.05 ± 0.33	2.79
2SNI	3p8f	0.44	-57.30 ± 0.26	1.11
2W9E	5glh	0.11	-40.54 ± 0.34	4.85
3AAA	5glh	0.09	-43.83 ± 0.39	1.35
3BP8	3p8f	0.47	-36.02 ± 0.23	3.41
3EOA	4k8y	0.23	-41.55 ± 0.28	1.27
3L5W	5glh	0.09	-30.03 ± 0.27	1.02
3V6Z	5glh	0.17	-36.90 ± 0.52	3.38

4.4. Conclusions

One possibility to create cyclic peptide variants based on known protein-protein interface structures is to identify a linear interface motif with an appropriate distance of terminal side chains or backbone atoms such that cyclization is possible. Ideally, such motifs give stable cyclic peptide binders. However, it can be difficult to predict the backbone structure of a cyclized peptide if one introduces a chemical backbone or side chain crosslink. The crosslink can destabilize a desired geometry such that the cyclic peptide does not resemble the structure of the interface motif. An advantage of our strategy is that we base the construction of a desired cyclic peptide on known stable (high resolution) cyclic template structures. Hence, the approach avoids the uncertainty on how well a select cyclization of a given motif resembles a desired backbone structure. Most related to our method is an approach by Duffy et al. [40] to identify both protein-protein interactions suitable for inhibition by cyclic peptides and the accompanying cyclic peptides that represent promising lead compounds. This approach is, however, based on a pharmacophore matching of the PPI interface region with corresponding side chains on sets of cyclic peptides whereas our method is entirely based on structural matching of the interface backbone onto segments of cyclic peptides with known structure. It is complementary also to other approaches to identify cyclic peptides that target PPI [80, 97, 82, 83, 84, 85, 41].

The initial screen of our approach (without explicit solvent refinement) is extremely rapid; hundreds of PPIs can be screened within a few seconds for cyclic peptides that match to backbone structures at the PPI interface. Interestingly, with a relatively small set of cyclic peptide templates, it is possible to identify cyclic peptide-protein complexes that are stable during explicit solvent MD simulations and indicate very favorable interaction energy scores comparable or better than for known stable peptide-protein complexes. The putative cyclic peptide modulators could be further improved by individual post-modifications and by studying specific cases experimentally. Our cPEPmatch method, although at an early stage, has demonstrated the ability to be an easy automatic step towards identifying putative cyclic peptide templates for PPI modulators that could form the basis for subsequent experimental testing. In the future, a larger and even more diverse set of cyclic peptides will be constructed, and some of the putative cyclic peptides will be tested experimentally.

5. Cyclic Peptides Mimicking Protein-Protein Hot Spots

In this chapter, we build upon our foundational approach with a new extension, providing a detailed, step-by-step description of the cPEPmatch method, particularly its application in matching to hot regions within protein-protein interfaces, before proceeding to the matching stage. This enhancement refines the process, reducing the number of putative cyclic peptides to be evaluated and enabling the specific design of peptides that directly compete with the most critical protein-protein binding regions. The utility of this method is exemplified through its application to an MHC class I protein complex.¹

5.1. Introduction

Typical interfaces of PPIs tend to be large, flat, and mainly hydrophobic, and only a few interface residues are crucial for protein-protein binding [30, 98, 99, 31]. These residues, often referred to as hot spots, are major determinants of affinity and specificity [33, 35]. It has also been observed that such hot spots interact cooperatively and tend not to be uniformly spread across the interface but grouped within tightly packed regions, known as hot loops [33, 34, 32]. Most notably, hot loop regions have shown secondary structural features, such as α -helices, β -strands, and turns [100, 101], establishing them as strong candidates for peptidomimetics drug design approaches. The key to peptidomimetics strategies is to graft suitable side chains onto stable backbone scaffolds. In the case of PPIs, a detailed analysis of hot spots at the native interface could serve as a guide for the mimetic rational design. An important step that determines the affinity and specificity of a peptidomimetics drug is the appropriate selection of the scaffold, which is influenced by the location of the hot spots and structural details [102]. Moreover, this understanding is not limited to PPIs alone; it extends and holds relevance for protein-target structures as well, emphasizing the ubiquity and importance of these structural insights across the

¹This chapter includes reproduced material © 2022 The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature.

biological spectrum.

Utilizing the cPEPmatch method, we target these hot spot regions with cyclic peptides [103, 40]. Our case study involves the MHC class I complex, pivotal in immune system regulation [95, 104]. MHC class I complexes bind small antigenic peptides and present them to the immune system at the cell surface, controlling which fragments of a pathogen or cancer antigen are presented to cytotoxic T cells for immune recognition [105]. Peptide binding is strongly coupled to the complexation of the heavy chain part with the β 2-microglobulin (β 2m) partner. Dissociation of β 2m leads also to loss of peptide binding [104]. Hence, the design of cyclic peptides that interfere with this interaction can potentially be used to control and modulate the immune response (including many undesired auto-immune reactions). For our application, we use the pdb-entry 5fa3 (a human MHC class I molecule). In the present example, our goal is to target the heavy chain partner chain by replacing/mimicking the β 2m partner with a cyclic peptide. In **Figure 5.1**, two main contact regions between both segments (chains A and C in the original 5fa3.pdb) are indicated.

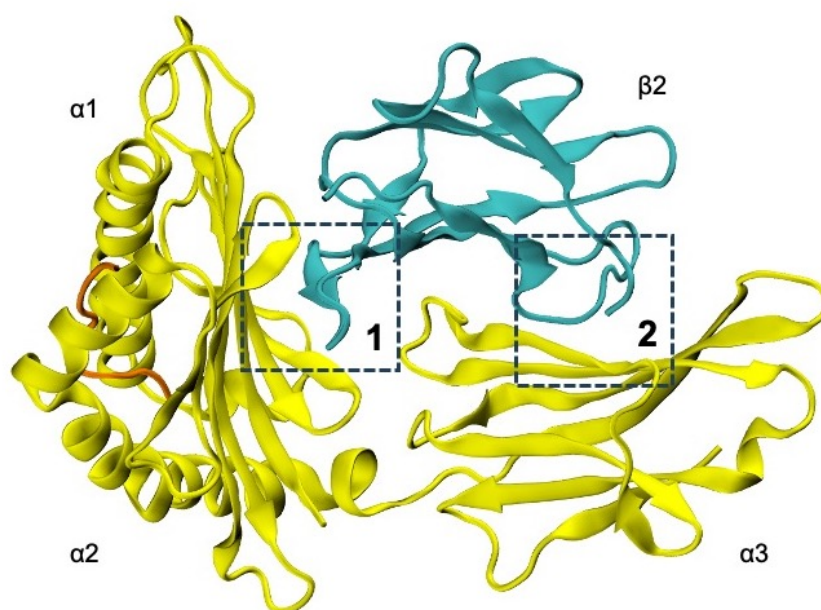


Figure 5.1.: Contact regions 1 and 2 between the heavy chain (yellow cartoon, with the marked sub-domains α 1, α 2, α 3) and the β 2m partner (cyan) of the Major Histocompatibility Complex (MHC) class I complex (pdb5fa3). The bound antigenic peptide (located in the binding cavity formed by the α 1, α 2 sub-domains) is indicated in orange. The dotted boxes mark the two main interface regions between the heavy chain and β 2m partner targeted by the cPEPmatch approach.

5.2. Materials and Methods

The cPEPmatch method requires a known protein-protein complex structure as input: The MHC class I complex PDB:5fa3 was downloaded from the Protein Data Bank (www.rcsb.org). For all energy minimization and MD simulations, we use the Amber18 package [106] that can be obtained from <https://ambermd.org/> (requires an academic license). The code for performing the cPEPmatch calculations can be downloaded freely from <https://github.com/briandasantini/cPEPmatch> (for academic use). It runs on any PC with a Linux operating system and includes installation instructions. Another important input for the cPEPmatch approach is a collection of known cyclic peptide structures that can be used for superposition with structural motifs found at the interface of PPIs. For this work, we have expanded our library from the one presented in the previous chapter to a set of 72 cyclic peptides that vary in cyclization type and size. The cyclic peptide structure types have been chosen to represent common portions of protein-protein interaction hot-spots such as α -helices, β -strands, turns, and loops and are included in the download from <https://github.com/briandasantini/cPEPmatch>.

5.2.1. Identification of hot loops for cPEPmatch

Since there is more than one contact region between chain A and C in 5fa3.pdb, and each includes multiple contact motifs, our direct application of cPEPmatch resulted in a large number of matches. In order to reduce the number of potential complexes that need to be evaluated and to design only the strongest competitors, we used an extension of our original approach. In this extension, we first perform a short MD simulation and MMGBSA application to predict the interface segments that contribute most to the interaction in the heavy chain/ β 2m complex (hot loops). We then focus the search for cyclic peptide binders to these hot loop interface regions. The setup and MD simulation on the MHC complex are performed exactly in the same way as described below for the scoring of the cyclo-peptide/protein complexes.

5.3. System preparation

Structures are processed for EM and MD simulations using the tleap module of Amber18 following standard procedures (see also Amber18 manual). Note, for the simulation of disulfide-bonded cyclic peptides, the input PDB for tleap preparation must have the amino acid name CYX instead of the regular CYS for the cysteine

residues participating in disulfide bonds. The `pdb4amber` tool of the package can add this automatically. Special steps must be added when preparing head-to-tail or similar cyclized peptides, as described in the Methods section, chapter 2.

Following the setup in Amber18, protein parameters are retrieved from the `ff14SB` force field [95]. The complexes are then neutralized by adding Na^+ or Cl^- ions and are solvated in an orthorhombic box with a minimum distance to box boundaries of 10 Å using explicit TIP3P water molecules [96]. First, all simulation systems are energy minimized with the steepest descent method in 2000 steps by using the Amber18 Sander module. Every subsequent MD simulations can be performed with the `pmemd.cuda` module allowing more rapid simulations than the Sander program. Initially, the systems are heated up to 310 K in three stages (100 K, 200K, 310K). Each stage is simulated for 100 ps with positional restraints on all non-hydrogen atoms with respect to the starting conformation. Subsequently, positional restraints are gradually reduced from the initial 25 to $0.5 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ in five consecutive simulations of 100 ps at 310 K and at a constant pressure of 1 bar. The equilibrated structures serve as input for the production runs for each system, with no restraints. Data gathering simulations are carried out for 10 ns. Coordinates are set to be written out every 500 steps. A time step of 2 fs is used and all bonds involving hydrogens are constrained to the optimal length using Shake [54].

5.3.1. Trajectory analysis to score the matches

Stable binding of the cyclic peptide can be assessed first by visual analysis of the MD trajectory. We use the MM/GBSA (Molecular Mechanics Generalized Born Surface Area) method for analyzing the mean interaction energy following the well-established single trajectory method [107] as implemented in the `MMPBSA.py` module of Amber18. Calculations are carried using 500 snapshots retrieved from the last 5 ns of the MD simulation production employing the modified GB model (`igb=5`) with `mbondi2`, and α , β , and γ values of 1.0, 0.8, and 4.85, respectively. Dielectric constants for the solvent and the solute are 80 and 5, respectively. As an output the approach gives the mean interaction energy between cyclic peptide and protein partner.

5.4. Results and Discussion

5.4.1. Trajectory analysis and binding hot spot identification

Residues that contribute most to the interaction in the MHC class I complex are identified using the MM/GBSA method as described in section 4.2 but employing the option to include a per residue interaction energy decomposition (ΔG_{res}), according to the single trajectory method as implemented in the MMPBSA.py module of Amber18 [107, 108]. As output one obtains the mean interaction energy contribution for each residue along the sequence (**Figure 5.2**). In our case, we use all residues belonging to the $\beta 2\text{m}$ partner because our designed cyclic peptide should superimpose on a backbone segment of the $\beta 2\text{m}$ partner. We define hot spot residues potentially important for binding as those with a total $\Delta G_{\text{res}} < -k_B T = -0.6 \text{ kcal}\cdot\text{mol}^{-1}$ (kB: Boltzmann constant, T: temperature: 300K) and with the majority of its interaction energy contribution due to side-chain interactions. Hot loops are then segments of 8 to 10 residues that include at least 4 hot spot residues. For the present example, two hot segments can be clearly identified (**Figure 5.2**). Loop 1 is chosen from residues 278 to 287, and it contains 4 hot spots: Lys279, Gln281, Tyr283, and Arg285. Loop 2 is selected from residue 326 to 335, which is comprised of 5 hot spots: Asp326, Leu327, Phe329, Trp 333, and Phe335. The side-chain and backbone contributions for all the hot spots are shown in **Table 5.1**. The coordinates of both loops were extracted from the last frame of the MD simulation and used as input for cPEPmatch. It should be noted that alternative methods to identify important interaction regions can also be used at this step.

Table 5.1.: MM/GBSA free energy decomposition (kcal/mol) for the two chosen hot loops. Three free energy values are shown for each residue: total (ΔG_{res}), side-chain ($\Delta G_{\text{res-ss}}$), and backbone ($\Delta G_{\text{res-bb}}$) contributions.

Loop 1				Loop 2			
Residue	ΔG_{res}	$\Delta G_{\text{res-ss}}$	$\Delta G_{\text{res-bb}}$	Residue	ΔG_{res}	$\Delta G_{\text{res-ss}}$	$\Delta G_{\text{res-bb}}$
PRO 278	0.02	-0.06	0.08	ASP 326	-9.11	-9.03	-0.08
LYS 279	-2.37	-2.49	0.12	LEU 327	-1.88	-2.38	0.49
ILE 280	0.20	-0.10	0.30	SER 328	-0.92	-0.42	-0.51
GLN 281	-3.26	-3.60	0.35	PHE 329	-5.35	-5.54	0.20
VAL 282	0.10	-0.14	0.24	SER 330	0.70	0.18	0.51
TYR 283	-6.89	-6.99	0.10	LYS 331	0.64	0.28	0.36
SER 284	-1.05	0.04	-1.10	ASP 332	0.87	0.70	0.17
ARG 285	-3.00	-3.77	0.77	TRP 333	-8.35	-7.28	-1.07
HIE 286	-0.32	-0.18	-0.14	SER 334	-0.06	-0.01	-0.05
PRO 287	-1.16	-0.80	-0.36	PHE 335	-2.83	-2.77	-0.06

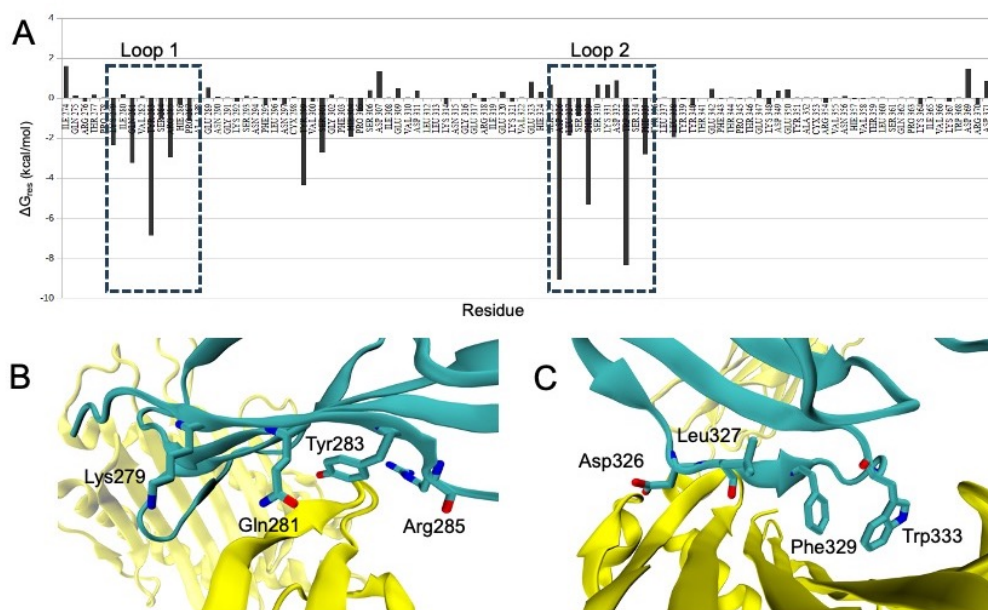


Figure 5.2.: Hot segment selection for each contact point is depicted in the corresponding figures. Figure A shows per-residue contributions to the binding effective energy as calculated by MM/GBSA decomposition, with the chosen hot loops highlighted inside blue rectangles. Figure B illustrates Loop 1 interface with labeled hot spots, and Figure C does the same for Loop 2. In all representations, domains are rendered as cartoons with $\alpha1$, $\alpha2$, $\alpha3$ in yellow, and $\beta2$ in cyan.

5.4.2. Application of cPEPmatch to hot loop regions

The application of the cPEPmatch approach follows the same procedure as described in Chapter 4, using a threshold and a cutoff of 0.5 and 7.0 Å, respectively. However, instead of the original MHC class I complex file, we use as input a modified complex file that contains the heavy chain coordinates (receptor) and just the identified hot loop structures (in two separate complex pdb files).

5.4.3. Selected results for targeting the MHC α chain

Three stable cyclic peptide matches were found for each hot loop (**Table 5.2**). Although it is not possible to include every single hot spot for each structure, as many as possible were mutated on each match. **Figure 5.3** shows one representing match for each loop after 10 ns of MD simulations. In both cases, we observe that the mutated hot spots have similar orientations than those found on the $\beta 2$ domain that they are mimicking. Also, the decomposition of the binding free energy shows similar behavior for all of the matches. A total of 6 putative cyclic peptides have been found to target the α chain of the MHC class I.

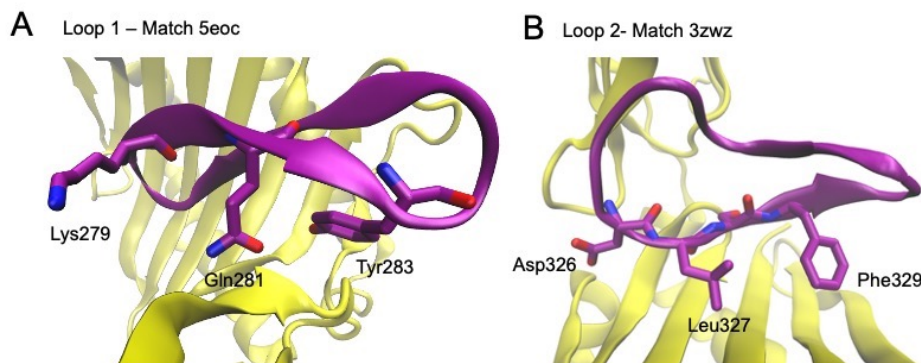


Figure 5.3.: Representative matches and modeled structures of protein-cyclic-peptide complexes for the MHC class I heavy chain (yellow) targeting the interaction with $\beta 2m$ are presented. In panel (A), the cyclic peptide match pdb5eoc mimics Loop 1. Panel (B) displays the cyclic peptide match pdb3zwz emulating Loop 2. For both instances, the heavy chain (yellow) and matched cyclic peptides (pink) are depicted as cartoons. The labeled residues (sticks) correspond to the $\beta 2m$ residues in the native MHC class I complex that are substituted in the complex with the cyclic peptides.

This targeted approach was particularly pivotal in the context of our cPEPmatch search for the MHC complex. By focusing on specific structural aspects unique to the MHC, we effectively honed our search criteria within the vast array of potential

matches. This was instrumental in not only reducing the computational load but also in ensuring that our search was intensely focused on the most viable regions of interest in the MHC structure. The application of cPEPmatch, tailored specifically to the nuances of the MHC complex, significantly enhanced the efficiency and precision of our identification process for potential cyclic peptide binders.

In the initial phase of our investigation, particularly in the application of the cPEPmatch method to the MHC complex, we encountered a significant challenge due to the vast number of matches generated. This complexity necessitated an extensive and computationally expensive molecular dynamics (MD) search to sift through these possibilities. However, by focusing specifically on the structural characteristics of the MHC system and learning from its unique interaction patterns, we directed our cPEPmatch efforts more effectively towards identifying relevant matches. This structure-based, targeted approach not only accelerated the discovery process but also proved to be computationally more efficient. Focusing on hot spots unique to the MHC complex, we were able to narrow down our search significantly and quickly zero in on the most promising cyclic peptide candidates, optimizing both our resources and time in the process.

Table 5.2.: Best matches found for both hot loops identified for the MHC class I system.

Loop	Match ^a	Substitutions in the cyclic peptide	F-RMSD (Å)	ΔG_{total} (kcal/mol)
1	1ebp	9-VAL, 10-TYR, 11-SER, 12-ARG	0.24	-25.17
	5eoc	2-LYS, 3-ILE, 4-GLN, 5-VAL, 6-TYR	0.04	-15.75
	4w4z	3-LYS, 5-GLN, 6-VAL, 7-TYR	0.09	-24.62
2	3zwz	8-ASP, 9-LEU, 10-SER, 11-PHE	0.16	-39.53
	3avb	1-ASP, 2-TRP, 3-SER, 4-PHE	0.06	-20.88
	1ebp	14-SER, 15-ASP, 16-LEU	0.14	-51.63

a) indicates the pdb-entry of the matching cyclic peptide

5.5. Conclusions

In this study, we have demonstrated that the key to finding a stable binder lies in the adaptation of the cyclic peptide sequence to mimic the essential protein-protein interface interactions closely. We described an extension of our original cPEPmatch approach, which is particularly effective for targeting protein-protein interactions (PPIs) that feature multiple and/or large binding sites. This method involves a short molecular dynamics (MD) simulation combined with MMGBSA application, facilitating the identification of 'hot loops' in the PPI prior to the matching process.

This step is crucial for pinpointing putative cyclic peptides that specifically target such regions. Importantly, the identification of hot loops substantially narrows down the number of potential cyclic peptides to be evaluated. It allows for the focused design of cyclic peptides that are tailored to compete with the strongest protein-protein binding regions. Applying the cPEPmatch hot loop extension to the heavy chain of an MHC class I complex, we successfully identified six putative cyclic peptide binders. These findings underscore the utility of our approach in the rational design of peptide-based inhibitors for complex PPIs.

6. Design of Glycosaminoglycan-Binding Cyclic Peptides

This chapter addresses the design of cyclic peptides targeting glycosaminoglycans (GAGs), crucial for numerous biological processes and various disease mechanisms. We discuss the structural complexities of GAG-protein interactions and our strategy to create cyclic peptides that mimic these interactions effectively, offering potential therapeutic and diagnostic uses. The chapter introduces an adapted version of the cPEPmatch approach for designing peptides that replicate the binding capabilities of GAG-interacting proteins. Additionally, it is noted that the binding efficacy of the tested peptide match has been experimentally validated.¹

6.1. Introduction

Glycosaminoglycans (GAGs) are linear, negatively charged periodic polysaccharides that play critical roles in various physiological and pathological processes in the extracellular matrix. Due to their biological functions and their role in the onset of a number of diseases, GAGs have become attractive targets for developing therapeutic and diagnostic tools [109]. However, the structural complexity and heterogeneity of GAGs pose significant challenges to designing molecules that can selectively target them [110]. Recently, the ability of cyclic peptides to bind to GAGs has been exploited for various applications, including the development of therapeutics and diagnostic agents. For example, cyclic peptides that bind to heparin (HP), a class of GAGs, have been developed for use as anticoagulants [111].

HP functions as an anticoagulant through its interaction with the serine protease inhibitor antithrombin-III (ATIII). ATIII's anticoagulant activity is caused by the selective inhibition of procoagulant proteases (mainly thrombin, factor XIa, and factor Xa) in the blood clotting cascade [112]. HP and heparan sulfate (HS) act

¹This chapter includes material from the article 'Rational Design of Glycosaminoglycan Binding Cyclic Peptides using cPEP-match,' currently under review in Computational and Structural Biotechnology Journal.

as ATIII cofactors, and in their presence, the inhibition activity of ATIII towards its target proteases is increased up to 4000 times [113, 114]. Unlike many GAG-protein interactions, predominantly mediated by unspecific electrostatic interactions, ATIII requires highly specific GAG sulfation patterns, including an uncommon 3-O sulfation group, for the effective mediation of its activity [115, 116]. Taking all this into account, the idea of engineering cyclic peptides with a similar GAG binding affinity as ATIII offers a promising direction in therapeutic advancements. Moreover, these peptides could also contribute to the development of therapeutics targeting other GAG-modulated processes, such as cell adhesion and inflammation [117].

In this chapter, we report an adaptation of the cPEPmatch approach to design cyclic peptides that specifically target GAGs by mimicking the hot spots of GAG-binding proteins. Mimicking the interfaces of GAG-binding proteins with cyclic peptides is a challenging task, and it requires a deep understanding of the structure and function of the protein binder interface. GAG-binding regions are often characterized by clusters of positively charged amino acid residues, which can form electrostatic interactions with the negatively charged GAGs. However, such hot spot residues do not generally belong to a single specific strand or chain but rather from a cluster of non-consecutive exposed regions of the binding proteins. In this work, we present a modified version of the cPEPmatch code that allows for matching cyclic peptides that target sequentially non-consecutive hot spots. Our approach provides a potential avenue for the development of novel GAG-targeted therapeutics and diagnostic tools that can potentially modulate GAG functions in various biological processes.

6.2. Materials and Methods

6.2.1. Structures

In this study, ATIII was modeled based on the crystal structure in which this protein is bound to a pentameric heparan sulfate (HS) substrate (PDB ID 1E03, [118]). This dimeric structure consists of ATIII in two different conformations, and the active conformation corresponding to chain A was chosen as it has a higher affinity towards the GAG-ligand [118, 119].

The sequence of the GAG used in simulations with cPEPs is based on the pentamer from the ATIII complexed crystal structure (1E03). In this oligosaccharide, the glucosamine nitrogens are replaced by oxygens. However, in the study, we used this structure with the natural N-sulfation corresponding to GlcNS(6S)–GlcA–GlcNS(3,6S)–IdoA(2S)–GlcNS(3,6S), where GlcA denotes glucuronic acid and IdoA

denotes iduronic acid, GlcNS is sulfoglucosamine, and additional sulfate groups are marked by their position in the respective sugar chain. This ligand is termed as HS in the manuscript. Additionally, the classical HP sequence GlcNS(6S)–IdoA(2S)–GlcNS(6S)–IdoA(2S)–GlcNS(6S) was modeled.

6.2.2. Cyclic Peptide Matching Update

The original cPEPmatch workflow consists of four main steps: (1) Cyclic peptide database construction. (2) Identification and characterization of the protein interface to mimic. (3) Interface backbone matching against the cyclic peptide database and conducting structure superimposition. (4) Mutation of the matched amino acids to the side chains corresponding to the protein to mimic. All matches are then refined and evaluated using Molecular Dynamics (MD) simulations and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) [107] free energy calculations, respectively. The original methodology underwent several modifications, which include:

All matches are then refined and evaluated using Molecular Dynamics (MD) simulations and MM-GBSA free energy calculations. The original methodology underwent several modifications, which include

- Migrating from Fortran to Python to improve ease of use, readability, and maintainability.
- In the first and second steps, we characterized the backbone structure of cyclic peptides and protein to mimic, respectively, by calculating distances between backbone atoms of four consecutive residues – based on the notion that the backbone structure of these motifs plays a crucial role in binding interactions [86]. In the GAG binding protein, hot spot residues are in general not present as sequentially consecutive motifs. Thus, we have added a modification for the search for non-sequential motifs.
- In step two, the user can now specify which particular amino acids to target, enhancing the selection process for the interface mimicry.
- In step three, a range of motif lengths from four to seven amino acids can be selected for matching.
- In step four, the matched side chains are replaced and minimized using Modeller [120] to reduce steric clashes and optimize the overall stability of the structure.

6.2.3. cPEPmatch Software Availability

The source code for cPEPmatch is publicly accessible in the GitHub repository at <https://github.com/briandasantini/cPEPmatch>. This resource provides insights into the software's functionality and usage guidelines.

6.2.4. ATIII Cyclic Peptide Matching Study

In our ATIII cyclic peptide matching study, we fine-tuned the search parameters to identify the most suitable matches for this complex system. A 'good match' was determined through visual inspection using VMD software, focusing on cyclic peptides that ensure correctly positioned matched side chains without interference with other receptor strands. The most effective results were achieved by targeting hotspot motifs composed of 5 residues, setting a Fit-RMSD Threshold of 3 Å. The hotspot entries for ATIII were defined as residues contributing less than -5 kcal/mol to the binding energy, according to MM-GBSA calculations, which include nine residues listed in descending order of their favorable free energy contributions.

In our ATIII cyclic peptide matching study, we fine-tuned the search parameters to identify the most suitable matches for this complex system. A 'good match' was determined through visual inspection using VMD [121], focusing on selecting cyclic peptides that ensure correctly positioned matched side chains, thereby avoiding interference with other ATIII strands. The most effective results were achieved by targeting hotspot motifs composed of 5 residues, along with setting a Fit-RMSD threshold of 3 Å. The Fit-RMSD threshold is pivotal as it dictates how closely the cyclic peptide's backbone must align with the targeted protein interface residues to ensure an optimal fit. In cPEPmatch, the hotspot entries for ATIII were defined as residues contributing less than -5 kcal/mol to the binding energy, according to MM-GBSA calculations. These include nine residues: K101, R34, R8, K112, N32, K6, R116, R33, T31, listed in descending order of their favorable free energy contributions [122].

Following the cyclic peptide matching, MD simulations and free energy calculations were conducted to evaluate the binding of the generated cyclic peptides with both analyzed ligands 1 and 2. Subsequently, an optimization phase was carried out for these cyclic peptides, involving a series of strategic mutations to enhance their binding. This refinement step was followed by additional MD simulations to assess the impacts of these modifications on the peptide-ATIII interactions.

6.2.5. MD Simulations

MD simulations involved up to four distinct versions of each cyclic peptide matched. The initial configuration (version 1) included up to 5 residues mutated based on the matching results. Alternative mutations were introduced in subsequent versions, informed by the outcomes of the initial simulations. Simulations with the wild-type sequences (version 0) were also performed for comparative analysis. The configurations of the various versions, including specific mutations and sequence variations, are detailed in Table 6.1.

Table 6.1.: Versions of the cPEPs investigated in this study. AA denotes amino acid residues, V. denotes version. Version 0 is the wild-type sequence. Bold residues are the ones that mimic the residues in the binding site of ATIII. Residues in italics are introduced cysteines for disulfide bridges.

ID	AA	V.	Sequence	Matching Residues
5EOC	13	0	CQLINTNGSWHIC	
		1	CNRRNTKGSWHIC	N32, R33, R34, K101
2LWT	18	0	GRYRRCIPGMFRAYCYMD	R33, R34
		1	GTNRRCKPGMFRAYCYMD	T31, N32, R33, R34, K101
		2	GTNARCKPGMFRAKCKMR	T31, N32, R34, K101, K6, R112, R116
		3	GTNARCRPGKFRAKCKMR	T31, N32, R33, R34, K101, K6, R112, R116
2JRW	23	0	CAEPMTLPENYFSERPYHPPPC	
		1	CAEPMTLPEKYFSENRRHPPKPC	K6, N32, R33, R34, K101
		2	CAEPRTLPEKYFKTNRHPPKPC	R8, K6, K112, T31, N32, R33, R34, K101
		3	CARPKTLPKRCFSENRCRPKPPC	R8, K6, K112, R116, N32, R33, R34, K101

Notes: Version 3 of 2JRW is structurally reinforced with an internal disulfide bridge introduced by mutating two residues to cysteine, enhancing peptide stability.

