DEMOCRITUS UNIVERSITY OF THRACE

SCHOOL OF HEALTH SCIENCES

DEPARTMENT OF MOLECULAR BIOLOGY & GENETICS

Modeling the interaction between a non-specific DNA-binding peptide and canonical B-DNA

MSc THESIS

MSc: Applied Bioinformatics & Data Analysis

Evangelos Marios Panagopoulos, Registration Number: 476

Members of the Advisory Committee:

Supervisor: Nicholas Glykos, Associate Professor of the Department of MB&G of DUTh

Member 2: Vasiliki Fadouloglou Associate Professor of the Department of MB&G of DUTh

Member 3: Aristotelis Papageorgiou Professor of the Department of MB&G of DUTh

Contents

Περίληψη

Αυτή η διατριβή διερευνά τις αλληλεπιδράσεις μεταξύ ενός 12-μερούς πεπτιδίου (SVSVGMKPSPRP) και ενός κανονικού μορίου B-DNA, εστιάζοντας στους μηχανισμούς που διέπουν τη μη ειδική σύνδεσή του με αυτό, χρησιμοποιώντας ένα συνδυασμό προηγμένων υπολογιστικών προσεγγίσεων. Για το σκοπό αυτό, 100 πιθανές διαμορφώσεις του πεπτιδίου, οι οποίες ήταν διαθέσιμες μέσω προσομοιώσεων μοριακής δυναμικής αναδίπλωσης, εξετάστηκαν για την ικανότητά τους να δεσμεύονται σε ένα κανονικό μόριο B-DNA, με χρήση του εργαλείου docking HDOCKlite. Οι ενεργειακές βαθμολογίες κάθε πρόσδεσης μελετήθηκαν και οδήγησαν στους εκπροσώπους του πεπτιδίου που εμφάνιζαν την ευνοϊκότερη ενεργειακά πρόσδεση. Τα αποτελέσματα αυτά αξιολογήθηκαν περεταίρω με άλλα δύο ακόμη προγράμματα docking, τα HADDOCK και PyDockDNA. Τα προκύπτοντα σύμπλοκα οπτικοποιήθηκαν με το πρόγραμμα Pymol. Δομική ανάλυση των αποτελεσμάτων αποκάλυψε τα 3 πιο αποδοτικά μοντέλα πρόσδεσης. Ο μηχανισμός πρόσδεσης βρέθηκε ότι κυριαρχείται από ηλεκτροστατικές αλληλεπιδράσεις, κυρίως μεταξύ των θετικά φορτισμένων καταλοίπων του πεπτιδίου και των αρνητικά φορτισμένων φωσφορικών ομάδων του φωσφοδιεστερικού σκελετού του DNA. Οι υδρόφοβες αλληλεπιδράσεις και οι δυνάμεις Van der Waals σταθεροποιούν περαιτέρω τα μοντέλα, ενώ οι δεσμοί υδρογόνου φαίνεται να παίζουν λιγότερο σημαντικό ρόλο. Παρόλο που τα εργαλεία docking δεν ήταν δυνατό να υποδείξουν μία και μοναδική θέση πρόσδεσης, καθώς το πεπτίδιο δεσμεύεται μη-ειδικά, φαίνεται ότι συγκλείνουν ως προς τις αλληλεπιδράσεις που εμπλέκονται στη μη ειδική σύνδεση του πεπτιδίου με το B-DNA. Η μελέτη υπογραμμίζει τους περιορισμούς της προσομοίωσης του πεπτιδίου στην ελεύθερη μορφή του (απουσία του μορίου DNA) και τονίζει την ανάγκη για μελλοντική έρευνα που να λαμβάνει υπόψη τη δυναμική φύση της αλληλεπίδρασης πεπτιδίου-DNA, αλλά και την ανάγκη πειραματικής επιβεβαίωσης των αποτελεσμάτων.

Abstract

This thesis investigates the interactions between a 12-mer peptide (SVSVGMKPSPRP) and canonical B-DNA, focusing on the mechanisms underlying its non-specific binding, using a combination of advanced computational approaches. For this purpose, 100 potential peptide conformations, which were available through folding molecular dynamics simulations, were docked to a canonical B-DNA molecule using the HDOCKlite tool. The docking energy scores were analyzed, identifying five promising conformations that exhibited significant energy drops. Further docking experiments were conducted using HADDOCK and PyDockDNA tools to refine these models, with the resulting complexes visualized using the Pymol program. Detailed structural analysis revealed the 3 most efficient models that demonstrate the interactions that are formed in the peptide-DNA complex. The binding mechanism was found to be dominated by electrostatic interactions, particularly between the peptide's positively charged residues and the DNA backbone's negatively charged phosphates. Hydrophobic interactions and Van der Waals forces stabilize further the models, along with hydrogen bonding which seem to play a less significant role. While the docking tools did not converge on a specific binding site, they consistently identified the interactions that drive the peptide's non-specific binding to B-DNA. The study highlights the limitations of simulating the peptide in its free form (in the absence of the DNA molecule) and emphasizes the need for future research to consider the dynamic nature of the peptide-DNA complex and also the need for experimental confirmation of the results.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Professor Nicholas Glykos, for his continuous support, guidance, and encouragement throughout the course of my MSc research. His expertise and insightful suggestions have been invaluable in shaping the direction and outcome of this thesis. I am incredibly grateful for the countless hours he devoted to discussions and for his patience in helping me navigate through the complexities of modeling the interaction between a non-specific DNA-binding peptide and canonical B-DNA.

I would also like to extend my sincere thanks to the members of my advisory committee, Associate Professor Vasiliki Fadouloglou and Professor Aristotelis Papageorgiou. Their constructive feedback, thought-provoking questions, and critical insights have significantly contributed to the refinement and completion of this work. I am particularly appreciative of their time and commitment in reviewing my progress and providing helpful suggestions that have enhanced the quality of this thesis.

Furthermore, I am grateful to the entire faculty and staff of the Department of Molecular Biology and Genetics at Democritus University of Thrace (DUTh) for providing an intellectually stimulating environment and access to the resources necessary for my research.

Lastly, I would like to thank my family and friends for their continuous support, understanding, and encouragement throughout this journey. Their belief in my abilities and their constant encouragement have been a source of motivation and strength during challenging times.

This thesis is a culmination of the support, guidance, and collaboration of many individuals, to whom I am deeply indebted. Thank you all for making this accomplishment possible

1.Introduction

A wide range of short peptides are found to exert significant biological activity, since they are found to intervene in the neuro-immuno-endocrine system and regulate crucial cellular processes like the apoptosis, proliferation and differentiation. Among others, they can have antioxidant, antibacterial, antitumor, anti-ageing and anti-inflammatory activities. It appears that the study of the mechanism of action of these short peptides and specifically the way in which they interact with the biological macromolecules like DNA, is a field of great interest and it can possibly serve as a fertile ground in the research and development of new pharmaceutical products (1).

1.1 DNA-binding peptides: a short introduction

It is well established that the double-stranded DNA, in its B-form, can interact with proteins and peptides (2). This may take place in the major or minor groove of dsDNA. In terms of the binding mechanism, two different types can be identified. The first one incorporates the recognition of a unique sequence of the DNA bases (base readout), in this case, hydrogen bonds and hydrophobic contacts are formed between the side chains of the amino acids and the functional groups found in the DNA bases (3). The second type incorporates the recognition of the shape of DNA, by the ligand, which is defined by the structural features of the binding site (shape readout) (5). The B-DNA, the most common form of DNA, is characterized by a major and a minor groove in its structure. In the majority of cases only the major groove can host sequence-specific interactions. In the minor groove A/T and T/A base pairs have a similar pattern of hydrogen bond acceptor and donor groups and this can make it difficult for a ligand to differentiate between them (4). "Shape readout", consists of the identification of the structural features of the binding site in the DNA molecule (core binding site and flanking regions). It depends on the groove's width and the electrostatic

potential, which should "match" with the equivalent characteristics of the ligand. In the minor groove of DNA, this comes down to the recognition of its geometry by positively charged amino acids (Arginine, Lysine, Histidine), that exist in the binding peptide. Biological macromolecules such as proteins, tend to bind in the major groove of ds-DNA, through hydrophobic contacts and by forming hydrogen bonds (3). Several motifs of DNA-protein interactions have been identified, but all of them incorporate globular proteins in a specific conformation (5). On the contrary, small flexible peptides tend to bind to dsDNA non-specifically. However, the minimum number of residues that a peptide should have, in order to be able to recognize a specific DNA sequence, is not clear.

There's a plethora of short peptides exerting several crucial biological activities concerning living organisms, since they can easily travel through cell membranes, including the nucleus (6,7). They can have antibacterial, antitumor, anti-inflammatory, analgesic, antioxidant, and anti-aging activities. They can also play a role in the immune and digestive systems, as well as in the regulation of blood pressure and appetite (8–11). These properties combined with the fact that their metabolites are in general non-toxic, make this a tempting field for the development of new drugs. However, there are some restrictions that should be addressed, such as their low bioavailability especially when they are administered orally, as they tend to be metabolized quickly (12).

1.2 SVSVGMKPSPRP : A non-specific DNA binding peptide

For the purpose of this research, a 12-mer peptide was used, the sequence of which is "SVSVGMKPSPRP". Wolcke J. and Weinhold E. conducted an experiment to demonstrate that this specific peptide has the ability to bind to DNA. To identify the peptide, they employed a

randomized phage display 12-mer peptide library, employing competitive elusion with DNA methyltransferase M TaqI, and subsequently confirmed its binding ability through ELISA testing (13).

Among the confirmed characteristics of this peptide, is that it can specifically recognize tumor-related neovasculature in mice with human tumors, while not affecting normal blood vessels. Additionally, when combined with liposome-carrying doxorubicin, it improved the survival rates of mice with human lung and oral neoplasms. Furthermore, the peptide has shown an affinity for certain antibodies such as those against the envelope protein of Japanese encephalitis virus and monoclonal antibody 2G12, which neutralizes HIV-1. It also binds specifically to carboxypeptidase B from pig pancreas, endothelial cells, cationic amino acid transporter 1, human prostate cancer cell line DU145, Torpedo acetylcholinesterase, and HIV-1 Virion infectivity factor protein. The peptide was found to interact with the Eph family of receptor tyrosine kinases, although the strength of this binding was not evaluated. Numerous publications have indicated that the peptide can adhere to inorganic targets. Notably, it demonstrated a strong and specific affinity for GaN and hydroxyapatite and acted as a template for the synthesis of cobalt-platinum nanoparticles. It was also discovered to bind to single-walled carbon nanotubes (SWNTs) and function as a pigment-binding peptide, potentially serving as a seed for metal nucleation. Moreover, researchers found that the peptide exhibited an ability to bind with different semiconductor materials, including gallium arsenide, gallium antimonide, zinc telluride, zinc selenide, and cadmium selenide. These findings suggest that small polypeptides, like this one with distinct properties, hold potential for integration into the field of materials science and engineering (14).

1.3 Aim and objectives

The purpose of this essay is to model the interaction between a non-specific DNA-binding peptide (SVSVGMKPSPRP) and canonical B-DNA. This study aims to explore and predict the binding mechanisms of this peptide, addressing the current gap in understanding their non-specific interactions with DNA. By utilizing advanced computational docking techniques and incorporating experimental data, this research seeks to reveal the structural and energetic properties of the peptide-DNA complexes. Ultimately, this work will contribute to a more comprehensive understanding of peptide-DNA interactions and their implications in biological processes.

2. Methods

To complete this essay, several steps were undertaken. The initial step involved obtaining the complete sequence of the 12-mer peptide, along with the atomic coordinates necessary for its formation, which were available via folding Molecular Dynamic (MD) simulations. These coordinates are essential as they provide the three-dimensional structures which is important for docking simulations. Consequently, automated docking was performed with HDOCKlite, to investigate the binding of the peptide to a canonical B-DNA molecule. The most promising models were identified and these results were verified further with HADDOCK and PyDockDNA, two other molecular docking tools. Finally, the most efficient models were illustrated with Pymol tool and their binding structure was analyzed.

