ABSTRACT

TAPDIYA, ARPAN. Computational Design of Peptide ligands as Biosensors for Early Detection of Ovarian Cancer. (Under the direction of Dr. Carol K. Hall).

Ovarian cancer (OvCa) diagnosis remains a challenge due to the lack of specific and sensitive methods for detecting the presence of OvCa cells in the body. This often leads to delayed treatment and poor prognosis. Approximately 72% of OvCa cases are diagnosed at advanced stages (II and beyond), with a corresponding 5-year survival rate of only around 31%. Early detection, however, provides a much brighter outlook, with 5-year survival rates exceeding 90% for localized cancers (stage I). Extracellular vesicles (EVs), particularly exosomes, are emerging as promising minimally-invasive biomarkers for OvCa diagnosis. EVs contain a variety of biomolecules that reflect the originating cell's state. Analyzing EV protein profiles allows for the identification of tumor-specific antigens, offering a potential tool for early cancer detection.

We aim to computationally design peptide ligands as a novel approach to improve OvCa early detection. These peptide probes bind to transmembrane proteins like tetraspanins (CD81, CD9) and EpCAM, a biomarker of malignant OvCa-derived EVs. Peptide probes potentially offer various advantages over traditional antibody-based immunoassays: (1) their smaller size allows better sensitivity, and (2) the computational design minimizes cross-reactivity.

Towards this goal, we employ PepBD (Peptide Binding design), a Monte Carlo-based algorithm developed in our lab that iteratively searches for peptide sequences with high affinity and specificity toward target OvCa EV membrane proteins. The top-scoring sequences (~10-20) identified through PepBD are further evaluated using explicit-solvent molecular dynamics simulations to assess binding affinity. This rigorous approach ensures the selection of highly specific peptide candidates. Our experimental collaborators are synthesizing the designed peptides and characterizing them for their binding affinity and selectivity toward the target proteins. Using

established biochemical techniques such as Bio-layer Interferometry and single EV imaging techniques, we aim to achieve high analytical accuracy and precision in EV identification and biomarker analysis.

Computational Design of Peptide ligands as Biosensors for Early Detection of Ovarian Cancer

by
Arpan Ghansham Tapdiya

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Chemical Engineering

Raleigh, North Carolina
2024

APPROVED BY:

_____
Dr Carol K. Hall
Committee Chair

_____
Dr. Jason M. Haugh

_____
Dr. Kirill M. Efimenko

DEDICATION

This thesis is dedicated to my family and especially to my grandfather, Dr. Omprakash Samdani, who tried his best to beat cancer.

BIOGRAPHY

Arpan Tapdiya, born in June 1998 in Buldhana, Maharashtra, India, is the son of Ghansham and Deepali Tapdiya and the elder brother of Poorva Tapdiya. He pursued his undergraduate education at Visvesvaraya National Institute of Technology (VNIT), Nagpur, India, from 2016 to 2020, earning a Bachelor of Technology in Chemical Engineering. During his undergraduate years, Arpan interned in Dr. Sanjay Mahajani's reaction engineering lab at the Indian Institute of Technology, Bombay, in the summer of 2018, focusing on kinetic data generation and esterification process development for salicylic acid with methanol. For his final year project in 2019-2020 at VNIT, he delved into molecular dynamics simulations under the guidance of Dr. Piyush Wanjari, investigating the structural properties of single-walled carbon nanotubes. After completing his undergraduate studies in the summer of 2020, Arpan served as a Senior Process Engineer at Larsen & Toubro Hydrocarbons Ltd. until June 2022. The following August, he embarked on his graduate studies in the Department of Chemical and Biomolecular Engineering at North Carolina State University, Raleigh, USA. In the summer of 2023, he undertook an internship as a Process Engineering Intern at Qorvo Inc., focusing on photolithography process optimization and development. Arpan is currently completing his Thesis for Master of Science in Chemical Engineering under the mentorship of Dr. Carol K. Hall.

ACKNOWLEDGMENTS

First and foremost, I extend my deepest gratitude to my advisor, Dr. Carol K Hall, for entrusting me with this remarkable project. Dr. Hall's unwavering support and guidance have been indispensable throughout my Master's program. Her mentorship has not only facilitated the completion of my thesis but also shaped my academic journey profoundly. I am immensely grateful for the opportunity to be a part of the Hall research group.

A heartfelt thank you to Dr. Sudeep Sarma for introducing me to the captivating research within the lab. His expertise in computational peptide design and insightful guidance on the PepBD algorithm have been pivotal in shaping my thesis. Our discussions have not only deepened my understanding of the research but also served as a constant source of inspiration, driving me to push the boundaries further.

I also wish to express my sincere appreciation to all my professors at NC State University for their dedication and support, which have been instrumental in my academic growth and success. I especially wish to thank the members of my graduate committee, Dr. Jason Haugh and Dr. Kirill Efimenko, for graciously providing me with their invaluable time and insightful guidance during my research.

I'd like to extend my thanks to all the members of the Hall research group for the wonderful times that we had together. Special thanks to Michael Bergman for helping me understand the PepBD code better and helping me debug the code whenever I had some errors. I would also like to thank Corey Febo for exchanging ideas regarding peptide design. Thanks to Matthew Dorsey, Xin Dong, and Van Nguyen for making the Hall lab experience memorable.

To all the friends I made at NC State – Oluwatobi Ojuade, Mann Verlekar, Karthik Sinha Moritz Woelk, Rakshit Jain, Aryan Razdan, Sarvesh Nadkarni, Mihir Khara, Kush Jain, Dhrunil

Timbadi, Shubham Loya, Amarnath Shinde, Mukul Wagdarikar, Nishant Raman, Anurag Chollangi , Sarjak Kanani Snehashree Reddy, Vaishnavi Kolluru, Diksha Kaul, Ashika Verma – thank you for making my time at Raleigh memorable. Special thanks to Shivani Sutrave and my roommates, Aditya Venikar, Archis Gune, Parth Tulapurkar, for making NC State feel like home and for the countless moments of shared experiences, from traveling to cooking and discussing life on many occasions.

Finally, a heartfelt thank you to my family for their unwavering love and support. To my parents, Ghansham and Deepali Tapdiya, your encouragement and sacrifices have paved the way for my educational pursuits. I am grateful for your unwavering support and decisions that have led me to NC State University. A special mention to my grandparents, Dr. Omprakash and Vijaya Samdani, whose unwavering encouragement has been a source of strength. To my younger sister, Poorva Tapdiya, your ability to lift my spirits and positivity have been a constant source of solace and joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

.

1. Introduction

Cancer is a major health problem, causing about 1 in every 6 deaths worldwide[1]. In 2020, cancer resulted in nearly 19.3 million new cases and almost 10 million deaths worldwide[2]. Cancer can be classified according to distinct stages (stages 0, I, II, III and IV) based on the extent of its spread from the primary organ. Of all the types of cancers, approximately 50% of the cancers are diagnosed at an advanced stage[3,4]. Advanced cancer (stage II and beyond) is very difficult to cure and control as the number of treatment options are limited. The five-year survival period for stage I cancers of different types varies from 70% to 90% but for stage IV cancer the five-year survival period drops down to 10% to 20%[3]. Worrying about the probability of losing one's life and the economic burden of cancer treatment increases rapidly with the progression of the disease, exacting a huge toll on patients and their families. Clearly, the identification of cancerous tumors at an early stage is important.

Currently, cancer diagnosis involves a multifaceted approach that integrates clinical evaluation with laboratory testing, imaging studies and other procedures. General laboratory tests, including blood work and urinalysis, provide insights into the body's biochemical and molecular makeup. Abnormal lab results, however, are not a sure sign of cancer. Therefore, imaging techniques such as PET, MRI and CT scans, and surgical procedures such as biopsies, are generally needed to identify the type of cancer and its rate of progression. Though these methods are necessary to confirm the presence of a tumor, they have some drawbacks. Imaging methods help in identifying suspicious tissues in the body but, without biopsies or further investigation, one cannot be sure if the change in the body is caused by cancer. Such tests are generally limited to identifying only large groups of cancer cells. By the time imaging techniques are used for diagnosis, the malignancy may have progressed to become a large tumor. Also, the cost of such

imaging techniques is often very high, making it less feasible for patients with limited resources. For example, the average cost of a full body CT or PET scan is around $3275[5] or $5750[6], respectively, in the United States. Tissue biopsies, though considered to be the "**gold standard**"[7] for cancer diagnosis, are limited by the amount of tissue sample that can be removed for examination. Tumor heterogeneity, i.e. the diversity of malignant cells that make up the tumor, may not be captured effectively in small biopsy samples[8]. This could lead to misdiagnosis or ineffective treatment, causing more harm than good. The option of removing a large tissue sample also carries a risk for the patient, such as bleeding, damage to surrounding tissues, and infections. Additionally, obtaining large samples might not always be feasible, especially in sensitive or hard-to-reach areas of the body. Thus, to avoid complications and provide the ideal treatment options, it is best to detect cancer at an early stage.