100 ns MD simulations were performed for all complexes with the cyclic peptides and the pentameric ligand HS, in addition to ATIII and 2JRW version 2 with both ligands 1 and 2. For the cyclic peptide complexes, the simulations were repeated up to three times. The simulations were performed in AMBER20 with ff14SBonlysc force field for protein and GLYCAM06 for GAG parts of the system [123, 124, 125]. To neutralize the system, sodium ions were added, and the systems were solvated with a 15 Å truncated octahedron box using TIP3P water [96]. Energy minimization was performed in two steps: 500 steepest descent cycles with 103 conjugate gradient cycles and 100 kcal/mol/Å² harmonic force restraints on solutes, followed by 3x10³ steepest descent cycles and 3x10³ conjugate gradient cycles without restraints. Subsequent heating to 300 K was performed for 10 ps using the same restraints before equilibration for 500 ps at 300 K and 105 Pa in isothermic isobaric ensemble. The production runs were performed at the same conditions as equilibration. Trajectories were evaluated with cpptraj module of AMBER20 and visualized in VMD.

Molecular mechanics-generalized Born surface area (MM-GBSA) [126] in AMBER 20 with mode igb = 2 was performed over the MD-trajectories to assess the affinities of the complexes. Per-residue decomposition allowed for a detailed investigation of the contribution of interacting residues.

6.3. Results and Discussion

6.3.1. Adapting cPEPmatch to Complex Binding Sites

The original proof-of-principle cPEPmatch code was designed for target structures with cyclic peptides featuring consecutive motifs of four residues. It defines the binding site of a receptor by including all residues within a specific distance cutoff from the binding interface with the ligand. However, for protein-GAG interactions, such as the ATIII binding site with heparan sulfate pentasaccharide (PDB ID 1E03, [118]), modifications to the matching algorithm were necessary. MD simulations and decomposed energies for hot spot residues from MM-GBSA energy calculations are depicted in **Figure 6.1 A**. The binding site, illustrated in **Figure 6.1B**, highlights nine residues with free binding energies below -5 kcal/mol, primarily non-consecutive. To address this, cPEPmatch was updated to allow matching with non-consecutive residues. The code now permits manual definition of binding sites, using nine hot spot residues identified from MM-GBSA calculations. **Figure 6.1 C** shows three distinct peptides selected for further investigation, and **Figure 6.1 D** illustrates the potential for additional mutations in these peptides. Details of mutated residues are listed in Table 1.

6.3.2. MD and Optimization

The potential binding of cyclic peptides to GAG was analyzed with MD simulations. The complexes of 2LWT and 2JRW, both derived from cPEPmatch and manual refinements, were examined, along with their wild-type sequences. MM-GBSA calculations determined the free energies of binding, with results suggesting stronger binding in peptides with additional ATIII-mimicking residues (**Figure 6.2**). Per-residue decomposed binding energies (**Figure 6.3**) and the RMSD analyses of GAG ligands and peptides (**Figure 6.4**) further characterized the stability of these complexes. The 2LWT complexes demonstrated the highest stability during simulations. Given its demonstrated strong binding affinity and stability in MD simulations, as shown in Figures 2-4, alongside the consistent stability observed in MD trajectories (**Figure 6.4**),

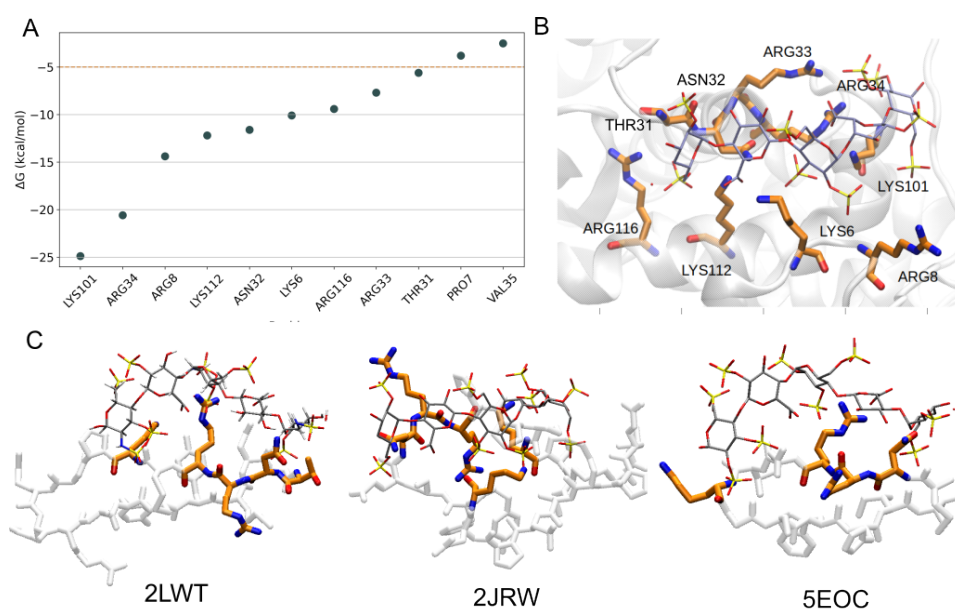


Figure 6.1.: A) Per-residue decomposition of MM-GBSA binding energy for ATIII, highlighting residues with binding energy stronger than -5 kcal/mol. Residues below the dotted blue line are included in panel B. B) Visualization of the ATIII binding site (PDB ID 1E03) as a white cartoon. The key GAG-binding residues are depicted in orange sticks, while the GAG molecule is represented in thin gray sticks. C) Illustration of three non-consecutive cPEPs that match the binding site.

2LWT version 2 was selected for further investigation into its binding specificity.

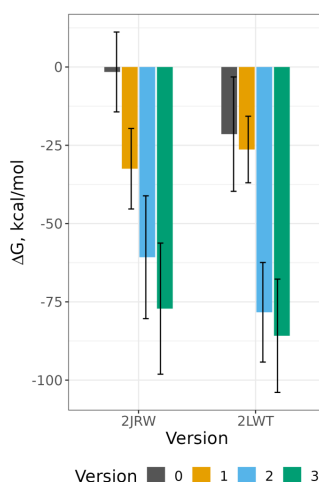


Figure 6.2.: Total ΔG of all tested cPEP versions of 2JRW and 2LWT in complex with the HS GAG ligand, as calculated by MM-GBSA. Error bars represent the average over all frames from three independent runs, totaling 300 ns.

6.3.3. Specificity of the ATIII-GAG Interactions

Antithrombin III (ATIII) exhibits high specificity when binding to the GAG ligand, as observed in the crystal structure (PDB ID 1E03), with significantly reduced affinity for GAG ligands having even minor alterations in sulfation patterns. The precise specificity observed among GAG-binding proteins and carbohydrates in the extracellular matrix is critical for various biologically relevant functions in mammals. To emulate a successful mimetic, our selected cyclic peptide must similarly exhibit specificity to the same GAG ligand as ATIII. To assess this, we compared the peptide's binding to heparan sulfate (HS) with its interaction with a heparin (HP) peptamer. Notably, the 3-O sulfate at position 5 is a key determinant for ATIII specificity [115, 116]. Additionally, we evaluated the effect of the ligand's initial binding pose, comparing two different interaction profiles with the cyclic peptide.

The binding free energies of 2LWT version 2 interacting with the GAG, calculated from trajectories of three MD simulations for each ligand in both initial positions, are displayed in **Figure 6.5 B**. The binding affinity for the HS ligand was slightly more favorable compared to the HP ligand. Interestingly, the HS ligand progressively shifted to a more energetically favorable interaction pose in the second position, as shown in **Figure 6.5 C**. This transition was not observed with the HP ligand, implying a distinct specificity of the cyclic peptide for the HS ligand. These findings indicate

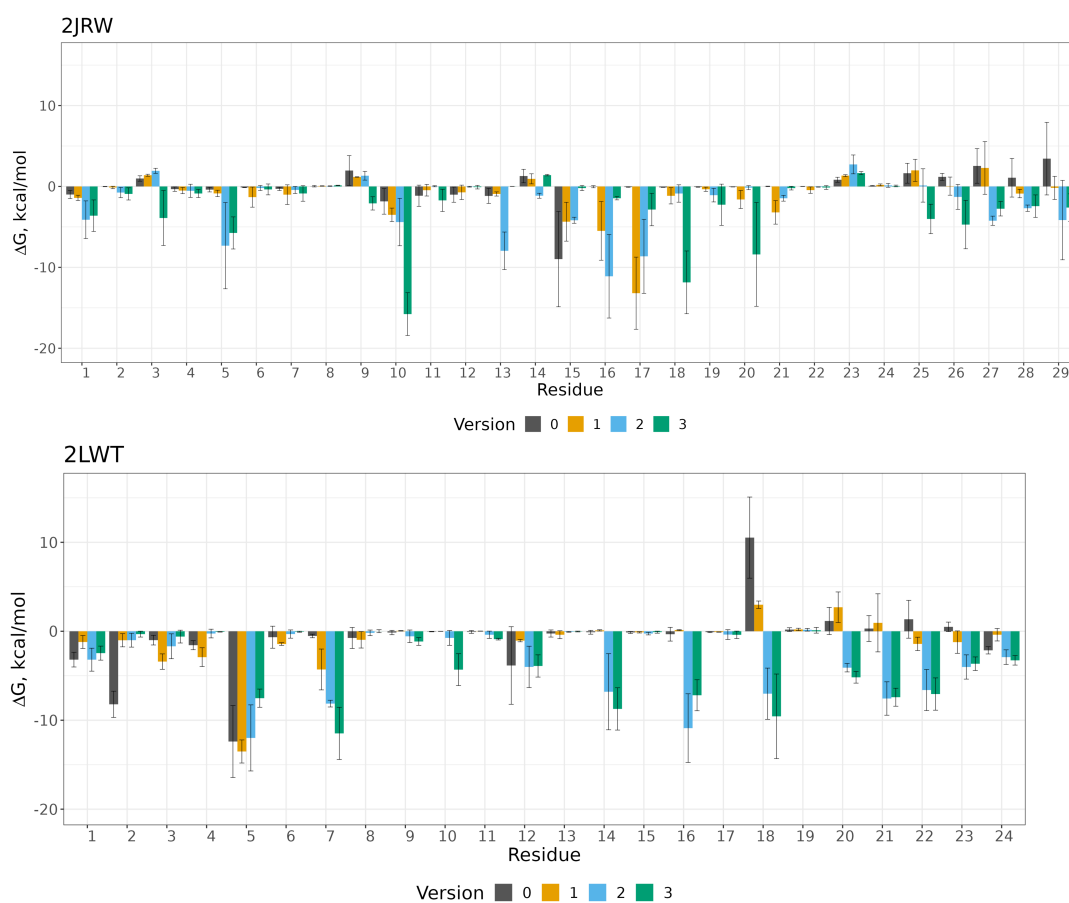


Figure 6.3.: MM-GBSA Per-residue ΔG decomposition for all tested cPEP versions of 2JRW and 2LWT in complex with the HS GAG ligand. Error bars indicate the average over all frames from three 300 ns runs.

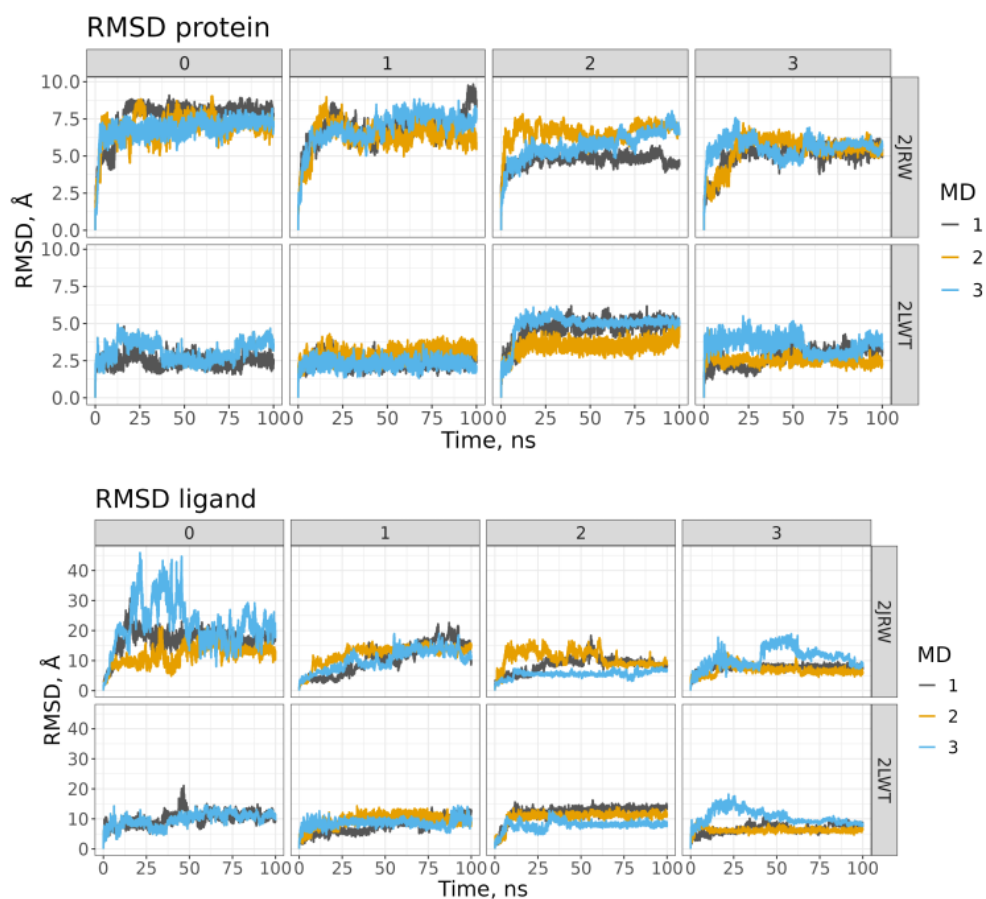


Figure 6.4.: RMSD plots for the ligand and protein over the entire simulation trajectory. The data is shown for cPEP matches 2JRW and 2LWT, illustrating the stability of these complexes over time.

that the cyclic peptide exhibits a certain degree of binding specificity towards the HS ligand.

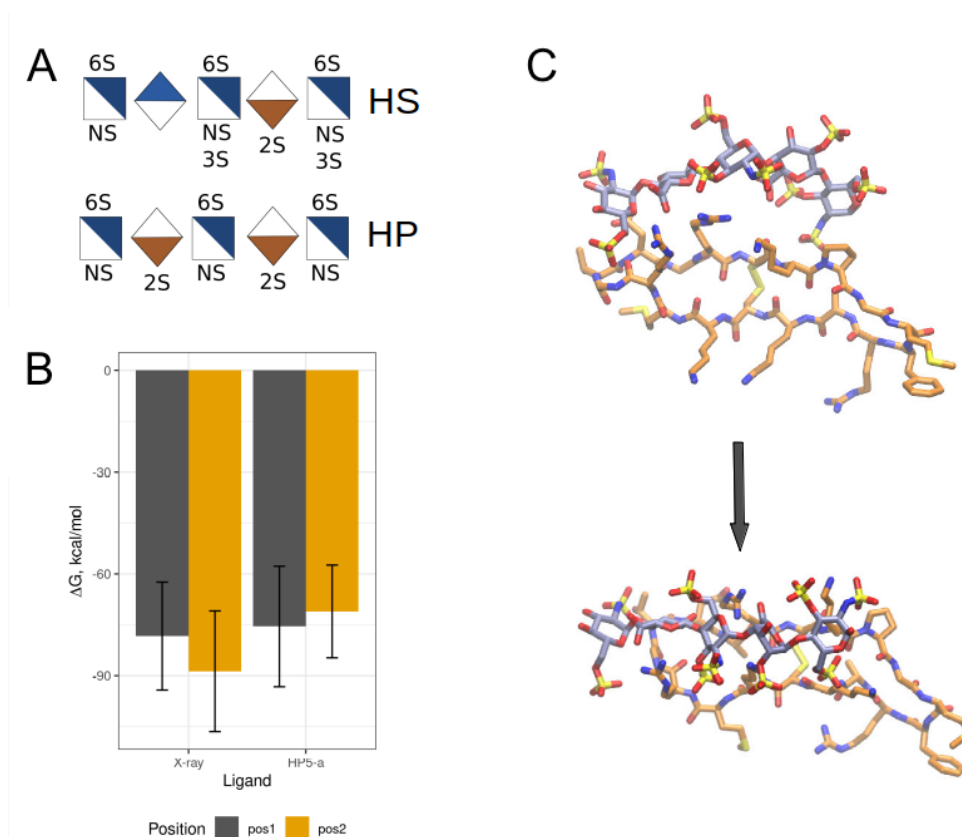


Figure 6.5.: A) Structures of HS and HP. B) ΔG values from MM-GBSA based on three MD simulations of the 2LWT v2 cyclic peptide with HS and HP, each starting from two distinct initial positions. C) Trajectory of the HS ligand's movement in the MD simulation starting from position 2, showcasing the initial and final frames.

6.3.4. Experimental Validation

The promising results from our computational predictions, encompassing cPEPmatch, MM-GBSA, and specificity studies, led to the experimental validation of the 2LWT version 2 cyclic peptide's binding. NMR and ITC techniques were employed to assess the binding affinity of this peptide to the GAG ligand, aiming to validate the insights gained from our computational studies. Experimental validation confirmed the binding of cPEP 2LWT V2 to Fondaparinux, as evidenced by NMR and Isothermal Titration Calorimetry (ITC) assessments, revealing a dissociation constant of approximately 64 μM and an entropy-driven binding process.

6.4. Conclusions

This study represents a significant stride in the computational analysis of glycosaminoglycan (GAG)-binding cyclic peptides, demonstrating the effectiveness of cPEPmatch in the discovery of peptide-based therapeutics. We have refined the cPEPmatch algorithm to effectively identify and characterize complex GAG-protein interaction sites, setting a precedent for methodical investigations into cyclic peptides as GAG binders. A notable outcome of this study is the identification of promising peptide binders, particularly 2LWT version 2, underscoring the strength of our computational strategy. This research not only enriches our comprehension of peptide-GAG interactions but also fosters new pathways for further investigation and development in cyclic peptide therapeutics aimed at GAGs. The methods and insights gained from this study are significant for the ongoing discovery and enhancement of novel GAG-binding cyclic peptides, potentially leading to breakthroughs in therapeutic applications.

7. Design of ICOS-ICOSL Cyclic Peptide Modulators

The interaction between the inducible costimulator (ICOS) and its ligand (ICOSL) plays a key role in T-cell differentiation and activation. ICOS/ICOSL inhibition results in a reduction in the promotion of immunosuppressive regulatory T cells (Tregs) in both hematologic malignancies and solid tumors. In this chapter, we implement a new cyclic peptide database and everything we've learned up to now about the computational cPEPmatch approach to design the first examples of cyclic peptides that inhibit ICOS/ICOSL interaction. The top cyclic peptide from our approach possessed an IC_{50} value of $1.87 \pm 0.15 \mu M$ as an ICOS/ICOSL inhibitor and exhibited excellent in vitro pharmacokinetic properties as a drug candidate. This work will lay the groundwork for future endeavors in cancer drug discovery, with the goal of developing cyclic peptides that target the ICOS/ICOSL interaction.¹

7.1. Introduction

Immune checkpoint blockade (ICB) has transformed the treatment of various cancer types, delivering remarkable survival outcomes for a subset of cancer patients [127, 128]. Unfortunately, the majority of cancer patients do not exhibit a positive response to ICB, or their responses are fleeting due to both inherent and acquired resistance [129, 130]. ICB resistance can be attributed to mechanisms within the cancer cells themselves, as well as factors originating from the tumor microenvironment (TME). These factors encompass the increased expression of alternative immune checkpoints and the presence of tumor-infiltrating regulatory T cells (Tregs).[129, 130] A promising approach to boost antitumor T-cell function and overcome ICB resistance is targeting co-stimulatory immune checkpoints expressed on the surface of T cells [131, 132, 133]. However, targeting co-stimulatory checkpoints with monoclonal antibodies (mAbs) has led to severe side effects that hinder their clinical application [134, 135]. For instance, phase I trials evaluating the effectiveness of anti-CD28 mAb were halted due

¹This chapter includes material published at © 2024 Elsevier Ltd. All rights reserved.

to adverse events such as cytokine storms, resulting in acute respiratory distress [134]. Current research efforts are dedicated to developing effective therapeutic strategies that target co-stimulatory receptors in a manner that allows for the management of these side effects.

Inducible T cell co-stimulator (ICOS) is a co-stimulatory receptor that enhances T-cell responses to foreign antigens [136, 137]. This receptor is predominantly found on T lymphocytes, appearing on various T cell subsets [138, 139]. The initiation of ICOS signaling occurs when it binds to its ligand, ICOSL, which is expressed on antigen-presenting cells such as macrophages, dendritic cells (DCs), and B-cells [140, 141]. Disrupting the interaction between ICOS and ICOSL presents a promising strategy for cancer immunotherapy by inhibiting the activity of immunosuppressive Tregs or preventing Tfh-cell differentiation. ICOS/ICOSL signaling is essential for the generation of Tregs as well as potentiating the immunosuppressive functions of Tregs [142, 143]. In addition, ICOS signaling is responsible for sustaining prolonged T-cell migration at the interface between the T-cell zone and the B-cell follicle [144]. ICOSL expressed on B cells serves as a molecular bridge that facilitates the interaction between T cells and B cells, and plays a pivotal role in ensuring a continuous engagement between Tfh cells and B cells within germinal centers (GCs) [145]. Angioimmunoblastic T-cell lymphoma (AITL) and peripheral T-cell lymphoma with T follicular helper phenotype (PTCL-Tfh) are a group of complex clinicopathological entities that originate from the proliferation of Tfh cells highly expressing ICOS [146, 147].

Targeting ICOS/ICOSL interaction in clinical trials is currently restricted to mAbs, which are associated with several drawbacks, such as high immunogenicity resulting in severe side effects, low tumor penetration, high manufacturing costs, and minimal flexibility in structural modification [148, 149, 150]. Unlike mAbs, peptides and small molecules possess smaller molecular weights, reduced immunogenicity, improved tissue and tumor penetration, and lower manufacturing costs [151, 152]. Thus, peptides hold significant promise as a complement to mAb-based therapy, offering the potential for enhanced synergistic effects [153].

Gabr et. al have pioneered the discovery of first-in-class small molecules targeting immune checkpoints, including ICOS [154, 155]. However, ICOS-targeted hits identified from our screening efforts either failed to inhibit ICOS/ICOSL interaction [150] or possessed weak inhibition potency [156] (greater than 10 μ M). Our goal is to identify peptide-based first-in-class ICOS antagonists that can be further optimized as monotherapy or in combination in certain T-cell lymphomas and solid tumors characterized with overexpression of ICOS and ICOSL. Such ICOS-targeted peptides can potentially be restricted to a short period of treatment to abrogate Treg amplification

or Tfh proliferation with minimal impact on the potential restoration of effector cells expressing ICOS. In this work, we employ cPEPmatch with pivotal modifications to target ICOS/ICOS-L interaction. Notably, the method was refined to focus on key binding hot spots within the PPI interface, enhancing both the specificity and effectiveness of the identified cyclic peptides.

7.2. Materials and Methods

7.2.1. Structure Model

The cPEPmatch method requires a pre-existent protein-protein complex structure as input for analysis. In this context, we utilized the X-ray crystallographic structure of the ICOS/ICOS-L complex, PDB:6x4g [157]. This structure served as the principal template for our subsequent simulations and analyses, and it was also the basis for hot spot identification. These identified hot spots guided the selection of residues for targeted matching via the cPEPmatch approach.

7.2.2. Matching

The methodology applied in this study mirrors the conditions explored in the GAG chapter, with a pivotal update to our computational resources. Specifically, we have expanded our cyclic peptide library to incorporate the 400 cyclic peptide structures analyzed in Chapter 2. These structures have now been integrated into the cPEPmatch method, enhancing our capability to identify and design potent inhibitors that target protein-protein interactions with increased precision and efficiency.

7.2.3. Molecular Dynamics (MD) Simulations

For the initial assessment of ICOS/ICOS-L binding interface hot spots, as well as the evaluation and ranking of protein-cyclic peptide complexes, we employed the Amber19 package [87]. We prepared the structures for Energy Minimization (EM) and Molecular Dynamics (MD) simulations using the tLeap module from Amber19, adhering to established protocols. Protein parameters were sourced from the ff14SB force field [95]. To neutralize the complexes, we added either Na⁺ or Cl⁻ ions and solvated them in an orthorhombic box, maintaining a minimum 10 Å distance to the box boundaries, using explicit TIP3P water molecules [96].

Initially, we performed energy minimization on all simulation systems using the steepest descent method for 2000 steps via the Amber19 Sander module. For subsequent MD simulations, we utilized the pmemd.cuda module. We incrementally heated the systems to 310 K through three stages (100 K, 200 K, 310 K), each lasting 100 ps, while applying positional restraints on all non-hydrogen atoms relative to their starting conformations. Following this, we systematically reduced the positional restraints from an initial value of 25 to 0.5 kcal·mol⁻¹·Å⁻² over the course of five successive 100 ps simulations at a constant pressure of 1 bar and a temperature of 310 K. The equilibrated structures then served as the starting points for the production runs, which were conducted without any restraints. We employed a 2 fs time step and constrained all bonds involving hydrogens to their optimal lengths using the Shake algorithm[54]. Data collection simulations were executed for 300 ns, and coordinates were saved every 30000 steps. All simulations were conducted in triplicates.

7.2.4. Trajectory Analysis

For the evaluation of stable cyclic peptide binding, we examined the molecular dynamics (MD) trajectories through visual inspection. Subsequently, we applied the Molecular Mechanics Generalized Born Surface Area (MM/GBSA) methodology to quantify the mean interaction energy[158, 107]. This analysis was conducted using the well-validated single trajectory approach, as facilitated by the MMPBSA.py module within Amber18. We utilized 200 snapshots extracted from the last 12 ns of the production MD simulation for these calculations. The modified GB model (igb=5) with mbondi2 parameters was employed, along with specific α , β , and γ values of 1.0, 0.8, and 4.85, respectively. Dielectric constants were set to 80 for the solvent and 5 for the solute. The outcome of this analysis yielded the mean interaction energy between the cyclic peptide and its protein partner. Per-residue decomposition allowed for a detailed investigation of the contribution of interacting residues.

In addition to energy calculations, our analytical framework extended to the characterization of hydrogen bonds and Root Mean Square Deviation (RMSD) calculations using the CPPTRAJ module from the Amber suite [159]. We considered a hydrogen bond to be significant if it met specific geometric criteria: a donor-acceptor distance of less than 3.5 Å and an angle exceeding 135°.

Furthermore, we conducted RS-REMD for our best three matches. Simulations were conducted in explicit TIP3P water12 at 300 K, together with neutralizing Na⁺ counterions in water and no additional salt. The prepared system structures were placed within periodic octahedral solvent boxes and a minimum distance between peptide and boundary of 15 Å. The systems were then minimized via Steepest Descent

for 10 k steps, heated up step-wise to a target temperature of 300 K within 60 ps and equilibrated within NVT and NPT conditions for 1 ns each. The RS-REMD simulations were executed with 25 ns per replica and exchange attempts every picosecond. These simulations were carried out using 16 replicas, each exhibiting a progressive increase in the bias applied to the ligand and receptor Lennard-Jones parameters (see Table 3.1). The trajectories were analyzed using pytraj and the associated relative binding free energies were calculated with MBAR as proposed by Siebenmorgen and Zacharias, 2020 [71].

7.3. Results and Discussion

7.3.1. Identification of ICOS/ICOS-L Hot Spots

Following a 1-microsecond MD simulation of the ICOS/ICOS-L complex, we observed stable binding characteristics and visually inspected the contact residues involved in the interaction, as illustrated in **Figure 7.1 B**. Using the MM/GBSA method, we pinpointed residues that play a pivotal role in the ICOS/ICOS-L interaction (**Figure 7.1 C**). This approach allowed us to discern the mean interaction energy contribution for each residue. Residues that demonstrated a ΔG value less than $-1 \text{ kcal}\cdot\text{mol}^{-1}$ were designated as hot spots. These residues, namely TYR132, TYR134, VAL143, TYR146, ILE148, GLN150, ASN151, LEU195, LEU197, GLY202, and PHE203, were identified as crucial components of the motif sequence and subsequently became the input for cPEPmatch.

7.3.2. cPEPmatch Selected Matches

In our quest to identify potential cyclic peptide binders, cPEPmatch was employed with specific parameters that included a motif size of 5 residues, interface cutoff of 10 Å, Fit-RMSD threshold of 2 Å, and a non-consecutive motif. Out of the 114 matches identified, 21 were handpicked based on their steric compatibility. These candidates were subjected to our analysis protocol that included triplicate MD simulations spanning 300 ns and MM/GBSA per-residue energy decomposition. Out of this cohort, three cyclic peptides stood out as potential stable binders (**Figure 7.2**).

In evaluating the efficacy of the matched cyclic peptides, we employed multi-faceted criteria for what constitutes an effective binder. This included not only stability during molecular dynamics (MD) simulations but also a negative total free binding energy, as gauged by MM/GBSA calculations. Additionally, we considered the free binding

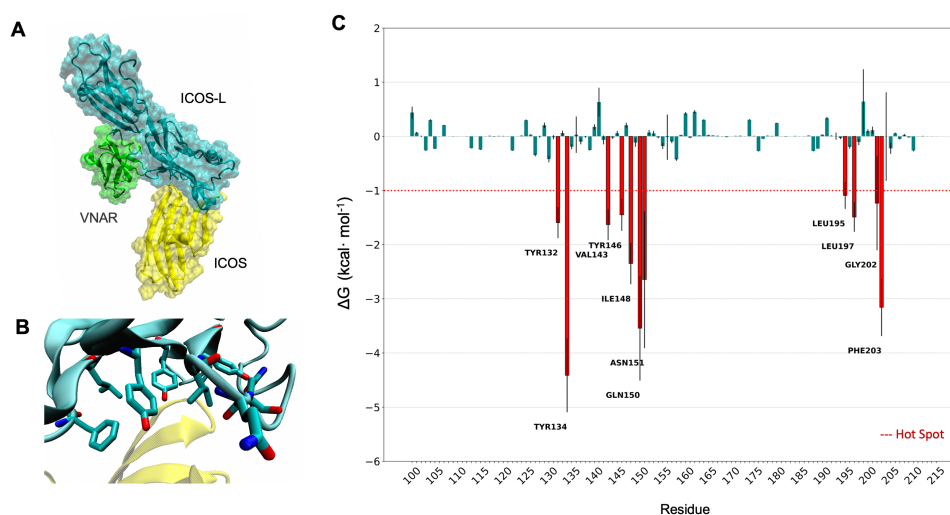


Figure 7.1.: (A) X-ray crystallographic structure depicting the complex of ICOS (yellow) with ICOS-L (cyan) and an anti-ICOS-L VNAR domain (green), derived from PDB entry 6x4G. (B) The binding interface between ICOS and ICOS-L is highlighted, with interacting residues represented in stick form. (C) Per-residue contributions to the binding free energy between ICOS and ICOS-L, calculated via MM/GBSA decomposition. Residues identified as hot spots are accentuated in red and labeled.