2.1 Obtaining initial peptide structures: folding molecular dynamics simulations

The initial peptides were acquired via folding molecular dynamics (MD) simulations.

Accurately determining the three-dimensional structures of peptides is crucial for understanding their biological functions and interactions. One of the key challenges in this process is obtaining reliable initial peptide structures. The problem is that there are too many possible conformations for every protein, which rends it impossible to find its native structure (thermodynamically stable state) by random searching, as it is expressed in Levinthal's Paradox. This indicates that the folding process is realized by specific "folding pathways", which include intermediate states (15). The comprehension of how proteins fold, necessitates the characterization of the underlying energy landscapes and the dynamics of the polypeptide chain during the entire folding process. Significant progress has been made towards these objectives due to a surge in interdisciplinary research and advancements in both experimental techniques and theoretical approaches. The development of MD simulations gave insight into the behavior of macromolecules over time, allowing researchers to observe and analyze their folding processes in a more feasible manner (16).

MD simulations consist a computational method that enables the prediction of the position and movement of atoms in a biomolecular system. It is based in the principle that by knowing the initial position of all atoms in a biomolecular system (e.g., protein in water), one can compute the forces each atom exerts on the others. Furthermore, using Newton's laws of motion, a prediction of the future positions of the atoms over time can be made. This is achieved by taking small time steps, continually recalculating the forces on each atom, and updating their

position and velocity accordingly. The outcome is a detailed, three-dimensional "movie" that depicts the atomic configuration of the system at every moment during the simulation period (17).

Folding MD simulations is a specific application of molecular dynamics used to study the folding process of proteins and other biomolecules. These simulations provide insights into how these complex structures form, from their linear polypeptide sequences and how their three-dimensional shapes relate to their biological functions (18).

Simulation Setup:

- Force Fields: The simulations rely on force fields, which are mathematical models used to calculate the forces acting on atoms and molecules. These can be, CHARMM (Chemistry at Harvard Molecular Mechanics), AMBER (Assisted Model Building with Energy Refinement), GROMOS (Groningen Molecular Simulation) and CVFF(Consistent Valence Force Filed). These models determine how atoms within the peptide interact with each other and with their environment (19).
- Initial Configuration: The process begins with an initial configuration of the peptide, which can be obtained from experimental data, homology modeling, or built de novo. Ideally, having a starting structure that closely matches the equilibrium configurations would significantly reduce the time required to stabilize the system during the equilibration phase.(20,21).
- Minimization: As the simulation progresses, the peptide samples different conformations. The goal is to identify stable, low-energy structures that correspond to the peptide's natural folding states. To achieve that it is necessary that the system's energy is minimized using the steepest descent method. The energy should be lowered by adjusting the protein's coordinates in a way that prevents clashes that would lead to an increase of the system's

energy. The minimization process consists only one part of the overall optimization and grands a potential energy minimum without considering kinetic energy. Thus, further analysis is required to fully optimize the system before running the MD simulation. The initial velocities of the atoms should also be addressed.

- Equilibration: this is the first part of an MD experiment. The system starting from the initial configuration is brought to an equilibrium, through monitoring various parameters. This involves gradually bringing the system to a stable state where temperature, pressure, and other parameters are steady. The primary metric which should be assed in order to determine if the system has reached equilibrium the Root Mean Square Deviation (RMSD). RMSD should be stabilized, fluctuating around constant values(20,21).
- Production Phase: This is the actual data collection phase where the dynamics of the system are recorded over a longer period, providing a detailed trajectory of the folding process. The primary distinction of production phase from the equilibration phase is that the main pursuit now is gathering data about the molecule of interest, which is the final aim of the whole experiment. At the end, a trajectory that describes the motion of the molecule is produced (21)

2.2 Docking: programs and scripts used for the analysis

The field of molecular docking has developed significantly over the past decades, propelled by the demands of structural molecular biology and structure-based drug discovery. This progress has been greatly aided by the exponential increase in computer power and the enhanced accessibility to small molecule and protein databases. The aim of automated molecular docking software is to identifying probable binding modes and estimating binding affinity. Typically, molecular docking involves the interaction between a small molecule and a target

macromolecule. Molecular docking has numerous applications in drug discovery, such as conducting structure–activity studies, optimizing lead compounds and identifying potential leads through virtual screening (22).

While the "ab initio" protein-protein docking has advanced significantly, as the collaborative CAPRI (Critical Assessment of Predicted Interactions) experiment indicates, the development of docking methods for modeling protein-DNA interactions has not kept pace, since it is characterized by some challenges. This has to do with the limited information available to define the DNA-binding interface, coupled with the intrinsic flexibility of DNA. There is still no general recognition code and the global conformation of the DNA can play an important role in modulating the eventual interaction surface. DNA frequently undergoes significant conformational changes when binding to a protein, which can substantially modify the shape of the interaction surface. As a result, the total conformational space that must be explored to identify favorable conformations expands significantly. DNA flexibility can be divided into global and local components. Global flexibility in DNA primarily involves two main movements: bending and twisting. This flexibility arises from the conformational changes in the flexible base pairs and the sugar-phosphate backbone. Incorporating both global and local flexibility into DNA during docking, while preserving the relevant conformation, poses a significant challenge in protein-DNA docking. Both of these two challenges can be addressed, using docking softwares like HADDOCK (High Ambiguity Driven DOCKing), which was used in this research as well (23).

The docking process consists of two fundamental steps: predicting the conformation, position, and orientation of the ligand within binding sites (commonly called the pose), and evaluating these conformations via a scoring function. The primary goal of molecular docking is to predict the structure of the ligand-receptor complex. Identifying the binding site

location prior to the docking process greatly enhances its efficiency. In this case the binding site isn't known so a blind docking is performed (24).

2.2.1 HDOCKlite

The docking program which was first utilized was HDOCKlite. HDOCKlite is a streamlined and efficient version of the HDOCK protein docking server, which is widely used for molecular docking simulations. It is designed to provide a fast and user-friendly interface for predicting protein-protein and protein-peptide, protein-DNA interactions. HDOCKLite uses scoring functions to evaluate the binding affinity and stability of the predicted molecular complexes (25).

The workflow is realized in the following steps. The data are inputted, while for proteins both sequences and structures are accepted, for DNA/RNA molecules only structure are applicable up until now, since modeling DNA/RNA structures from their sequence consists a challenge. Subsequently, based on the data inserted a sequence similarity search is realized, in which the PDB sequence database is investigated in search of homologous sequences for the receptor and ligand molecules respectively. To ensure computational efficiency and minimize service interruptions, a local copy of the PDB database is kept on the HDOCK web server and it is updated on a monthly basis. HHSuite package is employed for proteins, while for DNA/RNA this is realized by the FASTA program. This procedure leads to two homologous templates (receptor and ligand). Subsequently, the two sets of templates are contrasted and if they have the same PDB code, a common template will be chosen for the set of receptor and ligand. When there's no overlap and multiple templates are available, the one with the greatest sequence coverage, highest sequence similarity, and best resolution will be chosen. Using the selected templates, models are constructed with MODELLER, while sequence alignment is performed with ClustalW. Now, that the structures

are modeled, the next and final step is the actual docking process. This process leads to the best binding models (25).

The structures of the molecules to be docked were provided in PDB format. HDOCKlite simulates the binding process, predicting how the molecules will interact and form a stable complex. It generates output files that include the predicted docked complexes, binding scores, and detailed information about the interactions. It is a useful program in interpreting experimental data by providing structural models of molecular complexes, helping to elucidate the mechanisms of molecular interactions.

The analysis was done in the following steps:

- The pdb file for the B-DNA molecule was generated through NAMOT software, which is a tool used for designing DNA nanostructures
- Initial peptide structures were obtained in PDB form, from a previous reported folding molecular dynamic simulation performed by Alexiadou D. (26). The structures correspond to the representative members of 100 clusters identified via dihedral PCA (dPCA), which is used to analyze the conformational dynamics of the peptide during MD simulations. During the MD simulation the dihedral angles (φ, ψ) of the peptide are tracked over time, which creates a large dataset of angles corresponding to the peptide's motion, then dPCA identifies the main modes of motion that capture the most significant structural variations during the simulation, this reduces the dimensionality of the data while retaining the most important dynamic information (27–29). The study incorporated CARMA (30), which is developed to assist in the analysis of molecular dynamics trajectories, and GRCARMA (Graphical CARMA)(31) which is designed to fully automate tasks

in this process. These tools led to the interpretation and evaluation of the dPCA results.

- Prepare/Input Files: the structures in their PDB form, were placed in the same directory and sorted numerically.
- The docking parameters were configured and the files in their PDB form were inputted in HDockLite. Subsequently, the docking process begun and every PDB file yielded a corresponding .dat file for every docking model that was created.

```
##

#Iterate over each PDB file matching the pattern

for file in /home/ubuntu/DNAbp/running/peptide/representative*.pdb;
do
        #Run the Hdock program with B-DNA.pdb and the current file

         ./hdock B-DNA.pdb "$file" -out Hdock.out

        #Extract the base name of the current file without path and
        #extention
         name=$(echo "$file" | awk -F "." '{print$1}' | awk -F "/"'{pr
        int$NF}')

#Print the extracted name

        echo "$name"
##
```

```
##

#Extract the 7th column from Hdock.out starting from the 6th line
#onward and save it to the output directory with a .dat extension

tail --lines=+6 Hdock.out | awk '{print$7}' >
/home/ubuntu/DNAbp/running/HDOCKlite-v1.1/res/"${name}.dat"

done

##
```

The output files were consequently reviewed to interpret the docking results and evaluate the predicted complexes. The output file of HDOCKlite contains a table 7 columns, the first 6 columns correspond to the translational/rotational parameters which are used to define the

position and orientation of the peptide in 3D space, the 7th represents the binding affinity of each pose (binding free energy). Only the 7th column was chosen to be saved in order to use this data for the creation of energy score plots, that are useful to evaluate the models that have been generated and lead the research to the peptide representatives with the most efficient binding poses.

2.2.2 HADDOCK (High Ambiguity Driven DOCKing)

HADDOCK is an information-driven flexible docking software designed to model the complex structure of biomolecules. Unlike traditional docking programs that rely solely on geometric and energetic considerations, HADDOCK incorporates experimental data to guide the docking process. It is particularly useful for docking proteins, nucleic acids, and small molecules. Among all the docking methods participating in CAPRI, only HADDOCK employs a genuine data-driven strategy. It is driven by experimental data, which can be derived from mass spectrometry and NMR experiments. In the case that this experimental information isn't available or sparce, bioinformatic techniques can be employed (25).

In contrast to many other docking programs, HADDOCK permits conformational changes in the molecules during complex formation, concerning not just the side chains but also the backbone (25). Global and local DNA flexibility is incorporated into the docking process by permitting the DNA sugar-phosphate backbone and base pairs to explore different conformations during a semi-flexible refinement phase. Additionally, docking is initiated from a library of pre-generated DNA structures that exhibit various levels of conformational flexibility, which facilitates the exploration of a broader conformational space (23).

The docking protocol involves three stages: rigid body energy minimization, the semi-flexible optimization in torsion angle space and the optimization in explicit solvent. Non-structural experimental data are

used to guide the docking process throughout the rigid-body energy minimization, as well as the semi-flexible and water refinement stages. After each stage, structures are evaluated and ranked based on their scores, with the top structures advancing to the next stage. The HADDOCK score combines Van der Waals, electrostatic, desolvation, and restraint violation energies, along with buried surface area, into a weighted sum. The user can adjust various parameters for each stage, including the number of structures, scoring weights, and aspects of the docking protocol like temperature and force constants (25).