For early detection, understanding the underlying biology of cancer development and progression is essential. Biomarkers are measurable indicators that quantitatively capture the biological characteristics of different processes in the body[9]. These can be used as tools that help to predict the onset, cause, and progression of a disease, and the likely outcome of a treatment[10]. Many proteins have been recognized as critical biomarkers in cancer diagnostics. For example, EpCAM (epithelial cell adhesion molecule), a transmembrane glycoprotein, is found to be overexpressed in tumors associated with epithelial cells. Around 85% of the patients diagnosed with adenocarcinoma (malignant tumors formed from glandular structures in epithelial tissue) and 72% of the patients with squamous cell carcinoma have high and consistent expression of EpCAM[11]. Another protein biomarker, EGFR (epidermal growth factor receptor), has been found to be associated with the progression of tumor cells[12]. Several methods such as liquid biopsies

and immunoassays have been developed to detect such biomarkers, thereby aiding in early detection of cancer.

Liquid biopsies are lab tests done on a sample of a bodily fluid[13] (blood, urine, salvia) to look for (1) circulating tumor cells (CTCs)[14], (2) extracellular vesicles (EVs) which may contain molecular information from tumor cells[14], and (3) pieces of cell-free DNA (cfDNA) plausibly originating from cancerous cells[15]. Compared to tissue biopsies, liquid biopsies are minimally invasive, therefore reducing the risk of complications. They also are more accessible than tissue biopsies as the samples can be easily collected and analyzed. These can also be conducted multiple times, thereby enabling dynamic monitoring of a tumor[16]. Liquid biopsies, though, tend to have lower sensitivity and specificity to certain cancer types and biomarkers than tissue biopsies, potentially leading to false-positive or false-negative results, and therefore inaccurate diagnosis[17]. For example, prostate-specific antigen (PSA) is a biomarker to identify prostate cancer in the early stages. However, the PSA level fluctuates with age, prostate size, etc[18]. Thus, relying solely on liquid biopsies to measure PSA levels may yield incorrect results, potentially leading to overdiagnosis and overtreatment. Despite some limitations, ongoing research in cancer biomarkers and advancements in liquid biopsies make it an increasingly valuable tool for the early detection of cancer.

Immunoassay is another analytical method based on biosensing that is used for early detection of cancer. It depends on antibody-antigen interactions, where it relies on the ability of antibodies to recognize and specifically bind to target antigen proteins, thereby making the antibodies biological recognition elements (BREs). With the integration of a signal transducer, the binding between antibodies and antigens generates analytical outputs such as optical signals like fluorescence and electrochemical signals like potentiometry[19]. These signals are subsequently

compared against standard curves or reference ranges to determine the affinity and specificity of antibody binding with the antigen. Immunoassays are simple, fast, and cost-effective, and can be extended to detect multiple biomarkers by using multiple BREs simultaneously[20]. Nonetheless, when using antibodies as BREs, controlling the quality of immunoassays can be challenging. Antibodies, due to their large size, may exhibit accidental cross-reactivity with non-targeted proteins, resulting in background noise in the analytical signals[21]. Variability in the binding affinity of antibodies is also observed, potentially leading to inconsistent and non-reproducible results[22].

Short peptides are appealing alternative biological recognition elements (BREs) that show promise of overcoming the challenges associated with traditional antibody-based assays[23]. Short peptides containing 8 to 15 amino acid residues offer several technical advantages over antibodies due to their small size. Immunoassays with short peptides have the potential to produce more robust and consistent results than antibodies, while minimizing cross-reactivity and variability[24]. Their small size enables them to be more densely bound to target proteins than antibodies, creating a multivalency effect for enhanced assay signals. Additionally, their low cost, ease of synthesis, and enhanced environmental stability make them highly attractive choices for biosensors[23]. Several methods have been developed to discover short peptides for their use as BREs. In-vitro methods such as phage display library screening[25] and enzyme-linked immunosorbent assay (ELISA)[20] have successfully identified short peptides that bind a particular target antigen. While these methods are powerful tools, they have certain disadvantages that limit their effectiveness. Phage display libraries may have limited library diversity due to constraints on library size and sequence complexity[26]. This restricts the number of peptide sequences that can be screened, thereby reducing the probability of finding high-affinity peptide binders. Phage display libraries

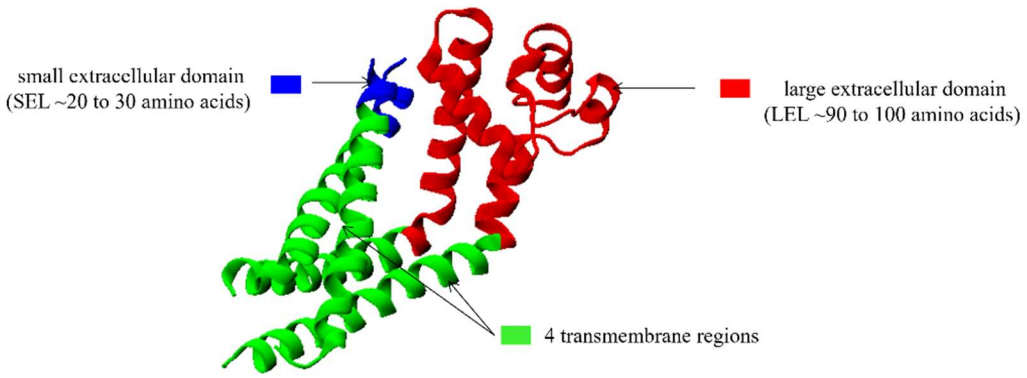
may also have a selection bias towards the peptide sequences that are more efficiently displayed on the phage surface[27].

Computational design of peptides is a promising approach to discover short peptides with specific biological functions. This methodology leverages computer algorithms and simulations to predict and design peptides tailored for desired purposes, such as targeting disease biomarkers or modulating cellular pathways. One key advantage of computational design is the ability to rapidly explore vast sequence spaces and identify peptides with optimized properties[28]. Computational methods can thereby save time and resources, compared to traditional experimental methods. Additionally, advances in computational power and algorithms have significantly enhanced the accuracy and efficiency of peptide design, enabling researchers to tackle increasingly complex design challenges. Importantly, this in-silico approach can complement experimental techniques like phage display by guiding the selection of peptide candidates for further validation[29]. By integrating computational predictions with in-vitro methods such as phage display, researchers can streamline the peptide discovery process, leading to the development of novel therapeutics and diagnostic tools with enhanced efficacy and specificity[28].

In this study, we aim to utilize the Hall group's computational peptide design algorithm, PepBD (Peptide Binding Design)[29–32] to discover synthetic peptide ligands that can detect cancer biomarkers. The discovered peptide ligands could serve as BREs for new biosensor devices to detect cancer at an early stage. We specifically focus on designing peptide ligands that recognize membrane proteins in extracellular vesicles. Extracellular vesicles (EVs) are relatively new targets for bioassays and have unique physical and biological traits[7]. Although initially considered as "cell dust", and a means to dispose of cellular components, they are now recognized as a circulating source of biomarkers for different diseases[33]. These vesicles carry molecular cargo

derived from tumor cells and are found circulating in readily accessible body fluids. With regards to cancer, tumor-associated EVs have been used effectively as biomarkers to define tumor type and stage of malignancy. They are generally classified into three groups based on their biogenesis and their sizes[7]: (1) exosomes (size 40 nm to 200 nm), (2) microvesicles (200 nm to 2000 nm), and (3) apoptotic bodies (500 nm to 2000 nm). Exosomes and microvesicles are carriers of membrane proteins such as tetraspanins like CD81, CD63, and CD9[34]. These tetraspanin proteins have emerged as effective biomarkers for detecting circulating EVs, leading to their widespread adoption in exosome identification, quantification, and characterization[7]. Tetraspanins belong to a superfamily of transmembrane proteins (about 230 to 300 residues) featuring 4 transmembrane alpha-helices and 2 extracellular domains: the larger extracellular loop (LEL) spanning 90 to 100 residues, and the shorter extracellular loop (SEL) consisting of 20 to 30 residues[35] (for structure of a tetraspanin protein refer Figure 1.1). The larger extracellular loop (LEL) plays an important role in mediating protein interactions and facilitates various cellular processes, including cell adhesion, migration, and signaling[36]. The distinctive structural and functional attributes of the LEL domain highlight its potential as a prime target for various biomedical applications, ranging from diagnostics to therapeutics.

Figure 1.1 Structure of tetraspanin protein CD81.