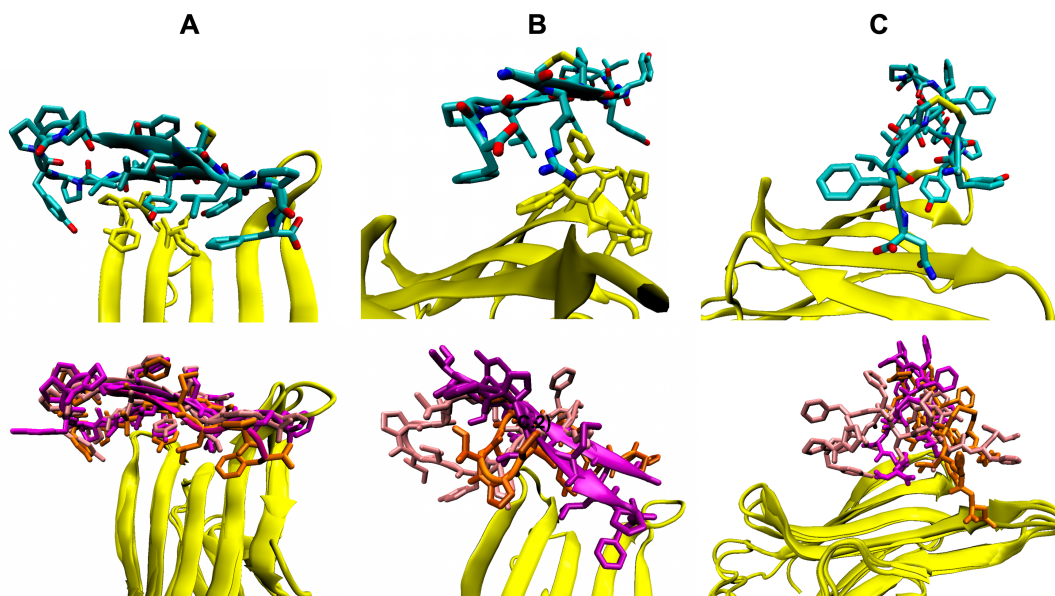


Figure 7.2.: Binding profiles and simulation outcomes for (A) Match 42, (B) Match 8, and (C) Match 41. (Top) Representative binding pose after a 300-ns unrestrained molecular dynamics simulation. (Bottom) Superimposed binding conformations from three independent replicas after 300 ns of unconstrained molecular dynamics simulations.

energy decomposition, specifically looking for a profile that closely mimics the energy hot spots identified in the ICOS/ICOS-L binding interface. The presence of hydrogen bonds similar to those in the native ICOS/ICOS-L interaction also served as a marker for effective binding. A comprehensive analysis of our top three peptide matches, encompassing their sequences, structural characteristics, and performance metrics, is provided to substantiate our evaluations. It should be noted that in the instances that follow, the residue name was altered from CYS to CYX to indicate its involvement in a disulfide bond.

The first analyzed match was labeled as Match 42 (PDB:2NB6). Comprising a 16-amino-acid sequence —TYR-CYX-TYR-PRO-VAL-PRO-TYR-PRO-PRO-PHE-PHE-LEU-CYX-LEU-PRO-PHE— this cyclic peptide is characterized by a disulfide-bond cyclization. Over the course of 300 ns simulations, each trajectory exhibited remarkable binding stability, as depicted in Figure 7.2, with binding also demonstrated with MBAR from rs-remd with a total value of -4.13 ± 0.67 kcal/mol. A. The peptide displayed structural rigidity, with an RMSD of $0.94 \text{ \AA}$. Most of the mutated residues contributed to the binding, as indicated in **Figure 7.3 A**, and three of the four most frequent hydrogen bonds involved our mutated residues (**Table 7.1**). Interestingly, a proline, PRO4, residue from the original 2NB6 structure made significant contribu-

tions to binding free energy and formed the second most frequent hydrogen bond. However, the peptide's elevated hydrophobicity emerged as a point of concern.

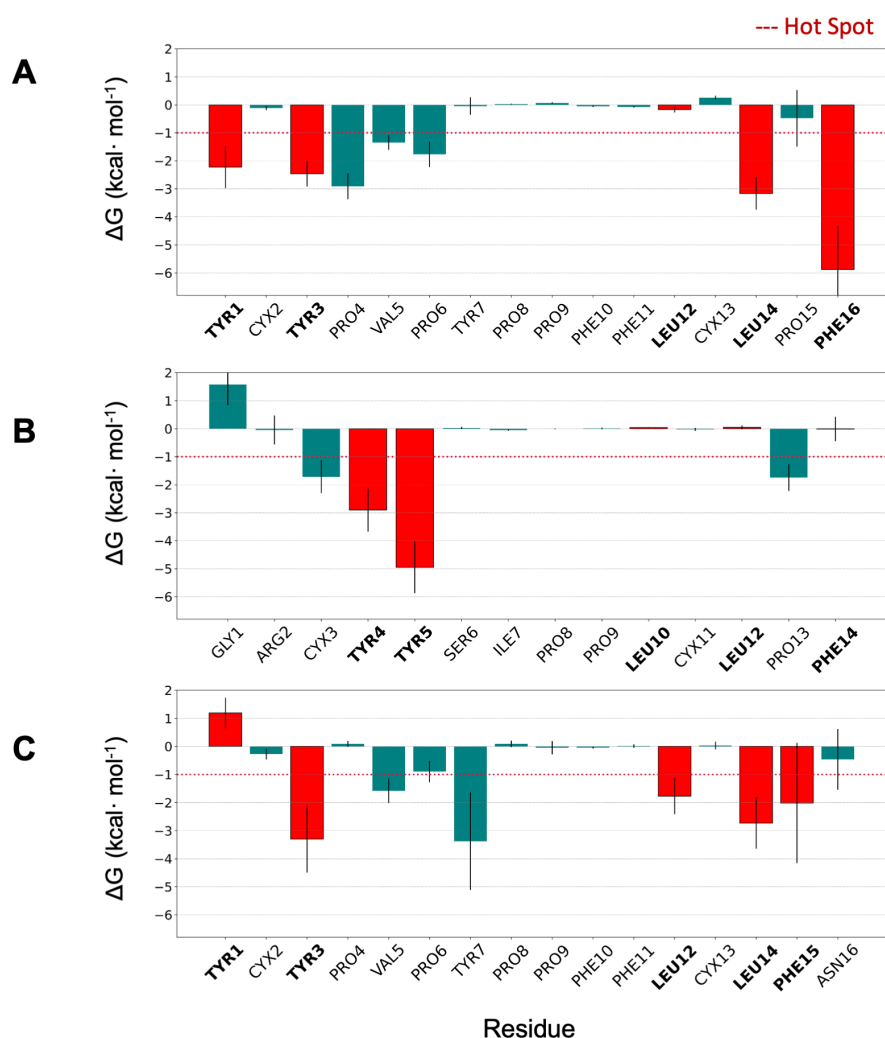


Figure 7.3.: Per-residue contributions to the binding free energy between ICOS and the respective cyclic peptides, as calculated through MM/GBSA decomposition for (A) Match 42, (B) Match 8, and (C) Match 41.

A second match of interest was Match 8, derived from PDB:1JBN, a 14-amino-acid cyclic peptide with the sequence —GLY-ARG-CYX-TYR-TYR-SER-ILE-PRO-PRO-LEU-CYX-LEU-PRO-PHE—, also exhibiting a disulfide-bond cyclization. Although the peptide remained in close proximity to the binding site throughout the simulations, **Figure 7.2 B**, displayed a higher RMSD value of 3.687 Å, suggesting reduced structural rigidity. This was consistent with its moderate binding affinity as determined by MBAR from rs-remd, which was -2.7 ± 0.38 kcal/mol. Among the residues mutated

Table 7.1.: Prevalence of hydrogen bonds formed during molecular dynamics simulations for Matches 42, 8, and 41. The bonds are categorized as either acceptors or donors, with residues that were specifically mutated to mimic ICOS-L are bold.

Match	Acceptor	Donor	Fraction
42 (2NB6)	PHE 101	TYR 1	0.67
	PRO 4	PHE 106	0.34
	PHE 16	SER 63	0.23
	PRO 15	THR 55	0.11
8 (1JBN)	PHE 104	TYR 5	0.59
	CYX 3	PHE 104	0.39
	TYR 4	VAL 106	0.09
41 (2NB6)	ASN 16	THR55	0.48
	ASN 16	SER 63	0.23
	GLN 37	TYR 3	0.18
	GLN 41	TYR 7	0.17

to mimic the interaction, only TYR4 and TYR5 made significant contributions to binding free energy, as seen in **Figure 7.3 B**. These residues also participated in the most frequent hydrogen bonding events, akin to the original interactions observed in the native ICOS/ICOS-L interface (Table 7.1)(Rujas et al., 2020). Additionally, two residues from the original template, CYX3 and PRO13, were strategically positioned to contribute to the free energy of binding, with CYX3 also engaging in a hydrogen bond for a considerable duration of the simulation.

Lastly, Match 41 (PDB:2NB6) presented an interesting case. This peptide, TYR-CYX-TYR-PRO-VAL-PRO-TYR-PRO-PRO-PHE-PHE-LEU-CYX-LEU-PHE-ASN, is a variant of the template peptide from Match 42, differing in just one matched residue. Despite this similarity, the peptide exhibited distinct free-binding energy decomposition and hydrogen-bonding patterns, as shown in **Figure 7.3 C** and **Table 7.1**. The simulation consistently displayed stable binding characteristics despite exploring a variety of binding orientations, as illustrated in **Figure 7.2, C-bottom**. Nonetheless, the structure remained relatively rigid, RMSD value of 1.87 Å, and a moderate binding affinity was observed, with MBAR from rs-remd calculating it to be -2.6 ± 0.14 kcal/mol. Four of the matched residues—TYR3, LEU12, LEU14, and PHE15—contributed favorably to binding free energy. Of these, only TYR3 participated in hydrogen bonding, although the overall structure formed a substantial number of hydrogen bonds. Intriguingly, the mutated residue TYR1 had an adverse effect on binding. Additionally, the peptide’s exploration of different binding sites allowed TYR7 from the original template to make a significant contribution to binding—a phenomenon not observed in Match 42.

This example underscores the potential for further optimizations to enhance binding affinity.

Additionally, two predictions, Match 14 1SFL and Match 45 2NDL, were identified as potential candidates; however, they have not yet been experimentally tested due to synthesis limitations. Despite this, computational analyses suggested that these matches demonstrated an energy per residue similar to those of the native mimicking interactions they were designed to replicate, as detailed in Figure Supplementary C.1. Moreover, during molecular dynamics (MD) simulations, these peptides were observed to be stable binders, with Match 14 1SFL maintaining stability in two out of three MD simulation replicates and Match 45 2NDL exhibiting stability across all three replicates. The calculated ligand RMSD values were at 1.86 Å and 2.13 Å, respectively. In contrast, other studied matches were unstable; a recurrent theme was observed where an arginine positioned on the opposite side of the interacting site led to an unfavorable 'pull' that extracted it from the binding pocket, or caused a deformation in the native structure of the cyclic peptide template, as depicted in Figure Supplementary C.2 A and B, respectively.

7.4. Experimental Validation

In the experimental validation conducted by Abdel-Rahman et al., [155] the three predicted cyclic peptides, Match 41, Match 42, and Match 8, were evaluated for their ability to inhibit the interaction between ICOS and ICOSL. Using a Time-Resolved Förster's Resonance Energy Transfer (TR-FRET) assay (**Figure 7.4 A**), it was observed that both Match 41 and Match 42 exhibited dose-dependent reductions in the TR-FRET signal, with IC_{50} values of $3.68 \pm 0.24 \mu M$ and $4.48 \pm 0.51 \mu M$, respectively. In contrast, Match 8 showed negligible effects on the ICOS TR-FRET signal. The competitive ELISA results (**Figure 7.4 B**) aligned with the TR-FRET assay, where Match 41 demonstrated an IC_{50} value of $1.87 \pm 0.15 \mu M$ as an ICOS/ICOSL inhibitor, while Match 42 failed to block the inhibition. Further analysis revealed that Match 42 may have resulted in false positives due to potential aggregation, as confirmed by dynamic light scattering (DLS). Match 41, on the other hand, displayed no detectable aggregation. Moreover, Match 41 exhibited selectivity for the ICOS/ICOSL interaction, as it had negligible effects on multiple other immune checkpoint interactions. The in vitro metabolic stability and physicochemical properties of Match 41 further supported its potential as a novel ICOS/ICOSL inhibitor with excellent selectivity for the ICOS/ICOSL interaction, metabolic stability, balanced lipophilicity, and high permeability (Supplementary Table C.1 and Figure C.3). These findings highlight the promising nature of Match 41 for further optimization and evaluation as a cancer

therapeutic.

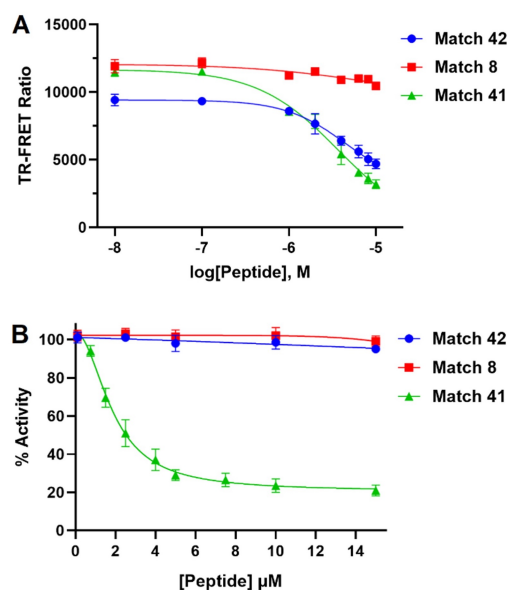


Figure 7.4.: (A) Dose-response curves of the three cyclic peptides in the ICOS/ICOSL TR-FRET assay. (B) Dose-response curves of the inhibitory profile of the peptides in ICOS/ICOSL ELISA. Error bars represent standard deviation (n = 3).

7.4.1. Conclusions

In this study, we introduced cyclic peptides as a novel class of efficient ICOS/ICOSL inhibitors. Our implementation of the established cPEPmatch on key binding hot spots within the ICOS/ICOSL interface resulted in the identification of Match 41 as the first example of cyclic peptides that function as ICOS/ICOSL inhibitors. The ICOS/ICOSL inhibitory profile of Match 41 was validated using orthogonal platforms and assessed key in vitro pharmacokinetic properties. We anticipate that future modifications of Match 41 will enable further evaluation of this novel class of immunomodulators using both in vitro and in vivo platforms.

8. cPEPmatch Code and Webserver: Application to Medically Relevant Systems

As we end our exploration in the field of cyclic peptide design, in this chapter, we focus on the comprehensive presentation of the cPEPmatch tool. Developed as a Python script, cPEPmatch has been crafted to automate the process of identifying and mutating cyclic peptide templates, aligning them with targeted protein interfaces. A critical aspect of our research involves using Molecular Dynamics simulations and free energy calculations, including Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and Hamiltonian Replica Exchange Molecular Dynamics (HREMD), to rigorously evaluate the binding efficacies of peptides. Furthermore, the transition of cPEPmatch from a Fortran-based program to a Python script has significantly enhanced its usability and maintainability. This development not only makes the tool more user-friendly but also prepares it for future advancements in computational drug design.

In this chapter, we will explore the technical aspects of cPEPmatch, detailing its operational framework and discussing the launch of its webserver platform. Additionally, we will demonstrate its application through a series of case studies involving medically relevant protein systems. The goal is to provide a comprehensive overview of cPEPmatch's capabilities and limitations and underscore its potential as a valuable tool in drug design and the study of protein interactions.

8.1. Development and Features of cPEPmatch

Here, we delve into the technical details of the cPEPmatch tool, exploring its various components, capabilities, and limitations.

8.1.1. Cyclic Peptide Database Creation

At the time of writing this thesis, the foundational database of *cPEPmatch* comprises 432 unique cyclic peptide structures sourced from the Protein Data Bank. This database is continually updated with new cyclic peptides as they are published in the Data Bank, thanks to a separate database update module. This module assesses new PDB files, determining their eligibility as cyclic peptides based on specific criteria such as the presence of head-to-tail bonds, disulfide bridges, or other cyclization types.

For each cyclic peptide in the database, comprehensive information is provided, including its cyclization characteristics and original PDB source. This level of detail is readily accessible to the user through our source code. When employing the *cPEPmatch* tool, users have the flexibility to specify the types of cyclic peptides they wish to incorporate into their studies, such as restricting the selection to only those peptides that are head-to-tail cyclized.

The backbone structure of these cyclic peptides is characterized using CA carbon distance matrices, tailored according to user-defined parameters like motif size (ranging from 4 to 7 amino acids) and the type of cyclization. This personalized approach allows for a more targeted and effective matching process, catering to the specific needs and objectives of each research endeavor.

8.1.2. Characterization of Protein Interfaces

cPEPmatch characterizes protein-protein or protein-ligand interfaces also by using a CA distance matrix. Users can define the interface cutoff () and include optional parameters for protein-specific residues, ensuring the relevancy and specificity of cyclic peptide matches.

cPEPmatch also characterizes protein-protein and protein-ligand interfaces by utilizing the same CA distance matrix method as applied to cyclic peptides. This approach allows users to define the interface with precision and specificity. The interface cutoff distance, measured in Å, determines which amino acids are considered part of the interface, thus influencing the matching process with cyclic peptides.

Users have two primary options for defining the interface:

1. **Proximity-Based Interface:** Users can define the interface based on proximity. By selecting a distance cutoff, *cPEPmatch* identifies amino acids in the vicinity of

the receptor as part of the interface. This option is suitable for general interface identification where specific binding hotspots are not known or not the primary focus.

2. **Hot Spot-Based Interface:** Alternatively, users can define the interface by specifying key amino acids known as hotspots, as discussed in Chapter 5. Selecting this option allows cPEPmatch to focus on these specific residues, enhancing the relevance and specificity of the cyclic peptide matches and reducing the output to more tailored matches

This dual approach in interface characterization ensures that cPEPmatch can be adapted to a wide range of research needs, from general exploratory studies to targeted investigations focusing on key interaction sites.

8.1.3. Backbone Match Motif

cPEPmatch executes a crucial step by matching the CA distance motifs between the characterized protein interfaces and the cyclic peptide database. This matching is based on a fit-RMSD (Root Mean Square Deviation) threshold, defined by the user, to ensure structural similarity within an acceptable range. This threshold determines the precision of matching, thus balancing accuracy with computational efficiency.

One limitation of cPEPmatch emerges when matching non-consecutive hot spots. In such cases, the tool must assess numerous combinations of non-consecutive residue distance matrices. This process can substantially slow down both the initial matching steps and the subsequent screening through a large number of potential matches. This is particularly pertinent when dealing with complex interfaces where too many hot spots are dispersed.

To mitigate this limitation, the user can focus on the matching search to hot spots that are in close proximity to each other. Users can group hot spots into clusters of 5-7 amino acids. This strategy simplifies the matching process by reducing the complexity of the distance matrices and the number of combinations to consider. This approach is especially beneficial in cases where a series of adjacent residues forms a critical part of the binding interface, allowing for a more efficient and targeted matching process.

This dual approach in hot spot matching — either targeting non-consecutive hot spots for comprehensive coverage or focusing on proximal clusters for efficiency — demonstrates cPEPmatch's adaptability to different research scenarios and its capability to handle complex protein interfaces.

8.1.4. Superimpose and Mutate

Upon identifying suitable matches, *cPEPmatch* superimposes these onto the target protein structure and mutates the side chains of the cyclic peptide to mimic the protein, employing Modeller software for this purpose. This process not only ensures functional mimicry of the target protein interface but also places these side chains accurately using homology modeling. Following this, a minimization step is carried out to ensure optimal positioning of the side chains, providing a more reliable starting structure for subsequent evaluations. This refinement is crucial for enhancing the accuracy and effectiveness of the matched cyclic peptides in potential binding and interaction studies.

A notable limitation of the *cPEPmatch* process arises when dealing with non-standard amino acids, including methylated variants. Modeller software, in its current capacity, does not process these non-standard amino acids during the mutation step. As a result, when such matches are encountered, *cPEPmatch* outputs the matched structure without mutating it to resemble the target protein. In such scenarios, users are required to manually perform the mutations. Tools like *pdb4amber*, part of the AmberTools suite, can be utilized for this purpose. The process typically involves mutating the residues using *pdb4amber*, followed by adding missing atoms and further processing using *tleap*. This manual intervention allows researchers to overcome the limitation posed by Modeller and ensure that the cyclic peptides accurately mimic the target protein, even when non-standard amino acids are involved.

An additional step in this *cPEPmatch* process involves the analysis of overlap between the matched cyclic peptide and the receptor. This analysis is crucial as it identifies and eliminates matches that result in significant overlap with the receptor, which would be counterproductive for potential binding studies. Matches that exhibit such overlaps are not presented to the user, ensuring that only feasible and sensible matches are considered for further analysis and simulations. This step is integral to maintaining the practicality and relevance of the cyclic peptide matches proposed by *cPEPmatch*.

8.1.5. Result Analysis and Output

The output from *cPEPmatch* includes a detailed match list and all mutated or aligned structures. The output table contains comprehensive information, including match number, PDB name, distance RMSD, matched cyclic peptide residues, and fit-RMSD values.

Post-generation of matches, a crucial step involves the visual analysis of these matches. Tools like PyMOL or VMD are recommended for this purpose. This visual inspection is pivotal in evaluating the steric fit of the matches. Although there is potential for further development in automating this selection process, the current necessity for visual inspection provides an opportunity for users to meticulously analyze and choose the best-fit cyclic peptides for subsequent MD and free energy analysis phases.

A notable aspect of using *cPEPmatch* is the flexibility for parameter adjustment. Users might find that the initial recommended parameters may yield too many or too few matches for their specific system. In such cases, adjusting parameters like the fit-RMSD threshold is possible. Reducing the threshold can help in narrowing down the number of matches, whereas increasing it may broaden the search results. This adjustable nature of the parameters allows users to fine-tune the tool to their specific requirements, ensuring optimal matching for their system.

The match list and the output structures—superimposed, mutated, and minimized—provided by *cPEPmatch* are instrumental in the easy inspection of potential lead binders. This feature of the tool simplifies the initial phase of selecting promising cyclic peptides, setting the stage for their further evaluation through molecular dynamics simulations and free energy calculations. The user-friendly nature of the output format contributes significantly to the efficiency of the research process, aiding users in their pursuit of identifying viable cyclic peptide modulators.

8.1.6. Installation and Setup

cPEPmatch requires a dedicated Conda environment, installation of necessary Python packages, and additional Conda modules like *vmd-python*, *biopython*, and *Modeller*. Detailed installation instructions are provided for user convenience.

8.2. *cPEPmatch* Webserver and Source Code

8.2.1. Source Code Structure

The source code for *cPEPmatch* is publicly available on GitHub, offering a comprehensive platform for users. As of the time of writing this thesis, the repository can be accessed via the following link: <https://github.com/briandasantini/cPEPmatch>.

The repository is organized to facilitate ease of use and understanding. It includes a README file that provides a general description of the tool, along with guidelines for installation and usage. The structure of the repository is as follows:

```
cPEPmatch
|-- cPEPmatch.py
|-- cpep_modules
|   |-- backbone_match.py
|   |-- cpep_database.py
|   |-- protein_target_characterization.py
|   |-- superimpose_mutate.py
|   '-- __init__.py
|-- database
|   |-- cyclo_peg.csv
|   '-- *.pdb
|-- lib
|   |-- old_databases
|   |-- test_system
|   '-- update_database
|-- README.md
'-- requirements.txt
```

Key components of the repository include:

- **cPEPmatch.py:** The main Python script for the cPEPmatch tool.
- **cpep_modules:** This directory contains individual modules for different aspects of the tool, such as backbone matching, database handling, protein target characterization, and the superimposition and mutation processes.
- **database:** This directory hosts the cyclic peptide database, including a CSV file detailing each peptide's characteristics and the cleaned PDB files used for matching.
- **lib:** Contains supplementary resources such as an example system for testing the tool, and scripts for updating or generating the cyclic peptide database.

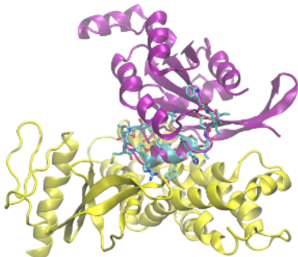
It's important to note that as the development of cPEPmatch continues, the structure and contents of the repository may evolve to incorporate new features or improvements.

8.2.2. Webserver for cPEPmatch

The webserver for cPEPmatch (Figure 8.1), hosted at the Technische Universität München (TUM) web services, presents an accessible and user-friendly interface for researchers and scientists to utilize the cPEPmatch tool. Below is an overview of the key components of the website:

Zacharias Webservices Home Submission User Guide Contact

cPEPmatch



Cyclic Peptide Matching Algorithm

cPEPmatch is a computational tool created to assist in the rational design of cyclic peptide mimetics capable of potentially modulating interactions within protein-protein or protein-ligand complexes.

You can find the details of the method [here](#).

Impatient and want to get started right away?

[Toggle our Quick Guide](#)

Or, for a more comprehensive guide with a practical example, check out our full [tutorial](#).

Access the Source Code and Database: [cPEPmatch GitHub Repository](#).

Please cite these references if you use the server in your research:
[Rapid in silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces](#) , Brianda L. Santini and Martin Zacharias .

Submit a Job

Job Name*

PDB File*
 No file chosen

Protein Chain*

Receptor Chain*

Cutoff

Threshold

Motif

☐ I want to target specific hot-spots.

Email*

Figure 8.1.: Screenshot of the cPEPmatch Webserver Homepage. This image captures the user interface of the cPEPmatch webserver as it appears at the time of writing, showcasing the layout, navigation options, and the initial information presented to the user for starting their cyclic peptide matching process.

- **URL:** The cPEPmatch webserver is currently accessible at <https://t38webservices.nat.tum.de/cpepmatch/>. It's important to note that the specific URL may change in the future due to updates in the university's server infrastructure.
- **Description and Interface:** The website features a concise description of the cPEPmatch tool. This provides visitors with a quick understanding of what the tool does and how it can be beneficial for their research.

- **Job Submission Section:** The core functionality of the webserver lies in its job submission feature. Users can submit their protein structures and specify parameters for the cPEPmatch analysis directly through the website. This section is designed to be intuitive, ensuring ease of use even for those new to the tool.
- **Quick Guide and Full-Length Tutorial:** To assist users in effectively utilizing the webserver, a quick guide is provided for rapid reference. Additionally, a comprehensive tutorial is available at <https://t38webservices.nat.tum.de/tutorial/>. This tutorial includes a detailed example to demonstrate the process of using cPEPmatch.
- **Technology and Queuing System:** The webserver is developed using Django, a high-level Python web framework, ensuring a robust and scalable platform. It also features a queuing system, allowing for multiple job submissions simultaneously.

This webserver represents a significant step in making the cPEPmatch tool widely accessible to the scientific community, simplifying the process of designing cyclic peptides for protein-protein and protein-ligand interactions.

8.3. cPEPmatch Webserver Tutorial

This tutorial demonstrates the use of the cPEPmatch tool through its webserver.

8.3.1. Step 1: Identify Your Target Complex

Download PDB Structure: Start by identifying the protein-protein or protein-ligand complex you want to target. Download the crystal structure of your chosen complex from the Protein Data Bank (PDB).

S100P Example: For this tutorial, we'll use the S100P - V Domain of the receptor for advanced glycation end products (RAGE) complex with PDB ID: 2MJW. S100P is a calcium-binding protein that interacts with the RAGE receptor, and disrupting this interaction has therapeutic potential for various diseases.

8.3.2. Step 2: Inspect and Prepare the PDB File

Inspect the_ Take a close look at the downloaded PDB file. If necessary, perform any cleanup, such as removing unnecessary chains, renumbering residues, and ensuring the correct PDB format. Ensure there are no missing residues in the binding site.

Tip: If there are missing residues outside the binding site, you can use tools like Modeller to fill them in for later simulations.

8.3.3. Step 3: Select cPEPmatch Parameters

- **Cutoff:** This parameter defines the distance between the receptor and the interface. Residues within this cutoff distance will be considered part of the interface. (Default: 6 Å)
- **Threshold:** Set the fit-RMSD maximum for a match to fit into the targeted protein interface residues. It determines how well the cyclic peptide backbone should fit those of the matched protein residues. (Default: 0.7 Å- Recommended: 0.3-1.5)
- **Motif Size:** This is the length of the motif residues to target. (Default: 5 amino acids, Recommended: Adjust based on the nature of your binding site—lower for pocket-like or larger for flat interfaces. Lower limit is 4, upper limit is 7)
- **Optional Residue Numbers:** Residue Numbers: Specify specific residues if you know the binding hot spots. Hot spots are residues contributing the most to binding free energy. Selecting this option allows non-consecutive motif searches. You can use experimentally known hot spots or calculate them computationally. When hot spot residues are selected, the cutoff is ignored.

Note: * Hot spot searches may take more time. * A higher threshold is recommended to accommodate non-consecutive residues - but no higher than 2. * Ensure your hot spot count is at least as large as the motif size.

S100P Example: In our case, we identified hot spots for chain A binding to chain B as residues '28 32 34 37 41 43 49 78 90 92'. We set threshold=2 and motif=5. Interface and cPEmatch parameter input selection are depicted in Figure 8.2.

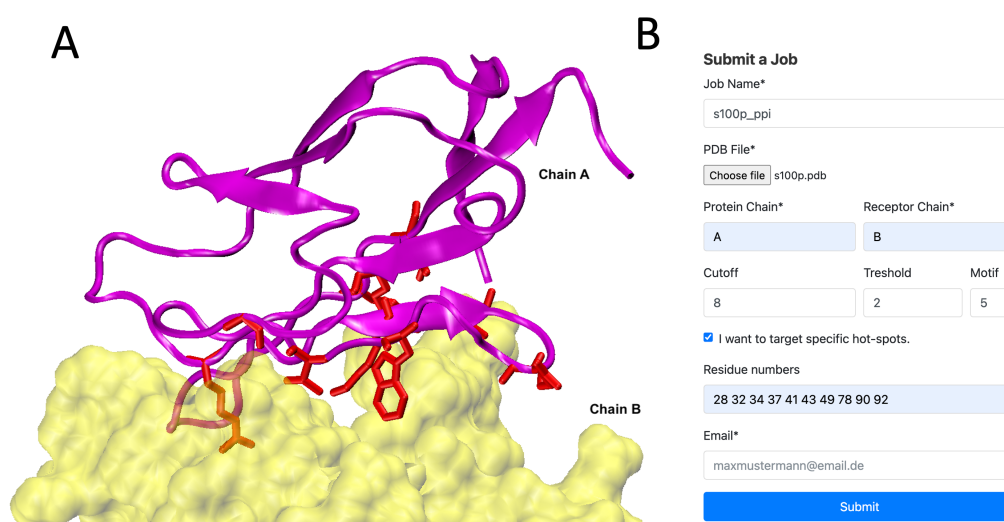


Figure 8.2.: Visual Interface and Parameter Selection on cPEPmatch Webserver. Panel A illustrates the binding interface of the S100P protein complex with Chain A in magenta and Chain B in yellow, highlighting the interaction hotspots in red. Panel B displays the cPEPmatch job submission form, where parameters such as cutoff, threshold, and motif are specified for the analysis, along with targeted hotspots for Chain A binding to Chain B.