2.2.3 PyDockDNA

PyDockDNA is a specialized computational docking tool designed for predicting the interactions between proteins and DNA. It extends the capabilities of the PyDock framework, focusing specifically on the unique challenges and characteristics of protein-DNA docking. It incorporates unique features to handle the flexibility and complexity of DNA structures during the docking process. It incorporates specific considerations for electrostatic interactions and desolvation energy. Its efficient and accurate predictions make it an invaluable resource for researchers studying protein-DNA interactions, gene regulation, and related fields in molecular biology (32).

The files that contain the coordinates of the peptide and the DNA molecule are inserted in PDB format. As standard, the program employs a 0.7 Å grid cell size, 1.3 Å surface thickness, 12° rotation sampling, and retains the top three poses for each rotation. For every target, there are 10,000 docking poses that are generated.

The docking poses are evaluated and ranked using a scoring function that incorporates electrostatic, desolvation, and van der Waals energy components. This function is derived from the original PyDock scoring function used for protein-protein docking but it incorporates atom types

for nucleotides from the Amber94 force field, which is appropriate for scoring models of protein-DNA complexes.

The output includes several formats of docking models:

- the 3D structures of the top 10 best-scoring docking models
- the PDB files that correspond to the top 100 models, which are available for download
- the rotation vectors are available and can be used to create up to 10,000 docking poses.

The outcome of the docking can be presented as a plot that shows the distribution of energy values across all docking poses

2.3 Other programs

Pymol

PyMOL is a powerful and widely used open-source molecular visualization tool in the field of computational biology and bioinformatics. It allows users to create high-quality 3D images of small molecules and macromolecules such as proteins and nucleic acids. PyMOL provides detailed visualization of molecular structures, enabling to explore the 3D configurations of molecules, view atomic details, and understand molecular interactions. It can also generate high-resolution images and animations. It supports a Python-based scripting interface, allowing the automation of tasks, complex analyses, and customized visualizations. One of the major advantages of Pymol is its adaptability via scripts and plugins, created by external developers, in order to enhance its functionality. As a result, plugins have been created for PyMOL over the years to support various well-known structural bioinformatics methods, including sequence analysis, molecular docking, molecular dynamics, structure-function relationship analysis, protein structure prediction, and virtual screening (33).

3. Results & Discussion

3.1 Initial structures and automated docking with HDock

A total of 100 potential structures of the12-mer peptide, were acquired in their PDB form via folding MD simulations. Each one of these contains the atomic coordinates that correspond to a different potential conformation of the peptide. Their ability to bind to B-DNA was investigated via automated docking with the HDocklite tool. Every structure was docked to the B-DNA molecule and the data concerning the structure and the energy score of the model created was then collected, in order to be analyzed.

3.2 Identification of promising solutions

The docking analysis was contacted and now the results should be evaluated. For this purpose, it's very useful to generate energy score plots (fig. 1), that are used to evaluate the docking results between the B-DNA and each conformation of every peptide representative, as they were created by HDOCKlite.

These plots are crucial for understanding the binding interactions and affinities. They illustrate the binding affinity of each binding pose of the peptide. More specifically the x-axis represents the different docking poses or conformations, while the y-axis represents the energy score, which is often a combination of various energy components (van der Waals, electrostatic, desolvation, etc.).

```
##
src="/home/ubuntu/DNAbp/running/res/"
for file in "$src"*.dat; do
        echo $(basename "$file")
         head -20 "$file" | plot
done
##
```

Since lower energy scores indicate more favorable poses, from each docking which was conducted only 20 docking poses with the lowest energy scores were chosen to be incorporated in the created plots. Lower energy scores indicate more favorable and stable docking poses. These poses are likely to have better binding affinity and interactions. On the contrary higher energy scores suggest less favorable interactions and are usually less relevant for identifying potential binding modes. Furthermore, a sharp decrease in energy scores usually points to a high-affinity binding pose, while multiple local minima can suggest several potential binding sites. Through analyzing all the plots that are generated by the energies calculated by HDOCKlite, the most promising plots can be quickly identified. These are the plots that correspond to representatives 9, 24, 42, 59 and 73 as presented bellow.

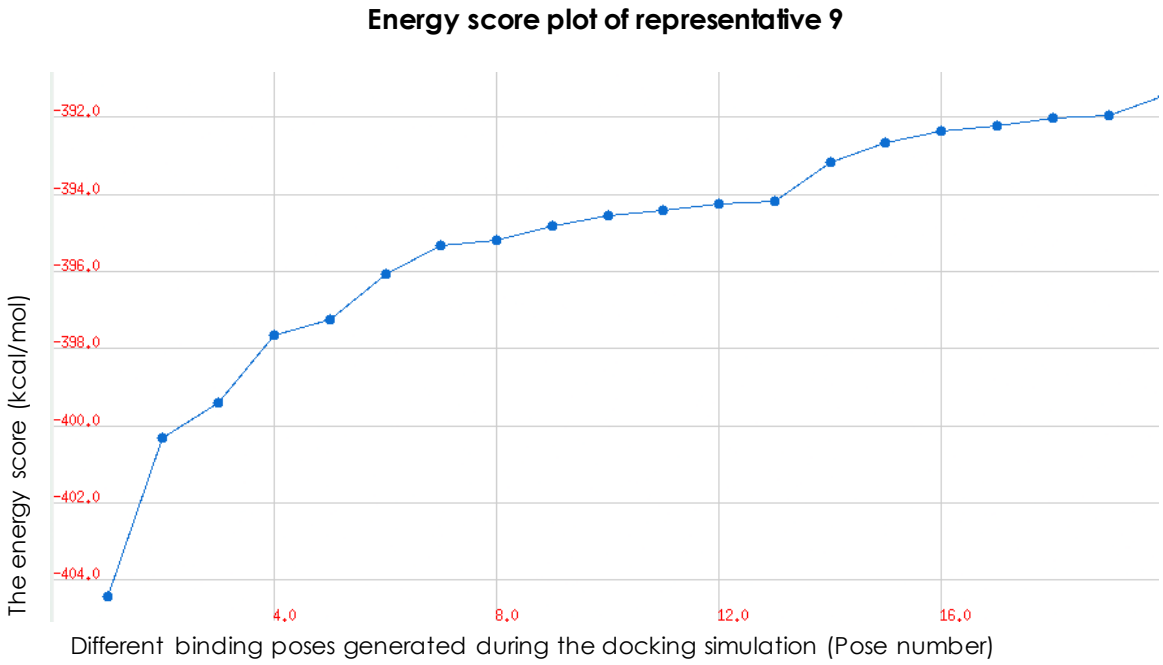

Figure 1. This is the plot that refers to representative 9, in this diagram a sharp decrease in binding free energy can be spotted. The first pose has an energy score -404 kcal/mol, indicating a very favorable interaction. As the pose number increases, the energy score rises, indicating less favorable interactions.

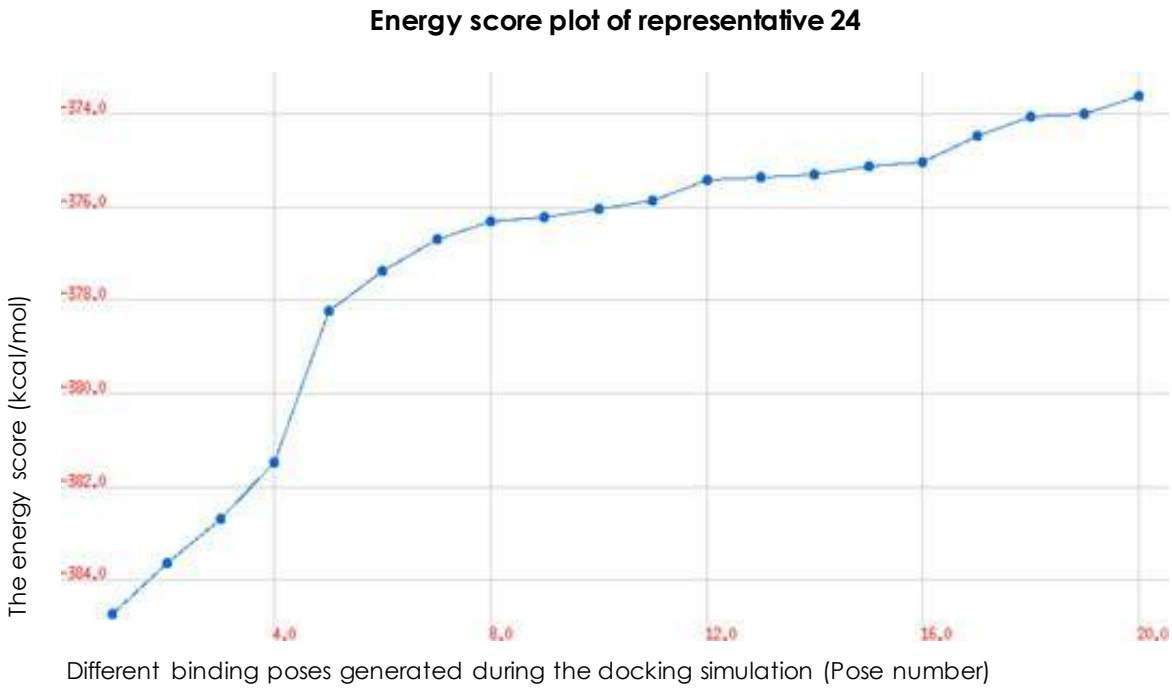

Figure 2. This is the energy score plot of representative 24, a sharp decrease in energy can be spotted, which leaves the 4 most favorable binding poses in the left part of the diagram.

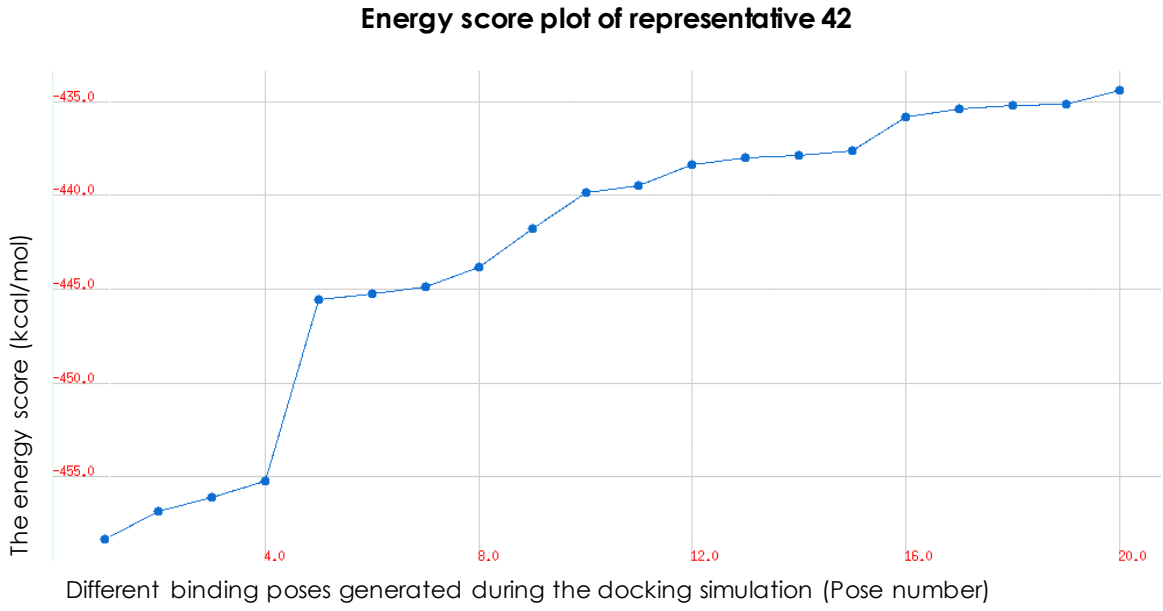

Figure 3. The energy score plot of representative 42, a sharp decrease in energy can be observed and the 4 most favorable poses, with the lower energy can be spotted at the left of the diagram.

