POLITECNICO DI TORINO

Master of Science in Biomedical Engineering

Politecnico
di Torino

Master's Degree Thesis

Comprehensive evaluation of docking algorithms for therapeutic peptides

Supervisors

J. A. TUSZYNSKI

G. K-S. WONG

Candidate

Pietro COMOGLIO

Academic Year 2022/2023

Summary

Therapeutic peptides are a unique class of pharmaceutical agents composed of a series of well-ordered amino acids, usually with molecular weights of 500-5000 Da. Peptide drug development has made great progress in the last decade and a wide variety of natural and modified peptides have been obtained and studied; remarkable achievements have been made resulting in the approval of more than 80 peptide drugs worldwide. The development of peptide drugs has thus become one of the hottest topics in pharmaceutical research. A performance evaluation of the current state-of-the-art peptide-protein docking algorithms on approved peptide drugs is therefore crucial to assess the quality of computational predictions. In this study, a set of FDA-approved peptide drugs was tested against three of the best-performing peptide-protein docking algorithms, and the quality of the predictions was evaluated. To further analyze the performance of the docking methods, the peptide conformations were altered using molecular dynamics simulation and docked again. Finally, the results were compared to obtain valuable insights into the predictive peculiarities of the examined protocols.

Acknowledgements

I would like to express my sincere gratitude to Professor Tuszynski for his helpful contributions and for the opportunity he gave me.

A special thanks to Professor Wong for the valuable advice throughout my work.

Table of Contents

List of Tables VII

List of Figures VIII

Acronyms XI

1 Computational methods 1
- 1.1 Molecular Docking . . . 1
- 1.2 Docking Approaches . . . 2
  - 1.2.1 Shape complementarity . . . 2
  - 1.2.2 Simulation . . . 2
- 1.3 Mechanics of docking . . . 2
  - 1.3.1 Search algorithm . . . 2
  - 1.3.2 Scoring function . . . 3
- 1.4 Docking assessment . . . 4
- 1.5 Molecular Mechanics . . . 5
  - 1.5.1 Potential energy . . . 5
  - 1.5.2 Environment and solvation . . . 7
  - 1.5.3 Energy minimization . . . 7
  - 1.5.4 Molecular Dynamics . . . 8

2 An overview on therapeutic peptides 10
- 2.1 Properties . . . 11
- 2.2 Market trends . . . 12

3 Peptides Selection 13
- 3.1 Sincalide . . . 15
- 3.2 Semaglutide and Liraglutide . . . 17
- 3.3 Salmon calcitonin . . . 21
- 3.4 Secretin . . . 23
- 3.5 Arginine vasopressin . . . 25

3.6 Oxytocin 27

4 Computational methods 30
4.1 Peptide-protein docking 30
4.2 FRODOCK 31
4.3 AutoDock CrankPep 34
4.4 HPepDock 37

5 Molecular Dynamics 38
5.1 Introduction 38
5.2 Molecular Dynamics simulation 39
5.2.1 Structure preparation 39
5.2.2 Minimization, heating, and equilibration 39
5.2.3 Energy Minimization and Molecular Dynamics run 40
5.3 Comparison with the native structures 40

6 Results 45
6.1 Native peptides 45
6.2 Altered peptides 68

7 Conclusions 83

A Additional resources 86

Bibliography 88

List of Tables

3.1 Peptide structures methods and resolutions 13
3.2 Selected peptides 29

4.1 Protein-protein docking categories 31

7.1 i-RMSD [Å] of native peptides docked on their receptors 84
7.2 L-RMSD [Å] of native peptides docked on their receptors 84
7.3 i-RMSD [Å] of altered peptides docked on their receptors 85
7.4 L-RMSD [Å] of altered peptides docked on their receptors 85