8.3.4. Step 4: Run cPEPmatch

Submit your job via the webserver. You will receive an email with a download link for the results. The zip file contains the following files:

match_list.txt file: List of all matches, their fit values, and which cyclic peptide residues matched with the protein.

match#_PDB.pdb files: All your matches' cyclic peptides, which have been superimposed, mutated, and refined using Modeller and MD to fit the matched residues. Note: If the peptide contains non-standard residues, mutations will not be executed, and the output file will have "(_non-mutated)" in the name. Tip: You can perform the mutations using Amber Tools or Modeller.

interface.pdb file: This is a PDB file with the residues that were selected as the interface.

S100P Example: cPEPmatch identified 16 matches using the provided example parameters. The output file, named (match_list.txt), is depicted in Figure 8.3.

List of matches found by cPEPmatch.
Motif Size: 5
Interface Cutoff: 8
Fit-RMSD Threshold: 2.0

Match	PDB	Dist-RMSD	cPep Residues					PP-Interface Residues					Fit-RMSD
1	1kal	1.7036	4	6	13	17	19	32	41	43	78	92	0.8242
2	1znu	1.4826	18	20	23	25	27	32	41	78	90	92	0.8980
3	2lwt	1.8484	1	6	11	14	16	28	32	37	41	43	1.9148
4	2lwt	1.9618	1	6	10	14	16	28	32	37	41	43	2.5561
5	2lvw	1.9023	3	5	8	15	17	32	43	49	78	92	1.5827
6	2lye	1.8721	1	5	9	13	15	28	32	37	41	43	1.3881
7	2lyf	1.9537	1	5	9	13	15	28	32	37	41	43	2.0695
8	2nb5	1.8391	3	5	8	15	17	32	43	49	78	92	1.2936
9	3pp4	1.8039	5	8	14	20	24	34	37	41	78	90	1.2305
10	4ttm	1.5786	8	15	17	21	23	34	41	43	78	92	2.5484
11	4ttn	1.5989	8	15	17	21	23	34	41	43	78	92	2.5233
12	4tto	1.5585	8	15	17	21	23	34	41	43	78	92	2.4890
13	7k7x	1.6365	4	6	9	27	29	32	34	37	41	43	1.1528
14	7lhc	1.8767	12	15	17	21	23	34	41	43	78	92	0.8075
15	7k7x	1.6365	4	6	9	27	29	32	34	37	41	43	1.1527
16	7lhc	1.8767	12	15	17	21	23	34	41	43	78	92	0.8075

Figure 8.3.: Output of cPEPmatch showing a list of 16 matches identified for the S100P example. The table presents match numbers alongside corresponding PDB codes, Distance-RMSD values, matched cyclic peptide residues, protein-protein interface residues, and Fit-RMSD values, demonstrating the precision of the cyclic peptides' alignment with the target protein interface.

8.3.5. Step 5: Post cPEPmatch Evaluation

Visual Inspection: Examine the matches for the best steric fit.

MD Simulations and HREMD: Evaluate the matches using MD simulations or HREMD to calculate binding free energy.

S100P Example: The top two scoring matches, confirmed through MD and HREMD simulations, are illustrated in Figure 8.4. Further details of this analysis, including stability during MD simulations and free energy values, will be discussed in the subsequent section.

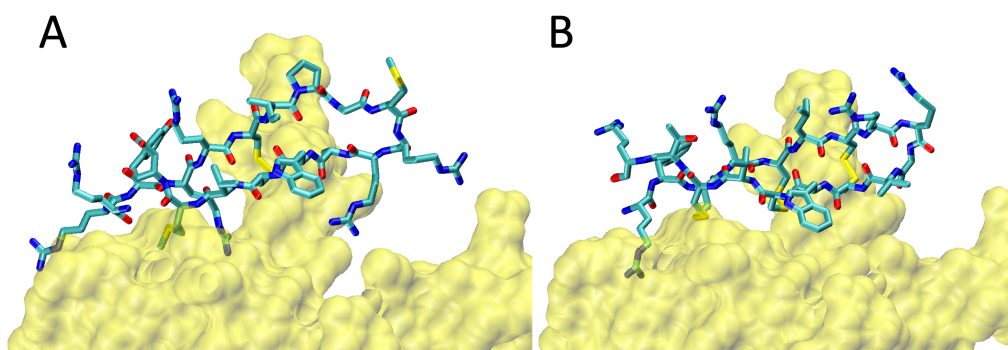


Figure 8.4.: (A) Match 3 (PDB:2lwv) and (B) Match 7 (PDB:2lyf) with the receptor in yellow surface representation and the cyclic peptide in atom-named colored sticks.

This tutorial aims to guide users through the cPEPmatch process, from identifying target complexes to evaluating potential lead structures for further optimization.

8.4. Application of cPEPmatch to Medically Relevant Case Studies

In our quest to develop novel therapeutic predictions, we are employing cPEPmatch to identify cyclic peptides that can mimic critical protein-protein interactions. These

interactions are medically significant due to their roles in various diseases such as viral infections, cancer, autoimmune disorders, and more. The following PPI complexes were tested:

ACE2 & SARS-CoV-2 Spike (PDB:6M0J): The interaction between the SARS-CoV-2 spike protein and the ACE2 receptor is a key mechanism for viral entry into host cells, making it a critical target for therapeutic interventions against COVID-19. The spike protein's engagement with ACE2 triggers a series of events that facilitate the fusion of the viral and cellular membranes, essential for the virus's entry and infection. Understanding this interaction is crucial for developing effective vaccines and neutralizing antibodies, which aim to block this critical step of the viral lifecycle [160].

BCR & Abl (PDB:1K1F): The interaction between the Breakpoint Cluster Region (BCR) and the Abelson (Abl) kinase is significant in chronic myelogenous leukemia (CML). The BCR-Abl complex results from a chromosomal translocation, creating a constitutively active tyrosine kinase that drives the proliferation of leukemic cells. Targeting this interaction has been pivotal in treating CML, with inhibitors like Imatinib revolutionizing patient outcomes [161].

ER & Coactivator (PDB:5DZ1): The interaction between the Estrogen Receptor (ER) and its coactivators plays a crucial role in regulating gene expression, particularly in the context of breast cancer. ER, a nuclear hormone receptor, upon binding estrogen, recruits coactivators to modulate transcription and influence cell proliferation and survival. Disruption of this interaction has therapeutic potential in hormone-dependent breast cancers [162].

GP120 & CD4 (PDB:1GC1): The GP120 protein of the Human Immunodeficiency Virus (HIV) interacts with the CD4 receptor on host T-cells, a critical step for viral entry and infection. GP120's binding to CD4 is the initial event in the virus's entry process, leading to subsequent steps that allow the virus to fuse with and enter host cells. This interaction is a major target for antiretroviral therapies aimed at preventing HIV infection [163].

IL12 & IL12Rb1 Subunit (PDB:6WDQ): Interleukin-12 (IL12) binding to its receptor subunit IL12Rb1 is essential in the regulation of immune responses. IL12 is a cytokine that plays a critical role in the differentiation and activation of T-cells and natural killer cells. Targeting this interaction is explored in therapies for autoimmune diseases and cancer, where modulation of the immune response is crucial [164].

MDM2 & p53 (PDB:4HFZ): The interaction between MDM2 and the tumor suppressor protein p53 is a focal point in cancer biology. MDM2 negatively regulates p53,

a critical protein for cell cycle control and apoptosis. Inhibition of the MDM2-p53 interaction can reactivate p53's tumor-suppressing functions, offering a strategy for cancer therapy [165].

PD-1 & PD-L1 (PDB:3BIK): The interaction between Programmed Death-1 (PD-1) on T cells and its ligand PDL1 is a key immune checkpoint in cancer immunotherapy. PDL1 expression on tumor cells inhibits T cell activity via PD-1, allowing cancer cells to evade the immune response. Inhibitors targeting PD1 or PD-L1 restore T-cell function and have shown remarkable efficacy in treating various cancers [166].

S100P & RAGE (PDB:2MJW): S100P's interaction with the Receptor for Advanced Glycation Endproducts (RAGE) is significant in cancer biology. S100P, a member of the S100 family of proteins, binds to RAGE, influencing cellular processes linked to tumor growth and metastasis. Targeting this interaction is investigated for its potential to hinder cancer progression and metastasis [167].

8.4.1. Materials and Methods

The cPEPmatch method requires the use of established protein-protein complex structures for its analyses. For this study, we utilized structures corresponding to the pdb codes listed in Table 8.1. These structures are critical, serving both as templates for simulation and analysis and for hot spot identification, guiding residue selection within the cPEPmatch framework. For instances of missing residues, Modeller [168] was used for addition, except in the BCR system, which utilized an AlphaFold2 [169] prediction. Additionally, Table 8.1 encompasses the chosen input or inputs, in cases where multiple matchings were performed, for the matching process.

For the initial assessment of our case studies binding interface hot spots, as well as the evaluation and ranking of protein-cyclic peptide complexes, we employed the Amber19 package[87]. We prepared the structures for Energy Minimization (EM) and Molecular Dynamics (MD) simulations using the tLeap module from Amber19, adhering to established protocols. Protein parameters were sourced from the ff14SB force field[95]. To neutralize the complexes, we added either Na⁺ or Cl⁻ ions and solvated them in an orthorhombic box, maintaining a minimum 15 Å distance to the box boundaries, using explicit TIP3P water molecules[96].

For all simulations, we performed energy minimization using the steepest descent method for 2000 steps via the Amber19 Sander module. For subsequent MD simulations, we utilized the pmemd.cuda module. We incrementally heated the systems to 310 K through three stages (100 K, 200 K, 310 K), each lasting 100 ps, while

applying positional restraints on all non-hydrogen atoms relative to their starting conformations. Following this, we systematically reduced the positional restraints from an initial value of 25 to 0.5 kcal·mol⁻¹·Å⁻² over the course of five successive 100 ps simulations at a constant pressure of 1 bar and a temperature of 310 K. The equilibrated structures then served as the starting points for the production runs, which were conducted without any restraints. We employed a 2 fs time step and constrained all bonds involving hydrogens to their optimal lengths using the Shake algorithm[54]. Data collection simulations were executed for 25 ns, and coordinates were saved every 30000 steps.

For the evaluation of hot spot residues and stable cyclic peptide binding, we examined the molecular dynamics (MD) trajectories through visual inspection. Subsequently, we applied the Molecular Mechanics Generalized Born Surface Area (MM/G-BSA) methodology to quantify the mean interaction energy[158, 107]. This analysis was conducted using the well-validated single trajectory approach, as facilitated by the MMPBSA.py module within Amber18. We utilized 1000 snapshots extracted from the last 5 ns of the production MD simulation for these calculations. The modified GB model (igb=5) with mbondi2 parameters was employed, along with specific α , β , and γ values of 1.0, 0.8, and 4.85, respectively. Dielectric constants were set to 80 for the solvent and 5 for the solute. The outcome of this analysis yielded the mean interaction energy between the cyclic peptide and its protein partner. Per-residue decomposition allowed for a detailed investigation of the contribution of interacting residues to identify the hot spots.

Furthermore, we conducted RS-REMD for our best three matches. Simulations were conducted in explicit TIP3P water¹² at 300 K, together with neutralizing Na⁺ counterions in water and no additional salt. The prepared system structures were placed within periodic octahedral solvent boxes and a minimum distance between peptide and boundary of 15 Å. The systems were then minimized via Steepest Descent for 10 k steps, heated up step-wise to a target temperature of 300 K within 60 ps and equilibrated within NVT and NPT conditions for 1 ns each. The RS-REMD simulations were executed with 25 ns per replica and exchange attempts every picosecond. These simulations were carried out using 8 and 16 replicas, each exhibiting a progressive increase in the bias applied to the ligand and receptor Lennard-Jones parameters (see Table 3.1). The trajectories were analyzed using pytraj and the associated relative binding free energies were calculated with MBAR as proposed by Siebenmorgen and Zacharias, 2020 [71].

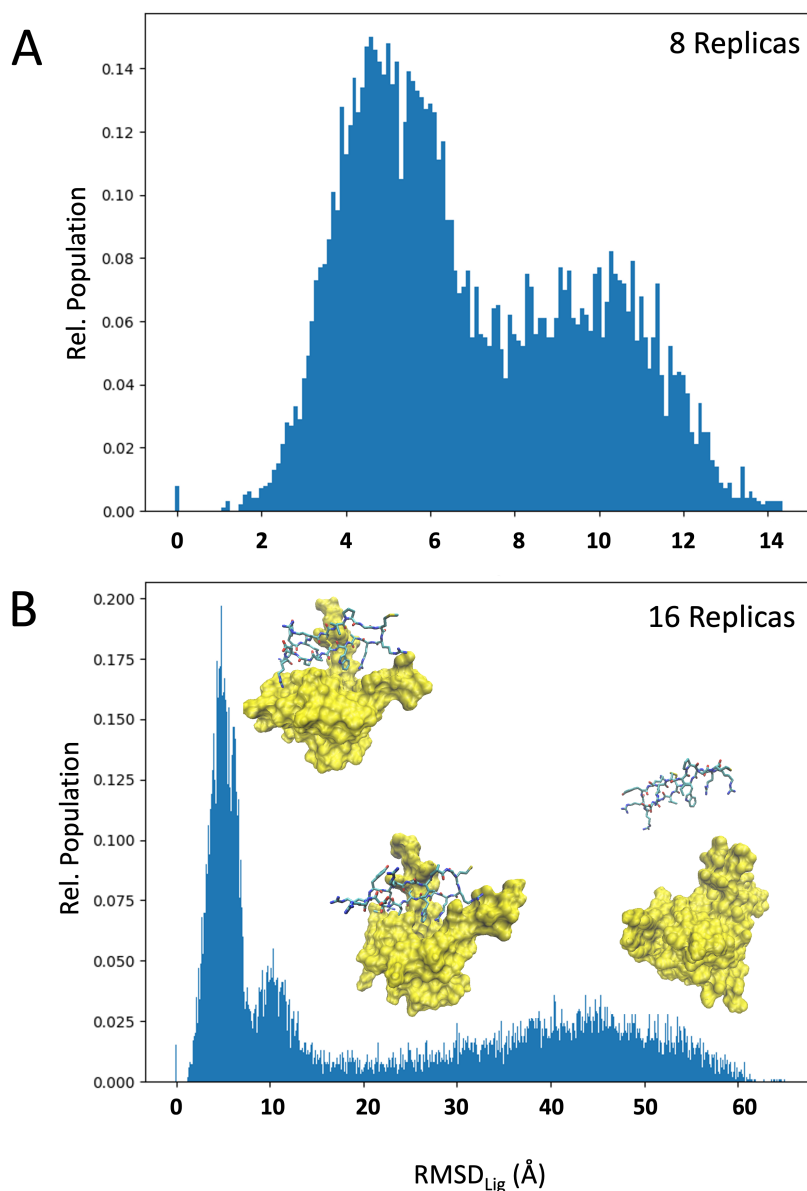


Figure 8.5.: Ligand RMSD Population Histograms for RS-REMD Simulations. Panel A represents the relative population distribution of ligand RMSD for 8-replica simulations, indicating the diversity of conformations sampled during the simulation. Panel B shows the corresponding distribution for 16-replica simulations alongside representative structures at key RMSD intervals, illustrating the enhanced sampling of bound and unbound states. The insets in Panel B depict the ligand in complex with the receptor (left), an intermediate state (middle), and a fully unbound state (right).

Table 8.1.: Input parameters and configurations for cPEPmatch algorithm. This table enumerates the PDB codes from the Protein Data Bank representing the target and its corresponding ligand. Listed hotspots, derived from MMGBSA analysis of MD simulations, indicate key interface regions; a dash signifies utilization of the entire interface. Columns 'C', 'T', and 'M' represent the cutoff, threshold, and motif type parameters, respectively, with 'T' for true (consecutive) and 'F' for false (non-consecutive). Where multiple parameter sets are provided, they correspond to distinct matching searches conducted.

Target	Ligand	PDB	Hot Spots	C	T	M
ACE2	Spike	6M0J	6 9 10 12 13 19 20 23 61 64 65 337	5;6	0.5;0.5	F;T
BCR	Abl	1K1F/AF2	198 201 204 218 222 223 226 259; 146 147 154 158 259 270 273 276 277	8;8	2;2	F;F
ER	Coactivator	5DZ1	242 243 244 245 246 247 248 249 250	8	1	T
GP120	CD4	1GC1	322 324 337 339 340 341 345 356 357	8	0.7	F
IL12	IL12Rb1	6WDQ	17 19 21 62 86 88 196 197 198 199	8;8	1;2	F;F
MDM2	p53	4HFZ	-	6	0.5	T
PD1	PDL1	3BIK	2 37 39 49 51; 96 98 104 105 106 108	5;5	1;0.5	F;F
S100P	RAGE	2MJW	28 32 34 37 41 43 49 78 90 92	8	2	F

8.4.2. Results and Discussion

The evaluation of stable matches in protein-protein interaction studies entails a multifaceted approach. We employed a three-pronged evaluation method: assessing stability during Free Molecular Dynamics (MD), calculating Molecular Mechanics Generalized Born Surface Area (MMGBSA) from free MD trajectories, and employing Multistate Bennett Acceptance Ratio (MBAR) analysis from Repulsive Scaling - Hamiltonian Replica Exchange Molecular Dynamics (RS-REMD) simulations, utilizing both 8 and 16 replicas. Our criteria for a stable match included sustained binding during the free MD trajectory, an MMGBSA value less than 25 kcal/mol, and an MBAR value lower than -2 kcal/mol (Figure D.11). These criteria were informed by previous studies: the MMGBSA threshold was based on our initial studies examining the binding affinity of small binders (see Chapter 2), and the MBAR criterion was derived from our investigation of the ICOS/ICOS-L system in Chapter 7, which is a recognized inhibitor, presenting values in a similar range.

Focusing on our tutorial-highlighted system, S100P, we identified hotspots (28, 32, 34, 37, 41, 43, 49, 78, 90, 92; refer to Table 8.1) through MD and MMGBSA analysis. cPEPmatch identified 16 potential matches (specifically matches 2, 3, 5, 7, and 15; see Table 2). After eliminating candidates with sterical clashes via visual inspection using VMD, five were selected for further simulation. According to Table 8.2, all selected

Table 8.2.: Comparative assessment of binding free energies change (ΔG) in kcal/mol across tested systems, calculated using MBAR from RS-HREMD trajectories, with 16 and 8 replicas, and MMGBSA methods.

System	Match	16 Replicas		8 Replicas		MMGBSA
		ΔG (kcal/mol)	+/-	ΔG (kcal/mol)	+/-	ΔG (kcal/mol)
ACE2	match2_2ll5	-1.81	0.74	-1.93	0.48	-12.17
ACE2	match12_3uc7	-3.0	0.32	-3.31	0.4	-32.46
ACE2	match145_6ve9	-2.36	0.2	-2.41	0.63	-27.41
BCR	match1_1jdp	-4.81	0.9	-4.07	0.26	-33.69
BCR	match3_4ttl	-5.87	0.77	-3.82	0.23	-30.71
BCR	match4_5otx	-6.26	0.79	-5.76	0.38	-48.31
ER	match1_3p72	-7.42	0.4	-3.99	0.12	-92.82
ER	match5_4ttl	-7.94	0.7	-4.65	0.35	-64.53
ER	match6_5wtt	-8.93	0.54	-5.95	0.35	-77.12
ER	match7_7m7x	-4.47	0.54	-3.11	0.21	-61.93
ER	match8_7m7x	-5.87	0.34	-5.37	0.74	-74.50
GP120	match5_1im7	-1.62	0.6	-3.99	0.61	-40.22
GP120	match9_1znu	-3.31	0.2	-3.7	0.66	-21.96
GP120	match18_3uc7	-2.4	0.91	-2.05	0.78	-38.09
GP120	match29_4tto	-3.47	0.22	-3.63	0.54	-12.38
IL12	match1_1ebp	-0.64	0.38	-1.11	0.22	-32.92
IL12	match2_2nb6	-1.2	0.85	-0.88	0.61	-24.32
IL12	match3_1im7	-1.9	0.16	-2.12	0.24	-33.97
IL12	match10_2ll5	-1.13	0.56	-0.94	0.47	-21.39
IL12	match14_2ndm	-0.78	0.48	-2.21	0.52	-17.01
IL12	match16_3pp4	-1.52	0.13	-1.22	0.24	-20.23
IL12	match30_5otx	-2.85	0.49	-1.58	0.29	-5.36
IL12	match31_6dhr	-2.05	0.24	-1.79	0.32	-38.19
MDM2	match6_3p72	-4.99	0.59	-4.32	0.46	-39.18
MDM2	match10_3pp4	-13.36	0.98	-7.0	0.43	-52.95
MDM2	match13_3uc7	-12.75	0.8	-6.02	0.24	-57.02
PD1	match9_2m7i	-0.17	0.18	-0.13	0.16	-30.10
PD1	match12_2ns4	-5.68	0.37	-2.49	0.26	-52.40
PD1	match13_2ns4	-3.57	0.4	-3.47	0.39	-43.12
PD1	match25_6q1u	-2.07	0.88	-2.71	0.17	-48.08
S100P	match2_1znu	-3.31	0.41	-2.83	0.46	-33.93
S100P	match3_2lwt	-8.65	0.7	-6.35	0.54	-66.98
S100P	match5_2lww	-2.76	0.56	-2.31	0.21	-43.25
S100P	match7_2lyf	-9.33	0.78	-4.19	0.28	-57.66
S100P	match15_7k7x	-8.59	0.79	-5.0	0.6	-41.60

Table 8.3.: Summary of stable matches identified through cPEPmatch and their corresponding point mutations. This table details the specific amino acid substitutions performed on the cyclic peptide templates to emulate the binding sites of their target proteins. Each row represents a unique match with its PDB code and a list of mutations that were introduced to achieve the desired structural and functional mimicry

System	Match	PDB	Mutations
ACE2	Match 12	3uc7	6-THR; 7-PHE; 9-ASP; 10-LYS
ACE2	Match 145	6ve9	11-PHE; 12-ASN; 13-HIS; 14-GLU; 15-ALA
BCR	Match 4	5otx	3-PHE; 5-ILE; 14-ARG; 24-PHE; 27-ILE
BCR	Match 1	1jdp	7-PHE; 11-GLN; 15-VAL
BCR	Match 3	4ttl	11-GLN; 17-VAL; 19-GLY; 20-PHE
ER	Match 1	3p72	4-HIS; 5-ARG; 6-LEU; 8-GLN
ER	Match 6	5wtt	3-ILE; 4-LEU; 6-ARG; 7-LEU
ER	Match 8	7m7x	4-ARG; 5-LEU; 7-GLN
ER	Match 5	4ttl	6-LEU; 7-HIS; 9-LEU; 10-LEU
ER	Match 7	7m7x	4-LEU; 5-HIS; 7-LEU
GP120	Match 5	1im7	7-GLN; 8-PHE; 12-ARG; 13-SER
GP120	Match 18	3uc7	9-HIS; 15-GLN; 17-PHE; 20-ARG
IL12	Match 31	6dhr	11-TRP; 15-PHE; 18-LYS; 20-LYS
IL12	Match 3	1im7	3-GLY; 11-THR
MDM2	Match 13	3uc7	1-PHE; 2-SER; 3-ASP; 4-LEU; 5-TRP
MDM2	Match 10	3pp4	15-PHE; 17-ASP; 18-LEU; 19-TRP
MDM2	Match 6	3p72	3-PHE; 4-SER; 5-ASP; 6-LEU; 7-TRP
PD1	Match 12	2ns4	2-ARG; 7-ALA; 8-ASP; 9-TYR
PD1	Match 25	6q1u	1-ARG; 3-MET; 10-ASP; 11-TYR; 13-ARG
PD1	Match 13	2ns4	2-ARG; 4-MET; 7-ALA; 8-ASP; 9-TYR
S100P	Match 3	2lwt	1-ARG; 11-ARG; 14-TRP; 16-VAL
S100P	Match 7	2lyf	1-ARG; 13-TRP; 15-VAL
S100P	Match 15	7k7x	4-LYS; 6-ASN; 27-TRP; 29-VAL
S100P	Match 5	2lwv	5-VAL; 8-GLY; 15-ARG; 17-ASN
S100P	Match 2	1znu	18-LYS; 20-TRP; 23-ARG; 25-LYS; 27-ASN

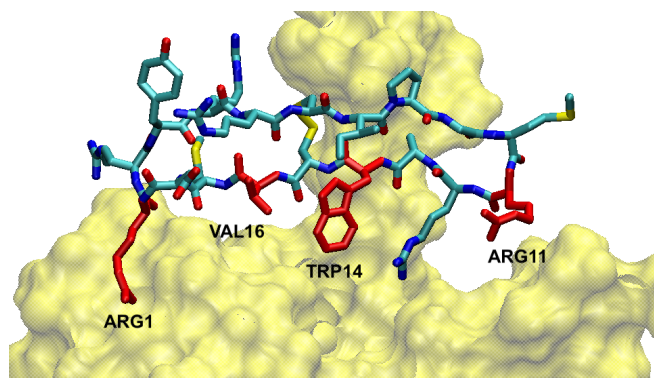


Figure 8.6.: Visualization of the strongest cPEPmatch predicted cyclic peptide binder, Match 3 (PDB: 2lwt), for the S100P system. The target protein S100P is rendered as a yellow surface, while the cyclic peptide binder is shown in licorice representation with atoms colored by element. Key hotspots on the cyclic peptide, corresponding to matched and mutated residues essential for binding, are highlighted in red.

matches exceeded the established stability criteria, as evidenced by their MDBAR (8 and 16 replicas) and MMGBSA values. Moreover, they demonstrated stability throughout the MD simulations. Detailed mutation information for these matches is available in Table 8.3. Notably, depicted in Figure 8.6, Match 3 exemplifies how effective mimicry of mutated hotspots can resemble those of the native binding site.

To refine our binding evaluation methods, we conducted RS-REMD simulations using 8 and 16 replicas across all systems. A notable Pearson correlation coefficient of 0.85 was observed between these two setups (Figure D.10). Typically, values from 16-replica simulations were lower compared to those from 8 replicas. A potential explanation for this discrepancy might lie in the sampling dynamics; 16-replica simulations are more effective at sampling a broader range of bound and unbound states between the cyclic peptide and its receptor, due to the stronger repulsion in the higher replicas. For instance, in the case of a strong binder like Match 3, the 16-replica simulations facilitated a comprehensive sampling of bound versus unbound states, while the 8-replica simulations failed to sample any completely unbound states, showing only the cyclic peptides in various conformations (Figure 8.5). This observation was consistent across other systems as well (refer to Supplementary Figures D.1-D.9).

Similar evaluation criteria were applied to other systems, resulting in the identification of stable matches with detailed mutation information presented in Table 8.2. Notably, while stable matches were observed in all systems during MD simulations, IL12 and GP120 predominantly exhibited weak to moderate binders. These

binders demonstrated lower MBAR binding scores compared to the ICOS/ICOS-L cyclic peptides, which are known experimentally validated binders or modulators, as discussed in Chapter 7. Importantly, our findings in Chapter 7 suggest that a higher binding affinity does not invariably translate into more effective modulation, indicating potential viability for the predicted binders in this study. In a systematic application of cPEPmatch, we identified a varying number of stable matches across systems: 2 for ACE2, 3 for BCR, 5 for ER, 2 for GP120, 2 for IL12, 3 for MDM2, and 3 for PD1 (See Table 8.3). These results underscore cPEPmatch’s efficacy as a tool in identifying lead molecules for therapeutic interventions.

8.5. Conclusions

This chapter has provided an in-depth understanding of the capabilities and applications of the cPEPmatch tool in the realm of medically relevant protein-protein interactions. From the development of the cPEPmatch tool, encompassing its transition from a Fortran-based program to a more accessible Python script, to its practical application in analyzing protein interfaces, this chapter has detailed each step of the process. The tool’s ability to automate the identification and mutation of cyclic peptide templates has been demonstrated, along with its application in a variety of medically significant systems.

The use of Molecular Dynamics simulations and free energy calculations, including MM/GBSA and Hamiltonian Replica Exchange Molecular Dynamics, has been pivotal in evaluating the binding efficacies of the peptides. This approach has allowed for a more rigorous assessment of the potential of these peptides as therapeutic agents. Furthermore, the launch of the cPEPmatch webserver has made this powerful tool accessible to a broader scientific community, simplifying the process of designing cyclic peptides for protein-protein and protein-ligand interactions.

Through case studies, we have showcased the tool’s efficacy in identifying viable cyclic peptides that can mimic critical protein-protein interactions, crucial in various diseases such as viral infections, cancer, and autoimmune disorders. The versatility and potential of cPEPmatch as a valuable tool in drug design and the study of protein interactions have been highlighted, underscoring its significance in the field of computational drug discovery.

9. Conclusions and Outlook

In this work, we have focused on studying cyclic peptides, a class of compounds renowned for their distinct structures and promising therapeutic applications. Our work has led to the development of the cPEPmatch algorithm and a virtual database of cyclic peptides, beginning with an initial set of 30 structures and expanding to encompass 432 unique structure entries. This growing database, sourced from the Protein Data Bank, is a testament to the broad scope of our research. It not only offers insights into the structural diversity and bioactivity of cyclic peptides but also provides a valuable resource for future studies, including potential applications in machine learning methods for drug discovery.

The emphasis on molecular dynamics (MD) simulations and free energy calculations throughout this study highlights the significance of computational methods in modern biomolecular research. These techniques have enabled us to model and analyze complex interactions at the molecular level, providing a deeper understanding of the dynamics of biological molecules.

Central to our investigation is the cPEPmatch method, a novel computational approach that we have developed for identifying cyclic peptide templates for protein-protein interaction (PPI) and protein-ligand modulators. Unique in its reliance on structural matching of protein-protein interfaces, cPEPmatch has proven its efficacy in suggesting potential cyclic peptide binders. This approach is complemented by our enhancement of the cPEPmatch method to specifically target hot loops within protein-protein interfaces. This refinement has proven instrumental in identifying putative cyclic peptides that precisely target critical regions in complex PPIs, as exemplified in our application to an MHC class I protein complex.

In addition to targeting protein-protein interfaces, our research has extended into designing cyclic peptides that mimic glycosaminoglycan (GAG)-binding proteins. By adapting the cPEPmatch algorithm to target these types of interfaces, we have identified promising peptide binders, notably 2LWT version 2. This discovery not only advances our understanding of peptide-GAG interactions but also opens new avenues for therapeutic interventions. Another significant achievement in our research has been the development of cyclic peptides as ICOS/ICOSL inhibitors. The discovery

of Match 41, identified through the cPEPmatch approach, marks a breakthrough in immunomodulation research, offering new perspectives in cancer therapy.

Furthermore, the recent establishment of the cPEPmatch available source code and webserver represents a major step forward in making our computational tools more accessible and user-friendly. This development is poised to facilitate how cyclic peptides are designed in future research. Significantly, we showcased its application in the identification of promising cyclic peptide predictions for medically relevant systems, showcasing the practical application of our tools in addressing real-world health challenges.

In conclusion, this thesis has laid a robust foundation for the future exploration of cyclic peptides. Our findings underscore the immense potential of these molecules in therapeutic applications, promising new horizons in drug discovery and biomedical research. The advancements made in this study are not only significant in their own right but also pave the way for further breakthroughs in the field of cyclic peptide therapeutics.