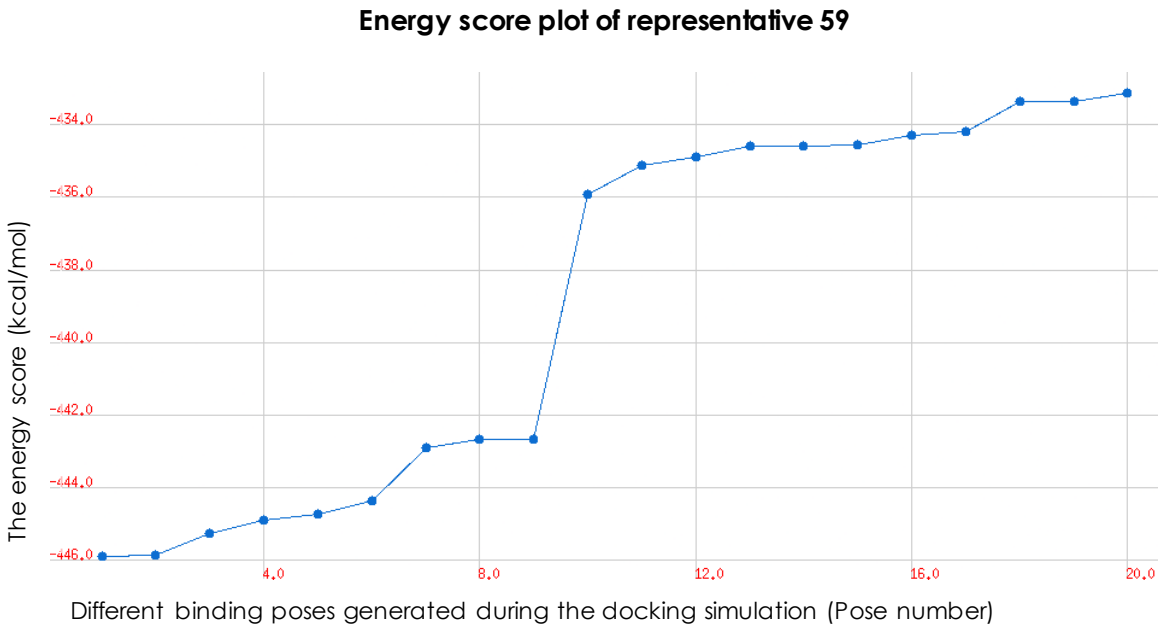

Figure 4. The energy score plot of representative 59, a significant drop in energy can be spotted, that separates the most favorable poses at the left side of the diagram.

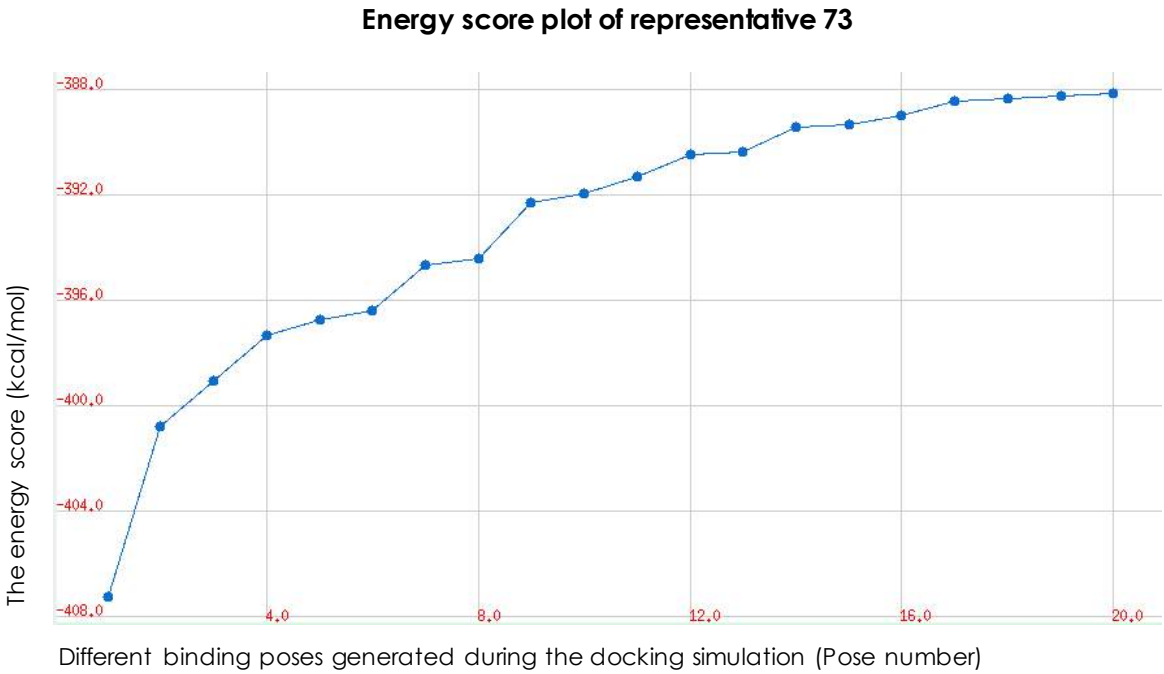

Figure 5. The energy score plot of representative 73, a sharp drop in energy can be observed, which leads to the one and only favorable pose at the bottom left of the diagram.

To put in contrast the following plot correspond to representative 22, it's slope doesn't present any sharp decrease which indicates no favorable binding poses are present in this binding model.

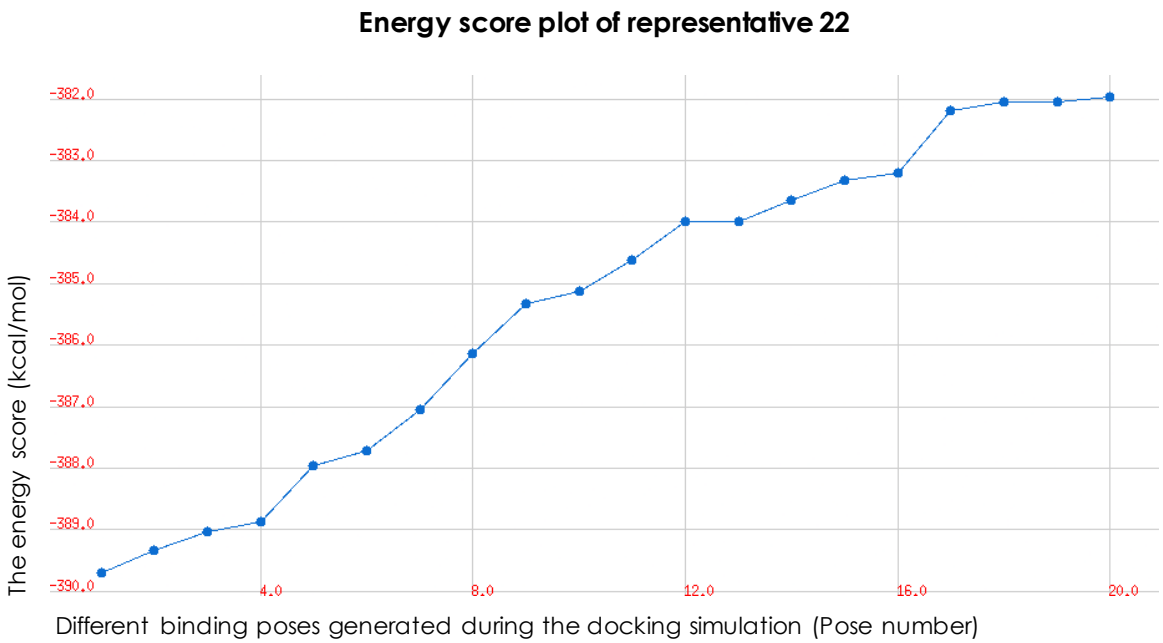

Figure 6. The energy score plot of representative 22 doesn't present any sharp decrease in the free binding energy, which indicates no favorable binding poses are present in this model.

3.3 Verification using HADDOCK & PyDockDNA

The peptide structures that were docked with HDOCKlite and presented the most efficient energy score plots, which are the most promising candidates for strong molecular interactions, should be now verified further by using two other molecular docking tools, HADDOCK and PyDockDNA.

Upon submitting the B-DNA.pdb file to the HADDOCK server, it was found out that some of the DNA atoms weren't recognized properly. More specifically, the hydrogen atoms in the form of H2'1, H2'2, H5'1 and H5'2

weren't compatible with this tool. They were replaced by H2A, H2B, H5A and H5B respectively.

```
##

input_file = 'B-DNA.pdb'
output_file = 'B-DNA2.pdb'

# Read the content of the input file

with open(input_file, 'r') as file:
    lines = file.readlines()

# Open the output file for writing

with open(output_file, 'w') as file:
    for line in lines:

        # Replace atom names as specified

        line = line.replace('H2\'1', 'H2A ')
        line = line.replace('H2\'2', 'H2B ')
        line = line.replace('H5\'1', 'H5A ')
        line = line.replace('H5\'2', 'H5B ')

        # Write the modified line to the output file

        file.write(line)

print(f"Atom names have been updated and saved to {output_file}.")

##
```

Subsequently, the atoms "HTER" and "HCAP", which are not essential for the DNA structure, aren't recognized and should be removed. At the same time the numbering of the file must be maintained.

```
##

input_file = 'B-DNA2.pdb'
output_file = 'B-DNA3.pdb'

def process_pdb_line(line, serial_number):
    """Process and format a line of the PDB file."""
    if line.startswith('ATOM'):

        # Keep the original formatting and just update the serial
        #number

        return f"{line[:6]}{serial_number:5}{line[11:]}"
    else:
        return line

def remove_and_renumber_pdb(input_file, output_file):
```

```
    """Remove lines with 'HTER' or 'HCAP' and renumber 'ATOM' records
without disturbing alignment."""

        with open(input_file, 'r') as infile:
        lines = infile.readlines()

    # Filter out lines with 'HTER' or 'HCAP' and prepare to renumber

       filtered_lines = [line for line in lines if not
       (line.startswith('ATOM') and ('HTER' in line or 'HCAP' in
       line))]

    # Prepare output with renumbered ATOM lines

        with open(output_file, 'w') as outfile:
        serial_number = 1
        for line in filtered_lines:
            if line.startswith('ATOM'):
                outfile.write(process_pdb_line(line, serial_number))
                serial_number += 1
            else:
                outfile.write(line)

    print(f"Filtered and renumbered PDB data has been written to
{output_file}")

##
```

As far as the peptide pdb file a problem with the residue sequence number of the atoms exists. It seems that there are atoms with the same residue number, something that should be corrected.

```
##

def correct_pdb(input_pdb, output_pdb):
    with open(input_pdb, 'r') as file:
        lines = file.readlines()

    corrected_lines = []
    residue_dict = {}

    for line in lines:
        if line.startswith('ATOM') or line.startswith('HETATM'):
            atom_name = line[12:16].strip()
            residue_num = int(line[22:26].strip())
            chain_id = line[21].strip()
            residue_id = (chain_id, residue_num)

            if residue_id not in residue_dict:
                residue_dict[residue_id] = {}

            if atom_name not in residue_dict[residue_id]:
                residue_dict[residue_id][atom_name] = 1
            else:
                suffix = residue_dict[residue_id][atom_name]
                new_atom_name = f"{atom_name}{suffix}"
```

```
                line = line[:12] + new_atom_name.ljust(4) + line[16:]
                residue_dict[residue_id][atom_name] += 1

        corrected_lines.append(line)

    with open(output_pdb, 'w') as file:
        file.writelines(corrected_lines)

input_pdb = 'representative9.pdb'  # Input PDB file
output_pdb = 'rep9.pdb'  # Output PDB file

##
```

3.4 Structural analysis of promising solutions and implications for non-specific DNA binding.

The most promising models subsequently underwent a comprehensive structural analysis, aiming to understand the mechanism of the nonspecific binding of the 12-mer to the B-DNA molecule. For this purpose, the Pymol software was used to illustrate the best fitting models based on the docking findings.