List of Figures

3.1 Sincalide . . . 15
3.2 Sincalide complex with CCK(1) receptor . . . 16
3.3 Sincalide complex, the *Interaction Surface* is highlighted on the receptor . . . 16
3.4 GLP-1 . . . 17
3.5 GLP-1 complex with GLP-1 receptor . . . 18
3.6 GLP-1 complex, the *Interaction Surface* is highlighted on the receptor 18
3.7 Semaglutide . . . 19
3.8 Semaglutide complex with GLP-1 receptor . . . 20
3.9 Semaglutide complex, the *Interaction Surface* is highlighted on the receptor . . . 20
3.10 Salmon calcitonin . . . 21
3.11 Salmon calcitonin complex with human calcitonin receptor . . . 22
3.12 Salmon calcitonin complex, the *Interaction Surface* is highlighted on the receptor . . . 22
3.13 Secretin . . . 23
3.14 Secretin complex with human secretin receptor . . . 24
3.15 Secretin complex, the *Interaction Surface* is highlighted on the receptor 24
3.16 AVP . . . 25
3.17 AVP complex with vasopressin receptor . . . 26
3.18 AVP complex, the *Interaction Surface* is highlighted on the receptor 26
3.19 Oxytocin . . . 27
3.20 Oxytocin complex with oxytocin receptor . . . 28
3.21 Oxytocin complex, the *Interaction Surface* is highlighted on the receptor . . . 28

5.1 Sincalide, RMSD = 3.07 Å . . . 41
5.2 GLP-1, RMSD = 5.35 Å . . . 42
5.3 Semaglutide, RMSD = 2.04 Å . . . 42
5.4 Salmon calcitonin, RMSD = 4.38 Å . . . 43
5.5 Secretin, RMSD = 3.44 Å . . . 43

5.6 AVP, RMSD = 1.57 Å 44
5.7 Oxytocin, RMSD = 1.64 Å 44

6.1 Sincalide i-RMSD 47
6.2 GLP-1 i-RMSD 48
6.3 Semaglutide i-RMSD 49
6.4 Salmon calcitonin i-RMSD 50
6.5 Secretin i-RMSD 51
6.6 AVP i-RMSD 52
6.7 Oxytocin i-RMSD 53
6.8 Sincalide L-RMSD 54
6.9 GLP-1 L-RMSD 55
6.10 Semaglutide L-RMSD 56
6.11 Salmon calcitonin L-RMSD 57
6.12 Secretin L-RMSD 58
6.13 AVP L-RMSD 59
6.14 Oxytocin L-RMSD 60
6.15 Sincalide RMSF 61
6.16 GLP-1 RMSF 62
6.17 Semaglutide RMSF 63
6.18 Salmon calcitonin RMSF 64
6.19 Secretin RMSF 65
6.20 AVP RMSF 66
6.21 Oxytocin RMSF 67
6.22 Altered sincalide i-RMSD 69
6.23 Altered GLP-1 i-RMSD 70
6.24 Altered semaglutide i-RMSD 71
6.25 Altered salmon calcitonin i-RMSD 72
6.26 Altered secretin i-RMSD 73
6.27 Altered AVP i-RMSD 74
6.28 Altered oxytocin i-RMSD 75
6.29 Altered sincalide L-RMSD 76
6.30 Altered GLP-1 L-RMSD 77
6.31 Altered semaglutide L-RMSD 78
6.32 Altered salmon calcitonin L-RMSD 79
6.33 Altered secretin L-RMSD 80
6.34 Altered AVP L-RMSD 81
6.35 Altered oxytocin L-RMSD 82

Acronyms

CAGR

Compound Annual Growth Rate

RMSD

Root Mean Square Deviation

ai-RMSD

Average increase of interface RMSD

aL-RMSD

Average increase of ligand RMSD

Chapter 1

Computational methods

The results presented in this work have been obtained adopting computational models and simulations, an exhaustive explanation is therefore required.