Bibliography

- [1] A. e. a. Atanasov. "Natural products in drug discovery: advances and opportunities". In: *Nature Reviews Drug Discovery* 20.3 (2021), pp. 200–216. doi: 10.1038/s41573-020-00114-z.
- [2] A. Grauer and B. König. "Peptidomimetics - A versatile route to biologically active compounds". In: *Eur. J. Org. Chem.* (2009), pp. 5099–5111.
- [3] J. Gavenonis, B. Sheneman, T. Siegert, M. Eshelman, and J. Kritzer. "Comprehensive analysis of loops at protein-protein interfaces for macrocycle design". In: *Nat. Chem. Biol.* 10 (2014), pp. 716–722.
- [4] L. Nevola and E. Giralt. "Modulating protein–protein interactions: the potential of peptides". en. In: *Chemical Communications* 51.16 (Feb. 2015), pp. 3302–3315. issn: 1364-548X. doi: 10.1039/C4CC08565E. URL: <http://pubs.rsc.org/en/content/articlelanding/2015/cc/c4cc08565e> (visited on 11/06/2022).
- [5] L. Walensky and G. Bird. "Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress". In: *Journal of Medicinal Chemistry* 57 (2014), pp. 6275–6288.
- [6] K. Deyle, X.-D. Kong, and C. Heinis. "Phage Selection of Cyclic Peptides for Application in Research and Drug Development". In: *Accounts of Chemical Research* 50.8 (2017), pp. 1866–1874. doi: 10.1021/acs.accounts.7b00184.
- [7] E. Driggers, S. Hale, J. Lee, and N. Terrett. "The exploration of macrocycles for drug discovery - An underexploited structural class". In: *Nat. Rev. Drug Discov.* 7 (2008).
- [8] E. Marsault and M. L. Peterson. "Macrocycles are great cycles: Applications, opportunities, and challenges of synthetic macrocycles in drug discovery". In: *Journal of Medicinal Chemistry* 54.7 (2011), pp. 1961–2004.
- [9] J. Mallinson and I. Collins. *Macrocycles in new drug discovery*. Future Science Ltd London, UK, 2012.
- [10] A. Zorzi, K. Deyle, and C. Heinis. "Cyclic peptide therapeutics: past, present and future". In: *Current Opinion in Chemical Biology* 38 (2017), pp. 24–29. doi: 10.1016/J.CBPA.2017.02.006.

- [11] N. London, B. Raveh, D. Movshovitz-Attias, and O. Schueler-Furman. "Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions?" In: *Proteins: Structure, Function, and Bioinformatics* 78.15 (2010), pp. 3140–3149.
- [12] M. e. a. Naylor. "Cyclic peptide natural products chart the frontier of oral bioavailability in the pursuit of undruggable targets". In: *Current Opinion in Chemical Biology* 38 (2017), pp. 141–147. DOI: 10.1016/j.cbpa.2017.04.012.
- [13] X. Ji, A. Nielsen, and C. Heinis. "Cyclic Peptides for Drug Development". In: *Angewandte Chemie International Edition* 63.3 (2024), e202308251. DOI: 10.1002/anie.202308251.
- [14] T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, and R. S. Lokey. "Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers". In: *Journal of the American Chemical Society* 128.8 (2006), pp. 2510–2511.
- [15] B. Khatri, V. Nuthakki, and J. Chatterjee. "Strategies to Enhance Metabolic Stabilities". In: *Cyclic Peptide Design*. Ed. by G. Goetz. New York, NY: Springer New York, 2019, pp. 17–40. DOI: 10.1007/978-1-4939-9504-2_2.
- [16] Z. Qian, P. G. Dougherty, and D. Pei. "Targeting intracellular protein-protein interactions with cell-permeable cyclic peptides". In: *Current Opinion in Chemical Biology* 38 (2017), pp. 80–86.
- [17] K. Klinker and S. Borgert. "Beyond Vancomycin: The Tail of the Lipoglycopeptides". In: *Clinical Therapeutics* 37 (2015), pp. 2619–2636.
- [18] K. Nicolaou, C. Boddy, S. Brase, and N. Winssinger. "Chemistry, biology, and medicine of the glycopeptide antibiotics". In: *Angewandte Chemie International Edition* 38 (1999), pp. 2096–2152.
- [19] D. Aguilar-Zapata, R. Petraitiene, and V. Petraitis. "Echinocandins: The Expanding Antifungal Armamentarium". In: *Clinical Infectious Diseases* 61 (2015), S604–S611.
- [20] I. Modlin, M. Pavel, M. Kidd, and B. Gustafsson. "Review article: somatostatin analogues in the treatment of gastroenteropancreatic neuroendocrine (carcinoid) tumours". In: *Alimentary Pharmacology Therapeutics* 31 (2010), pp. 169–188.
- [21] C. Bruns, I. Lewis, U. Briner, G. Meno-Tetang, and G. Weckbecker. "SOM230: a novel somatostatin peptidomimetic with broad somatotropin release inhibiting factor (SRIF) receptor binding and a unique antisecretory profile". In: *European Journal of Endocrinology* 146 (2002), pp. 707–716.

-
- [22] S. Jain and J. Zain. "Romidepsin in the treatment of cutaneous T-cell lymphoma". In: *J Blood Med* 2 (2011), pp. 37–47.
- [23] M. Corsetti and J. Tack. "Linaclotide: A new drug for the treatment of chronic constipation and irritable bowel syndrome with constipation". In: *United European Gastroenterology Journal* 1 (2013), pp. 7–20.
- [24] A. Bryant, R. Busby, W. Bartolini, E. Cordero, G. Hannig, M. Kessler, C. Pierce, R. Solinga, J. Tobin, S. Mahajan-Miklos, et al. "Linaclotide is a potent and selective guanylate cyclase C agonist that elicits pharmacological effects locally in the gastrointestinal tract". In: *Life Sciences* 86 (2010), pp. 760–765.
- [25] Y. Chang, B. Graves, V. Guerlavais, et al. "Stapled alpha-helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy". In: *Proceedings of the National Academy of Sciences* 110 (2013), E3445–E3454.
- [26] N. Srinivas et al. "Peptidomimetic Antibiotics Target Outer-Membrane Biogenesis in *Pseudomonas aeruginosa*". In: *Science* 327 (2010), pp. 1010–1013.
- [27] F. Fontaine, J. Overman, and M. François. "Pharmacological manipulation of transcription factor protein-protein interactions: Opportunities and obstacles". In: *BioMed Central Ltd.* (2015).
- [28] D. P. Ryan and J. M. Matthews. "Protein–protein interactions in human disease". In: *Current Opinion in Structural Biology* 15.4 (2005), pp. 441–446.
- [29] B. O. Villoutreix, M. A. Kuenemann, J.-L. Poyet, H. Bruzzoni-Giovanelli, C. Labbé, D. Lagorce, O. Sperandio, and M. A. Miteva. "Drug-like protein-protein interaction modulators: Challenges and opportunities for drug discovery and chemical biology". In: *Molecular Informatics* (2014).
- [30] R. Bahadur and M. Zacharias. "The interface of protein-protein complexes: Analysis of contacts and prediction of interactions". In: *Cell. Mol. Life Sci.* 65 (2008), pp. 1059–1072.
- [31] L. Conte, C. Chothia, and J. Janin. "The atomic structure of protein-protein recognition sites". In: *J. Mol. Biol.* 285 (1999), pp. 2177–2198.
- [32] M. Arkin, Y. Tang, and J. Wells. "Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality". In: *Chem. Biol.* 21 (2014), pp. 1102–1114.
- [33] O. Keskin, B. Ma, and R. Nussinov. "Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues". In: *J. Mol. Biol.* 345 (2005), pp. 1281–1294.

- [34] J. A. Wells and C. L. McClendon. "Reaching for high-hanging fruit in drug discovery at protein-protein interfaces". In: *Nature* 450.7172 (2007), pp. 1001–1009. DOI: 10.1038/nature06526.
- [35] A. Metz, C. Pfleger, H. Kopitz, S. Pfeiffer-Marek, K.-H. Baringhaus, and H. Gohlke. "Hot spots and transient pockets: Predicting the determinants of small-molecule binding to a protein-protein interface". In: *Journal of Chemical Information and Modeling* 52.1 (2012), pp. 120–133. DOI: 10.1021/ci200322s.
- [36] D. Fry. "Protein-Protein Interactions As Targets for Small Molecule Drug Discovery". In: *Biopolymers* (2006), pp. 535–552.
- [37] C. Bechara and S. Sagan. "Cell-penetrating peptides: 20 years later, where do we stand?" In: *Febs Letters* 587 (2013), pp. 1693–1702.
- [38] T. Passioura, T. Katoh, Y. Goto, and H. Suga. "Selection-Based Discovery of Druglike Macrocyclic Peptides". In: *Annual Review of Biochemistry, Vol 83* 83 (2014), pp. 727–752.
- [39] F. J. Duffy et al. "CycloPs: Generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids". In: *Journal of Chemical Information and Modeling* 51.4 (2011), pp. 829–836. DOI: 10.1021/ci100431r.
- [40] F. Duffy, D. O'donovan, M. Devocelle, N. Moran, D. O'connell, and D. Shields. "Virtual Screening Using Combinatorial Cyclic Peptide Libraries Reveals Protein Interfaces Readily Targetable by Cyclic Peptides". In: (2015).
- [41] Z. Zhang, Z. Lin, Z. Zhou, H. C. Shen, S. F. Yan, A. V. Mayweg, Z. Xu, N. Qin, J. C. Wong, Z. Zhang, et al. "Structure-based design and synthesis of potent cyclic peptides inhibiting the YAP-TEAD protein-protein interaction". In: *ACS Medicinal Chemistry Letters* 5.9 (2014), pp. 993–998.
- [42] D. R. Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 24.1 (1928), pp. 89–110. DOI: 10.1017/S0305004100011919.
- [43] V. Fock. "Nherungsmethode zur Lsung des quantenmechanischen Mehrkörperproblems". de. In: *Zeitschrift für Physik* 61.1 (Jan. 1930), pp. 126–148. ISSN: 0044-3328. DOI: 10.1007/BF01340294. URL: <https://doi.org/10.1007/BF01340294> (visited on 02/22/2023).
- [44] J. C. Slater. "Note on Hartree's Method". In: *Physical Review* 35.2 (Jan. 1930), pp. 210–211. DOI: 10.1103/PhysRev.35.210.2. URL: <https://link.aps.org/doi/10.1103/PhysRev.35.210.2> (visited on 02/22/2023).

-
- [45] E. R. Johnson and A. D. Becke. "A post-Hartree-Fock model of intermolecular interactions". In: *The Journal of Chemical Physics* 123.2 (July 2005), p. 024101. ISSN: 0021-9606. DOI: 10.1063/1.1949201. URL: <https://aip.scitation.org/doi/abs/10.1063/1.1949201> (visited on 02/22/2023).
- [46] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory*. en. Google-Books-ID: qm5cCwAAQBAJ. John Wiley & Sons, Nov. 2015. ISBN: 978-3-527-80281-4.
- [47] V. W. D. Cruzeiro, M. Manathunga, K. M. Merz, and A. W. Götz. "Open-Source Multi-GPU-Accelerated QM/MM Simulations with AMBER and QUICK". en. In: *Journal of Chemical Information and Modeling* 61.5 (May 2021), pp. 2109–2115. ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.1c00169. URL: <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00169> (visited on 02/22/2023).
- [48] L. Monticelli and D. P. Tieleman. "Force fields for classical molecular dynamics". In: *Biomolecular simulations: Methods and protocols*. 2013, pp. 197–213.
- [49] D. Case et al. *AMBER 20*. University of California, San Francisco. 2020.
- [50] T. A. Halgren. "The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters". In: *Journal of the American Chemical Society* 114.20 (Sept. 1992), pp. 7827–7843. ISSN: 0002-7863. DOI: 10.1021/ja00046a032. URL: <https://doi.org/10.1021/ja00046a032> (visited on 02/22/2023).
- [51] T. A. Halgren and W. Damm. "Polarizable force fields". eng. In: *Current Opinion in Structural Biology* 11.2 (Apr. 2001), pp. 236–242. ISSN: 0959-440X. DOI: 10.1016/s0959-440x(00)00196-2.
- [52] G. Glashagen, S. de Vries, U. Uciechowska-Kaczmarzyk, S. A. Samsonov, S. Murail, P. Tuffery, and M. Zacharias. "Coarse-grained and atomic resolution biomolecular docking with the ATTRACT approach". en. In: *Proteins: Structure, Function, and Bioinformatics* 88.8 (2020), pp. 1018–1028. ISSN: 1097-0134. DOI: 10.1002/prot.25860. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.25860> (visited on 11/07/2022).
- [53] L. Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Physical Review* 159.1 (July 1967), pp. 98–103. DOI: 10.1103/PhysRev.159.98. URL: <https://link.aps.org/doi/10.1103/PhysRev.159.98> (visited on 02/22/2023).
- [54] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341.
-

- [55] C. W. Hopkins et al. "Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning". In: *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1864–1874.
- [56] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Elsevier, 2001.
- [57] T. Darden, D. York, and L. Pedersen. "Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems". In: *The Journal of Chemical Physics* 98.12 (June 1993), pp. 10089–10092. ISSN: 0021-9606. DOI: 10.1063/1.464397. URL: <https://aip.scitation.org/doi/10.1063/1.464397> (visited on 02/22/2023).
- [58] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". en. In: *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.448118. URL: <http://aip.scitation.org/doi/10.1063/1.448118> (visited on 05/22/2019).
- [59] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, and H. J. C. Berendsen. "Efficient Algorithms for Langevin and DPD Dynamics". In: *Journal of Chemical Theory and Computation* 8.10 (Oct. 2012), pp. 3637–3649. ISSN: 1549-9618. DOI: 10.1021/ct3000876. URL: <https://doi.org/10.1021/ct3000876> (visited on 06/28/2020).
- [60] S. Genheden and U. Ryde. "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities". In: *Expert Opinion on Drug Discovery* 10.5 (May 2015), pp. 449–461. ISSN: 1746-0441. DOI: 10.1517/17460441.2015.1032936. URL: <https://doi.org/10.1517/17460441.2015.1032936> (visited on 09/19/2022).
- [61] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang, and T. Hou. "End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design". In: *Chemical Reviews* 119.16 (Aug. 2019), pp. 9478–9508. ISSN: 0009-2665. DOI: 10.1021/acs.chemrev.9b00055. URL: <https://doi.org/10.1021/acs.chemrev.9b00055> (visited on 02/22/2023).
- [62] D. Bashford and D. Case. "Generalized Born Models of Macromolecular Solvation Effects". In: *Annual Review of Physical Chemistry* 51 (2000), pp. 129–152. DOI: 10.1146/annurev.physchem.51.1.129.
- [63] H. Fukunishi, O. Watanabe, and S. Takada. "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction". In: *The Journal of Chemical Physics* 116.20 (2002), pp. 9058–9067.

-
- [64] N. M. et al. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.
- [65] T. Siebenmorgen and M. Zacharias. "Origin of ion specificity of telomeric DNA G-quadruplexes investigated by free-energy simulations". In: *Biophysical Journal* 112.11 (2017), pp. 2280–2290.
- [66] K. Ostermeir and M. Zacharias. "Rapid alchemical free energy calculation employing a generalized born implicit solvent model". In: *The Journal of Physical Chemistry B* 119.3 (2015), pp. 968–975.
- [67] M. P. Luitz and M. Zacharias. "Protein–ligand docking using hamiltonian replica exchange simulations with soft core potentials". In: *Journal of Chemical Information and Modeling* 54.6 (2014), pp. 1669–1675.
- [68] F. Zeller and M. Zacharias. "Adaptive biasing combined with Hamiltonian replica exchange to improve umbrella sampling free energy simulations". In: *Journal of Chemical Theory and Computation* 10.2 (2014), pp. 703–710.
- [69] M. L. et al. "Exploring biomolecular dynamics and interactions using advanced sampling methods". In: *Journal of Physics: Condensed Matter* 27.32 (2015), p. 323101.
- [70] M. Kienlein and M. Zacharias. "Ligand binding and global adaptation of the GlnPQ substrate binding domain 2 revealed by molecular dynamics simulations". In: *Protein Science* 29.12 (2020), pp. 2482–2494.
- [71] T. Siebenmorgen and M. Zacharias. "Prediction of protein–protein complexes using replica exchange with repulsive scaling". In: *Journal of Computational Chemistry* 41.15 (2020), pp. 1436–1447.
- [72] T. Siebenmorgen and M. Zacharias. "Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling". In: *Journal of Chemical Information and Modeling* 60.11 (2020), pp. 5552–5562.
- [73] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw. "Atomic-level characterization of proteinprotein association". In: *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019), pp. 4244–4249.
- [74] F. L. Gervasio, A. Laio, and M. Parrinello. "Flexible docking in solution using metadynamics". In: *J. Am. Chem. Soc.* 127 (2005), pp. 2600–2607.
- [75] C. H. Bennett. "Efficient estimation of free energy differences from Monte Carlo data". In: *Journal of Computational Physics* 22.2 (1976), pp. 245–268.
- [76] M. R. Shirts and J. D. Chodera. "Statistically optimal analysis of samples from multiple equilibrium states". In: *The Journal of Chemical Physics* 129.12 (2008), p. 124105.
-

- [77] J. Träg and D. Zahn. "Improved GAFF2 parameters for fluorinated alkanes and mixed hydro-and fluorocarbons". In: *Journal of Molecular Modeling* 25.2 (2019), p. 39. doi: 10.1007/s00894-019-3927-y.
- [78] J. W. et al. "Development and testing of a general amber force field". In: *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174. doi: 10.1002/jcc.20035.
- [79] L. Nevola and E. Giralt. "Modulating protein-protein interactions: The potential of peptides". In: *Chemical Communications* 51.15 (2015), pp. 3302–3315.
- [80] M. Dechantsreiter, E. Planker, B. Mathä, E. Lohof, G. Hölzemann, A. Jonczyk, S. Goodman, and H. Kessler. "N-methylated cyclic RGD peptides as highly active and selective $\alpha(v)\beta3$ integrin antagonists". In: *J. Med. Chem.* 42 (1999), pp. 3033–3040.
- [81] S. L. Schreiber and G. R. Crabtree. "The mechanism of action of cyclosporin A and FK506". In: *Immunology Today* 13 (1992), pp. 136–142.
- [82] Z. Shi, C. Wei, K. Lee, H. Liu, M. Zhang, T. Araki, L. R. Roberts, K. M. Worthy, R. J. Fisher, B. G. Neel, et al. "Macrocyclization in the Design of Non-Phosphorus-Containing Grb2 SH2 Domain-Binding Ligands". In: *Journal of Medicinal Chemistry* 47.8 (2004), pp. 2166–2169.
- [83] W.-H. Shin, C. W. Christoffer, and D. Kihara. "In silico structure-based approaches to discover protein-protein interaction-targeting drugs". In: *Methods* 131 (2017), pp. 22–32.
- [84] T. R. Siegert, M. J. Bird, K. M. Makwana, and J. A. Kritzer. "Analysis of Loops that Mediate Protein-Protein Interactions and Translation into Submicromolar Inhibitors". In: *Journal of the American Chemical Society* 138.39 (2016), pp. 12876–12884.
- [85] G. A. G. Sulyok, C. Gibson, S. L. Goodman, G. Hölzemann, M. Wiesner, and H. Kessler. "Solid-phase synthesis of a nonpeptide RGD mimetic library: New selective $\alpha v\beta3$ integrin antagonists". In: *Journal of Medicinal Chemistry* 44.16 (2001), pp. 1938–1950.
- [86] J. Kallen, V. Mikol, P. Taylor, and M. D.walkinshaw. "X-ray structures and analysis of 11 cyclosporin derivatives complexed with cyclophilin A". In: *J. Mol. Biol.* 283 (1998), pp. 435–449.
- [87] D. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods. "The Amber biomolecular simulation programs". In: *J. Comput. Chem.* 26 (2005), pp. 1668–1688.

-
- [88] R. Huigens, K. Morrison, R. Hicklin, T. Timothy, M. Richter, and P. Hergenrother. "A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products". In: *Nat. Chem.* 5 (2013), pp. 195–202.
- [89] M. A. Kuenemann, O. Sperandio, C. M. Labbé, D. Lagorce, M. A. Miteva, and B. O. Villoutreix. "In silico design of low molecular weight protein-protein interaction inhibitors: Overall concept and recent advances". In: *Progress in Biophysics and Molecular Biology* 119.1 (2015), pp. 20–32.
- [90] M. B. Martins and I. Carvalho. "Diketopiperazines: biological activity and synthesis". In: *Tetrahedron* 63.40 (2007), pp. 9923–9932.
- [91] S. Milhas, B. Raux, S. Betzi, C. Derviaux, P. Roche, A. Restouin, M.-J. Basse, E. Rebuffet, A. Lugari, M. Badol, et al. "Protein-Protein Interaction Inhibition (2P2I)-Oriented Chemical Library Accelerates Hit Discovery". In: *ACS Chemical Biology* 11.8 (2016), pp. 2140–2148.
- [92] Y. Che. "Design of Cyclic Peptides as Protein Recognition Motifs". In: *Cyclic Peptide Design*. Ed. by G. Goetz. 2019, pp. 97–106.
- [93] F. Duffy, N. Maheshwari, N.-V. Buchete, and D. Shields. "Computational Opportunities and Challenges in Finding Cyclic Peptide Modulators of Protein-Protein Interactions". In: *Cyclic Peptide Design*. Ed. by G. Goetz. 2019, pp. 73–95.
- [94] D. Tikk, P. Thomas, P. Palaga, J. Hakenberg, and U. Leser. "A Comprehensive Benchmark of Kernel Methods to Extract Protein-Protein Interactions from Literature". In: *PLoS Computational Biology* 6.7 (2010), e1000837.
- [95] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". In: *Journal of Chemical Theory and Computation* 11.8 (2015), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.
- [96] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein. "Comparison of simple potential functions for simulating liquid water". In: *J. Chem. Phys.* 79 (1983), pp. 926–935.
- [97] S. L. Schreiber and G. R. Crabtree. "The mechanism of action of cyclosporin A and FK506". In: *Immunology Today* (1992).
- [98] J. K. Murray and S. H. Gellman. "Targeting protein-protein interactions: Lessons from p53/MDM2". In: *Biopolymers* 88.6 (2007), pp. 657–686. doi: 10.1002/bip.20741.
-

- [99] C. Corbi-Verge and P. M. Kim. "Motif mediated protein-protein interactions as drug targets". In: *Cell Communication and Signaling* 14.1 (2016), p. 8. DOI: 10.1186/s12964-016-0133-7.
- [100] Y. Qiu, X. Li, X. He, J. Wang, and X.-Q. Xie. "Computational methods-guided design of modulators targeting protein-protein interactions (PPIs)". In: *European Journal of Medicinal Chemistry* 207 (2020), p. 112764. DOI: 10.1016/j.ejmech.2020.112764.
- [101] D. E. Scott, A. R. Bayly, C. Abell, and J. Skidmore. "Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge". In: *Nature Reviews Drug Discovery* 15.8 (2016), pp. 533–550. DOI: 10.1038/nrd.2016.29.
- [102] S. A. Andrei, P. J. de Vink, E. Sijbesma, L. Han, L. Brunsveld, and C. Ottmann. "Rationally Designed Semisynthetic Natural Product Analogues for Stabilization of 14-3-3 Protein-Protein Interactions". In: *Angewandte Chemie International Edition* 130.44 (2018), pp. 13658–13662. DOI: 10.1002/ange.201806584.
- [103] B. L. Santini and M. Zacharias. "Rapid in silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces". In: *Frontiers in Chemistry* 8 (2020), p. 573259. DOI: 10.3389/fchem.2020.573259.
- [104] S. Montealegre, V. Venugopalan, S. Fritzsche, N. Kern, D. L. Tierney, T. J. Rutherford, S. M. Freund, G. Winter, A. van Hateren, and T. Elliott. "Dissociation of β 2-microglobulin determines the surface quality control of major histocompatibility complex class I molecules". In: *FASEB J* 29.6 (2015), pp. 2780–2788. DOI: 10.1096/fj.14-268094.
- [105] K. Maenaka and E. Y. Jones. "MHC superfamily structure and the immune system". In: *Current Opinion in Structural Biology* 9.6 (1999), pp. 745–753. DOI: 10.1016/S0959-440X(99)00037-4.
- [106] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. I. Cheatham, V. W. D. Cruzeiro, T. Darden, R. E. Duke, G. Giambasu, and et al. *AMBER 2018*. University of California, San Francisco, 2018.
- [107] C. Wang, D. Greene, L. Xiao, R. Qi, and R. Luo. "Recent developments and applications of the MMPBSA method". In: *Frontiers in Molecular Biosciences* 4 (2018).
- [108] H. Gohlke, C. Kiel, and D. A. Case. "Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes". In: *Journal of Molecular Biology* 330.4 (2003), pp. 891–913. DOI: 10.1016/S0022-2836(03)00610-7.
- [109] N. e. a. Karamanos. "Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics". In: *Chemical Reviews* 118.18 (2018), pp. 9152–9232. DOI: 10.1021/acs.chemrev.8b00354.

-
- [110] S. e. a. Perez. "Glycosaminoglycans: What Remains To Be Deciphered?" In: *JACS Au* 3.3 (2023), pp. 628–656. doi: 10.1021/jacsau.2c00569.
- [111] J. e. a. Wilbs. "Cyclic peptide FXII inhibitor provides safe anticoagulation in a thrombosis model and in artificial lungs". In: *Nature Communications* 11.1 (2020), p. 3890. doi: 10.1038/s41467-020-17648-w.
- [112] S. T. Olson, B. Richard, et al. "Molecular mechanisms of antithrombin–heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors". In: *Biochimie* 92.11 (2010), pp. 1587–1596.
- [113] A. Dementiev et al. "The ternary complex of antithrombin–anhydrothrombin–heparin reveals the basis of inhibitor specificity". In: *Nat Struct Mol Biol* 11.9 (2004), pp. 863–867.
- [114] S. T. Olson, I. Björk, et al. "Role of the antithrombin-binding pentasaccharide in heparin acceleration of antithrombin-proteinase reactions. Resolution of the antithrombin conformational change contribution to heparin rate enhancement". In: *J Biol Chem* 267.18 (1992), pp. 12528–12538.
- [115] J. Kreuger et al. "Interactions between heparan sulfate and proteins: the concept of specificity". In: *J Cell Biol* 174.3 (2006), pp. 323–327.
- [116] B. E. Thacker et al. "Heparan sulfate 3-O-sulfation: A rare modification in search of a function". In: *Matrix Biol* 35 (2014), pp. 60–72.
- [117] S. Vallet, O. Clerc, and S. Ricard-Blum. "Glycosaminoglycan–Protein Interactions: The First Draft of the Glycosaminoglycan Interactome". In: *Journal of Histochemistry & Cytochemistry* 69.2 (2021), pp. 93–104. doi: 10.1369/0022155420946403.
- [118] A. J. McCoy et al. "Structure of β -Antithrombin and the Effect of Glycosylation on Antithrombin's Heparin Affinity and Activity". In: *J Mol Biol* 326.3 (2003), pp. 823–833.
- [119] A. Zhou, J. A. Huntington, and R. W. Carrell. "Formation of the Antithrombin Heterodimer In Vivo and the Onset of Thrombosis". en. In: *Blood* 94.10 (Sept. 1999), pp. 3388–3396. issn: 0006-4971. doi: 10.1182/blood.V94.10.3388.422k20_3388_3396. URL: <https://www.sciencedirect.com/science/article/pii/S0006497120710062> (visited on 10/17/2022).
- [120] A. Sali and T. Blundell. "Comparative protein modelling by satisfaction of spatial restraints". In: *Journal of Molecular Biology* 234.3 (1993), pp. 779–815. doi: 10.1006/jmbi.1993.1626.
-

- [121] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". In: *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.
- [122] S. A. Samsonov and M. T. Pisabarro. "Computational analysis of interactions in structurally available protein–glycosaminoglycan complexes". en. In: *Glycobiology* 26.8 (Aug. 2016). Publisher: Oxford Academic, pp. 850–861. issn: 0959-6658. doi: 10.1093/glycob/cww055. URL: <https://academic.oup.com/glycob/article/26/8/850/2389058> (visited on 11/20/2020).
- [123] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods. "GLYCAM06: a generalizable biomolecular force field. Carbohydrates". In: *Journal of Computational Chemistry* 29.4 (Mar. 2008), pp. 622–655. issn: 1096-987X. doi: 10.1002/jcc.20820.
- [124] H. Nguyen, J. Maier, H. Huang, V. Perrone, and C. Simmerling. "Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent". In: *Journal of the American Chemical Society* 136.40 (Oct. 2014). Publisher: American Chemical Society, pp. 13959–13962. issn: 0002-7863. doi: 10.1021/ja5032776. URL: <https://doi.org/10.1021/ja5032776> (visited on 12/07/2021).
- [125] D. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. Cerutti, T. Cheatham III, G. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, C. Jin, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xue, D. York, S. Zhao, and P. Kollman. *Amber 2021*. University of California, San Francisco, 2021.
- [126] A. Onufriev et al. "Modification of the Generalized Born Model Suitable for Macromolecules". In: *J Phys Chem B* 104.15 (2000), pp. 3712–3720.
- [127] C. Robert. "A decade of immune-checkpoint inhibitors in cancer therapy". In: *Nature Communications* 11.1 (2020), p. 3801. doi: 10.1038/s41467-020-17670-y.
- [128] J. Jacob, M. Jacob, and P. Parajuli. "Review of immune checkpoint inhibitors in immuno-oncology". In: *Advances in Pharmacology* 91 (2021), pp. 111–139. doi: 10.1016/bs.apha.2021.01.002.