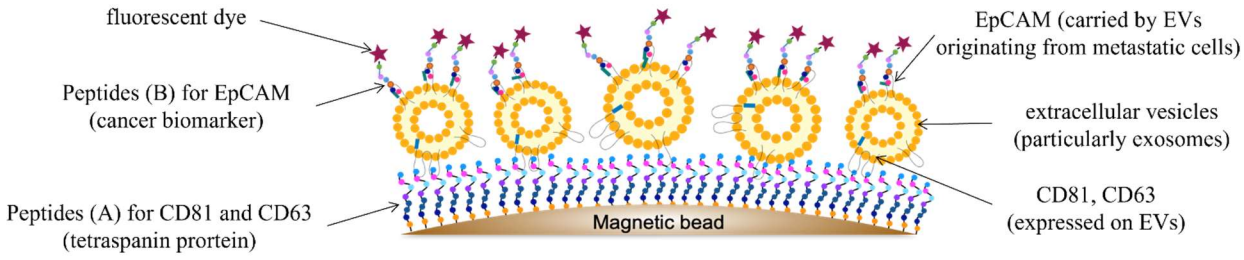
We employed the PepBD algorithm to design synthetic peptide ligands that have a specific affinity for the LEL domain of tetraspanins CD81 and CD63 for EV detection. To do so, we first searched for peptide ligands (called reference peptides) previously discovered through phage-display screening that bind specifically to the LEL domain of each tetraspanin. These reference peptides served as the starting input for the PepBD algorithm. Using the PepBD algorithm and explicit-solvent molecular dynamics simulations in AMBER[37], we then designed new peptide sequences that exhibit higher affinity and specificity toward our target tetraspanins than the reference peptide. Subsequently, by using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method[38], we calculated the binding free energies for the newly designed peptide-protein complexes and compared them with binding free energies of the reference peptide-protein complex. Five peptides demonstrating favorable binding free energies (i.e. $|\Delta G_{new\ peptide\text{-}protein}| > |\Delta G_{reference\ peptide\text{-}protein}|$) have been selected for synthesis, experiments are now being performed to validate our computational findings.

We collaborate with Dr. Hakho Lee and his team at Massachusetts General Hospital and Harvard Medical School to characterize the computationally designed peptide ligands for their specificity and affinity toward EV biomarkers. Immunoassays employing single EV-imaging techniques are adapted to characterize these peptide ligands. Dr. Lee and his team have also researched the presence of ovarian cancer (OvCa) biomarkers like epithelial cell adhesion molecule (EpCAM), epidermal growth factor receptor (EGFR), and CD24 in tumor-associated EVs[39]. We leverage our previous efforts in designing peptide ligands that specifically bind to EpCAM and combine them with these newly designed peptide ligands to screen for metastatic-derived EVs. Our focus on ovarian cancer is driven by the high percentage of OvCa cases being diagnosed at advanced stages, and the high capacity of OvCa tumors to metastasize, making

Ovarian Cancer the most lethal gynecological cancer[40]. By conjugating two distinct peptide probes, one targeting tetraspanins for EV identification and the other targeting EpCAM, we aim to develop a robust method for early-stage ovarian cancer detection.

Figure 1.2 illustrates the proposed magnetic bead-based peptide immunoassay technique for detecting ovarian cancer. In this method, magnetic beads are coated with peptide probes (A) designed to selectively bind to tetraspanin proteins CD81 and CD63, which are commonly present on the surface of extracellular vesicles (EVs). Upon introduction of a patient's bodily fluid sample, EVs expressing these targeted proteins attach to the magnetic beads. A second set of peptide probes (B), labeled with a fluorescent dye, is also introduced with the fluid sample. These probes specifically target EpCAM which might be carried by these EVs if they were released from metastatic cells. If EpCAM is present on the captured EVs, the fluorescently labeled probes (B) will bind, producing a signal that can be detected through EV imaging analysis. The intensity of this fluorescent signal can then be measured to confirm the presence and potential abundance of ovarian cancer biomarkers within the isolated EVs.

Figure 1.2: Mechanism of peptide-based immunoassay to capture EVs and detect the presence of ovarian cancer biomarkers.



2. Methods

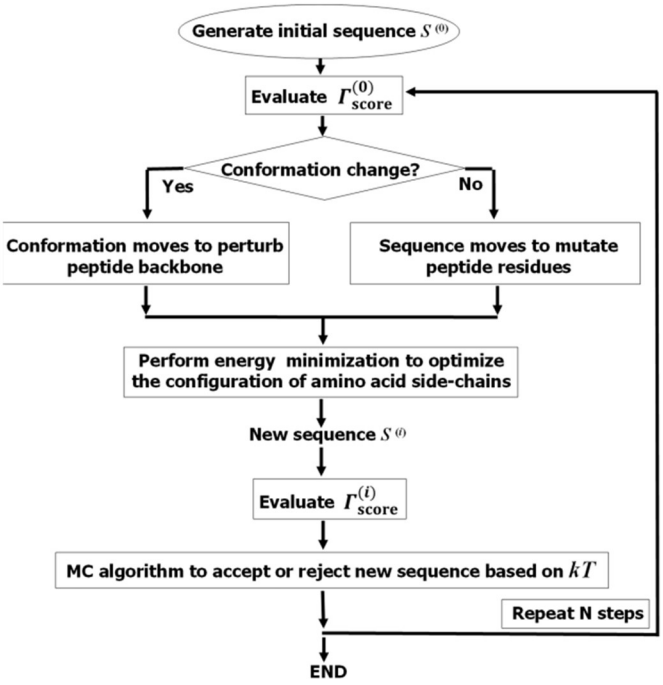
2.1 Overview of PepBD algorithm: Computational peptide design using the PepBD algorithm follows an iterative process to discover new peptide sequences that have greater binding affinity and specificity to a biomolecular target than a previously identified "reference peptide ligand". The PepBD algorithm is based on the Monte Carlo method which explores the vast conformational and sequence space of peptides over numerous iterations. These designed peptides have potential applications in biomedicine and drug development[29–32].

Hydration properties for peptides: For peptides to be effective as potential drugs or biosensors, they need to have specific properties such as good water solubility, charged residues to strengthen electrostatic interactions with the target receptor, and structural stability. PepBD incorporates this knowledge by categorizing each residue of the peptide sequence into 6 groups based on their hydrophobicity, charge (cationic or anionic), polarity, and size (refer to Table 2.1). Each category represents a specific "**hydration property**"[32]. Every charged residue is regarded as possessing one unit of charge. For instance, a positively charged residue represents +1 unit of charge, while a negatively charged residue signifies -1 unit of charge. By considering these properties, one can narrow down the sequence search space for PepBD to find peptides with specific characteristics.

Table 2.1 Classification of amino acid residues types in PepBD[32]

Hydration property type	Residues
Hydrophobic (H)	LEU (L); VAL (V); ILE (I); MET (M); PHE (F) , TYR (Y); TRP (W)
Positively charged / cationic (P)	ARG (R); LYS (K)
Negatively charged / anionic (N)	GLU (E); ASP (D)
Hydrophilic / Polar (L)	SER (S); THR (T); ASN (N); GLN (Q); HIS (H)
Others	ALA (A); CYS (C); PRO (P)
Glycine	GLY (G)

Figure 2.1 Flow diagram for the PepBD algorithm[32]



The workflow of PepBD (refer to Figure 2.1) can be broken down into five key steps:

Step 1 - Initial complex generation: This step, independent of PepBD, utilizes techniques like docking (detailed later) to generate a starting complex containing a known peptide ligand (referred to as the reference peptide) that has been previously identified for its ability to bind to the target receptor (generally a protein). The reference peptide-protein complex found through docking serves as the input structure for PepBD. The structure of the protein in the complex is kept fixed; the peptide sequence is optimized throughout the next steps of the design process.

Step 2 – Trial move selection: This step begins with a random change (called a "trial move") in either the peptide sequence or the peptide conformation (i.e. the backbone structure). A random number (***R***) is generated and is compared to a pre-defined probability ($P_{sequence}$). If $R < P_{sequence}$,

the algorithm proceeds with a sequence mutation. This can occur in two ways: **(1)** substitution of a random amino acid with a residue of similar hydration property at the same position in the peptide sequence or, **(2)** swapping two distinct random amino acids along the peptide within the sequence. A conformational change is selected when $\boldsymbol{R > P_{sequence}}$. There are three different ways in which a conformational change can be carried out: **(1)** concerted rotation (CONROT) that displaces three consecutive central residues of the peptide sequence while fixing the terminal residues[31], **(2)** rotation of either the N-terminus or the C-terminus of the peptide sequence plus two residues in the middle, and **(3)** rotation of the entire peptide backbone by selecting both the N and C-terminus and one residue in the middle.

Step 3 – Side-chain repacking: After completing a trial move, the side-chain configuration of the mutated amino acid is optimized. This optimization involves evaluating various conformational isomers, or rotamers, of the mutated amino acids. It ensures that the newly selected residues can be repacked without encountering atomic overlap or steric hindrance by the existing residues in the peptide sequence. To accomplish this, the algorithm employs rotamers defined in the well-established Lovell rotamer library[41]. This library is renowned for its reliability and versatility in protein design.

Step 4 – Energy Minimization: The energy of the new trial peptide-receptor complex is minimized to achieve an energetically favorable and physically realistic conformation. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm[42] is employed for this purpose. BFGS is a powerful iterative method that adjusts the coordinates of atoms in the peptide-receptor complex to minimize its energy while ensuring the removal of any rotamers that result in atomic overlap.