The analysis reveals that the peptide primarily binds to the DNA backbone through electrostatic interactions. Positively charged residues, such as lysine (K) and arginine (R), play a significant role in these interactions by forming electrostatic attractions with the negatively charged phosphate groups of the DNA. This binding mechanism is non-specific, driven by the uniform negative charge of the DNA backbone rather than specific base pair sequences. Additionally, hydrogen bonding, van der Waals forces and hydrophobic interactions contribute to the stabilization of the peptide-DNA complex.

Binding model of representative 9

The results of the docking analysis led to the most efficient binding model of representative 9, as illustrated in Figure 7. The 12-mer peptide binds to the B-DNA molecule non-specifically.

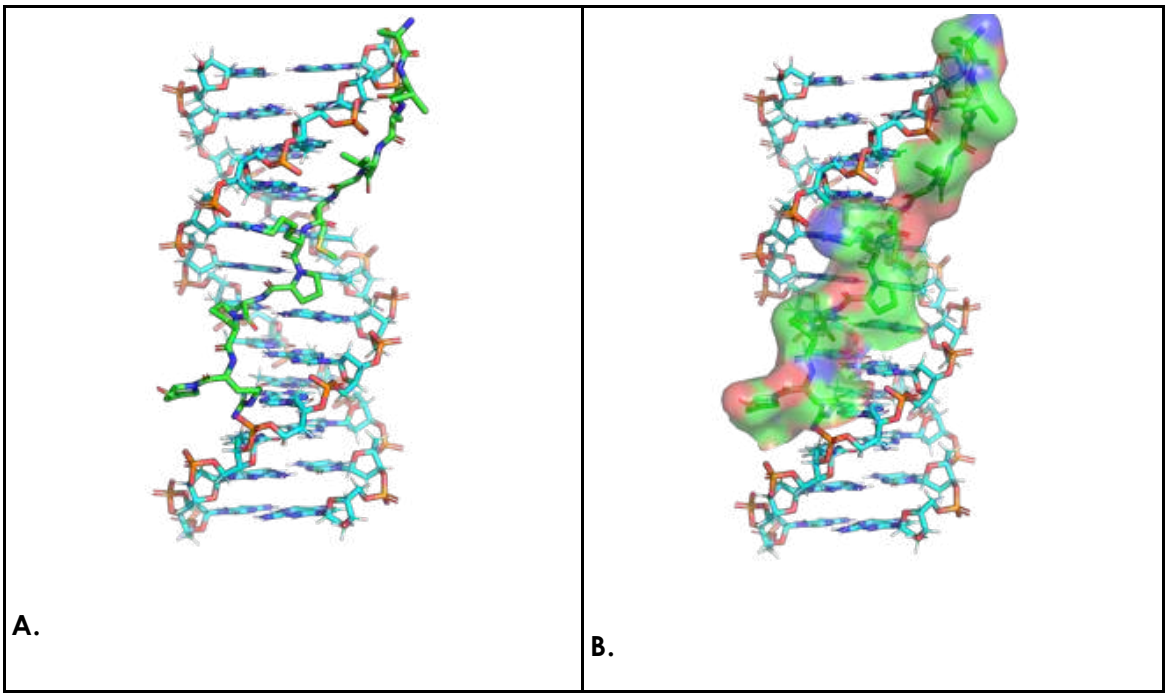

Figure 7. (A) This is the most efficient model of representative 9 illustrated using Pymol, the peptide can be identified in green color, while the B-DNA molecule in turquoise. **(B)** The same model illustrating the spatial arrangement of the peptide in the binding model.

Electrostatic attraction occurs between the positively charged amino group of Lys7 and the phosphate group of the DNA backbone. A similar electrostatic interaction is observed between the positively charged guanidinium group of Arg11 and a negatively charged phosphate group of the B-DNA. Furthermore, the hydrophobic side chain of Met6 forms Van der Waals interactions with the DNA bases in the core of the B-DNA molecule. Additionally, the non-polar ring of Pro8 enhances the stability of the peptide-DNA complex interacting hydrophobically with the methyl group of a Thymine residue of the DNA molecule.

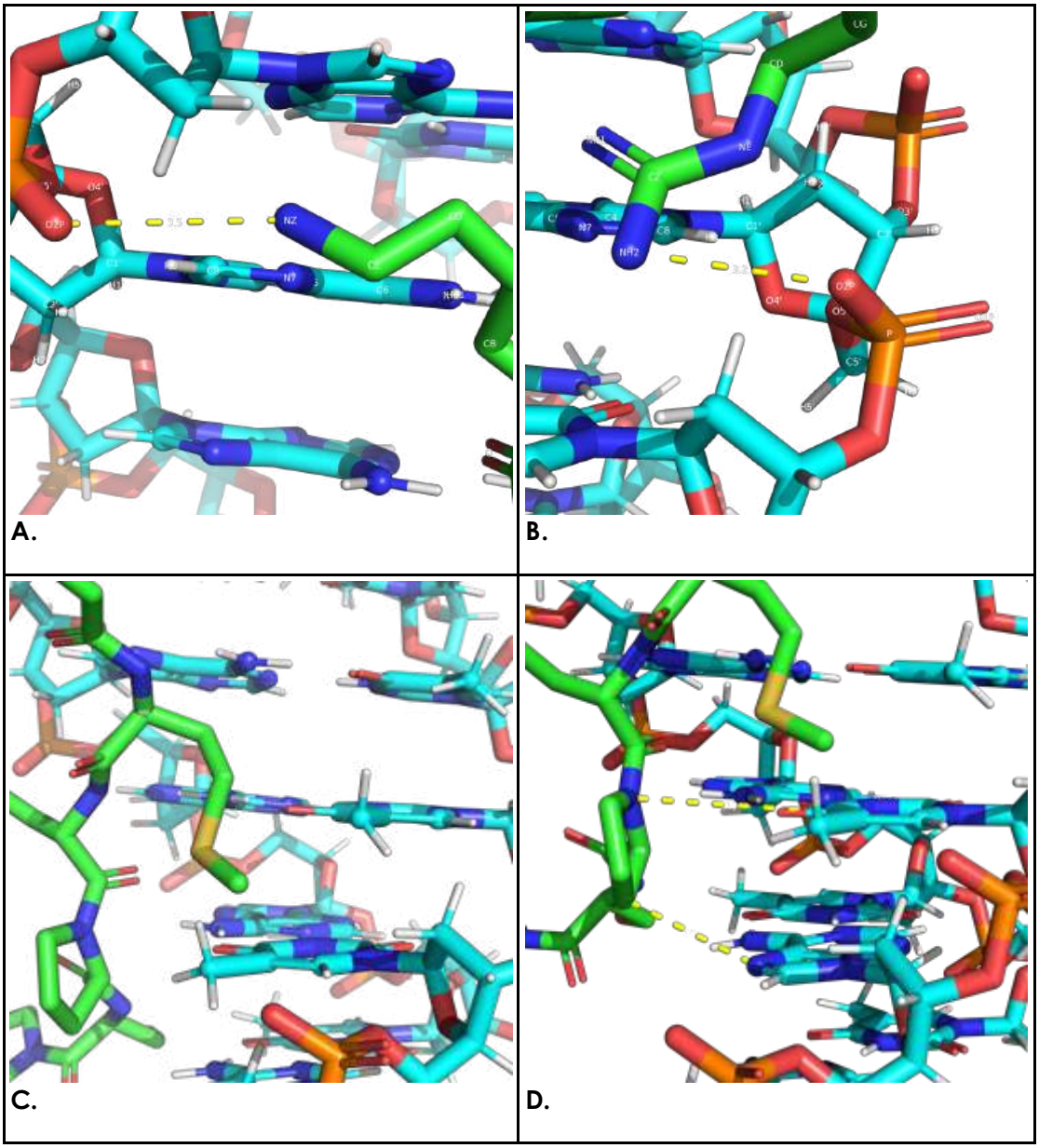

Figure 8. Interactions that stabilize the binding of representative 9 (12-mer) to the B-DNA **(A)** Electrostatic attractions between the positive charged amino group of Lys7 and the phosphate group of the DNA backbone. **(B)** Electrostatic interactions between the positive charged guanidinium group of Arg11 and a negatively charged phosphate group of B-DNA. **(C)** The hydrophobic side chain of Met7 forms Van der Waals interactions with the bases of DNA. **(D)** The non-polar ring of Pro8 contribute to the stabilization of the peptide-DNA complex interacting hydrophobically with the methyl-group of a thymine residue in the DNA molecule

Binding model of representative 24

The most efficient binding model of representative 24, is illustrated by Pymol tool in Figure 9.

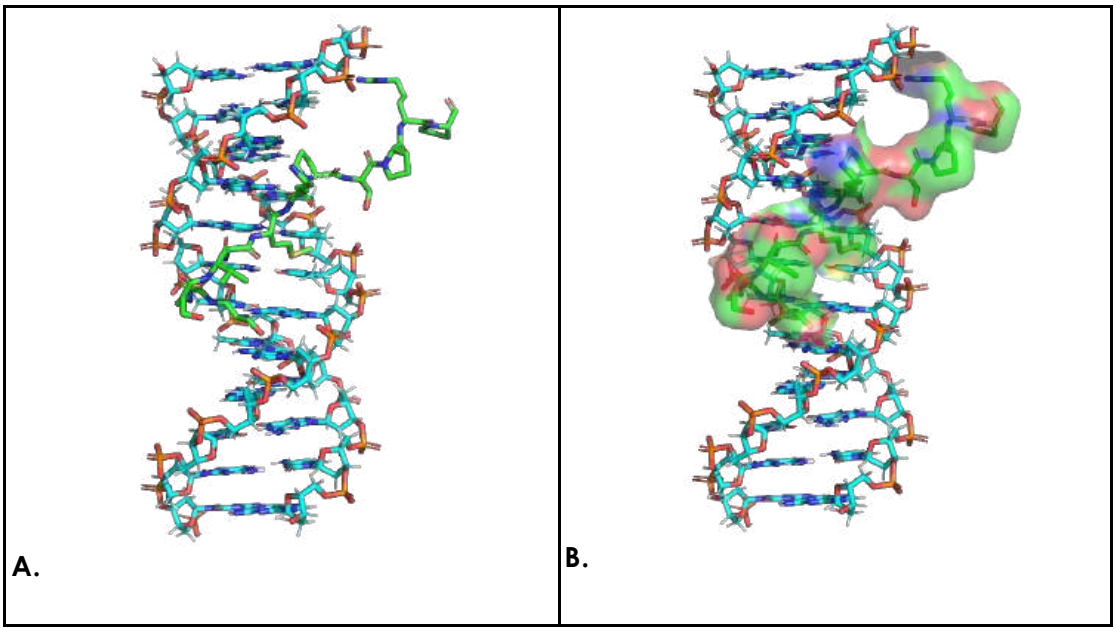

Figure 9. (A) This is the most efficient model of representative 24 illustrated using Pymol, the peptide can be identified in green color, while the B-DNA molecule in turquoise. **(B)** The same model illustrating the spatial arrangement of the peptide in the binding model.

The binding of the 12-mer peptide is mediated through electrostatic interactions between the positively charged amino group of Lys7 and the negatively charged oxygens of two phosphate groups in the DNA backbone. Similar electrostatic interactions are formed, as well, between the Arg11 residue and the negatively charged phosphate group of B-DNA. Furthermore, there are hydrophobic interactions between the Met7 residue and the B-DNA molecule, the nonpolar side chain of methionine interacts with the hydrophobic regions of the DNA (aromatic rings). These interactions exclude water molecules and are stabilized by van der Waals forces, contributing to the overall stability of the model. Another hydrophobic interaction forms between the nonpolar side chain of proline and the hydrophobic core domain of the DNA molecule.