1.1 Molecular Docking

Molecular docking is an essential tool in computer-aided molecular biology, as it attempts to predict the binding pose of two molecules where the ligand and the target are supposed to form a stable complex. It is of pharmaceutical interest the understanding of biologically relevant interactions such as ligand-protein, thus trying to predict the most likely binding mode(s) of a ligand with a known three-dimensional structure of a protein. Ligands input are usually sequences but 3D structures provide useful insights on the native conformation which can be computed together with the final results for a more complete analysis. Effective docking systems effectively explore high-dimensional spaces and employ a scoring function that appropriately ranks candidate dockings. Docking may be used to perform virtual screening on vast libraries of compounds, rate the results, and provide structural theories about how the ligands inhibit the target, which is extremely useful in lead optimization. The idea behind molecular docking could be simply put down with the lock-and-key model, where the ligand is the key and the procedure's purpose is to find the correct orientation (binding pose) in the lock (the receptor), identifying docking as an optimization problem. However, since both the ligand and the receptor are flexible another design could better represent their interaction, the so called *hand-in-glove* model allows for the ligand and receptor to adjust their conformation during the docking, aiming to minimize the free energy of the system. Two approaches are particularly renowned, one describes ligand and receptor as complementary surfaces, while the second reproduce the docking process by calculating the energies of the interaction. on-the-fly

1.2 Docking Approaches

1.2.1 Shape complementarity

Also known as geometric matching, these methods describe ligand and receptor as a set of features with a certain propensity to achieve a successful dock. A common example of these descriptors could be the molecular surface/complementary surface, where the receptor is characterized in terms of solvent-accessible surface area and the ligand's molecular surface is characterized in terms of its matching surface description, the complementarity between the two descriptors is then used to find the best pose. Another approach is the modeling of hydrophobic features with backbone atoms or Fourier transform descriptors. While being robust and allowing rapid screenings of large libraries, this methods are often too simplistic and fail to accurately represent the ligand flexibility.

1.2.2 Simulation

Simulating the docking process is a more challenging work. In this approach both the ligand and receptor are represented as accurate entities and the ligand tries multiple binding poses performing a certain number of moves in its conformational space. The set of moves is comprehensive of external moves such as rotation and translation as well as internal changes to the ligand conformation. The energy of the system is calculated after every move and is the parameter assessing the quality of each pose. Other than the obvious advantage of representing the ligand flexibility, this approach also provides faithful models of reality, with the drawback of being computationally expensive and slower.

1.3 Mechanics of docking

The structures of ligand and receptor are a vital part of the docking process and are usually determined with biophysical methods such as X-ray crystallography, Nuclear magnetic resonance (NMR) spectroscopy or cryo-electron microscopy (cryo-EM), but can also be derived from computational models. The effectiveness of a docking search is determined by two main factors: a search algorithm and a scoring function.

1.3.1 Search algorithm

Searching for the most favorable interaction means to explore the conformational space of both the ligand and the receptor. This term includes all of the possible internal (ie. bond rotations) and external (ie. rotation, translation) arrangements

of the structures in exam. Most docking algorithms employ this definition for the ligand only, while usually the receptor is kept still and with a certain degree of flexibility. The conformational space search is usually implemented with one of three main methods:

1. Molecular dynamics simulations
2. Stochastic search
3. Systematic search
4. Genetic algorithms

Ligand flexibility

The ligand conformation may be determined in absence of the receptor and docked afterwards or selected from a pool of configurations assumed inside of the binding pocket. The best pose selection is usually selected according to the minimum free energy but in specific cases *a priori* knowledge can be functional.

Receptor flexibility

Regardless of the drastic increase in computational capacity the degree of receptor flexibility remains a challenging task due to the sheer number of degrees of freedom that have to be taken into account. Many approaches have been presented without reaching an all-purpose solution, from a fully rigid receptor, expanding the flexible region around binding pockets until obtaining fully flexible structures.