The energy values during this process are computed by a scoring function, as described by Equation 1[43]

$$\boldsymbol{\Gamma(score) = \Delta E_{binding} + \lambda(E^{bound}_{peptide-VDW} + E^{bound}_{peptide-ELE} + E^{bound}_{peptide-EGB})} \quad (1)$$

The first term of Eq. (1), $\Delta E_{binding}$, is the difference in the energy of the complex and the energies of the peptide and target protein prior to binding. $\Delta E_{binding}$ is calculated according to Equation 2

$$\boldsymbol{\Delta E_{binding} = E^{complex}_{TOT} - E^{ligand}_{TOT} - E^{receptor}_{TOT}} \quad (2)$$

where $E^{complex}_{TOT}$, E^{ligand}_{TOT} and $E^{receptor}_{TOT}$ are total free energies of the complex, ligand and receptor. The free energy E^{x}_{TOT} is the sum of the internal energy, van der Waals energy (VDW), electrostatic energy (ELE), and polar solvation energy (EGB) $E^{x}_{TOT} = U_{INT} + U_{VDW} + U_{ELE} + U_{EGB}$. The second term is the peptide stability term and accounts for the energy of the free peptide in the bound-state configuration. It is the sum of the van der Waals energy (VDW), electrostatic energy (ELE), and polar solvation energy (EGB) of the peptide prior to binding.

The parameter $\boldsymbol{\lambda}$ is a user-defined weighting factor and is used to account for the importance of the folding stability term in the scoring function[30]. A large $\boldsymbol{\lambda}$ value can change the outcome of the algorithm from binding a better peptide binder to finding a stable folded peptide. Based on previous studies, a value of $\boldsymbol{\lambda = 0.010}$ has been found to provide a good balance between optimizing the peptide's binding ability and its capacity to fold into a stable structure. In Equation 2, the van der Waals energy is calculated using the 6-12 Lennard Jones potential; the electrostatic energy calculations follow Coulomb's Law and the polar solvation energy is computed using the generalized Born model.

Step 5 – Accepting or rejecting the trial move: After evaluating the score of the new trial sequence, the algorithm decides whether or not to keep the changes it has made to the peptide

sequence. Since the algorithm is based on the Monte Carlo method, it utilizes the Metropolis criteria and accepts or rejects the trial move based on the calculated probability $\boldsymbol{P}$

$$\boldsymbol{P = min\{1, exp\,(\ (\Gamma_{old\ score}\ -\ \Gamma_{new\ score}\ )/kT)\}} \quad (3)$$

The new move is accepted under the following conditions: (1) If the new score is lower (i.e., more favorable) than the old score, the probability $\boldsymbol{P}$ will be greater than 1, and therefore, the new score will always be accepted; (2) If the new score is higher (i.e., less favorable) than the old score, the probability $\boldsymbol{P}$ will be calculated as $\boldsymbol{P = \exp(\frac{\Gamma(old\ score) - \Gamma(new\ score)}{kT_{(x)}})}$. In this case, the probability of acceptance depends on the energy difference between the old and new peptide sequences, and the system's temperature $\boldsymbol{T}$.

The parameter $\boldsymbol{kT_{(x)}}$ (where *x* can be either conformation or sequence) is adjustable and depends on the type of mutation (sequence or conformational) that the algorithm chooses while performing Step 2. For example, if a conformational change was performed during the trial move in Step 2, the algorithm uses $\boldsymbol{kT_{(conformation)}}$ to calculate the probability ***P;*** similarly it uses $\boldsymbol{kT_{(sequence)}}$ if the trial move was a sequence mutation. A high value of $\boldsymbol{kT_{(x)}}$ increases the probability that the algorithm accepts the mutations, which increases the probability of finding peptide sequences that might not have favorable scores. On the other hand, a small value of $\boldsymbol{kT_{(x)}}$ means that the algorithm rejects the mutations more frequently and thus the peptide search space is not explored efficiently. Therefore, choosing an appropriate value of $\boldsymbol{kT_{(x)}}$ is important as it enables PepBD to search through the peptide space efficiently. In this study, we initially ran trial simulations of 1000 steps with multiple values of $\boldsymbol{kT_{(x)}}$ varying from 0.6 kcal/mol to 1.4 kcal/mol, and based on the evolution of peptide scores and the ability of the algorithm to map the peptide space, we fixed our values as $\boldsymbol{kT_{(sequence)} = 0.9\ kcal/mol}$ and $\boldsymbol{kT_{(conformation)} = 1.2\ kcal/mol}$.

The root-mean square deviation δ_{RMSD}, between the new trial peptide sequence and the original peptide sequence is calculated to ensure that δ_{RMSD} is not too large ($\delta_{RMSD} < \delta_{MAX}$) and not too small ($\delta_{RMSD} > \delta_{MIN}$). In this study, we choose $\delta_{MAX} = 4.0Å$ and $\delta_{MIN} = 0.3Å$. To explore the peptide space comprehensively, we repeat Step 2 to 5 over 10,000 cycles. Each cycle involves mutating the peptide, optimizing its energy, and deciding whether or not to accept the changes using the Metropolis criteria. We further explore the peptide space by executing each 10,000-step cycle with three random seeds, generating three distinct random numbers $\boldsymbol{R}$ that are used for comparison with $P_{sequence}$ as described in Step 2.

2.2 Initial reference peptide and docking: In-silico peptide design using PepBD requires a starting input structure for the peptide-protein complex. To identify a suitable starting point, we first search for a known peptide ligand, called the reference peptide (RP), that binds to the target receptor (generally a protein). Techniques like phage display and ELISA library screening[25] can be valuable resources for finding such RPs. After finding potential RPs through literature review, we use computational docking simulations to generate the three-dimensional structure of the RP-protein complex.

Docking simulations are employed to predict the structure of a molecular complex that depicts the interaction between a specific ligand molecule and a target receptor[44]. The tools that perform docking simulations require an independent structures for the reference peptide ligand and the target protein receptor. The structure of the target protein is generally available on the Protein Data Bank (PDB), which is a repository of experimentally determined protein structures.

Through literature review, we can identify multiple short peptide sequences that have been discovered to bind to the target receptor through methods such as peptide assay-based screening.

Once the reference peptide is identified, we generate its structure using the tLEaP program within the AMBER molecular dynamics suite[45]. The ff14SB force field[46] within AMBER defines parameters such as bond length, torsion angles etc. for each residue in the peptide sequence.

To account for the inherent flexibility of both the peptide and the protein, we perform two independent 100-nanosecond MD simulations, one on the generated structure of the peptide and one on the structure of the target protein obtained from the PDB. The MD simulation details are in the following sub-section. These simulations capture stable conformational ensembles that are then used as inputs for docking simulations. We leverage two popular docking platforms, Haddock[47,48] and Autodock Vina[49,50], to identify potential binding conformations. Haddock can utilize information about known interacting residues to guide docking, while Autodock Vina efficiently explores the conformational space to find favorable binding poses. By using two platforms, we increase confidence in the identified binding conformation for the RP-protein complex, which serves as the foundation for subsequent optimization steps within the PepBD algorithm.

2.3 Explicit-solvent molecular dynamics in AMBER: Explicit-solvent molecular dynamics (MD) simulations serve as a pivotal tool to analyze the dynamics of the binding process for the reference peptide and the top-scoring peptides obtained from PepBD to the target protein. The starting structure of the target protein complexed with the best-scoring peptides for the MD simulations were obtained from the design algorithm MD simulations are executed using the AMBER20 software suite. TIP3P water molecules[51] are introduced into the system. The system is enclosed in a periodically-truncated octahedral box with a size of 12 Å, accommodating approximately 8000 to 9000 water molecules.

The MD simulations start with a two-stage energy minimization process employing the steepest descent method. The initial stage comprises 1000 iterations focused solely on minimizing the energy of the peptide-protein complex. Subsequently, the second stage, comprising 2500 iterations, results in the energy minimization for the entire simulation box, encompassing both the complex and the modeled water molecules. Following energy minimization, the second phase of the simulation entails an NVT ensemble simulation conducted at a constant temperature of 298K for a duration of 100 ns. Hierarchical clustering analysis[53] is employed on the last 5 ns of the simulation trajectories to identify representative structures of the complexes in solution. For post-simulation analysis, the implicit-solvent molecular mechanics/generalized Born surface area (MM/GBSA) approach utilizing the uniform internal dielectric constant model is applied to the final 5 ns simulation trajectories. This method calculates the binding free energies for the peptide-protein complex. To validate the robustness and reproducibility of the simulations, each MD simulation is independently replicated three times, ensuring the attainment of an equilibrium state and bolstering the reliability of the obtained results.

3. Results and Discussion

3.1 Generating initial complex structures for PepBD: The PepBD algorithm requires the starting structure for the complex formed between the reference peptide and the target protein receptor. The structure of our target protein, the large extracellular loop (LEL) of tetraspanin CD81, was obtained from the Protein Data Bank; PDB ID: 1G8Q[54]. The reference peptide was obtained by considering work by Suwatthanarak et al.[55] , who discovered multiple 8-residue peptides that bind to the target protein CD81. The authors used a peptide library-based array screening method to discover these sequences. Among these, peptide P152 (sequence: **CFMKRLRK**) was selected as the reference peptide due to its strong binding affinity for CD81. Using spot fluorescence intensity method and the Hill equation, these authors reported the estimated dissociation constant (K_D) for the binding between the peptide P152 and CD81 as 0.91 μM[56]. We hypothesized that the entire LEL domain (approximately 90 amino acids in length) could serve as a potential binding site for the reference peptide P152.