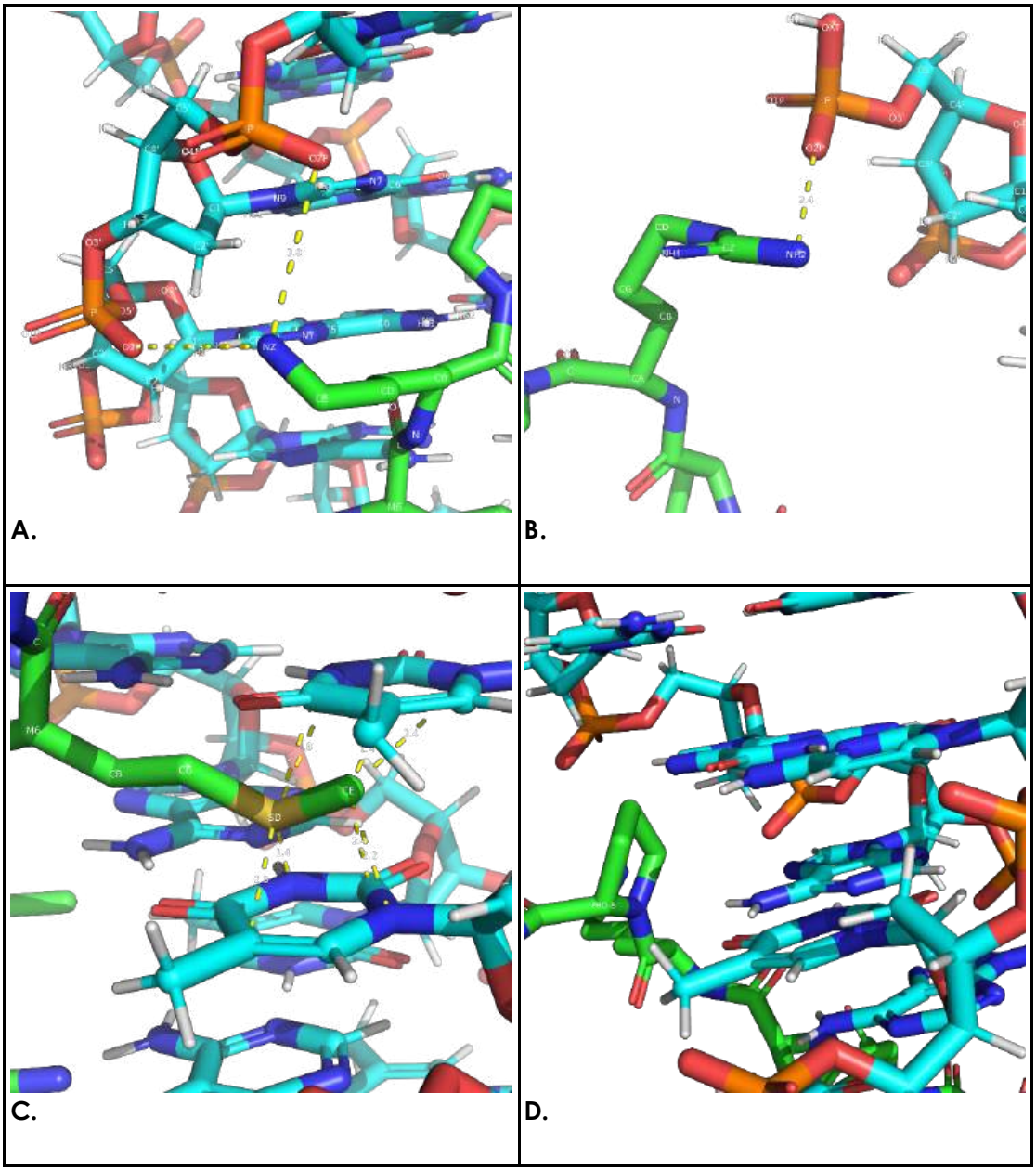

Figure 10. These are the interactions that mediate the non-specific binding of the 12-mer to the DNA molecule, for representative 24 **(A)** electrostatic interactions are developed between the positive charged amino-group of Lys7 and the negative charged oxygens of two consecutive phosphate groups of the DNA backbone. **(B)** Electrostatic interactions occur between the positively charged guanidinium group of Arg11 and a negatively charged phosphate group of B-DNA. **(C)** The non-polar side chain of Met7 interacts with the hydrophobic aromatic rings in the molecule of DNA **(D)** Van der Waals interaction are developed between the proline non-polar side chain and the bases of the DNA molecule.

Binding model of representative 73

The docking analysis led to the most efficient binding model of representative 73 (Figure 11).

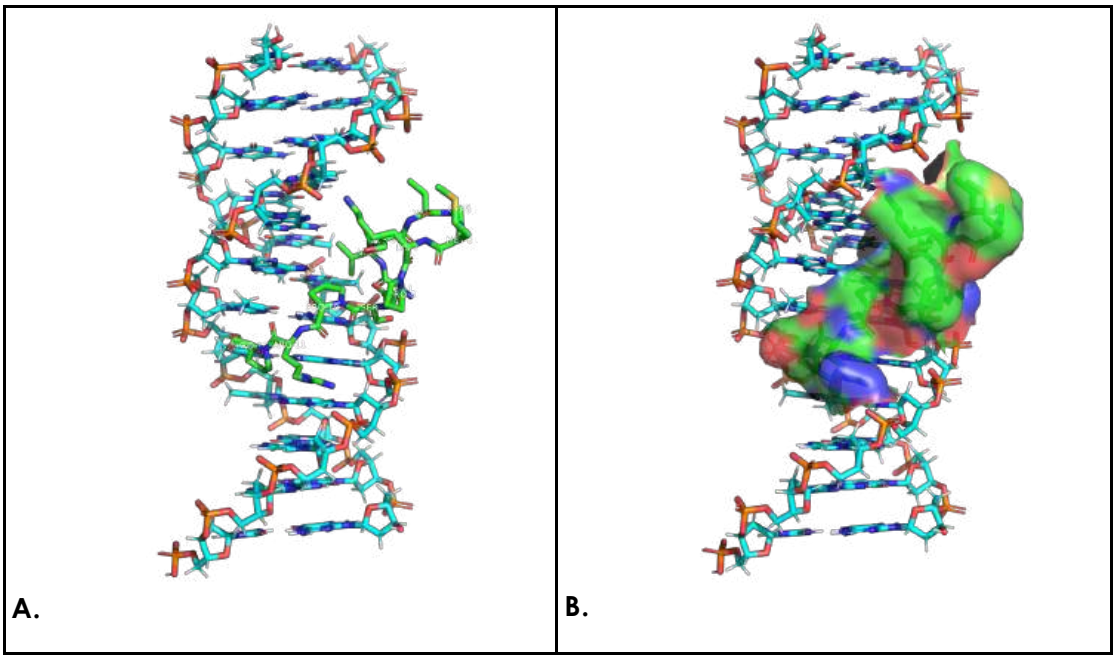

Figure 11. (A) This is the most efficient model of representative 73 illustrated using Pymol, the peptide can be identified in green color, while the B-DNA molecule in turquoise. **(B)** The same model illustrating the spatial arrangement of the peptide in the binding model.

This model is stabilized through electrostatic interaction which are formed between the positively charged amino group of Lys7 and the negatively charged oxygens of two neighboring phosphate groups in the DNA backbone. electrostatic interactions occur between the positively charged guanidinium group of Arg11 and a negatively charged phosphate group in B-DNA. Hydrophobic interactions, stabilize further the complex, taking place between the non-polar side chain of Val4 and the hydrophobic aromatic rings within the DNA molecule, while similar interactions occur between the non-polar ring of Pro8 and the hydrophobic aromatic domain of the DNA. Additionally, a hydrogen bond is developed between the side chain of Ser1 residue and the phosphate group of the backbone of the DNA.

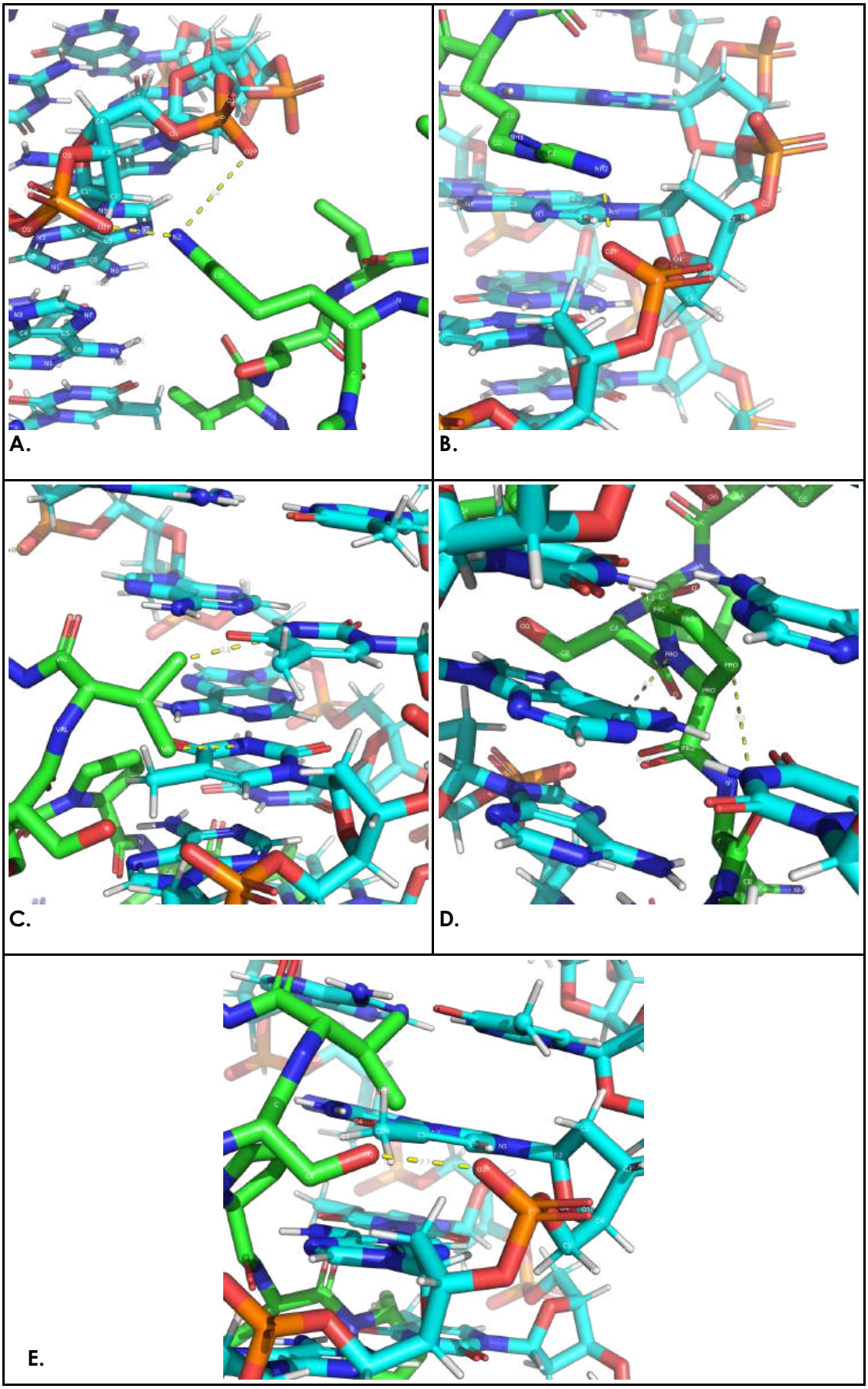

Figure 12. The binding interactions of representative 73. **(A)** Electrostatic interactions occur between the positively charged amino group of Lys7 and the negatively

charged oxygens of two adjacent phosphate groups in the DNA backbone. **(B)** Similar electrostatic interactions are formed between the positive charged guanidinium group of Arg11 and a negatively charged phosphate group of B-DNA. **(C)** Hydrophobic interactions between the non-polar side chain of Val4 and the hydrophobic aromatic rings formed by the DNA bases. **(D)** The non-polar side chain of Pro8 is engaging in hydrophobic interactions with the aromatic groups of the DNA. **(E)** The hydroxyl group of ser1 can donate a hydrogen atom to the oxygen atom in the phosphate group of the DNA backbone, leading to hydrogen bond formation.