-
- [129] A. Schoenfeld and M. Hellmann. "Acquired Resistance to Immune Checkpoint Inhibitors". In: *Cancer Cell* 37.4 (2020), pp. 443–455. DOI: 10.1016/j.ccell.2020.03.017.
- [130] L. Barrueto et al. "Resistance to Checkpoint Inhibition in Cancer Immunotherapy". In: *Translational Oncology* 13.3 (2020), p. 100738. DOI: 10.1016/j.tranon.2020.100738.
- [131] D. Capece et al. "Targeting costimulatory molecules to improve antitumor immunity". In: *J Biomed Biotechnol* 2012 (2012), p. 926321.
- [132] P. Greaves and J. Gribben. "The role of B7 family molecules in hematologic malignancy". In: *Blood* 121.5 (2013), pp. 734–744.
- [133] S. Jeong and S. Park. "Co-stimulatory receptors in cancers and their implications for cancer immunotherapy". In: *Immune Netw* 20.1 (2020), e3–e3.
- [134] P. Suntharalingam, M. R. Ward, B. Stephen, A. Castello-Cortes, M. D. Brunner, and N. Panoskaltsis. "Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412". In: *N Engl J Med* 355.10 (2006), pp. 1018–1028.
- [135] G. Nicholson, L. I. Carlesso, H. Cheng, K. Cook, J. DaCosta, K. Leininger, S. (McKeever, D. Scott, K. Taylor, S. Streicher, M. Eck, R. Reed, R. Faggioni, R. Herbst, P. C. Dixit, and Ryan. "Effects of ICOS T cell depletion via afucosylated monoclonal antibody MEDI-570 on pregnant cynomolgus monkeys and the developing offspring". In: *Reprod Toxicol* 74 (2017), pp. 116–133.
- [136] A. Hutloff, A. M. Dittrich, K. C. Beier, B. Eljaschewitsch, R. Kraft, I. Anagnostopoulos, and R. A. Kroczeck. "ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28". In: *Nature* 397.6716 (1999), pp. 263–266.
- [137] V. Redoglia, U. Dianzani, J. M. Rojo, P. Portoles, M. Bragardo, H. Wolff, D. Buonfiglio, S. Bonisconi, and C. A. J. Jr. "Characterization of H4: a mouse T lymphocyte activation molecule functionally associated with the CD3/T cell receptor". In: *Eur J Immunol* 26.11 (1996), pp. 2781–2789.
- [138] A. Sharpe and G. Freeman. "The B7–CD28 superfamily". In: *Nat Rev Immunol* 2.2 (2002), pp. 116–126.
- [139] D. Wikenheiser and J. Stumhofer. "ICOS co-stimulation: Friend or foe?" In: *Front Immunol* 7 (2016), p. 304.
- [140] M. Swallow, J. Wallin, and W. Sha. "B7h, a Novel Costimulatory Homolog of B7.1 and B7.2, Is Induced by TNF". In: *Immunity* 11.4 (1999), pp. 423–432.
-

- [141] S. Khayyamian, A. Hutloff, K. Büchner, M. Gräfe, V. Henn, R. A. Kroczeck, and H. W. Mages. "ICOS-ligand, expressed on human endothelial cells, costimulates Th1 and Th2 cytokine secretion by memory CD4 T cells". In: *Proc Natl Acad Sci U S A* 99.9 (2002), pp. 6198–6203.
- [142] F. Amatore, L. Gorvel, and D. Olive. "Role of Inducible Co-Stimulator (ICOS) in cancer immunotherapy". In: *Expert Opin Biol Ther* 20.2 (2020), pp. 141–150.
- [143] J. Cinier, M. Hubert, L. Besson, A. D. Roio, C. Rodriguez, V. Lombardi, C. Caux, and C. Ménétrier-Caux. "Recruitment and Expansion of Tregs Cells in the Tumor Environment—How to Target Them?" In: *Cancers* 13.8 (2021).
- [144] H. Xu, X. Li, D. Liu, J. Li, X. Zhang, X. Chen, S. Hou, L. Peng, C. Xu, W. Liu, L. Zhang, and H. Qi. "Follicular T-helper cell recruitment governed by bystander B cells and ICOS-driven motility". In: *Nature* 496.7446 (2013), pp. 523–527.
- [145] D. Liu, H. Xu, C. Shih, Z. Wan, X. Ma, W. Ma, D. Luo, and H. Qi. "T-B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction". In: *Nature* 517.7533 (2015), pp. 214–218.
- [146] S. Yoon et al. In: *Exp Hematol Oncol* 10.1 (2021), p. 33.
- [147] S. Chiba and M. Sakata-Yanagimoto. "Advances in understanding of angioimmunoblastic T-cell lymphoma". In: *Leukemia* 34.10 (2020), pp. 2592–2606.
- [148] A. Sosa et al. In: *Ther Adv Med Oncol* 10 (2018), p. 1758835918764628.
- [149] I. Hernandez et al. "Pricing of Monoclonal Antibody Therapies: Higher If Used for Cancer?" In: *Am J Manag Care* 24.2 (2018), pp. 109–112.
- [150] L. Calvo-Barreiro, V. Talagayev, S. Pach, S. A. Abdel-Rahman, G. Wolber, and M. T. Gabr. In: *ChemMedChem* 2023 (2023), e202300305.
- [151] K. Fosgerau and T. Hoffmann. "Peptide therapeutics: current status and future directions". In: *Drug Discov Today* 20.1 (2015), pp. 122–8.
- [152] H. Weinmann. "Cancer Immunotherapy: Selected Targets and Small-Molecule Modulators". In: *ChemMedChem* 11.5 (2016), pp. 450–66.
- [153] G. I. Vladimer, B. Snijder, N. Krall, J. W. Bigenzahn, K. V. M. Huber, C.-H. Lardeau, K. Sanjiv, A. Ringler, U. W. Berglund, M. Sabler, O. L. de la Fuente, P. Knobl, S. Kubicek, T. Helleday, U. Jager, and G. Superti-Furga. "Global survey of the immunomodulatory potential of common drugs". In: *Nat Chem Biol* 13.6 (2017), pp. 681–690.
- [154] M. Gabr and S. Gambhir. "Discovery and Optimization of Small-Molecule Ligands for V-Domain Ig Suppressor of T-Cell Activation (VISTA)". In: *J Am Chem Soc* 142.38 (2020), pp. 16194–16198.
- [155] S. A. Abdel-Rahman, K. Świderek, and M. T. Gabr. In: *RSC Med Chem* 14.9 (2023), pp. 1767–1777.

-
- [156] S. A. Abdel-Rahman, L. Zhang, and M. T. Gabr. In: *SLAS Discov* 28.4 (2023), pp. 188–192.
- [157] E. Rujas et al. “Structural characterization of the ICOS/ICOS-L immune complex reveals high molecular mimicry by therapeutic antibodies”. In: *Nature Communications* 11.1 (2020), p. 5066. DOI: 10.1038/s41467-020-18828-4.
- [158] J. Wang, T. Hou, and X. Xu. “Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models”. In: *Current Computer-Aided Drug Design* 2.3 (2006), pp. 287–306. DOI: 10.2174/157340906778226454.
- [159] D. R. Roe and T. E. Cheatham. “PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data”. In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 3084–3095. DOI: 10.1021/ct400341p.
- [160] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and X. Wang. “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor”. In: *Nature* 581 (2020), pp. 215–220. DOI: 10.1038/s41586-020-2180-5.
- [161] X. Zhao, S. Ghaffari, H. Lodish, V. Malashkevich, and P. Kim. “Structure of the Bcr-Abl oncoprotein oligomerization domain”. In: *Nat Struct Biol* 9 (2002), pp. 117–120. DOI: 10.1038/nsb747.
- [162] J. Nwachukwu, S. Srinivasan, Y. Zheng, S. Wang, J. Min, C. Dong, Z. Liao, J. Nowak, N. Wright, R. Houtman, K. Carlson, J. Josan, O. Elemento, J. Katzenellenbogen, H. Zhou, and K. Nettles. “Predictive features of ligand-specific signaling through the estrogen receptor”. In: *Mol Syst Biol* 12 (2016), p. 864. DOI: 10.15252/msb.20156701.
- [163] P. Kwong, R. Wyatt, J. Robinson, R. Sweet, J. Sodroski, and W. Hendrickson. “Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody”. In: *Nature* 393 (1998), pp. 648–659. DOI: 10.1038/31405.
- [164] C. Glassman, Y. Mathiharan, K. Jude, L. Su, O. Panova, P. Lupardus, J. Spangler, L. Ely, C. Thomas, G. Skinotis, and K. Garcia. “Structural basis for IL-12 and IL-23 receptor sharing reveals a gateway for shaping actions on T versus NK cells”. In: *Cell* 184 (2021), 983–999.e24. DOI: 10.1016/j.cell.2021.01.018.
- [165] B. Anil, C. Riedinger, J. Endicott, and M. Noble. “The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant”. In: *Acta Crystallogr D Biol Crystallogr* 69 (2013), pp. 1358–1366. DOI: 10.1107/S0907444913004459.

- [166] D. Lin, Y. Tanaka, M. Iwasaki, A. Gittis, H. Su, B. Mikami, T. Okazaki, T. Honjo, N. Minato, and D. Garboczi. "The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors". In: *Proc Natl Acad Sci U S A* 105 (2008), pp. 3011–3016. DOI: 10.1073/pnas.0712278105.
- [167] S. Penumutthu, R. Chou, and C. Yu. "Structural insights into calcium-bound S100P and the V domain of the RAGE complex". In: *PLoS One* 9 (2014), e103947. DOI: 10.1371/journal.pone.0103947.
- [168] A. Fiser and A. Šali. "MODELLER: generation and refinement of homology-based protein structure models". In: *Methods in Enzymology* 374 (2003). Ed. by C. Carter and R. Sweet, pp. 463–493.
- [169] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

A. Supplementary Information for Chapter 2: Cyclic Peptides and Protein-Protein Interactions

Table A.1.: Comprehensive List of Cyclic Peptides in the cPEPmatch Database. This table presents the PDB codes of cyclic peptides as registered in the Protein Data Bank, along with their respective cyclization types, total amino acid count, and a breakdown of their secondary structures.

	PDB	Cyclization Type	A.A.	Secondary Structure
0	6PIO	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
1	4TWT	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
2	5VAV	head to tail	8	coil: 8
3	5XN3	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
4	6PIN	1x disulfide bond	10	coil: 10
5	4MNV	side-chain to backbone	13	hairpin: 6 / coil: 7
6	4X1N	1x disulfide bond	13	hairpin: 4 / coil: 9
7	4ZQW	2x disulfide bond	12	coil: 12
8	4IB5	1x disulfide bond	15	coil: 15
9	1CNL	2x disulfide bond	21	helix: 9 / coil: 12
10	1QX9	head to tail	8	hairpin: 2 / coil: 6
11	5GLH	side-chain to side-chain	6	coil: 6
12	3AVA	head to tail	8	hairpin: 2 / coil: 6
13	5LSO	side-chain to backbone	5	coil: 5
14	3AVB	head to tail	8	coil: 8
15	1URC	1x disulfide bond	19	helix: 4 / coil: 15
16	6AXI	1x disulfide bond / head to tail	17	hairpin: 6 / coil: 11
17	5XCO	2x disulfide bond / head to tail	22	helix: 12 / coil: 10
18	2LWV	side-chain to side-chain	13	helix: 3 / coil: 10
19	2AJW	ACE staple	17	helix: 4 / coil: 13
20	4KTS	side-chain to side-chain	14	coil: 14
21	3WMG	1x disulfide bond	23	coil: 23

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
22	4OS1	1x disulfide bond / head to tail	18	hairpin: 8 / coil: 10
23	2JRW	1x disulfide bond	16	hairpin: 8 / coil: 8
24	2LWT	1x disulfide bond / head to tail	15	hairpin: 4 / coil: 11
25	3OE0	0QZ staple / head to tail	6	coil: 7
26	2LWU	head to tail	6	coil: 6
27	4MQ9	head to tail	8	hairpin: 2 / coil: 6
28	2OX2	1x disulfide bond / head to tail	12	helix: 3 / coil: 9
29	3AVI	head to tail	5	coil: 5
30	2LWS	side-chain to side-chain	12	coil: 12
31	1L5G	side-chain to backbone	6	coil: 6
32	4KTU	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
33	6C1R	1x disulfide bond	16	hairpin: 6 / coil: 10
34	1JBL	head to tail	8	hairpin: 2 / coil: 6
35	3QN7	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
36	1EBP	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
37	3ZWZ	1x disulfide bond	8	helix: 3 / coil: 5
38	3AV9	head to tail	7	hairpin: 2 / coil: 5
39	2M2G	1x disulfide bond	9	helix: 3 / coil: 6
40	3P8F	1x disulfide bond	13	hairpin: 4 / coil: 9
41	4ZJX	head to tail	8	coil: 8
42	6AWM	1x disulfide bond	9	coil: 9
43	1NPO	2x disulfide bond	13	coil: 13
44	4Z0F	head to tail	10	hairpin: 4 / coil: 6
45	6AWK	1x disulfide bond	6	coil: 6
46	2MGO	2x disulfide bond	12	helix: 6 / coil: 6
47	4GLY	side-chain to side-chain	11	helix: 4 / coil: 7
48	1TK2	1x disulfide bond	13	hairpin: 4 / coil: 9
49	4JK5	BAL staple / 1x disulfide bond	11	coil: 12
50	1VWB	1x disulfide bond / head to tail	22	helix: 10 / coil: 12
51	2C9T	side-chain to backbone	11	coil: 11
52	4N7Y	head to tail	5	coil: 5
53	4K8Y	side-chain to backbone	8	coil: 8
54	4W4Z	1x disulfide bond	6	hairpin: 2 / coil: 4
55	5AFG	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
56	2N07	side-chain to side-chain	7	coil: 7
57	3PQZ	2x disulfide bond / head to tail	23	helix: 10 / coil: 13
58	5N99	ACE staple	16	hairpin: 4 / coil: 12

Continued on next page

Table A.1 – *Continued from previous page*

	PDB	Cyclization Type	A.A.	Secondary Structure
59	4I80	1x disulfide bond	8	coil: 8
60	3WNF	1x disulfide bond	13	hairpin: 6 / coil: 7
61	6PIP	1x disulfide bond	7	coil: 7
62	1FOZ	head to tail	6	coil: 6
63	2AK0	ACE staple	11	helix: 4 / coil: 7
64	5BXO	1x disulfide bond	6	coil: 6
65	5B4W	head to tail	10	coil: 10
66	1SLE	side-chain to backbone	16	hairpin: 8 / coil: 8
67	5EOC	1x disulfide bond / head to tail	12	hairpin: 2 / coil: 10
68	5LFF	1x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
69	3WNE	1x disulfide bond / head to tail	14	hairpin: 4 / coil: 10
70	5KGN	side-chain to backbone	5	coil: 5
71	1SLD	2x disulfide bond	26	helix:11/hairpin:6/coil:9
72	6TVJ	side-chain to backbone	16	hairpin: 8 / coil: 8
73	3ZXY	1x disulfide bond / head to tail	26	helix: 19 / coil: 7
74	4P4W	1x disulfide bond	10	hairpin: 6 / coil: 4
75	6BET	ACE staple	18	helix: 9 / coil: 9
76	5KX2	1x disulfide bond / head to tail	14	hairpin: 6 / coil: 8
77	6U7Q	side-chain to backbone	16	hairpin: 10 / coil: 6
78	4ZQV	1x disulfide bond / head to tail	22	helix: 14 / coil: 8
79	6YSA	head to tail	14	hairpin: 2 / coil: 12
80	1NBJ	1x disulfide bond / head to tail	14	hairpin: 4 / coil: 10
81	6TH7	side-chain to backbone	13	hairpin: 2 / coil: 11
82	3ZXX	side-chain to side-chain	29	helix: 21 / coil: 8
83	6YDX	head to tail	14	hairpin: 2 / coil: 12
84	5KWO	3x disulfide bond / head to tail	27	hairpin: 8 / coil: 19
85	4P4V	head to tail	14	hairpin: 4 / coil: 10
86	4PJ6	side-chain to side-chain	29	helix: 21 / coil: 8
87	6ZRC	head to tail	10	coil: 10
88	5KX0	1x disulfide bond	10	hairpin: 4 / coil: 6
89	1N0C	1x disulfide bond	10	coil: 10
90	5VA9	side-chain to side-chain	9	coil: 9
91	6U7S	1x disulfide bond	18	hairpin: 4 / coil: 14
92	4X0S	1x disulfide bond	7	helix: 3 / coil: 4
93	5KX1	head to tail	9	coil: 9
94	4P8Q	side-chain to side-chain	29	helix: 21 / coil: 8
95	5NF0	head to tail	7	hairpin: 2 / coil: 5

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
96	5U3A	1x disulfide bond / head to tail	11	hairpin: 6 / coil: 5
97	6U22	head to tail	14	hairpin: 4 / coil: 10
98	5UZW	side-chain to side-chain	10	coil: 10
99	6U7R	ESD staple / head to tail	8	coil: 8
100	8J3S	side-chain to side-chain	10	coil: 10
101	4ZQU	side-chain to side-chain	12	coil: 12
102	5N8W	head to tail	8	coil: 8
103	5TU5	1x disulfide bond / head to tail	16	coil: 16
104	5OTV	head to tail	14	hairpin: 8 / coil: 6
105	5N89	1x disulfide bond	28	helix: 21 / coil: 7
106	6U7U	head to tail	8	coil: 8
107	6L63	side-chain to side-chain	28	helix: 9 / coil: 19
108	2MSQ	side-chain to backbone	16	hairpin: 10 / coil: 6
109	5TU4	side-chain to backbone	13	hairpin: 8 / coil: 5
110	6U24	head to tail	8	coil: 8
111	5OTW	side-chain to backbone	16	hairpin: 10 / coil: 6
112	6BEQ	side-chain to side-chain	29	helix: 9 / coil: 20
113	6J32	side-chain to side-chain	16	helix: 9 / coil: 7
114	5NIJ	side-chain to backbone	16	hairpin: 10 / coil: 6
115	1N0D	side-chain to side-chain	14	helix: 11 / coil: 3
116	4X0W	side-chain to backbone	16	hairpin: 8 / coil: 8
117	6ZRD	head to tail	14	coil: 14
118	5OTT	side-chain to backbone	16	hairpin: 10 / coil: 6
119	7Y8X	side-chain to side-chain	20	helix: 18 / coil: 2
120	2KYA	side-chain to backbone	15	hairpin: 4 / coil: 11
121	1BMB	head to tail	8	coil: 8
122	6J31	2x disulfide bond / head to tail	28	hairpin: 14 / coil: 14
123	2M3N	staple / head to tail	2	coil: 5
124	7M7X	staple	8	coil: 8
125	6WI3	2x disulfide bond	17	hairpin: 14 / coil: 3
126	5N8T	side-chain to backbone	5	coil: 5
127	5OTU	head to tail	14	hairpin: 8 / coil: 6
128	5TU6	2x disulfide bond / head to tail	28	hairpin: 14 / coil: 14
129	6BES	2x disulfide bond	17	hairpin: 12 / coil: 5
130	6U7X	side-chain to side-chain	7	coil: 7
131	7R1W	HAO staple / HAO staple	10	hairpin: 2 / coil: 10
132	7RMQ	side-chain to backbone	15	hairpin: 2 / coil: 13

Continued on next page

Table A.1 – *Continued from previous page*

	PDB	Cyclization Type	A.A.	Secondary Structure
133	7Y8V	ACE staple	15	hairpin: 4 / coil: 11
134	7YF6	side-chain to side-chain	8	hairpin: 4 / coil: 4
135	7R1V	1x disulfide bond	10	coil: 10
136	4BFH	1x disulfide bond	10	hairpin: 6 / coil: 4
137	6VY8	1x disulfide bond	14	hairpin: 4 / coil: 10
138	7Y8U	2x disulfide bond / head to tail	28	hairpin: 16 / coil: 12
139	6OPJ	ACE staple	15	coil: 15
140	6XN9	side-chain to side-chain	7	coil: 7
141	7RMS	side-chain to side-chain	7	coil: 7
142	6WHN	side-chain to backbone	14	coil: 14
143	2BEY	2x disulfide bond / head to tail	28	hairpin: 8 / coil: 20
144	6D2U	head to tail	8	coil: 8
145	5OTX	1x disulfide bond / head to tail	26	helix: 20 / coil: 6
146	7RMR	ACE staple	11	coil: 11
147	6WHO	HAO staple	9	hairpin: 2 / coil: 8
148	2LJS	side-chain to backbone	16	hairpin: 10 / coil: 6
149	8CWX	ACE staple	10	coil: 10
150	4P4Z	head to tail	14	hairpin: 2 / coil: 12
151	6YCR	head to tail	7	coil: 7
152	7YA5	side-chain to side-chain	7	coil: 7
153	8EQI	side-chain to backbone	16	hairpin: 10 / coil: 6
154	5VF1	side-chain to side-chain	7	coil: 7
155	5EGF	side-chain to backbone	16	hairpin: 10 / coil: 6
156	1AY6	HAO staple	9	hairpin: 2 / coil: 8
157	8C17	side-chain to backbone	16	hairpin: 6 / coil: 10
158	4P4X	1x disulfide bond	10	coil: 10
159	7MX1	2x disulfide bond	26	helix:11/hairpin:8/coil:7
160	6AAW	side-chain to backbone	16	hairpin: 8 / coil: 8
161	6YW0	side-chain to backbone	16	hairpin: 10 / coil: 6
162	4P4Y	1x disulfide bond	10	coil: 10
163	2MN1	side-chain to side-chain	11	coil: 11
164	5UW3	side-chain to backbone	16	hairpin: 6 / coil: 10
165	5NEY	head to tail	8	coil: 8
166	2AOA	2x disulfide bond	27	helix:8/hairpin:10/coil:9
167	2M7J	side-chain to backbone	16	hairpin: 10 / coil: 6
168	4WC8	side-chain to backbone	5	coil: 5
169	6SBA	side-chain to backbone	8	coil: 8

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
170	8ALX	side-chain to backbone	5	coil: 5
171	6WSJ	side-chain to backbone	8	coil: 8
172	7L53	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
173	1TMB	ACE staple	14	hairpin: 6 / coil: 8
174	5N8B	1x disulfide bond	8	coil: 8
175	7S5G	side-chain to side-chain	11	coil: 11
176	6PI2	ACE staple	15	helix: 5 / coil: 10
177	5LRG	1x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
178	2M7I	side-chain to side-chain	9	coil: 9
179	2AOB	head to tail	14	hairpin: 2 / coil: 12
180	7L51	1x disulfide bond / head to tail	14	coil: 14
181	6PI3	side-chain to backbone	10	coil: 10
182	2MH1	0QZ staple / head to tail	6	coil: 7
183	5DIJ	side-chain to side-chain	28	helix: 3 / coil: 25
184	6YVX	BAL staple / 1x disulfide bond	10	coil: 12
185	7NRF	ACE staple	11	hairpin: 5 / coil: 6
186	5EJD	1x disulfide bond / head to tail	14	coil: 14
187	7YAA	1x disulfide bond	8	coil: 8
188	5WOV	1x disulfide bond	18	hairpin: 6 / coil: 12
189	5UW7	1x disulfide bond	24	helix: 11 / coil: 13
190	3NI3	ACE staple	12	helix: 5 / coil: 7
191	2MAU	side-chain to side-chain	13	coil: 13
192	5O4Y	0QZ staple / head to tail	6	coil: 7
193	5N8E	1x disulfide bond	9	coil: 9
194	6PV9	B3A BAL staple / head to tail	5	coil: 7
195	7Y99	head to tail	10	hairpin: 2 / coil: 8
196	5UW6	ACE staple	11	hairpin: 5 / coil: 6
197	7QUX	side-chain to side-chain	19	coil: 19
198	5WOW	0QZ staple / head to tail	6	coil: 7
199	6A8G	ACE staple	14	helix: 11 / coil: 3
200	1N09	BAL staple	10	coil: 11
201	1O8Y	ACE staple	15	helix: 10 / coil: 5
202	6YVZ	1x disulfide bond	8	coil: 8
203	7L54	side-chain to backbone	13	helix: 10 / coil: 3
204	5UW5	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
205	6LYC	BAL staple / head to tail	6	hairpin: 2 / coil: 5
206	6S35	1x disulfide bond	8	coil: 8

Continued on next page

Table A.1 – *Continued from previous page*

	PDB	Cyclization Type	A.A.	Secondary Structure
207	7NRE	1x disulfide bond	8	coil: 8
208	5O45	head to tail	14	hairpin: 10 / coil: 4
209	8CV7	ACE staple	12	hairpin: 2 / coil: 10
210	7L55	ACE staple	15	helix: 3 / coil: 12
211	6WHQ	1x disulfide bond	14	hairpin: 4 / coil: 10
212	5KWZ	48V staple	16	hairpin: 8 / coil: 8
213	7Y8D	1x disulfide bond	12	coil: 12
214	3VVR	head to tail	14	hairpin: 8 / coil: 6
215	6I7N	ABU staple / head to tail	4	coil: 4
216	2MSO	ACE staple	14	hairpin: 6 / coil: 8
217	3Q9J	1x disulfide bond	8	hairpin: 2 / coil: 6
218	5F1T	BAL staple / head to tail	6	hairpin: 2 / coil: 5
219	3VVS	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
220	7TB1	ACE staple	12	hairpin: 6 / coil: 6
221	6VXY	head to tail	6	coil: 6
222	6CN8	side-chain to side-chain	16	helix:3/hairpin:2/coil:11
223	6YVW	side-chain to backbone	11	coil: 11
224	7P1C	3x disulfide bond / head to tail	29	hairpin: 5 / coil: 24
225	5DLK	1x disulfide bond	8	coil: 8
226	5F1W	48V staple	16	hairpin: 8 / coil: 8
227	7NRI	1x disulfide bond / head to tail	18	hairpin: 12 / coil: 6
228	6CG3	head to tail	6	coil: 6
229	3Q9H	48V staple	13	coil: 13
230	5N8J	B3A staple / head to tail	6	coil: 7
231	6SRU	BAL staple	10	coil: 11
232	5N7X	2x disulfide bond / head to tail	22	helix: 13 / coil: 9
233	5TTY	ACE staple	14	coil: 14
234	5NES	side-chain to side-chain	21	coil: 21
235	7JQT	ACE staple	14	hairpin: 6 / coil: 8
236	4X1P	ACE staple	14	coil: 14
237	5KWP	head to tail	8	coil: 8
238	7JQS	ACE WCM staple	9	coil: 9
239	5JI4	ACE staple	10	coil: 10
240	6CG5	B3A staple / head to tail	6	coil: 7
241	7WMT	side-chain to backbone	7	coil: 7
242	7Y90	48V staple	13	coil: 13
243	6A8N	3x disulfide bond / head to tail	29	hairpin: 9 / coil: 20

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
244	2CV3	1x disulfide bond	16	hairpin: 6 / coil: 10
245	8AFI	1x disulfide bond / head to tail	14	hairpin: 4 / coil: 10
246	7JQR	side-chain to side-chain	11	helix: 4 / coil: 7
247	6WHZ	head to tail	8	coil: 8
248	6Z19	ACA staple / head to tail	6	coil: 6
249	5KVN	1x disulfide bond	14	hairpin: 4 / coil: 10
250	6CG4	48V staple	13	coil: 13
251	5LRK	NAG staple	0	NA: 5
252	8CUF	head to tail	8	helix: 3 / coil: 5
253	5LRJ	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
254	8CUG	head to tail	23	hairpin: 14 / coil: 9
255	5INZ	head to tail	18	hairpin: 6 / coil: 12
256	6U8G	48V staple	14	hairpin: 2 / coil: 12
257	8DEW	side-chain to side-chain	8	coil: 8
258	1VWP	3x disulfide bond / head to tail	29	hairpin: 9 / coil: 20
259	2RTN	head to tail	12	coil: 12
260	8Q1N	side-chain to side-chain	13	coil: 13
261	7E5E	3x disulfide bond / head to tail	18	hairpin: 10 / coil: 8
262	7DOC	head to tail	8	coil: 8
263	2M2S	1x disulfide bond	8	hairpin: 2 / coil: 6
264	5J31	1x disulfide bond	12	hairpin: 2 / coil: 10
265	6QZY	DAV staple / head to tail	6	coil: 6
266	1T9E	head to tail	23	hairpin: 14 / coil: 9
267	5GKM	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
268	6BVY	head to tail	14	hairpin: 8 / coil: 6
269	5VFW	side-chain to backbone	5	coil: 5
270	2RTO	head to tail	8	coil: 8
271	8DEV	side-chain to backbone	11	hairpin: 2 / coil: 9
272	4OIR	ACE staple	11	hairpin: 5 / coil: 6
273	1E68	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
274	1VWQ	ITZ staple / head to tail	5	coil: 6
275	5V1S	ACE staple	17	hairpin: 10 / coil: 7
276	6VJQ	3x disulfide bond / head to tail	29	hairpin: 9 / coil: 20
277	5JR2	1x disulfide bond	13	coil: 13
278	6U6K	48V staple	12	hairpin: 2 / coil: 10
279	6ANN	3x disulfide bond	18	hairpin: 8 / coil: 10
280	6BVX	2x disulfide bond	11	coil: 11

Continued on next page

Table A.1 – *Continued from previous page*

	PDB	Cyclization Type	A.A.	Secondary Structure
281	2M6C	48V staple	13	hairpin: 8 / coil: 5
282	6H41	1x disulfide bond	8	helix: 5 / coil: 3
283	3PP4	3x disulfide bond / head to tail	29	hairpin: 10 / coil: 19
284	6ULT	1x disulfide bond	14	coil: 14
285	7ROV	side-chain to side-chain	10	helix: 5 / coil: 5
286	7UOA	3x disulfide bond / head to tail	29	hairpin: 6 / coil: 23
287	2RTL	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
288	1VWR	1x disulfide bond	18	coil: 18
289	4OIQ	1x disulfide bond	7	coil: 7
290	8DEU	2x disulfide bond / head to tail	29	hairpin: 10 / coil: 19
291	8EHV	ACE staple	14	hairpin: 4 / coil: 10
292	1VSC	side-chain to backbone	11	coil: 11
293	4M6E	Y1V staple	15	hairpin: 10 / coil: 6
294	6U4A	ACE staple	12	helix: 9 / coil: 3
295	6VLJ	48V staple	13	hairpin: 8 / coil: 5
296	6QZZ	side-chain to side-chain	20	coil: 20
297	5V1Q	BAL staple	10	coil: 11
298	4OIP	ACE staple	12	helix: 9 / coil: 3
299	2RTM	1x disulfide bond / head to tail	14	hairpin: 2 / coil: 12
300	2MSR	head to tail	12	coil: 12
301	6ULV	1x disulfide bond	11	hairpin: 2 / coil: 9
302	5OL1	1x disulfide bond / head to tail	14	coil: 14
303	6TSC	48V staple	13	coil: 13
304	5W6I	ACE staple	15	helix: 6 / coil: 9
305	6ULQ	head to tail	8	coil: 8
306	1QFB	1x disulfide bond	6	coil: 6
307	3OE8	1ZN staple	6	coil: 7
308	2M6F	head to tail	16	hairpin: 4 / coil: 12
309	5OL6	head to tail	6	coil: 6
310	6MJF	1x disulfide bond	18	coil: 18
311	5TWW	1x disulfide bond	13	coil: 13
312	2M77	1x disulfide bond	12	coil: 12
313	5UZT	side-chain to side-chain	16	coil: 16
314	6DN7	head to tail	6	coil: 6
315	2RTJ	head to tail	7	coil: 7
316	6MJG	head to tail	14	hairpin: 4 / coil: 10
317	3OE9	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
318	2M6G	head to tail	8	coil: 8
319	5OL7	3x disulfide bond / head to tail	29	hairpin: 10 / coil: 19
320	1D7T	BE2 staple / head to tail	9	coil: 10
321	6DZ9	head to tail	11	coil: 11
322	6ULP	3x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
323	7OS8	head to tail	6	coil: 7
324	4UVQ	head to tail	16	hairpin: 6 / coil: 10
325	8DNQ	side-chain to side-chain	17	coil: 17
326	8AC7	head to tail	8	coil: 8
327	4M1D	AEA staple	6	coil: 6
328	2RTK	side-chain to side-chain	16	coil: 16
329	6XS7	ACE staple / II7 staple	12	coil: 13
330	4W50	head to tail	8	coil: 8
331	5J0M	BE2 staple / head to tail	9	coil: 10
332	6DN6	head to tail	16	hairpin: 6 / coil: 10
333	5OBT	3x disulfide bond / head to tail	29	hairpin: 2 / coil: 27
334	5ISX	head to tail	7	coil: 7
335	6U6L	head to tail	6	coil: 6
336	1STS	side-chain to side-chain	13	helix: 4 / coil: 9
337	2M6D	side-chain to side-chain	9	hairpin: 4 / coil: 5
338	1Y5C	1x disulfide bond / head to tail	17	hairpin: 6 / coil: 11
339	2RTH	head to tail	8	coil: 8
340	6DN5	2x disulfide bond	11	coil: 11
341	6PIO	ITZ staple / head to tail	5	coil: 6
342	8F1F	head to tail	14	hairpin: 8 / coil: 6
343	6AZF	head to tail	16	hairpin: 6 / coil: 10
344	5V1T	head to tail	7	coil: 7
345	4HY2	1x disulfide bond	8	coil: 8
346	1FQ8	head to tail	14	hairpin: 4 / coil: 10
347	7YFS	side-chain to backbone	16	hairpin: 9 / coil: 7
348	7YUZ	side-chain to side-chain	7	coil: 7
349	1E4X	head to tail	6	coil: 7
350	1KAL	1x disulfide bond	8	coil: 8
351	1STR	3x disulfide bond / head to tail	29	hairpin: 9 / coil: 20
352	5OL5	head to tail	14	hairpin: 8 / coil: 6
353	2M6E	1x disulfide bond	12	coil: 12
354	6XS5	head to tail	6	coil: 6