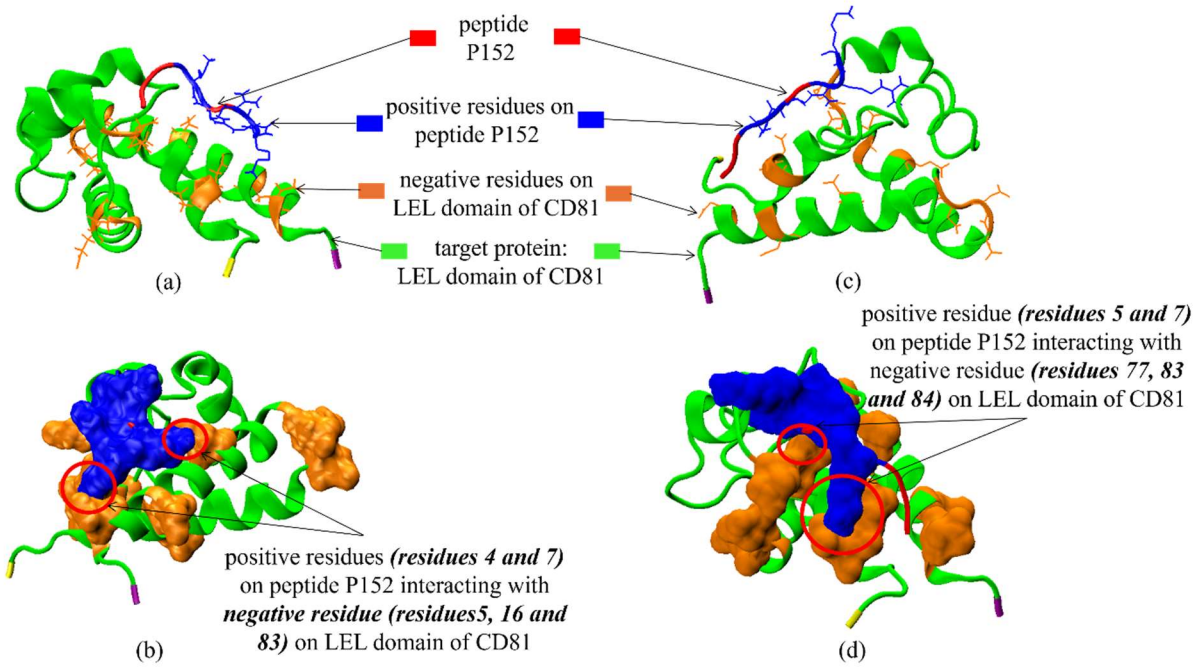
Docking simulations in which peptide P152 was docked on the LEL domain of CD81 were conducted using both Autodock Vina and Haddock. The conformation of P152 with the LEL domain of CD81 that had the best docking score from both software platforms was subjected to explicit-solvent molecular dynamics (MD) simulations. Free energies were computed, and are presented in Table 3.1. The binding conformations from Autodock Vina and from Haddock were analyzed. Interestingly, the two conformations exhibited similar interactions between the charged residues of the peptide and the protein (refer Figure 3.1) In both these conformations, two positively-charged residues on peptide P152 interact with three negatively-charged residues on the LEL domain of CD81. Based on the comparison of the free energies of the two conformations, we

selected the conformation from Autodock Vina for the first round of peptide design due to its lower free energy compared to that from Haddock.

Table 3.1 Free energy using the MM/GBSA method of the reference peptide P152 and the LEL domain of CD81 (peptide-protein complex) obtained from Autodock Vina and Haddock.

Energy type	***Energy values from docking software (kcal/mol)***	
	Autodock Vina	***Haddock***
Electrostatic ELE (1)	-270.2	-445.98
van der Waals VDW (2)	-31.01	-20.93
Non-polar solvation GBSUR (3)	-5.76	-3.51
Polar solvation GB (4)	272.15	438.55
Total entropy change TSTOT (5)	-23.16	-22.58
Total = 1+2+3+4-5	***-11.66 kcal/mol***	-9.29 kcal/mol

Figure 3.1: (a and b): Best conformation of P152 bound to the LEL domain of CD81 found through Autodock Vina with ΔG = -11.66 kcal/mol. This conformation serves as input for the first round of peptide design; (c and d): Best conformation of P152 bound to the LEL domain of CD81 found through Haddock with ΔG = -9.29 kcal/mol



3.2 First Round of Computational Peptide Design: Following the generation of the initial input structure from the docking simulations, we employed the PepBD algorithm to optimize the peptide sequence. The reference peptide, P152 (CFMKRLRK), comprises three hydrophobic residues (M – MET; F – PHE; L – LEU), one cysteine residue (C – CYS) and four positively charged residues (2 K's – LYS; 2 R's – ARG), resulting in a net charge of +4 units on the peptide P152 (Refer to Table 2.1 for the classification of each residue based on the hydration property). Analysis of the binding dynamics revealed that the positively-charged residues in peptide P152 play a crucial role in the strong peptide-protein binding. To look for peptide sequences with specific hydration properties that make them suitable biosensor candidates, we explored seven design cases with various hydration properties (see Table 3.2) , primarily focusing on altering the net charge and the hydrophobicity of the peptides. We examined peptides with net charge ranging from 0 to +4 units. While the reference peptide lacked specific hydrophilic residues (as classified in Table 2.1), we ensured that at least one hydrophilic residue was included in each design case to enhance peptide solubility.

Table 3.2: Seven different hydration property combinations explored in the first round of peptide designs for 8-mer sequences.

Hydration property	***Case number***						
	1	***2***	***3***	***4 (same as reference peptide)***	***5***	***6***	***7***
$N_{hydrophobic}$	3	3	2	3	3	2	2
$N_{negative}$	1	1	1	0	0	0	0
$N_{positive}$	2	1	1	4	3	4	3
N_{polar}	1	2	2	0	1	1	1
N_{others}	1	1	1	1	1	1	2
$N_{Glycine}$	0	0	1	0	0	0	0
Overall charge	+1	0	0	+4	+3	+3	+2

We performed 21 independent PepBD runs, with each of the seven hydration property constraints executed three times. In each repetition, a distinct randomly-generated number R was utilized. Each PepBD run started with a random trial peptide sequence and followed a unique search path. Key input parameters, such as $\boldsymbol{kT_{(sequence)}}$, $\boldsymbol{kT_{(conformation)}}$ and $\boldsymbol{\delta_{RMSD}}$ (as discussed in the methods section) remained consistent across all runs. PepBD scores were computed at each step using the previously-defined scoring function. As the search commenced with random sequences, the initial PepBD scores tend to be high. However, as the peptide space was efficiently explored, multiple minima appeared in the score profile, indicating potential peptides with favorable binding affinities for the target protein. Figure 3.2 illustrates a representative PepBD score and RMSD evolution profile with step number, while Table 3.3 lists the best-scoring 8-mer peptides identified in each case. The top-scoring peptides from the initial design round were predominantly positively charged. However, since the cell membrane possesses an overall negative charge, there was a concern by our collaborator, Dr. Lee, that these positively charged peptides could lead to non-specific binding with the cell membrane (which has a net negative charge), instead of with to the extracellular domain of the tetraspanin. This led us to perform an additional round of peptide design.

Figure 3.2: PepBD score versus the step number in blue. RMSD evolution is depicted in red. This is a typical score profile with high scores at the start of the PepBD search. The algorithm discovers new peptide sequences and generates corresponding PepBD scores, which are calculated by the score function. Low scores indicate potential peptide sequences with favorable binding free energy.