1.3.2 Scoring function

Many are the factors at play when it comes to finding a function that takes two positions as inputs and returns the plausibility of the pose. Most scoring functions are force fields that compute the free energy of the pose within the binding site, they could be roughly written as an additive equation:

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/t} + \Delta G_{vib} \qquad (1.1)$$

The free binding energy is therefore represented by a sum of the energies associated with solvent effects, conformational changes, internal rotations, associated energy from ligand and receptor to form a single complex and vibrational factors. There are currently four main classes of scoring functions:

1. Force field: the quality of a pose is assessed using a force field, hence the weight of intermolecular interactions (electrostatic and van der Waals) are evaluated accordingly.

2. Empirical: a specific interaction (ie. idrophobic, idrophilic, hydrogen) is chosen and a score is assigned based off the number of occurrences.

3. Knowledge-based: starting from statistical observations of large datasets of ligand-receptor complexes, statistical potentials are derived from the assumptions that frequent interactions are energetically favorable.

4. Machine learning: this unique method benefits from not assuming any binding affinity between the two structures, therefore the quality of a pose is determined directly from the data.

1.4 Docking assessment

The most common evaluation parameter in structural biology is without doubts the Root Mean Square Deviation (RMSD) of atomic positions, a measure of the distance between the atoms of superimposed structures.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \tag{1.2}$$

Where N is the number of atoms and δ_i is the distance between the i-atom and either a reference or the mean position of all the atoms in the structure. RMSD is generally calculated over the backbone atoms C_α , C , N , O. Although it is remarkably versatile, RMSD fails to highlight important details involved in a reliable docking pose, hence other metrics have received attention for their increased accuracy. Three are particularly valuable regarding of peptide-protein interactions:

1. Ligand RMSD (L-RMSD): the receptor is kept fixed and the RMSD is calculated between each atom of the ligand native structure versus the model, highlighting dissimilarities in the conformation or position in space of the ligands.

2. Interface RMSD (i-RMSD): first an interface region is defined as all the atoms of the receptor closer than 5 Åto the ligand, the RMSD of the examined region is therefore calculated.

3. Root Mean Square Fluctuation (RMSF): same procedure is applied to the L-RMSD but the analysis is performed on the time scale instead of the space scale. Once selected the ligand, each atom is compared to itself throughout the generated models, in this way each residue RMSD variation over the different models is emphasized.

1.5 Molecular Mechanics

Molecular mechanics (MM) is a computational method that adopts classical mechanics to reproduce molecular systems. It is based on the Born-Oppenheimer, where the wave functions of atomic nuclei and electrons are treated separately, and thus calculating the potential energy as a function of the nuclear coordinates. The set of parameters assigned to the coordinates are named a force field. The general properties of a MM simulation can summarized as:

1. Each atom is modeled by a particle and, if properly defined, a group of atoms can be considered as a unique particle
2. For each atom type a set of properties is defined (ie. mass, hybridization, charge)
3. Bonded interactions are modeled as springs, where the equilibrium point is defined experimentally

According to these rules it is possible to evaluate the behavior of a molecular system when interacting with an external perturbation or structure, or sample the space of conformations.

1.5.1 Potential energy

With knowledge of the coordinates of the atoms in the system, the force field calculates the total potential energy (U) as a sum of two factors: i) Covalent interactions and ii) Non-covalent interactions.

Covalent interactions

Described by the formula:

$$U_{covalent} = U_{bond} + U_{angle} + U_{dihedral} + U_{i_dihedral} \tag{1.3}$$

The U_{bond} term represents covalent bonds between two atoms and it is modeled as a spring. The potential is therefore usually an harmonic function:

$$U(l) = \sum_{bonds} \frac{1}{2} k_l [l - l_0]^2 \tag{1.4}$$

Where K_l is the force constant, l_0 is the reference bond length assumed when all the other terms are set to zero and l is the bond length at the equilibrium. The first two terms are derived empirically. The Bond interaction can be modeled in several other ways, for example using a Morse potential (allowing bond breaking)

or a Cubic potential (enhancing the accuracy of the model in the proximity of l_0 but poor performances when the bond is stretched). The U_{angle} term models the potential associated with the interaction of three atoms, once again modeled with an harmonic function:

$$U(\theta) \;=\; \sum_{angles} \frac{1}{2} k_\theta [\theta - \theta_0]^2 \tag{1.5}$$