Continued on next page

Table A.1 – *Continued from previous page*

	PDB	Cyclization Type	A.A.	Secondary Structure
355	2RTI	1x disulfide bond	8	coil: 8
356	2KX5	head to tail	8	coil: 8
357	4GUX	head to tail	22	helix: 12 / coil: 10
358	6AZG	head to tail	8	hairpin: 2 / coil: 6
359	6YW3	head to tail	6	coil: 6
360	3F58	3x disulfide bond / head to tail	18	hairpin: 10 / coil: 8
361	5ISW	head to tail	6	coil: 6
362	7K2S	ACE staple	11	hairpin: 5 / coil: 6
363	5W5S	side-chain to backbone	11	coil: 11
364	4TTL	1x disulfide bond	6	coil: 6
365	4MNY	head to tail	9	coil: 9
366	6UXS	side-chain to side-chain	21	helix: 13 / coil: 8
367	7JVF	BAL staple / head to tail	6	coil: 7
368	6U72	1x disulfide bond	16	hairpin: 6 / coil: 10
369	7BPH	B3S staple / 1x disulfide bond	12	coil: 13
370	2RTG	1x disulfide bond	6	coil: 6
371	1AMU	side-chain to backbone	5	coil: 5
372	6DKZ	2x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
373	7KFA	1x disulfide bond	8	coil: 8
374	7KEV	head to tail	14	hairpin: 2 / coil: 12
375	1FQ7	1x disulfide bond	7	coil: 7
376	6UFA	1x disulfide bond	12	hairpin: 6 / coil: 6
377	7K2R	head to tail	9	coil: 9
378	1E4W	ACE staple	13	helix: 4 / coil: 9
379	6YW2	V7M staple / CYS staple	9	coil: 9
380	2RTF	side-chain to side-chain	16	coil: 16
381	4MNX	1x disulfide bond / head to tail	16	coil: 16
382	2F58	MPT staple / DHL staple	8	coil: 8
383	4TTM	1x disulfide bond / head to tail	16	hairpin: 6 / coil: 10
384	2LWQ	head to tail	8	hairpin: 2 / coil: 6
385	1FQ6	ACE staple	14	helix: 3 / coil: 11
386	5JZA	1x disulfide bond / head to tail	18	hairpin: 2 / coil: 16
387	6NOX	1x disulfide bond / head to tail	18	hairpin: 14 / coil: 4
388	3ZL4	head to tail	10	coil: 10
389	5JM4	head to tail	14	hairpin: 8 / coil: 6
390	6DL0	BAL staple	10	coil: 11
391	7K2Q	1x disulfide bond / head to tail	13	coil: 13

Continued on next page

Table A.1 – Continued from previous page

	PDB	Cyclization Type	A.A.	Secondary Structure
392	1JBN	head to tail	8	hairpin: 2 / coil: 6
393	6YW1	ITZ staple / head to tail	5	coil: 6
394	8AC9	9WV staple	10	coil: 11
395	1FQ5	ACE staple	14	coil: 14
396	6XCI	1x disulfide bond / head to tail	16	hairpin: 8 / coil: 8
397	2M78	side-chain to backbone	5	coil: 5
398	8BI9	head to tail	22	helix: 12 / coil: 10
399	6HLB	1x disulfide bond	14	helix: 4 / coil: 10
400	2M2X	1x disulfide bond	6	coil: 6
401	2RTE	head to tail	9	coil: 9
402	6XS9	1x disulfide bond	12	hairpin: 6 / coil: 6
403	6DN8	head to tail	9	coil: 9
404	4TTN	48V staple	10	coil: 10
405	6UG3	head to tail	9	coil: 9
406	1JD8	head to tail	11	coil: 11
407	2LZI	1x disulfide bond	18	coil: 18
408	6DL1	1x disulfide bond	8	coil: 8
409	3OE6	head to tail	8	coil: 8
410	6NDY	9WV staple	10	coil: 11
411	2M6H	ACY staple	14	coil: 14
412	4HGC	1x disulfide bond	17	hairpin: 6 / coil: 11
413	7K2P	head to tail	22	helix: 11 / coil: 11
414	8BI8	head to tail	8	coil: 8
415	2M79	1x disulfide bond	11	helix: 6 / coil: 5
416	5UZZ	B74 staple / CYS staple	9	coil: 9
417	1FQ4	1x disulfide bond	13	coil: 13
418	1B6J	head to tail	7	coil: 7
419	6DKY	1x disulfide bond	11	coil: 11
420	6UG2	side-chain to backbone	6	coil: 6
421	6U71	head to tail	10	helix: 3 / coil: 7
422	1SFI	head to tail	7	coil: 7
423	6HZD	head to tail	8	coil: 8
424	7L9D	3x disulfide bond / head to tail	18	hairpin: 2 / coil: 16
425	6U8M	0QZ staple / head to tail	6	coil: 7
426	4TTO	1x disulfide bond	6	coil: 6
427	1A1P	head to tail	8	coil: 8
428	2RTD	48V staple	15	hairpin: 4 / coil: 11

Continued on next page

	PDB	Cyclization Type	A.A.	Secondary Structure
429	6XS8	1x disulfide bond	11	helix:3/hairpin:2/coil:6
430	2M2Y	head to tail	7	coil: 7
431	6HZC	48V staple	6	coil: 6
432	2RTC	head to tail	8	coil: 8

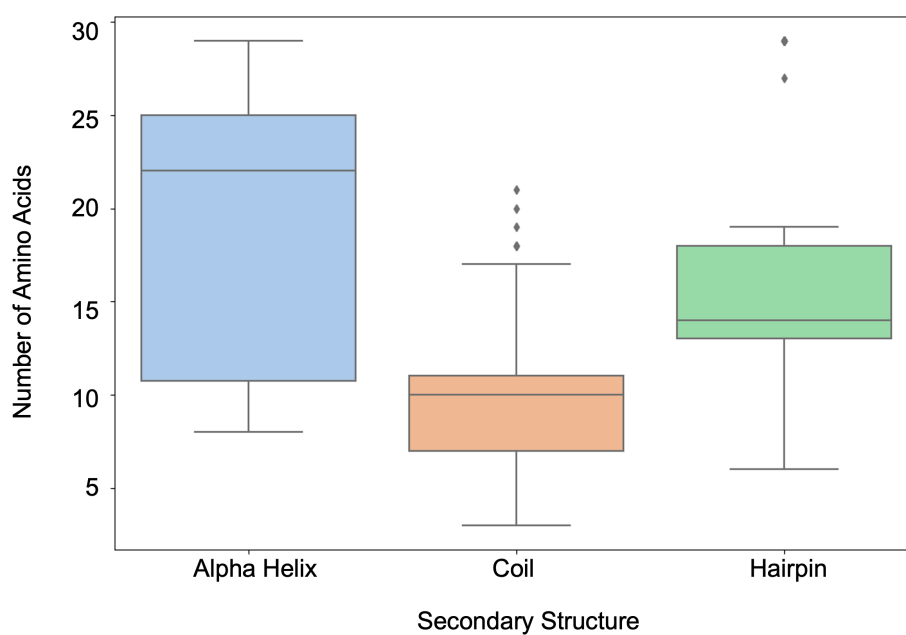


Figure A.1.: Range of Amino Acid Lengths for Alpha Helices, Coils, and Hairpin Secondary Structures in the cPEPmatch Database Cyclic Peptides.

B. Supplementary Information for Chapter 4: cPEPmatch

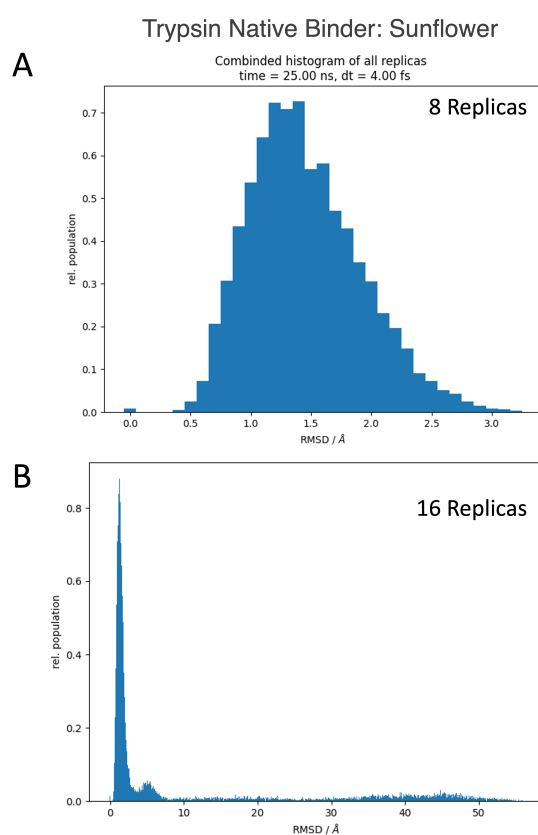


Figure B.1.: Combined Ligand RMSD Distributions for the Sunflower Trypsin inhibitor across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

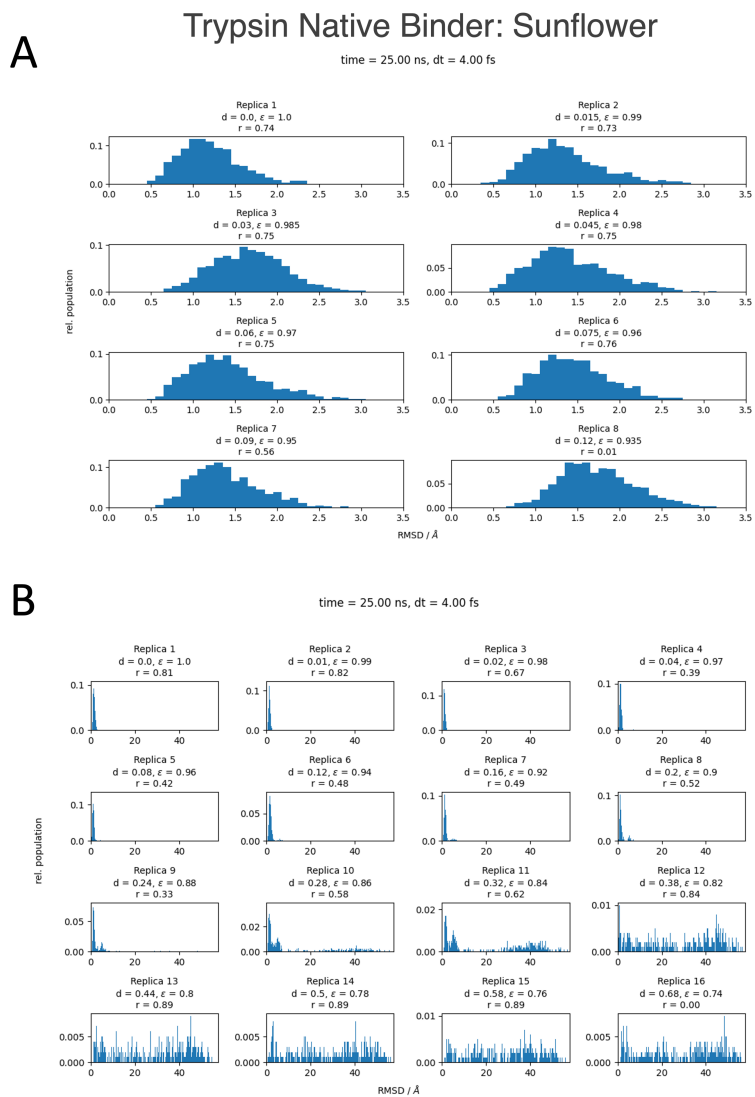


Figure B.2.: Ligand RMSD histograms per replica for the Sunflower Trypsin inhibitor across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

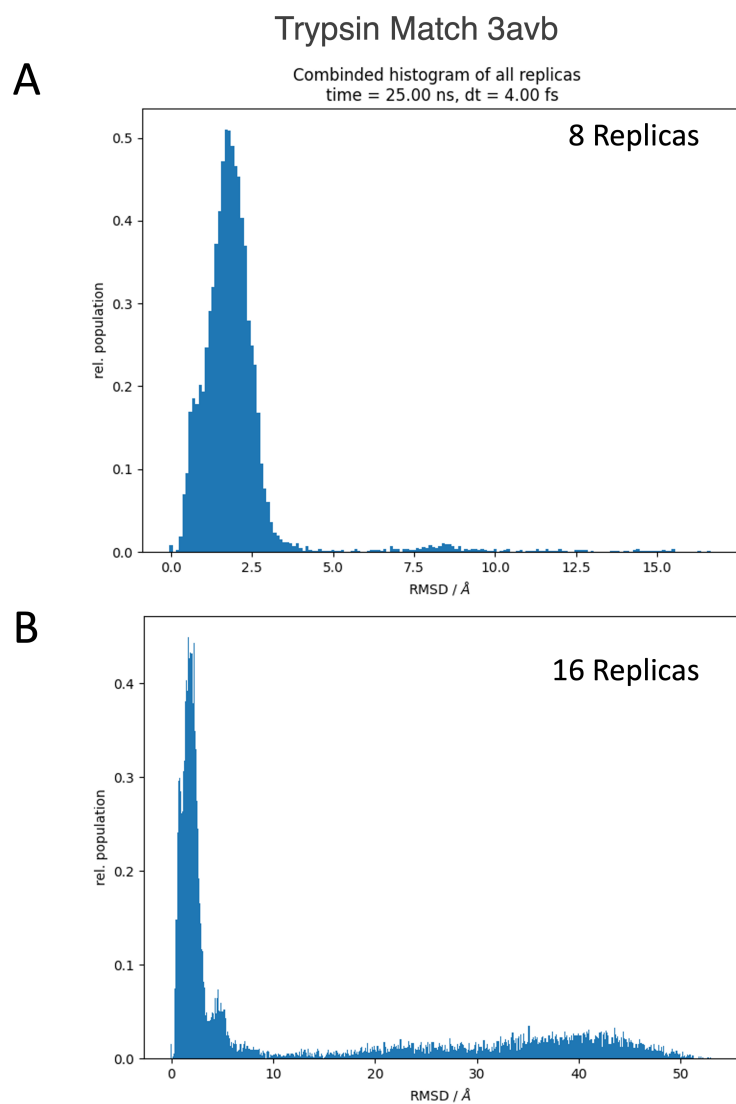


Figure B.3.: Combined Ligand RMSD Distributions for Trypsin Match 3avb across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

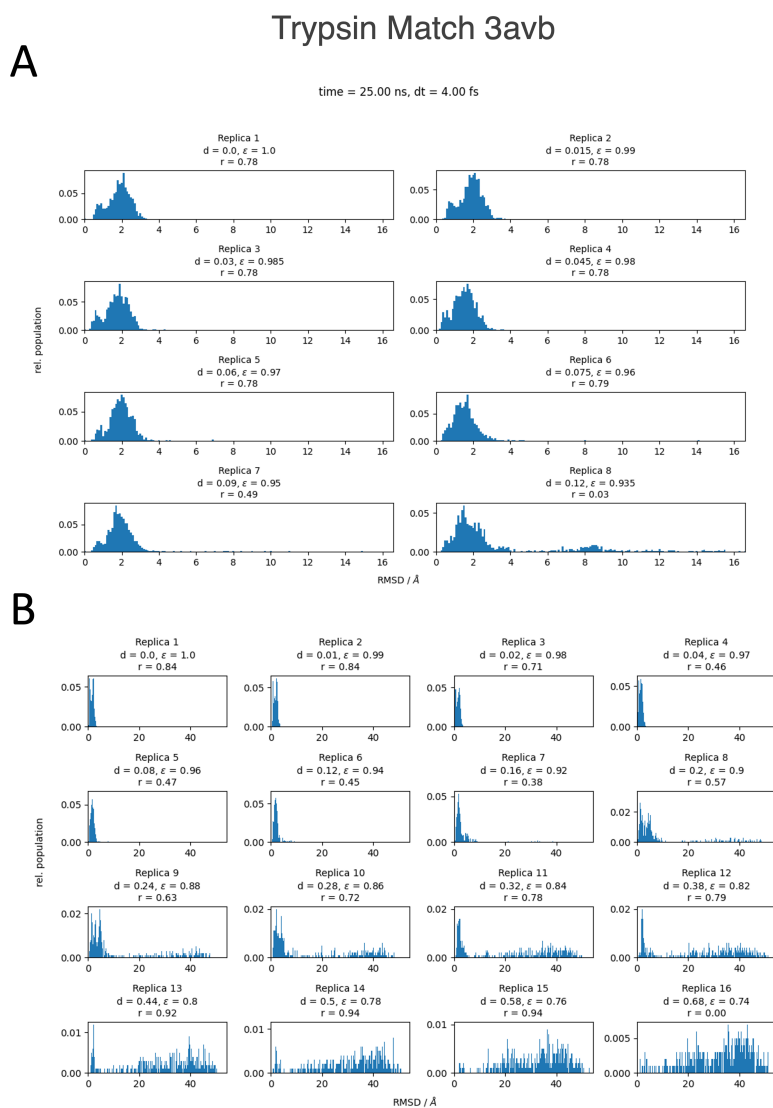


Figure B.4.: Ligand RMSD histograms per replica for Trypsin Match 3avb across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

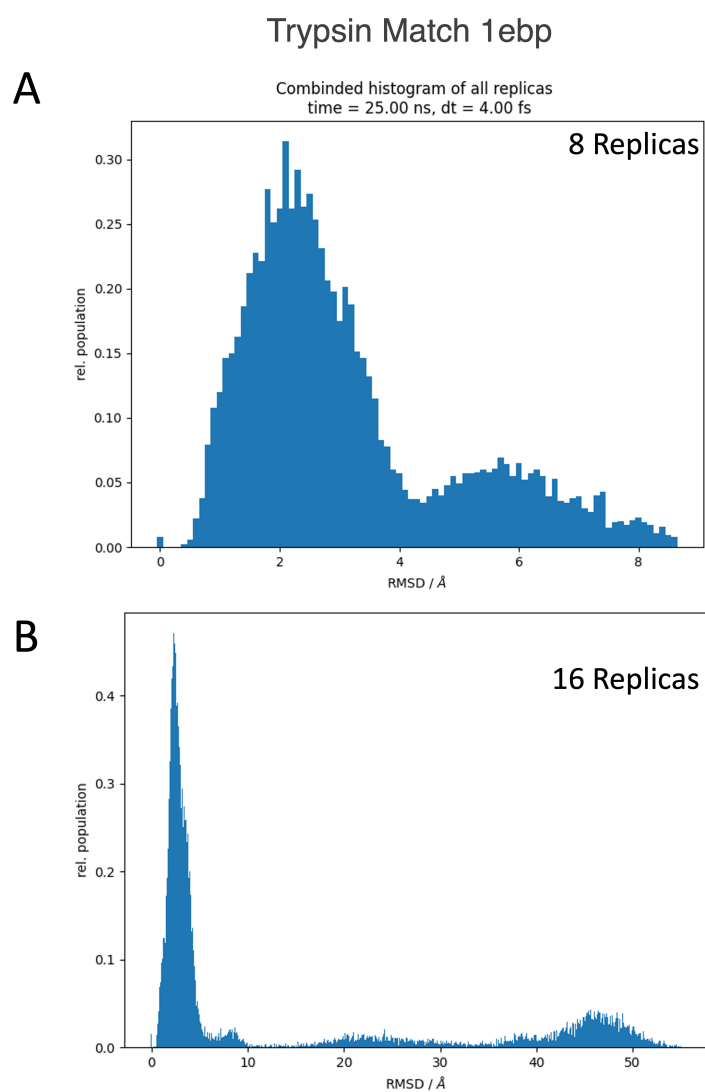


Figure B.5.: Combined Ligand RMSD Distributions for Trypsin Match 1ebp across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

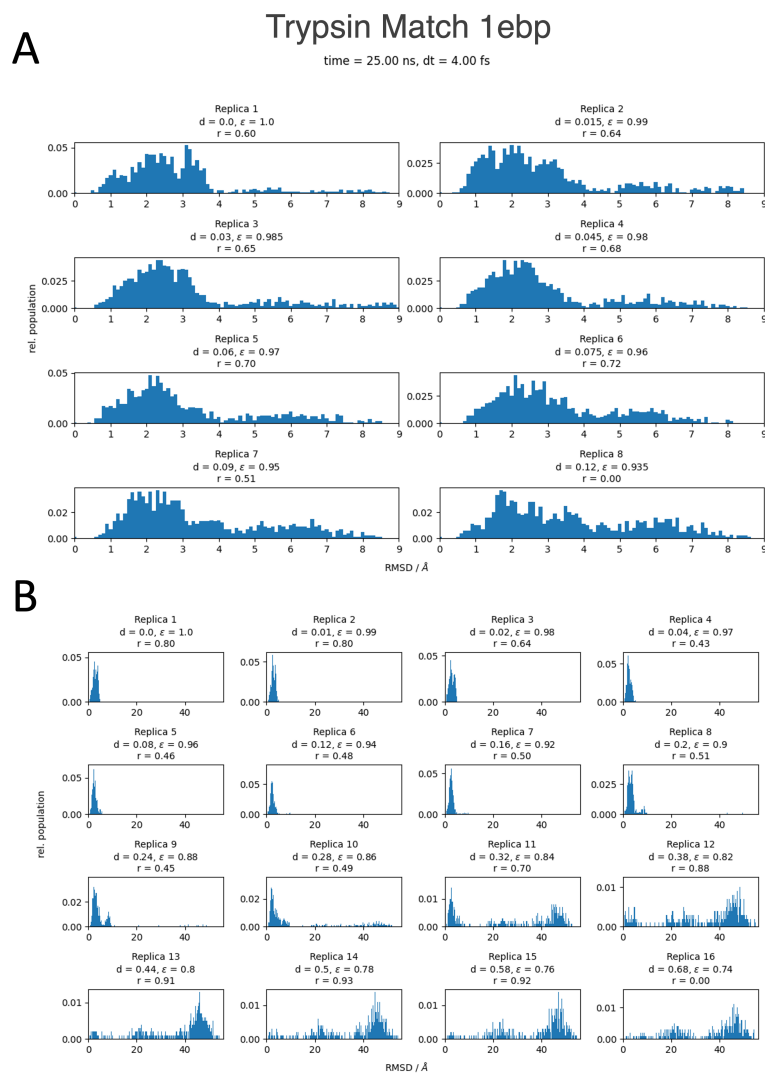
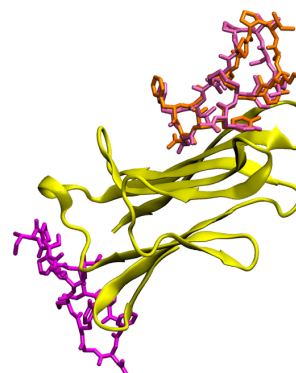
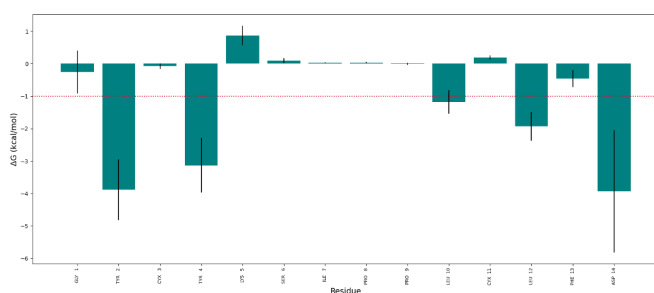


Figure B.6.: Ligand RMSD histograms per replica for Trypsin Match 1ebp across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

C. Supplementary Information for Chapter 7: Design of ICOS-ICOSL Cyclic Peptide Modulators

A) Match 14 1SFI

GLY-TYR-CYX-TYR-LYS-SER-ILE-PRO-PRO-LEU-CYX-LEU-PHE-ASP



B) Match 45 2NDL

GLY-TYR-CYX-TYR-PRO-MET-GLY-PRO-TRP-GLY-PRO-LEU-CYX-LEU-PRO-PHE

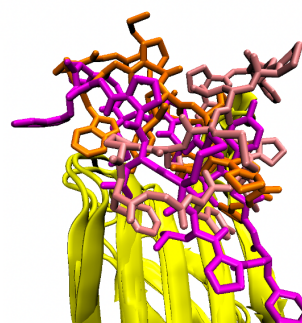
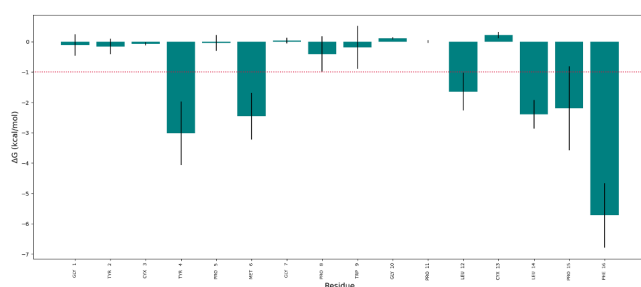


Figure C.1.: MM/GBSA per-residue free energy contribution and molecular conformations highlighting the structural orientation and interactions within the binding sites of Match 14 1SFL (A) and Match 45 2NDL (B).

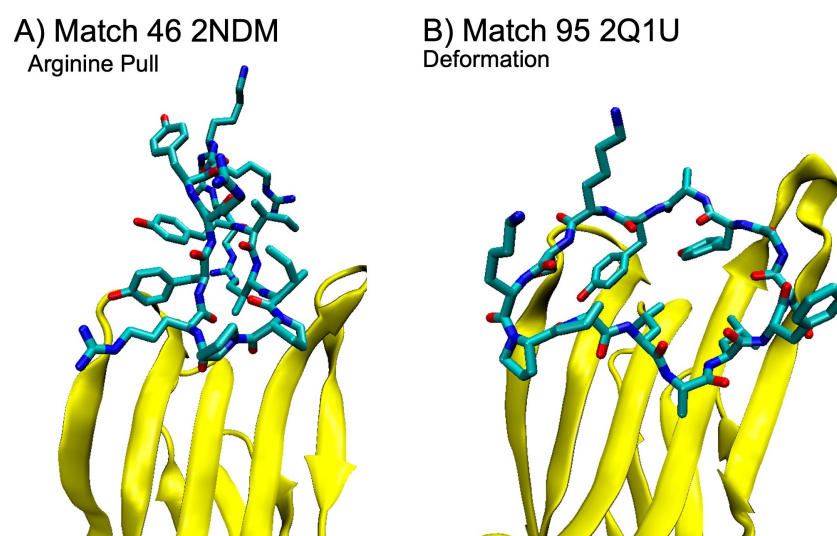


Figure C.2.: Depiction of structural instabilities in cyclic peptide interactions for Match 46 2NDM (A) showing an arginine pull effect, and Match 95 2Q1U (B) demonstrating deformation in the native structure.

Table C.1.: In vitro pharmacokinetic profile of Match 41.

Test	Match 41
Stability in simulated fluids ($t_{1/2}$, h)	
Gastric (pH 1.6)	6.5
Intestinal (pH 6.5)	20.4
Stability in human plasma (% remaining at 2 h)	98.2
LogD _{7.4} ^a	2.94
CACO cell permeability ^b	
P _{app} , A → B (10 ⁻⁶ cm/s)	6.72
P _{app} , B → A (10 ⁻⁶ cm/s)	10.8
Stability in rat liver microsomes ($t_{1/2}$, h)	18.3

^aDetermined by miniaturized shake-flask procedure at pH7.4.

^bPermeation through a Caco-2 cell monolayer was assessed in the absorptive apical to basolateral (a-b) and secretory basolateral to apical (b-a).