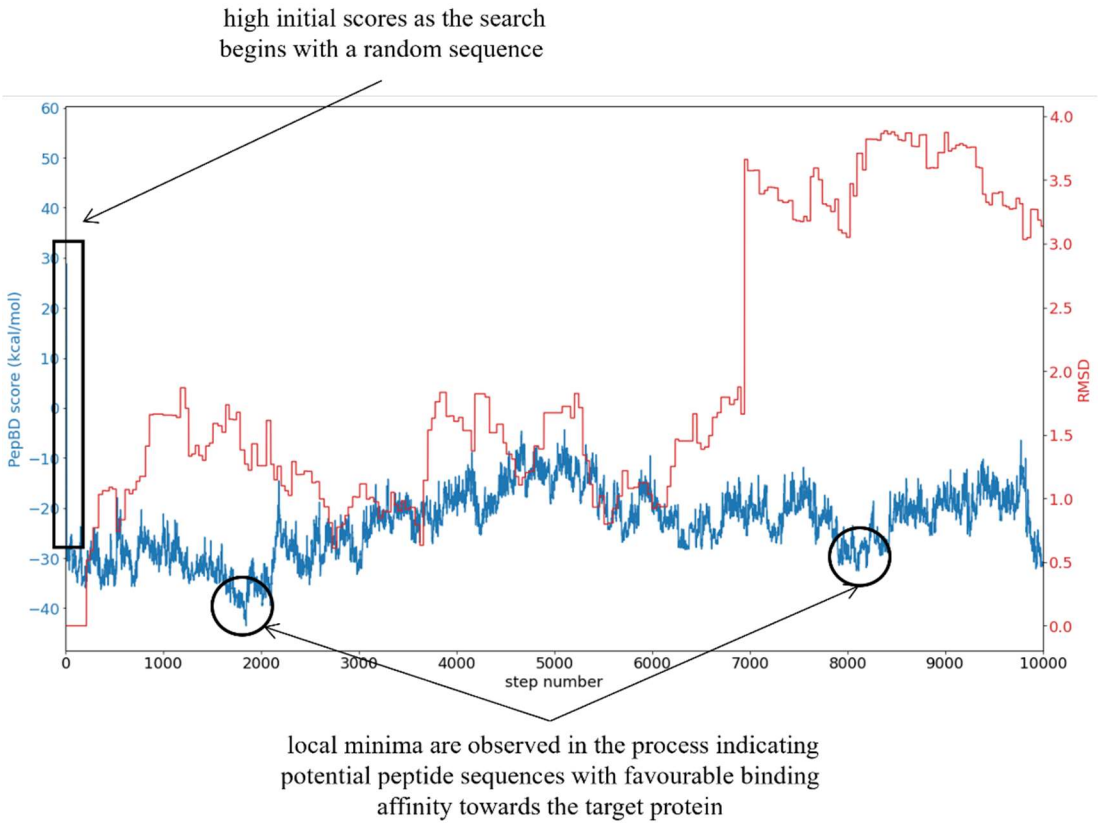


Table 3.3: Top scoring peptide from each hydration property case along with their PepBD scores in (kcal/mol) as calculated using the scoring function Γ_{score} for first round of peptide design.

Case number	***Best peptide sequence***	***PepBD score (kcal/mol)***
1	P1: DQWLRARW	-46.22
2	P2: DNWMRPQW	-38.66
3	P3: WWGLRPNE	-42.63
4	P4: RRWLRPRM	-47.82
5	P5: PRWRNWRW	-52.63
6	P6: QWWRRPRR	-49.06
7	P7: ARTQRPRW	-45.004

.

3.3 Second round of computational peptide design: To avoid non-specific binding of the designed peptides with the cell membrane, we initiated a second design round aiming for peptides that were negative or neutral in charge. Four new reference peptides, named SEEDs 1, 2, 3, and 4 with a net charge of -1 (see Table 3.4), were generated by adding two negative residues to the peptide sequence P1 (DQWLRARW). This peptide was one of the best scoring sequences discovered in the first round of design and had a net charge of +1. Adding two negative residues to P1 thus facilitated the creation of new reference peptides (SEED 1, 2, 3, and 4) with a net of charge -1.

Table 3.4: New reference peptides (SEEDs 1, 2 3 and 4) generated by adding negative residues to the peptide P1: DQWLRARW which was obtained from the first round of design. Results from free energy calculations performed using the MM/GBSA method for the binding conformations obtained from Haddock and Autodock Vina for each SEED.

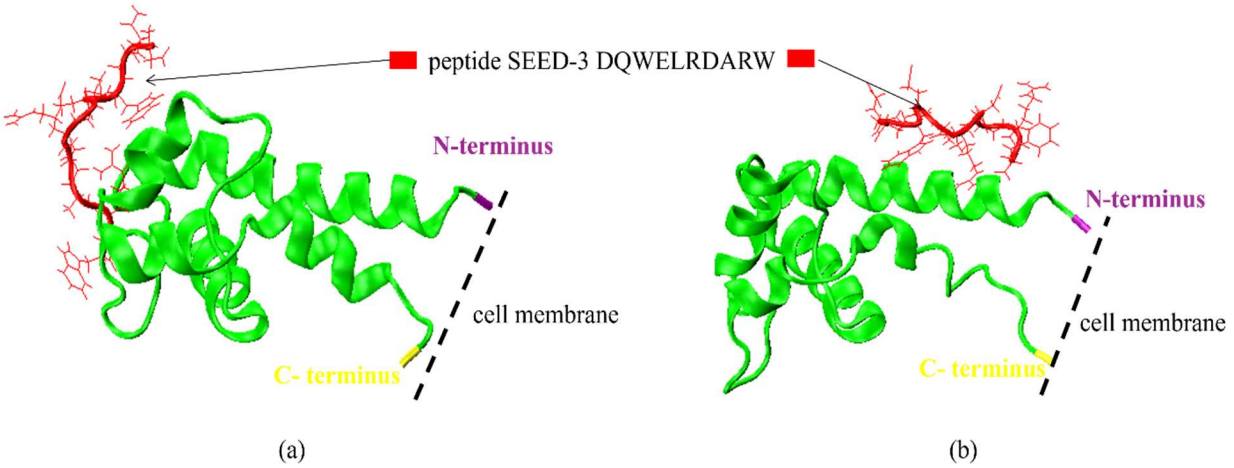
Reference peptide	***Energy values from docking software (kcal/mol)***	
	Haddock	***Vina***
DDQWLRARWE (SEED 1)	0.81	3.99
DQWLDERARW (SEED 2)	***-8.08***	-0.67
DQWELRDARW (SEED 3)	***-17.91***	2.09
DQEWLRADRW (SEED 4)	0.66	3.73

Binding conformations were generated using these new reference peptides (SEEDs 1, 2, 3, and 4) through docking simulations performed using Autodock Vina and Haddock. Since binding between the original reference peptide P152 and the target protein was predominantly due to the

interaction between the positive residues on the reference peptide and the negative residues on the target protein, the mutations to include negative residues in the reference peptides necessitated the exploration of new binding conformations and locations. Molecular dynamics simulations were performed for 100ns on the best scoring conformations for each reference peptide obtained from the two docking softwares and were followed-up by free energy calculations.

Binding conformations between the new reference peptides (SEEDs 1, 2, 3 and 4) and the LEL domain of CD81 that had negative free energies were identified as potential binding conformations for the second set of PepBD runs (Refer to Table 3.4). We only observed one favorable case of negative binding free energy from Autodock Vina for SEED 2 peptide (ΔG = -0.67 kcal/mol) whereas we observed two cases of negative binding free energy from Haddock for SEED 2 and SEED 3 peptide; these had ΔG = -8.08 kcal/mol and ΔG = -17.91 kcal/mol respectively. We selected the binding conformations from Haddock for the second round of peptide design because of their more negative free energy values. One advantage of the binding conformations from Haddock was that they bound further away from the cell membrane (as represented in Figure 3.3) than to those predicted by Autodock Vina. This had the further advantage that it minimized non-specific interactions with the cell membrane.

Figure 3.3: Binding conformations for reference peptide SEED-3 obtained from Haddock (a) and from Autodock Vina (b). The binding site identified from Autdock Vina is closer to the cell membrane (represented in dotted black line) than the binding site identified from Haddock.



For the second round of peptide design, we investigated seven different hydration property combinations, as outlined in Table 3.5, such that the net charge on the peptides discovered during this round of design was negative or neutral. The two selected binding conformations obtained from Haddock for reference peptide SEED-2 and SEED-3 underwent PepBD runs for each of the seven hydration property combinations. We performed three PepBD runs for each hydration property combination using three different random numbers. Therefore, a total of 42 independent PepBD runs were performed for the second round of peptide design. (7 design cases × 3 random numbers × 2 reference peptides = 42 PepBD runs).

Table 3.5: New seven cases of hydration properties explored in second round of design.

Hydration property	***Case number***						
	1	***2***	***3***	***4***	***5***	***6***	***7***
$N_{hydrophobic}$	3	4	4	3	3	3	4
$N_{negative}$	3	2	1	2	2	2	3
$N_{positive}$	2	1	1	1	2	1	2
N_{polar}	1	1	2	1	1	2	1
N_{others}	1	2	2	3	2	2	0
$N_{Glycine}$	0	0	0	0	0	0	0
Overall charge	-1	-1	0	-1	0	-1	-1

The lowest-scoring peptides from each PepBD run were then subjected to an explicit solvent molecular dynamics simulation for 100 ns to assess their binding dynamics with the LEL domain of tetraspanin CD81. PepBD scores and free energies for the top 10 peptide candidates are presented in Table 3.6. A more negative value of free energy (ΔG) indicates a stronger binding affinity of the peptide sequence to the target protein. Four peptide candidates from design cases SEED3-case7, SEED3-case5, SEED3-case2, and SEED2-case3 exhibited lower binding free energy compared to the reference peptide, and hence were selected for experimental validation. Solubility scores were also computed to predict the solubility of the discovered peptides. Table 3.6 includes these solubility scores, calculated using the CamSol Intrinsic method[58], which assesses the solubility profile for each residue in the peptide sequence and computes an overall solubility score. Scores greater than 1 indicate highly soluble peptides, while scores less than -1 denote poorly soluble peptides.