The 12-mer peptide binds non-specifically to the B-DNA, which means that it can intervene in any part of its molecule, without recognizing any pattern of the base sequence, using the same electrostatic interactions and its hydrophobic properties it could bind in various other parts of the DNA molecule. For this reason, the models that were produced cannot be conclusive in describing the binding of the 12-mer peptide.

Comparing the 3 binding poses (representatives 9, 24, 73), as it illustrated in Figure 13, a difference in the extension of each molecule can be spotted. Representative 9 seems to be the most stretched molecule, what's more, the molecule exhibits a curvature in its conformation, which allows it to interact with the major groove of the B-DNA molecule, forming Van der Waals interactions, while simultaneously interacting through electrostatic interactions with the phosphate backbone of the B-DNA. Representative 73 presents the most compact form, while representative 24 rests in between them in terms of extension.

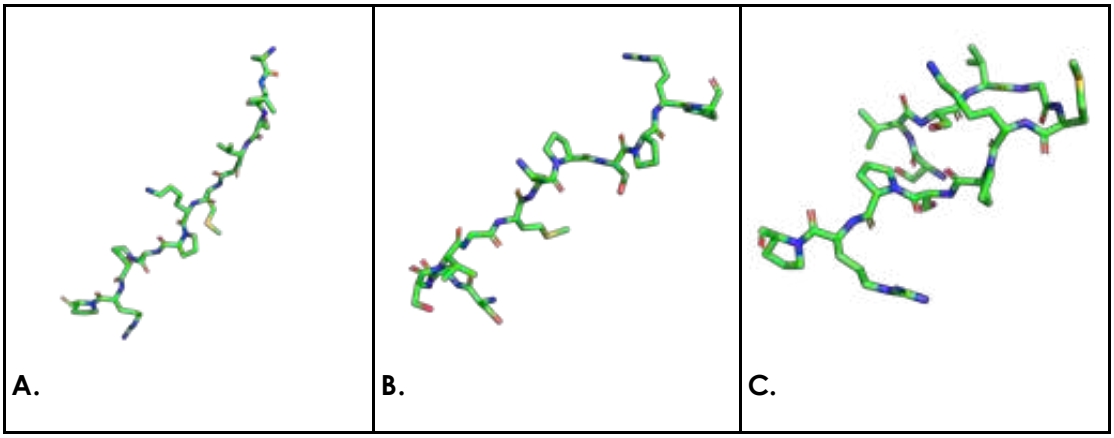

Figure 13. In this figure the peptide binding conformations are compared. **(A)** This is the binding pose of representative 9, it seems the most extended one compared to the

other 2. **(B)** This is the binding pose of representative 24 which presents a less extended conformation compared to representative 9. **(C)** This is the binding pose of representative 73 which is by far the most compact one compared to the rest.

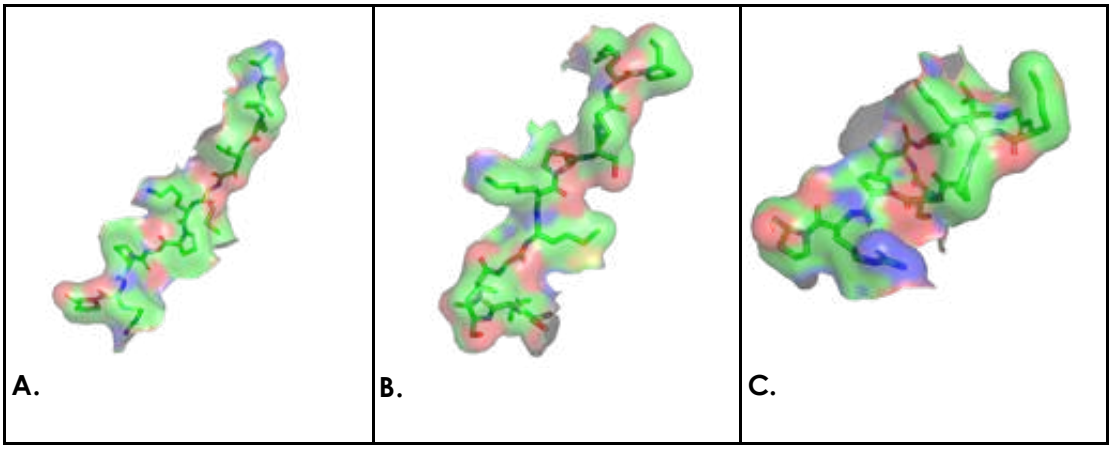

Figure 14. The 3 binding poses illustrated with their molecular surface, in a more "schematic view". **(A)** Representative 9, **(B)** representative 24, **(C)** representative 73.

The binding pose of representative 73 appears to form hairpin, a sharp turn in the backbone of the molecule can be observed between Gly5, Met6 and Lys7 residues. The unique conformation of representative 73 is stabilized through hydrophobic interactions between the side chain of Val2 and Pro10 which interact hydrophobically, facing each other. Furthermore, a non-linear hydrogen bond is formed between the hydroxyl group of the side chain of Ser9 and the free amino group of Ser1, while another non-linear hydrogen bond is formed between the hydroxyl group of the side chain of Ser3 and the carbonyl oxygen that belongs to the Pro8 residue (Figure 15).

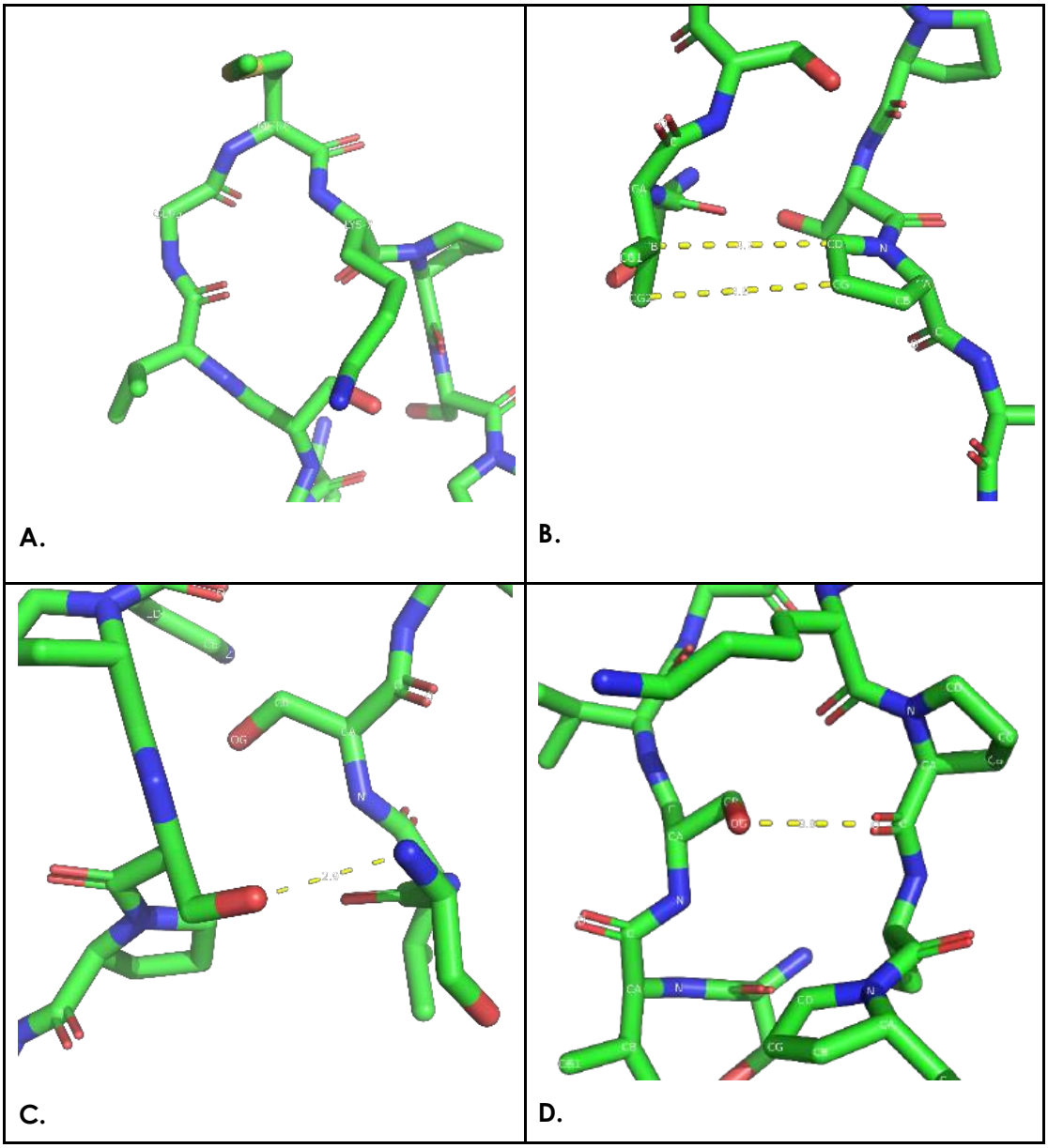

Figure 15. The binding pose of representative 73 which is characterized by a "compact" conformation and has some unique features: **(A)** a sharp turn in the backbone of the peptide chain between Gly5, Met6, and Lys7, leading to the formation of a hairpin. **(B)** The side chain of Val2 faces the side chain of Pro10 and interacts with it hydrophobically. **(C)** A non- linear hydrogen bond is formed between the hydroxyl group of the side chain of Ser9 and the free amino group of Ser1, since these two residues face each other. **(D)** Another non-linear hydrogen bond is formed between the hydroxyl group of side chain of Ser3 and the carbonyl oxygen at the Pro8 residue.

As shown in table 1 below, comparing the total binding energy of the three models, representative 73 with total binding energy of −73.811

kcal/mol is the most efficient in terms of binding, as it has the most negative energy value, indicating the strongest and most stable interaction with the B-DNA molecule. This suggests that the most compact conformation is the most favorable for binding.

Representative 9	Representative 24	Representative 73
-64.096 kcal/mol	-68.985 kcal/mol	-73.811 kcal/mol

Table 1. This table refers to the total binding energy of each binding model, excluding the desolvation energy. It is a combination of the electrostatic and Van der Waals energies.

4. Summary & Conclusions

This thesis demonstrates a comprehensive analysis that aims to model the interactions between a small 12-mer peptide (SVSVGMKPSPRP) of significant interest and canonical B-DNA. While Wocke and Weinhold had already proven that the peptide binds non-specifically to DNA, this work intends to delineated the key factors driving the non-specific binding of the peptide to DNA. This was achieved by leveraging advanced computational docking techniques and integrating experimental data.