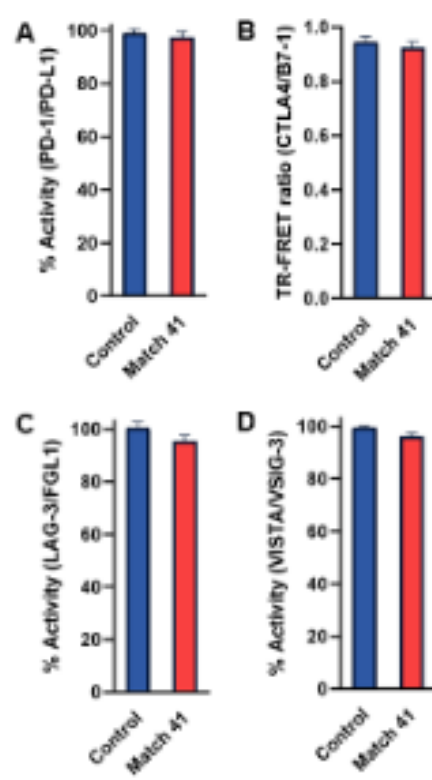


Figure C.3.: Assessment of the inhibitory effect of Match 41 using TR-FRET assays on (A) PD-1/PD-L1, (B) CTLA4/B-7, (C) LAG-3/FGL-1, and (D) VISTA/VSIG-3 interactions. All assays were performed in the presence of vehicle control or 50 μ M of Match 41. Error bars represent standard deviation (n = 3)

D. Supplementary Information for Chapter 8: cPEPmatch Code and Webserver

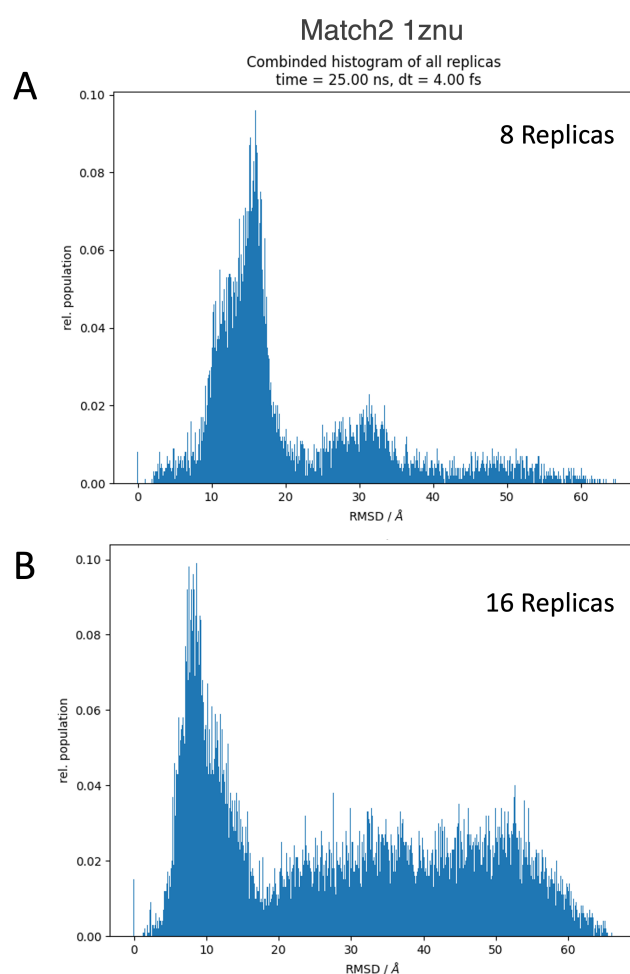


Figure D.1.: Combined Ligand RMSD Distributions for Match 2 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

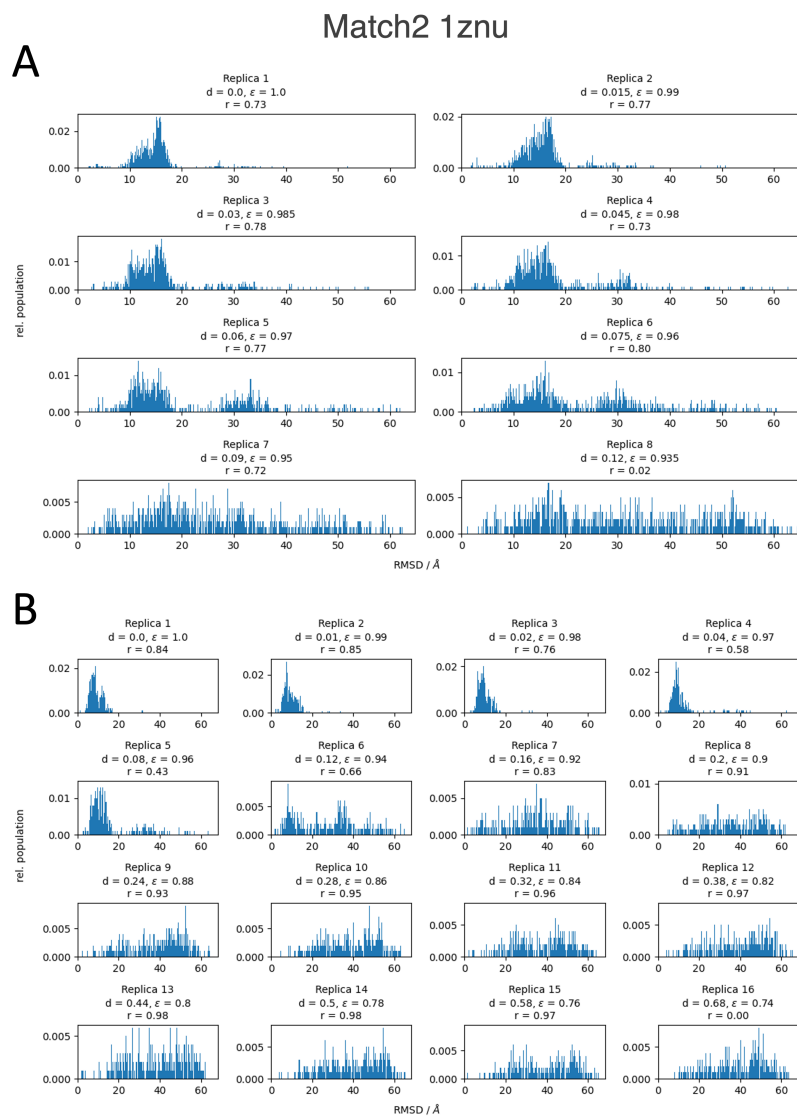


Figure D.2.: Ligand RMSD histograms per replica for Match 2 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

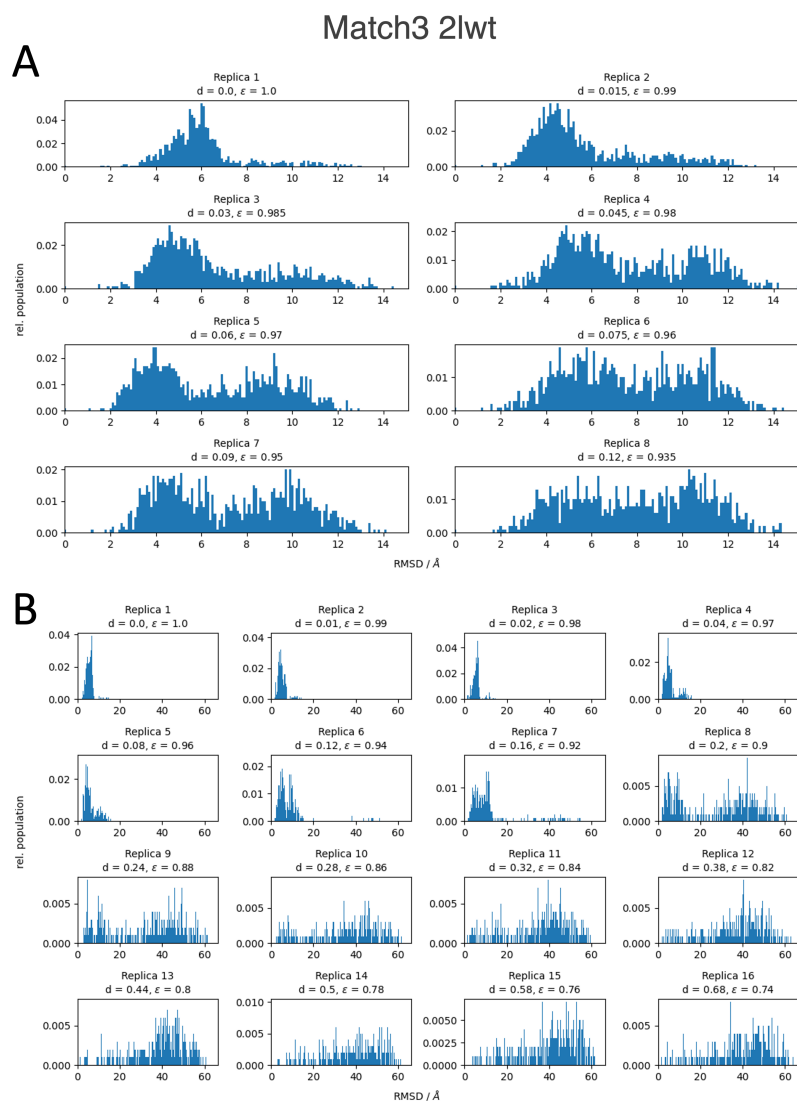


Figure D.3.: Ligand RMSD histograms per replica for Match 3 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

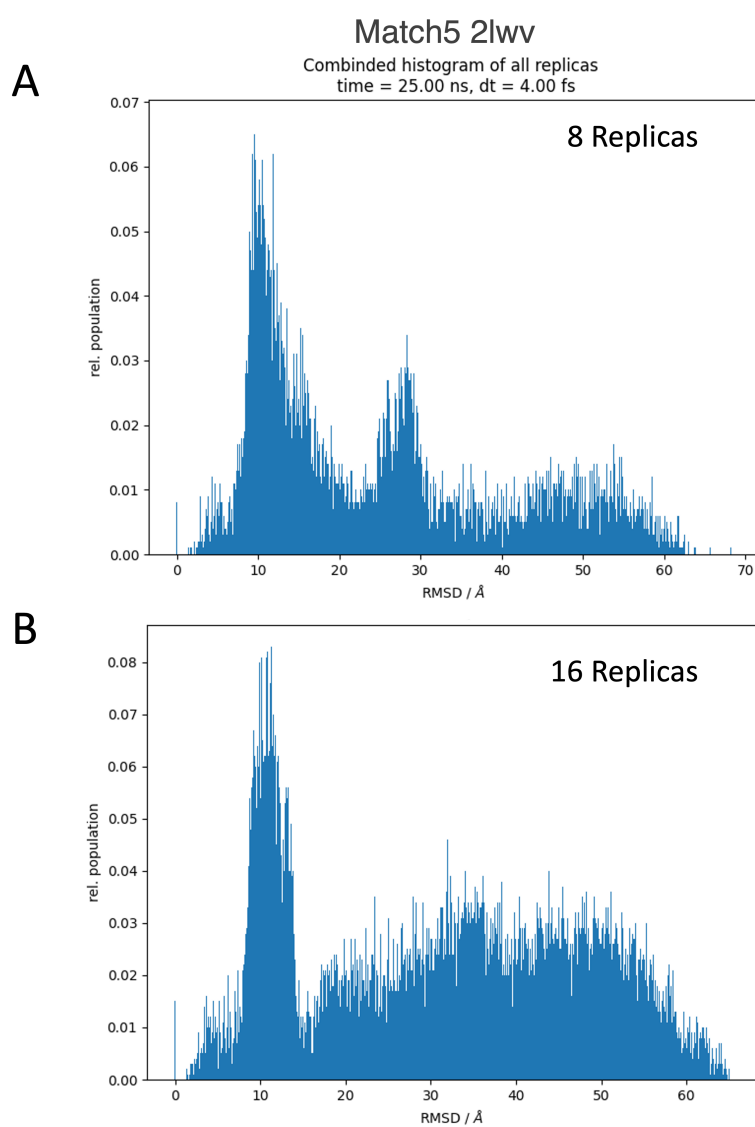


Figure D.4.: Combined Ligand RMSD Distributions for Match 5 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

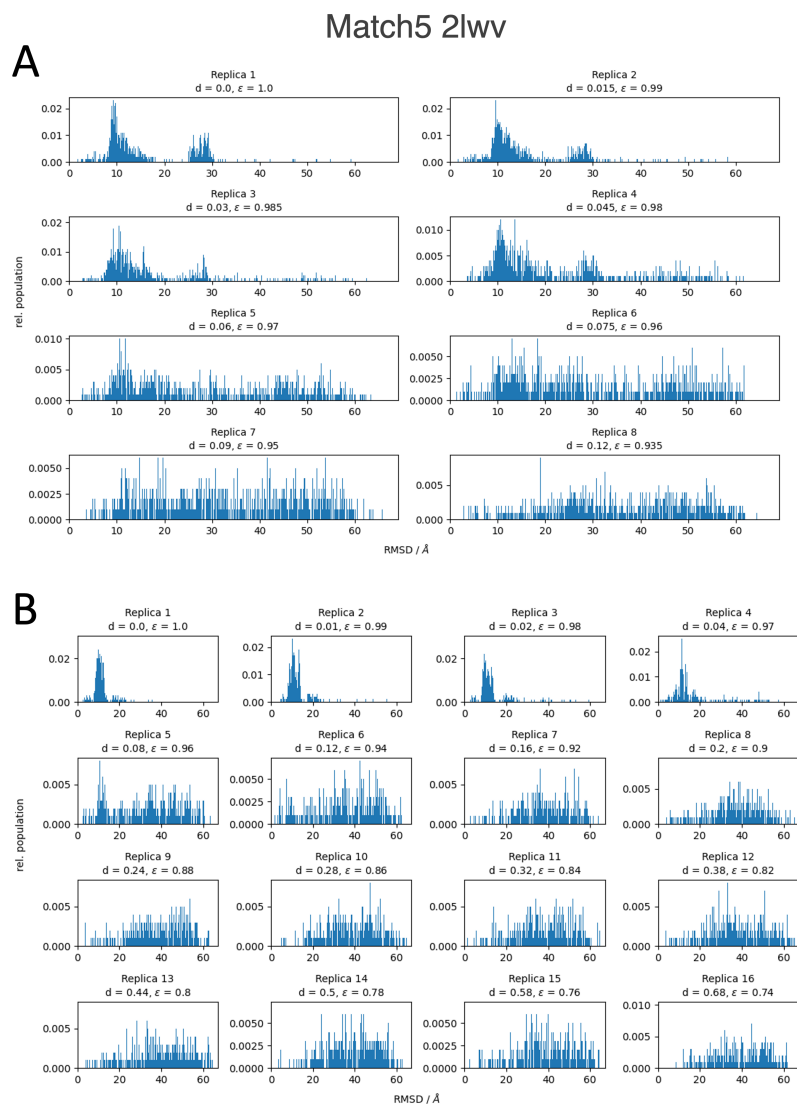


Figure D.5.: Ligand RMSD histograms per replica for Match 5 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

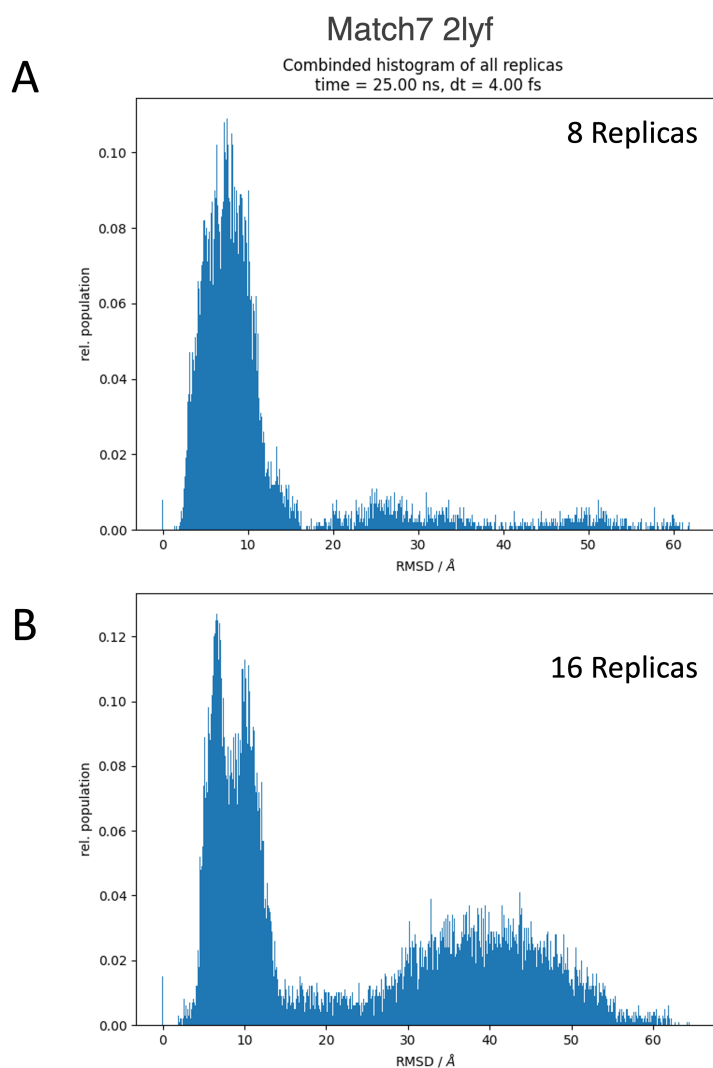


Figure D.6.: Combined Ligand RMSD Distributions for Match 7 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

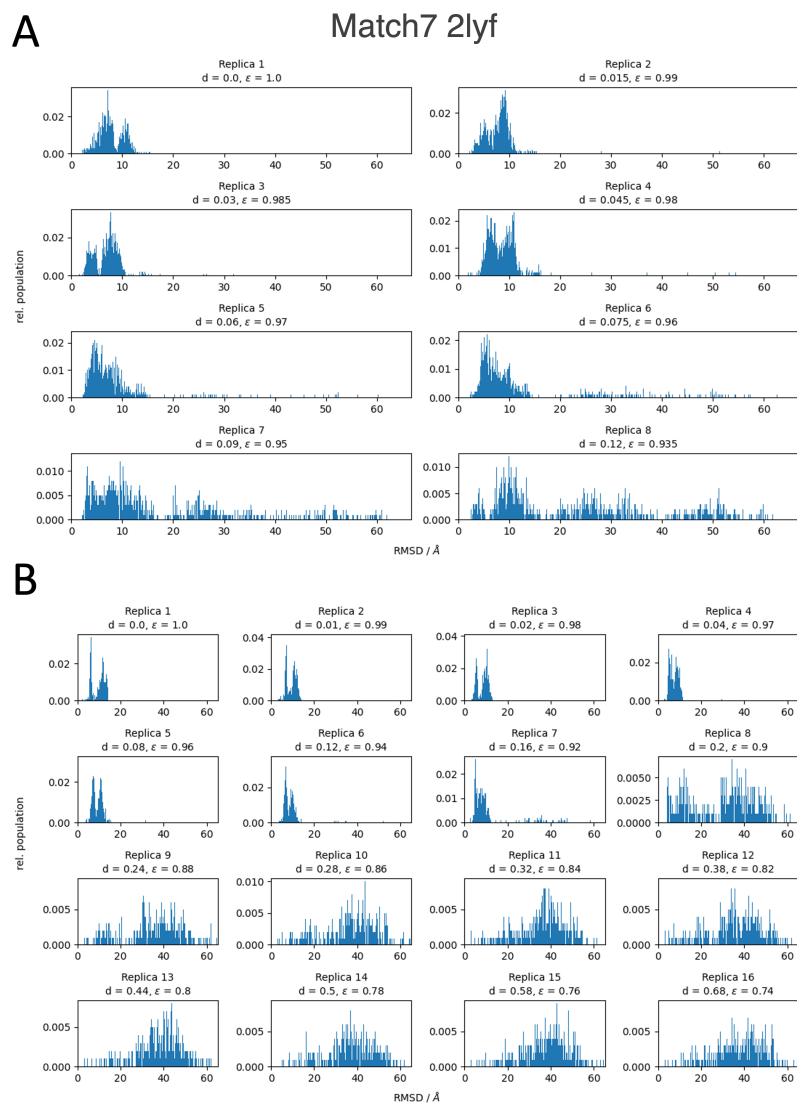


Figure D.7.: Ligand RMSD histograms per replica for Match 7 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

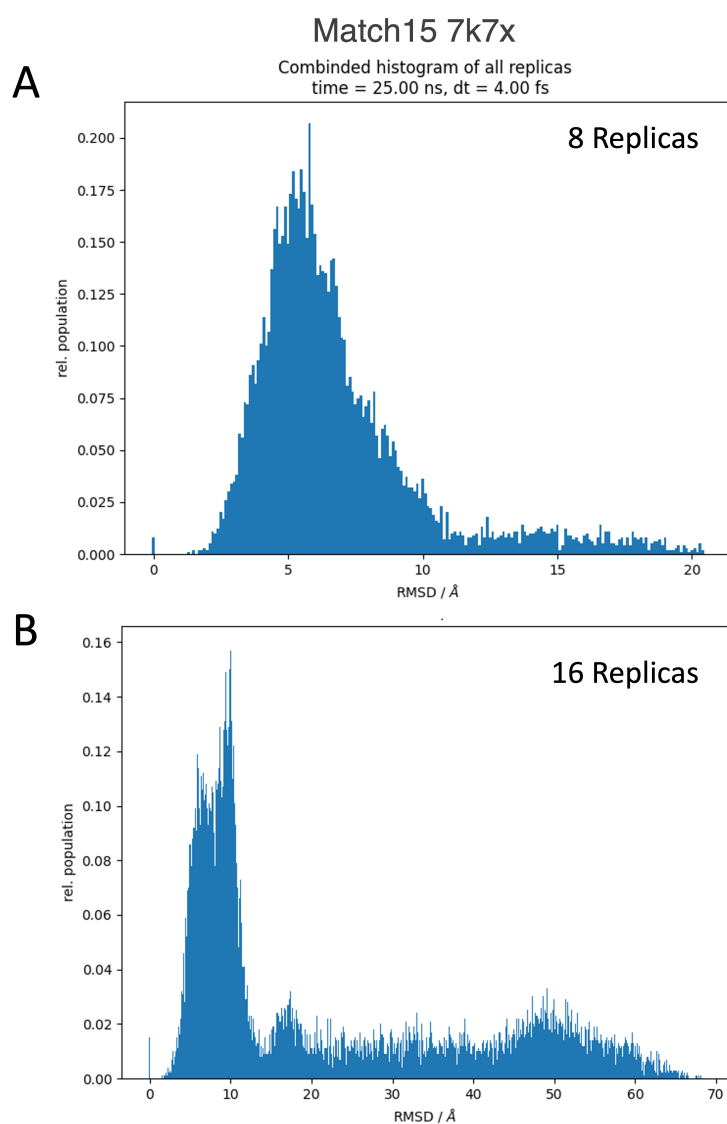


Figure D.8.: Combined Ligand RMSD Distributions for Match 15 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

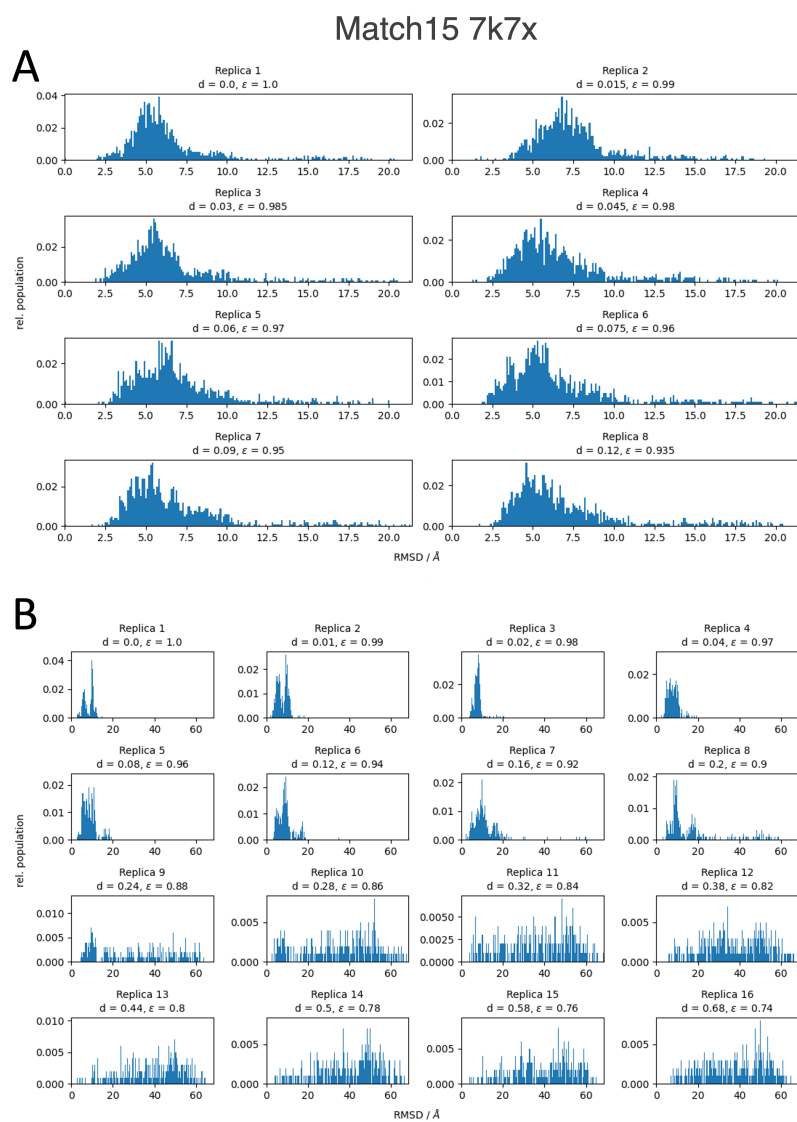


Figure D.9.: Ligand RMSD histograms per replica for Match 15 across 25 ns for (A) 8 replicas and (B) 16 replicas of RS-HREMD simulations.

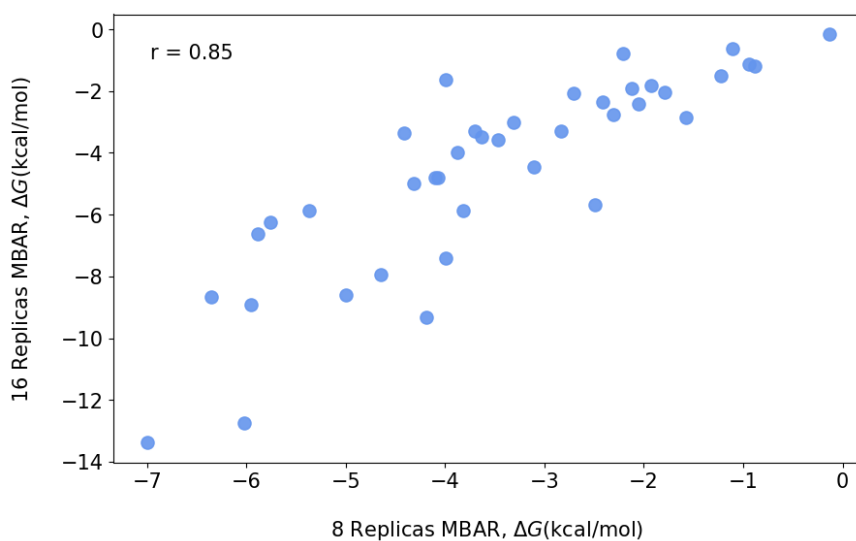


Figure D.10.: Correlation of RS-REMD MBAR Free Energy Calculations. Comparison of 25 ns replica exchange simulations with repulsive scaling Hamiltonian using 8 vs. 16 replicas of RS-HREMD simulations.

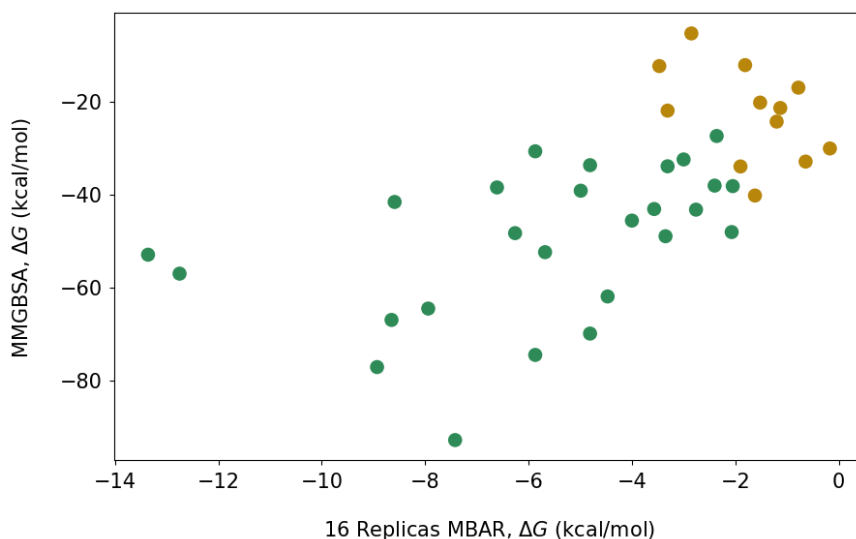


Figure D.11.: Selection criteria based on RS-REMD MBAR (x-axis) and MM-GBSA from regular MD (y-axis). Stable matches are identified where MBAR $\Delta G < -2$ kcal/mol and MM-GBSA $\Delta G < -25$ kcal/mol (green), with all other points considered non-matches (orange).

Acknowledgements

I'd like to start by thanking TUM for opening the doors to the most wonderful learning experience of my life. For being a hub filled with brilliant minds—people who are not only incredibly smart, but also interesting, enriching, and kind—at every level, in every hallway. To the Max Planck School Matter to Life—for teaching me what true passion and excitement for science really are. This community spills positivity and curiosity like nothing else I've ever known.

To my supervisor, Martin—always an open door and a ready smile. Thank you for being the greatest supervisor one could ask for. If I had to do my PhD all over again, I'd choose your guidance every time. I never left your office without a better mood, a positive outlook, and the feeling that I could do it all. I will never be able to thank you enough.

To my dearest T38—I cannot even describe what I feel for you all as a whole. You've changed me, shaped me, helped me grow—not just professionally, but as a person. These past years have been the most beautiful period of my life, and that is entirely because of you. I value you all, deeply, as friends. To Korbi, Asmar, Till, Max, Maria, Julian, and Paul—thank you for welcoming me into the group, easing my way into Germany, and showing me how much fun this journey could be. Patrick, for all the honest conversations. Simone, for your openness. Richard, for being someone I could always count on—and always the first to help me solve my research doubts. Luis, for the laughs, deep conversations, and the best hugs. Shu-yu, my man—how much can one person teach another? Turns out, infinitely. Thank you for all the life lessons. Carlitos—a little piece of home. The world needs more of your kindness and brilliance. You were always the calm I needed on a rough day.

To my favorite girl duo, Asha and Alice. Alice, my soulmate—we'll forever share a brain. I never imagined I'd meet a soulmate in my scientific journey, but then you walked in and lit up the whole place. Asha, my brat—your friendship means the world to me. You were the glue that held us all together. I expect great things from you, and I hope I get to witness them all.

Anna and Isabel—for being the most unwavering voices of encouragement a friend could ask for. You probably never noticed it, but whenever I doubt myself, it takes just two minutes with either of you to bring me back—chin up, back straight, and ready again. To the friends who were once my home and now live across different oceans—Cesar, Osu, Jurado, Cami, Vázquez, and Cubito—your support still reaches me. Even when daily conversations fade, I know you’re family, and always will be.

To my partner, Dennis—you walked into this wild chapter right at the end, and oh, how I needed your calm and your care. You were there with a coffee for the long nights, a hug for the tangled thoughts. And of course—for listening to my thesis defense presentation more times than anyone ever should. Have you heard the phrase “Protein-protein interactions are the heavy lifters of the cell”? Dennis has. Just about a million times. You are my rock. I love you.

To my entire family—thank you for always believing in me, standing by me, and lifting me up from thousands of kilometers away. I carry your spirit with me, always. To have a whole squad of cousins, aunts, uncles—and to have had the most amazing grandparents—supporting me in the way only a Mexican family can, has been one of the greatest wins of my life.

To my sister, Stefanie—my best friend, my strongest and forever support. The one I laugh with when life gets hard—and it does—but we joke it off together and keep going. I could never imagine moving forward without you. I’m stronger every day because of you. And you gave me the greatest joy of my life: Frida and Crucito. Everything I do, I do with them in mind. You three are my happiness.

And finally, a mis padres, Dina y Norberto. This work belongs to you too. To be the first academic in our family, across all the generations that came before, was only possible because you helped me break free. You gave everything for us. You moved us from place to place in search of a better future. You made sure that, deep down, I never doubted I could become anything I dreamed of. And even if science was something completely new to you, before anyone else did—you believed in me. To you, I owe everything I am. Gracias. Los amo.

List of Publications

1. Rapid in silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces, **BL Santini**, M Zacharias, *Frontiers in chemistry* 8, 573259 20 (2020)
2. Rapid Rational Design of Cyclic Peptides Mimicking Protein-Protein Interfaces, **BL Santini**, M Zacharias, *Computational Peptide Science: Methods and Protocols*, 231-244 (2022)
3. Design of cyclic peptides as novel inhibitors of ICOS/ICOSL interaction, SA Abdel-Rahman, **BL Santini** L Calvo-Barreiro, M Zacharias, M Gabr, *Bioorganic and Medicinal Chemistry Letters* 99, 129599 (2024)
4. Rational Design of Glycosaminoglycan Binding Cyclic Peptides using cPEPmatch, **BL Santini**, M Gaardl s, D Wyrzykowski, S Rothmund, A Penk, M Zacharias, SA Samsonov *Computational and Structural Biotechnology Journal* 23, 2985-2994 (2024)
5. cPEPmatch Webserver: A comprehensive tool and database to aid rational design of cyclic peptides for drug discovery, **BL Santini**, S Wendel, N Halbwedl, A Knipp, M Zacharias, *Computational and Structural Biotechnology Journal* 23, 3155-3162 (2024)