It is important to note that a low binding free energy ***(ΔG)*** doesn't always correspond to a low dissociation constant ***(K_D)***, which needs to be determined experimentally. One reason for this potential discrepancy is that the MM/GBSA method used for calculating binding free energies, relies on an implicit-solvent model. This method may not fully account for the influence of the solvent, water, thus failing to capture the true enthalpy and entropy changes during desolvation for both the LEL domain of CD81 and the peptides upon binding. Furthermore, the MM/GBSA method assumes that the conformational states of the peptide ligand and protein receptor remain unchanged upon binding to form a complex. This assumption may not accurately reflect the dynamic nature of protein-ligand interactions, potentially leading to a discrepancy in the predicted binding affinity and the experimentally calculated dissociation constant.

Table 3.6: Final set of designed peptide ligands for LEL domain of CD81. The top scoring peptide sequences along with their PepBD scores, intrinsic solubility scores as calculated through the CamSol method, and best binding free energies (ΔG in kcal/mol) computed from running three independent explicit-solvent MD simulations for 100ns. The peptide sequences in bold italics have been selected for synthesis and experimental validation because of their low scores as compared to the other peptides.

Sl. no.	***Search***	***Peptide sequence***	***Overall charge***	***Γ_{score} (kcal/mol)***	***Solubility score***	***ΔG (kcal/mol)***
1	seed3-RP	DQWELRDARW	-1	-	1.88	-17.91
2	seed2-RP	DQWLDERARW	-1	-	2.08	-8.08
3	***seed3-case7***	***EWYEWEHRRY***	***-1***	***-48.95***	***1.98***	***-24.93***
4	***seed3-case5***	***YHWPWEDRRA***	***0***	***-42.26***	***1.82***	***-24.81***
5	***seed3-case2***	***ADWQYDPRYW***	***-1***	***-44.68***	***1.68***	***-23.19***
6	***seed2-case3***	***APMMMDRQHW***	***0***	***-40.28***	***1.77***	***-19.74***
7	seed3-case1	WWAEDDYRRN	-1	-40.04	2.28	-17.23
8	seed2-case7	DEWMRQRMWD	-1	-40.55	1.88	-10.68
9	seed2-case2	EHWMLDRAWA	-1	-38.15	1.73	-9.21
10	seed3-case4	HAAWDEARWF	-1	-38.22	1.57	-9.09
11	seed3-case6	PAWDWEHRFN	-1	-40.49	1.83	-8.46
12	seed2-case1	AWWMEERQER	-1	-37.72	2.12	-6.49

4. Conclusion and Future scope of work

The study presents the application of the PepBD algorithm to discover peptide sequences with enhanced binding affinity compared to known reference peptide ligands for the large extracellular domain of the tetraspanin protein CD81. Leveraging computational methodologies such as docking simulations, explicit-solvent molecular dynamics, and MM/GBSA, we identified multiple novel peptide sequences predicted to exhibit lower binding free energy than the original reference peptide P152. Among these, four top-performing peptides are currently undergoing experimental validation to assess their binding affinity towards the target protein.

Ongoing efforts are directed towards designing peptides that could bind to our next target, CD63. Once the peptide designs for CD63 are finalized, these peptides, along with those designed for CD81, will be instrumental in the identification and quantification of extracellular vesicles. Subsequently, the designed peptide probes for tetraspanins will be combined with those targeting the ovarian cancer biomarker EpCAM, leading to the development of a microfluidic-based device for ovarian cancer detection. This interdisciplinary approach underscores the potential for peptide-based strategies in advancing diagnostic technologies for critical medical conditions.

5. REFERENCES

(1) *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer (accessed 2024-04-15).

(2) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **2021**, *71* (3), 209–249. https://doi.org/10.3322/caac.21660.

(3) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. Cancer Statistics, 2022. *CA. Cancer J. Clin.* **2022**, *72* (1), 7–33. https://doi.org/10.3322/caac.21708.

(4) *Cancer Statistics Review, 1975-2014 - SEER Statistics*. SEER. https://seer.cancer.gov/archive/csr/1975_2014/index.html (accessed 2024-04-05).

(5) *What is the cost of a CT scan in the U.S.?*. New Choice Health Blog. https://www.newchoicehealth.com/ct-scan/cost (accessed 2024-04-17).

(6) *How much should your PET scan cost?*. New Choice Health Blog. https://www.newchoicehealth.com/pet-scan/cost (accessed 2024-04-17).

(7) Shao, H.; Im, H.; Castro, C. M.; Breakefield, X.; Weissleder, R.; Lee, H. New Technologies for Analysis of Extracellular Vesicles. *Chem. Rev.* **2018**, *118* (4), 1917–1950. https://doi.org/10.1021/acs.chemrev.7b00534.

(8) Gerlinger, M.; Rowan, A. J.; Horswell, S.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; Tarpey, P.; Varela, I.; Phillimore, B.; Begum, S.; McDonald, N. Q.; Butler, A.; Jones, D.; Raine, K.; Latimer, C.; Santos, C. R.; Nohadani, M.; Eklund, A. C.; Spencer-Dene, B.; Clark, G.; Pickering, L.; Stamp, G.; Gore, M.; Szallasi, Z.; Downward, J.; Futreal, P. A.; Swanton, C. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **2012**, *366* (10), 883–892. https://doi.org/10.1056/NEJMoa1113205.

(9) Sarhadi, V. K.; Armengol, G. Molecular Biomarkers in Cancer. *Biomolecules* **2022**, *12* (8), 1021. https://doi.org/10.3390/biom12081021.

(10) Ding, Z.; Wang, N.; Ji, N.; Chen, Z.-S. Proteomics Technologies for Cancer Liquid Biopsies. *Mol. Cancer* **2022**, *21* (1), 53. https://doi.org/10.1186/s12943-022-01526-8.

(11) Baeuerle, P. A.; Gires, O. EpCAM (CD326) Finding Its Role in Cancer. *Br. J. Cancer* **2007**, *96* (3), 417–423. https://doi.org/10.1038/sj.bjc.6603494.

(12) Ganti, A. K. Epidermal Growth Factor Receptor Signaling in Nonsmall Cell Lung Cancer. *Cancer Invest.* **2010**, *28* (5), 515–525. https://doi.org/10.3109/07357900903476760.

(13) *Definition of liquid biopsy - NCI Dictionary of Cancer Terms - NCI*. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/liquid-biopsy (accessed 2024-04-15).

14) Martins, I.; Ribeiro, I. P.; Jorge, J.; Gonçalves, A. C.; Sarmento-Ribeiro, A. B.; Melo, J. B.; Carreira, I. M. Liquid Biopsies: Applications for Cancer Diagnosis and Monitoring. *Genes* **2021**, *12* (3), 349.

(15) McLarty, J.; Yeh, C. Circulating Cell-Free DNA: The Blood Biopsy in Cancer Management. *MOJ Cell Sci Rep* **2015**, *2* (2), 00021.

(16) Kwo, L.; Aronson, J. The Promise of Liquid Biopsies for Cancer Diagnosis. *Evid Based Oncol* **2021**, *27* (7), 261–262.

(17) Lone, S. N.; Nisar, S.; Masoodi, T.; Singh, M.; Rizwan, A.; Hashem, S.; El-Rifai, W.; Bedognetti, D.; Batra, S. K.; Haris, M.; others. Liquid Biopsy: A Step Closer to Transform Diagnosis, Prognosis and Future of Cancer Treatments. *Mol. Cancer* **2022**, *21* (1), 79.

(18) *Prostate-Specific Antigen (PSA) Test - NCI*. https://www.cancer.gov/types/prostate/psa-fact-sheet (accessed 2024-04-15).

(19) Bohunicky, B.; Mousa, S. A. Biosensors: The New Wave in Cancer Diagnosis. *Nanotechnol. Sci. Appl.* **2010**, *4*, 1–10. https://doi.org/10.2147/NSA.S13465.

(20) Tighe, P. J.; Ryder, R. R.; Todd, I.; Fairclough, L. C. ELISA in the Multiplex Era: Potentials and Pitfalls. *PROTEOMICS – Clin. Appl.* **2015**, *9* (3–4), 406–422. https://doi.org/10.1002/prca.201400130.

(21) Baker, M. Reproducibility Crisis: Blame It on the Antibodies. *Nature* **2015**, *521* (7552), 274–276. https://doi.org/10.1038/521274a.

(22) Bradbury, A.; Plückthun, A. Reproducibility: Standardize Antibodies Used in Research. *Nature* **2015**, *518* (7537), 27–29. https://doi.org/10.1038/518027a.

(23) Kim, S. S.; Kuang, Z.; Ngo, Y. H.; Farmer, B. L.; Naik, R. R. Biotic–Abiotic Interactions: Factors That Influence Peptide–Graphene Interactions. *ACS Appl. Mater. Interfaces* **2015**, *7* (36), 20447–20453. https://doi.org/10.1021/acsami.5b06434.