Through folding molecular dynamics simulations, 100 different potential conformations (representatives) of the peptide were available. An automated docking process was subsequently contacted, by using HDOCKlite tool, where each one of these structures were docked to a B-DNA molecule. The energy score of each docking session was used to determine promising binding poses. Only representatives which exhibit a sharp decrease in the energy score were considered as promising. These were representatives 9, 24, 42, 59 and 73. In order to investigate what caused this sharp drop of energy, another two rounds of docking were realized for these specific conformations of the peptide, using HADDOCK and PyDockDNA tools. The results of this process led to the most efficient models of the binding of the peptide to B-DNA and the

3D- structures of these complexes were illustrated with the help of Pymol tool. These models were then investigated closely, in order to find the interactions that are developed between the molecules that rends them the best solutions. This stage of the research led to representatives 9, 24, 73 that are the best models. They present the potential binding mechanism of the peptide which is dominated by electrostatic interactions, where positively charged residues (Lys, Arg) interact with the negatively charged phosphates of the DNA backbone. Furthermore, hydrophobic interactions (as well as Van der Waals) between the side chains of residues like Met, Val and Pro which interact with the non-polar core of the B-DNA molecule. Additionally, hydrogen bonding can also play a role in the binding of the peptide, for example between the hydroxyl group of Serine's side chain (hydrogen bond donor) and the phosphates groups of the DNA backbone.

The three docking tools that were used do not agree on the position of the peptide in the DNA molecule, as the peptide binds non-specifically and doesn't recognize any pattern of DNA bases. However, they do agree on the interactions that are developed between the two molecules, regardless of where exactly the 12-mer binds in the B-DNA molecule.

It's important to highlight that this research has limitations. First of all, the peptide was simulated in its free form, in the absence of the B-DNA molecule, which may reduce the accuracy of the predictions. Furthermore, the peptide should have been in a continuous interaction with the B-DNA molecule throughout the binding process. In a real system, its conformation would change and adapt to the environment and the DNA molecule, until a dynamic equilibrium was reached. This is the case rather than a rigid conformation of the peptide trying to adapt to the DNA molecule. What's more, the B-DNA molecule have been chosen arbitrarily, which can also influence the docking process. Different DNA molecules can have different electrostatic potentials,

different capacities for hydrogen bonding, as well as different overall shape (e.g., minor and major groove width), which may influence the peptide binding. If the study aims to model biologically relevant interactions, choosing an arbitrary sequence might not reflect the actual sequences the peptide would encounter in vivo. This could lead to results that are not applicable to real-world biological systems. It can also make it harder to compare results across different studies unless the sequence is standardized.

Future work may involve a molecular dynamics simulation of the peptide in the presence of a DNA oligomer. Such a simulation would allow the peptide to sample putative biologically relevant conformations that would be inaccessible/unstable in the absence of DNA. An experimental study of the peptide could also be conducted, for example, through Alanine Scanning (34). In such an experiment, each one of the peptide residues would be replaced by alanine successively, leading to 12 mutated versions of the 12-mer peptide, which would be examined for their ability to bind to the B-DNA molecule. The ultimate goal of these experiments could be the design of a peptide that binds more strongly and/or with specificity to the B-DNA molecule. Furthermore, the peptide could be designed to bind pharmaceutical molecules in order to transfer them to a specific DNA target.

5. References

1. Kolchina N, Khavinson V, Linkova N, Yakimov A, Baitin D, Afanasyeva A, et al. Systematic search for structural motifs of peptide binding to double-stranded DNA. Nucleic Acids Res [Internet]. 2019 Nov 18 [cited 2024 Jul 24];47(20):10553–63. Available from: https://dx.doi.org/10.1093/nar/gkz850

2. Branden CI, Tooze J. Introduciton to protein structure [Internet]. Garland Science, editor. Vol. 1. 1999 [cited 2024 Jul 23]. 159–166 p. Available from: https://www.routledge.com/Introduction-to-Protein-Structure/Branden-Tooze/p/book/9780815323051

3. Chiu TP, Rao S, Mann RS, Honig B, Rohs R. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein-DNA binding. Nucleic Acids Res [Internet]. 2017 Dec 1 [cited 2024 Jul 23];45(21):12565–76. Available from: https://pubmed.ncbi.nlm.nih.gov/29040720/

4. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. Annu Rev Biochem [Internet]. 2010 Jul 7 [cited 2024 Jul 23];79:233–69. Available from: https://pubmed.ncbi.nlm.nih.gov/20334529/

5. Huth JR, Bewley CA, Nissen MS, Evans JNS, Reeves R, Gronenborn AM, et al. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. Nat Struct Biol [Internet]. 1997 [cited 2024 Jul 23];4(8):657–65. Available from: https://pubmed.ncbi.nlm.nih.gov/9253416/

6. Guidotti G, Brambilla L, Rossi D. Cell-Penetrating Peptides: From Basic Research to Clinics. Trends Pharmacol Sci [Internet]. 2017 Apr 1 [cited 2024 Jul 23];38(4):406–24. Available from: https://pubmed.ncbi.nlm.nih.gov/28209404/

7. Fedoreyeva LI, Kireev II, Khavinson VK, Vanyushin BF. Penetration of short fluorescence-labeled peptides into the nucleus in HeLa cells and in vitro specific interaction of the peptides with

deoxyribooligonucleotides and DNA. Biochemistry (Mosc) [Internet]. 2011 [cited 2024 Jul 23];76(11):1210–9. Available from: https://pubmed.ncbi.nlm.nih.gov/22117547/

8. Sánchez A, Vázquez A. Bioactive peptides: A review. Food Quality and Safety [Internet]. 2017 Mar 1 [cited 2024 Jul 23];1(1):29–46. Available from: https://dx.doi.org/10.1093/fqsafe/fyx006

9. Kraskovskaya NA, Kukanova EO, Lin'kova NS, Popugaeva EA, Khavinson VK. Tripeptides Restore the Number of Neuronal Spines under Conditions of In Vitro Modeled Alzheimer's Disease. Bull Exp Biol Med [Internet]. 2017 Aug 1 [cited 2024 Jul 23];163(4):550–3. Available from: https://pubmed.ncbi.nlm.nih.gov/28853087/

10. Zamorskii II, Shchudrova TS, Lin'kova NS, Nichik TE, Khavinson VK. Nephroprotective Effect of EDL Peptide at Acute Injury of Kidneys of Different Genesis. Bull Exp Biol Med [Internet]. 2017 Jul 1 [cited 2024 Jul 23];163(3):389–93. Available from: https://link.springer.com/article/10.1007/s10517-017-3811-1

11. Anisimov VN, Khavinson VK. Peptide bioregulation of aging: results and prospects. Biogerontology [Internet]. 2010 Apr [cited 2024 Jul 23];11(2):139–49. Available from: https://pubmed.ncbi.nlm.nih.gov/19830585/

12. Renukuntla J, Vadlapudi AD, Patel A, Boddu SHS, Mitra AK. Approaches for Enhancing Oral Bioavailability of Peptides and Proteins. Int J Pharm [Internet]. 2013 Apr 4 [cited 2024 Jul 23];447(0):75. Available from: /pmc/articles/PMC3680128/

13. Wölcke J, Weinhold E. A DNA-binding peptide from a phage display library. Nucleosides Nucleotides Nucleic Acids [Internet]. 2001 [cited 2024 Jul 23];20(4–7):1239–41. Available from: https://pubmed.ncbi.nlm.nih.gov/11562993/

14. Estephan E, Dao J, Saab MB, Panayotov I, Martin M, Larroque C, et al. SVSVGMKPSPRP: a broad range adhesion peptide. Biomed Tech (Berl) [Internet]. 2012 Dec [cited 2024 Jul 23];57(6):481–9. Available from: https://pubmed.ncbi.nlm.nih.gov/23183721/

15. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nature Structural Biology 1997 4:1 [Internet]. 1997 Jan [cited 2024 Jul 24];4(1):10–9. Available from: https://www.nature.com/articles/nsb0197-10

16. Buchner GS, Murphy RD, Buchete NV, Kubelka J. Dynamics of protein folding: Probing the kinetic network of folding–unfolding transitions with experiment and theory. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2011 Aug 1;1814(8):1001–20.

17. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron [Internet]. 2018 Sep 19 [cited 2024 Jul 23];99(6):1129–43. Available from: https://pubmed.ncbi.nlm.nih.gov/30236283/

18. Galano-Frutos JJ, Nerín-Fonz F, Sancho J. Calculation of Protein Folding Thermodynamics Using Molecular Dynamics Simulations. J Chem Inf Model [Internet]. 2023 Dec 25 [cited 2024 Jul 23];63(24):7791–806. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jcim.3c01107

19. Scheraga HA, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. Annu Rev Phys Chem [Internet]. 2007 [cited 2024 Jul 24];58:57–83. Available from: https://pubmed.ncbi.nlm.nih.gov/17034338/

20. An Introduction to Molecular Dynamics Simulations [Internet]. [cited 2024 Jul 24]. Available from: https://portal.valencelabs.com/blogs/post/an-introduction-to-molecular-dynamics-simulations-C9nXeGJ8hbNYghL

21. Setting up a Molecular Dynamics simulation - Compchems [Internet]. [cited 2024 Jul 24]. Available from: https://www.compchems.com/setting-up-a-molecular-dynamics-simulation/#1-level-of-theory

22. Morris GM, Lim-Wilby M. Molecular docking. Methods Mol Biol [Internet]. 2008 [cited 2024 Jul 24];443:365–82. Available from: https://pubmed.ncbi.nlm.nih.gov/18446297/

23. Van Dijk M, Van Dijk ADJ, Hsu V, Rolf B, Bonvin AMJJ. Information-driven protein–DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Res [Internet]. 2006 [cited 2024 Jul 25];34(11):3317. Available from: /pmc/articles/PMC1500871/

24. Meng XY, Zhang HX, Mezei M, Cui M. Molecular Docking: A powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des [Internet]. 2011 Jun 6 [cited 2024 Jul 24];7(2):146. Available from: /pmc/articles/PMC3151162/

25. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res [Internet]. 2017 Jul 7 [cited 2024 Jul 23];45(Web Server issue):W365. Available from: /pmc/articles/PMC5793843/

26. Alexiadou D. Structural computational biology studies of a non-specific DNA-binding peptide. 2023 [cited 2024 Aug 12]; Available from: https://repo.lib.duth.gr/jspui/handle/123456789/15374

27. Mu Y, Nguyen PH, Stock G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins [Internet]. 2005 Jan 1 [cited 2024 Aug 28];58(1):45–52. Available from: https://pubmed.ncbi.nlm.nih.gov/15521057/

28. Altis A, Nguyen PH, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. J Chem Phys [Internet]. 2007 [cited 2024 Aug 28];126(24). Available from: https://pubmed.ncbi.nlm.nih.gov/17614541/

29. Altis A, Otten M, Nguyen PH, Hegger R, Stock G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. J Chem Phys [Internet]. 2008 [cited 2024 Aug 28];128(24). Available from: https://pubmed.ncbi.nlm.nih.gov/18601386/

30. Glykos NM. Software news and updates. Carma: a molecular dynamics analysis program. J Comput Chem [Internet]. 2006 Nov 15 [cited 2024 Aug 28];27(14):1765–8. Available from: https://pubmed.ncbi.nlm.nih.gov/16917862/

31. Koukos PI, Glykos NM. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. J Comput Chem [Internet]. 2013 Oct 5 [cited 2024 Aug 28];34(26):2310–2. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.23381

32. Rodríguez-Lumbreras LA, Jiménez-García B, Giménez-Santamarina S, Fernández-Recio J. pyDockDNA: A new web server for energy-based protein-DNA docking and scoring. Front Mol Biosci [Internet]. 2022 Oct 6 [cited 2024 Jul 23];9. Available from: /pmc/articles/PMC9582769/

33. Rosignoli S, Paiardini A. Boosting the Full Potential of PyMOL with Structural Biology Plugins. Biomolecules 2022, Vol 12, Page 1764 [Internet]. 2022 Nov 27 [cited 2024 Jul 23];12(12):1764. Available from: https://www.mdpi.com/2218-273X/12/12/1764/htm

34. Boersma MD, Sadowsky JD, Tomita YA, Gellman SH. Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein–protein recognition: Application to the Bim

BH3 domain. Protein Sci [Internet]. 2008 Jul [cited 2024 Aug 27];17(7):1232. Available from: /pmc/articles/PMC2442000/