(24) Tadepalli, S.; Kuang, Z.; Jiang, Q.; Liu, K.-K.; Fisher, M. A.; Morrissey, J. J.; Kharasch, E. D.; Slocik, J. M.; Naik, R. R.; Singamaneni, S. Peptide Functionalized Gold Nanorods for the Sensitive Detection of a Cardiac Biomarker Using Plasmonic Paper Devices. *Sci. Rep.* **2015**, *5* (1), 16206. https://doi.org/10.1038/srep16206.

(25) Molek, P.; Strukelj, B.; Bratkovic, T. Peptide Phage Display as a Tool for Drug Discovery: Targeting Membrane Receptors. *Molecules* **2011**, *16* (1), 857–887. https://doi.org/10.3390/molecules16010857.

(26) Derda, R.; Tang, S. K. Y.; Li, S. C.; Ng, S.; Matochko, W.; Jafari, M. R. Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification. *Molecules* **2011**, *16* (2), 1776–1803. https://doi.org/10.3390/molecules16021776.

(27) Ryvkin, A.; Ashkenazy, H.; Weiss-Ottolenghi, Y.; Piller, C.; Pupko, T.; Gershoni, J. M. Phage Display Peptide Libraries: Deviations from Randomness and Correctives. *Nucleic Acids Res.* **2018**, *46* (9), e52. https://doi.org/10.1093/nar/gky077.

(28) Suárez, M.; Jaramillo, A. Challenges in the Computational Design of Proteins. *J. R. Soc. Interface* **2009**, *6* (Suppl 4), S477–S491. https://doi.org/10.1098/rsif.2008.0508.focus.

(29) Monti, A.; Vitagliano, L.; Caporale, A.; Ruvo, M.; Doti, N. Targeting Protein–Protein Interfaces with Peptides: The Contribution of Chemical Combinatorial Peptide Library Approaches. *Int. J. Mol. Sci.* **2023**, *24* (9), 7842. https://doi.org/10.3390/ijms24097842.

(30) Xiao, X.; Kuang, Z.; Slocik, J. M.; Tadepalli, S.; Brothers, M.; Kim, S.; Mirau, P. A.; Butkus, C.; Farmer, B. L.; Singamaneni, S.; Hall, C. K.; Naik, R. R. Advancing Peptide-Based Biorecognition Elements for Biosensors Using in-Silico Evolution. *ACS Sens.* **2018**, *3* (5), 1024–1031. https://doi.org/10.1021/acssensors.8b00159.

(31) Xiao, X.; Agris, P. F.; Hall, C. K. Designing Peptide Sequences in Flexible Chain Conformations to Bind RNA: A Search Algorithm Combining Monte Carlo, Self-Consistent Mean Field and Concerted Rotation Techniques. *J. Chem. Theory Comput.* **2015**, *11* (2), 740–752. https://doi.org/10.1021/ct5008247.

(32) Xiao, X.; Agris, P. F.; Hall, C. K. Introducing Folding Stability into the Score Function for Computational Design of RNA-Binding Peptides Boosts the Probability of Success. *Proteins Struct. Funct. Bioinforma.* **2016**, *84* (5), 700–711. https://doi.org/10.1002/prot.25021.

(33) Xiao, X.; Wang, Y.; Leonard, J. N.; Hall, C. K. Extended Concerted Rotation Technique Enhances the Sampling Efficiency of the Computational Peptide-Design Algorithm. *J. Chem. Theory Comput.* **2017**, *13* (11), 5709–5720. https://doi.org/10.1021/acs.jctc.7b00714.

(34) Xingqing Xiao, C. K. H.; Agris, P. F. The Design of a Peptide Sequence to Inhibit HIV Replication: A Search Algorithm Combining Monte Carlo and Self-Consistent Mean Field Techniques. *J. Biomol. Struct. Dyn.* **2014**, *32* (10), 1523–1536. https://doi.org/10.1080/07391102.2013.825757.

(35) Johnstone, R. M.; Adam, M.; Hammond, J. R.; Orr, L.; Turbide, C. Vesicle Formation during Reticulocyte Maturation. Association of Plasma Membrane Activities with Released Vesicles (Exosomes). *J. Biol. Chem.* **1987**, *262* (19), 9412–9420.

(36) van Niel, G.; Charrin, S.; Simoes, S.; Romao, M.; Rochin, L.; Saftig, P.; Marks, M. S.; Rubinstein, E.; Raposo, G. The Tetraspanin CD63 Regulates ESCRT-Independent and -Dependent Endosomal Sorting during Melanogenesis. *Dev. Cell* **2011**, *21* (4), 708–721. https://doi.org/10.1016/j.devcel.2011.08.019.

(37) Andreu, Z.; Yáñez-Mó, M. Tetraspanins in Extracellular Vesicle Formation and Function. *Front. Immunol.* **2014**, *5*. https://doi.org/10.3389/fimmu.2014.00442.

(38) Kummer, D.; Steinbacher, T.; Schwietzer, M. F.; Thölmann, S.; Ebnet, K. Tetraspanins: Integrating Cell Surface Receptors to Functional Microdomains in Homeostasis and Disease. *Med. Microbiol. Immunol. (Berl.)* **2020**, *209* (4), 397–405. https://doi.org/10.1007/s00430-020-00673-3.

(39) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E. I.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. Jr. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183–6191. https://doi.org/10.1021/acs.jcim.3c01153.

(40) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10* (5), 449–461. https://doi.org/10.1517/17460441.2015.1032936.

(41) Im, H.; Shao, H.; Park, Y. I.; Peterson, V. M.; Castro, C. M.; Weissleder, R.; Lee, H. Label-Free Detection and Molecular Profiling of Exosomes with a Nano-Plasmonic Sensor. *Nat. Biotechnol.* **2014**, *32* (5), 490–495. https://doi.org/10.1038/nbt.2886.

(42) Menon, U.; Gentry-Maharaj, A.; Burnell, M.; Singh, N.; Ryan, A.; Karpinskyj, C.; Carlino, G.; Taylor, J.; Massingham, S. K.; Raikou, M.; Kalsi, J. K.; Woolas, R.; Manchanda, R.; Arora, R.; Casey, L.; Dawnay, A.; Dobbs, S.; Leeson, S.; Mould, T.; Seif, M. W.; Sharma, A.; Williamson, K.; Liu, Y.; Fallowfield, L.; McGuire, A. J.; Campbell, S.; Skates, S. J.; Jacobs, I. J.; Parmar, M. Ovarian Cancer Population Screening and Mortality after Long-Term Follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A Randomised Controlled Trial. *The Lancet* **2021**, *397* (10290), 2182–2193. https://doi.org/10.1016/S0140-6736(21)00731-5.

(43) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40* (3), 389–408.

(44) Head, J. D.; Zerner, M. C. A Broyden—Fletcher—Goldfarb—Shanno Optimization Procedure for Molecular Geometries. *Chem. Phys. Lett.* **1985**, *122* (3), 264–270. https://doi.org/10.1016/0009-2614(85)80574-1.

(45) Xiao, X.; Sarma, S.; Menegatti, S.; Crook, N.; Magness, S. T.; Hall, C. K. In Silico Identification and Experimental Validation of Peptide-Based Inhibitors Targeting Clostridium Difficile Toxin A. *ACS Chem. Biol.* **2022**, *17* (1), 118–128. https://doi.org/10.1021/acschembio.1c00743.

(46) Bitencourt-Ferreira, G.; de Azevedo, W. F. J. How Docking Programs Work. *Methods Mol. Biol. Clifton NJ* **2019**, *2053*, 35–50. https://doi.org/10.1007/978-1-4939-9752-7_3.

(47) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688. https://doi.org/10.1002/jcc.20290.

(48) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(49) Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. M. J. J. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8*. https://doi.org/10.3389/fmolb.2021.729513.

(50) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428* (4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

(51) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(52) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.

(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(54) Roe, D. R.; Cheatham, T. E. I. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(55) Kitadokoro, K.; Bordo, D.; Galli, G.; Petracca, R.; Falugi, F.; Abrignani, S.; Grandi, G.; Bolognesi, M. CD81 Extracellular Domain 3D Structure: Insight into the Tetraspanin Superfamily Structural Motifs. *EMBO J.* **2001**, *20* (1–2), 12–18. https://doi.org/10.1093/emboj/20.1.12.

(56) Suwatthanarak, T.; Thiodorus, I. A.; Tanaka, M.; Shimada, T.; Takeshita, D.; Yasui, T.; Baba, Y.; Okochi, M. Microfluidic-Based Capture and Release of Cancer-Derived Exosomes via Peptide–Nanowire Hybrid Interface. *Lab. Chip* **2021**, *21* (3), 597–607.

(57) Suwatthanarak, T.; Usuba, K.; Kuroha, K.; Tanaka, M.; Okochi, M. Screening of EWI-2-Derived Peptides for Targeting Tetraspanin CD81 and Their Effect on Cancer Cell Migration. *Biomolecules* **2023**, *13* (3), 510. https://doi.org/10.3390/biom13030510.

(58) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **2015**, *427* (2), 478–490. https://doi.org/10.1016/j.jmb.2014.09.026.