Where $K_[\theta]$ is the force constant, θ_0 is the reference bond angle assumed when all the other terms are set to zero and θ is the bond angle at the equilibrium. The dihedral angles are structures comprised of four bonded atoms describing the relative position of two planes, hence steric effects are meant to be taken into account. The general form is a series of cosines:

$$U(\varphi) \;=\; \sum_{dihedrals} C_n cos(\varphi)^n \tag{1.6}$$

Where n is the multiplicity (number of energetic minima over a full rotation) and C_n accounts for the energetic costs needed for the angle deformation and for the position of the energy minima. Thus several possible values of the dihedral angle are allowed. Improper dihedrals are a correction additional term required for limiting the rotation, for example keeping the benzene ring on the same plane.

Non-covalent interactions

Described by the formula:

$$U_{noncovalent} \;=\; U_{electrostatics} \;+\; U_{vdW} \tag{1.7}$$

Non-covalent interactions representation is a challenging task, their number in fact grows rapidly with the system size and a perturbation will require for all of the terms to be re-evaluated, making the calculation computationally costly. The two potential functions are modelled as inversely proportional to the distance between two atoms. The van der Waals (vdW) interactions can referred to any atom in the system and are divided into attractive (London forces) and repulsive (where two atoms overlap) according to the distance between the two atoms. The two contributions are modeled with different functions, with a distance minimum representing the equilibrium. The most employed function to model vdW interactions is the 6-12 *Lennard-Jones* (LJ):

$$U(r) \;=\; 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1.8}$$

Where σ is the collision diameter (the distance where the vdW potential equals zero) and ϵ is the well depth (vdW potential energy minima). The term on the left models the repulsive interaction of higher magnitude, while the one on the right

the weaker attractive contribution. No model perfectly depicts reality and other functions better model certain aspects of vdW interactions, for example the Halgren function manage to give a finite value of $U(r)$ when $r \to 0$. The Hydrogen bonds potentials can be calculated using a modified version of LJ. Electrostatic potentials are modeled using Coulomb's Law, which can be combined with the LJ function being dependent on r, adding the two charges of the atoms examined. Given the large number of interactions at play the calculations are preferably performed on electric moments (assigning partial charges) instead of single charges, allowing for a drastic decrease in the number of evaluations. Usually the molecule is split in fragments and the partial charges are obtained through various methods like Molecular Dynamic or Monte Carlo simulations.

1.5.2 Environment and solvation

Biological structures are physiologically always surrounded by an aqueous environment and an accurate model should take into account the representation of this feature. The representation is perilous for two major reasons: the sheer amount of water molecules and the unique properties of water. In order to limit the number of water molecules a solvation box is defined around the structure (ie. cubic, hexagonal, octahedron), virtually replicated multiple times and simulating an infinite environment, avoiding boundary effects. Two are the main methods to simulate the solvent effect:

1. Explicit solvent: high accuracy but computationally expensive. The water molecules are explicitly represented and evaluated as independent objects. TIP3P is the 3-site (three interaction points corresponding to the three atoms) water model implemented in the CHARMM force field (with the slight difference of placing LJ parameters on the hydrogens as well).

2. Implicit solvent: model the solvent as, a continuum with specific features, largely improving the computational speed and reducing errors in statistical averaging. This class of methods usually relies on solvent accessible surfa,ce area (SASA) calculations or continuum electrostatics models.

1.5.3 Energy minimization

Thinking of energy potential as a landscape (also called Potential Energy Surface or PES), the most feasible conformations of a structure are located in the minima. It is therefore crucial to develop functions that can explore the PES seeking for points of minimum. It is important to remark these points can be local (the lowest point in a certain area) or global (lowest possible energy for a system), Energy Minimization (EM) simulations are only able to explore local minima, lacking the

ability to efficiently sample large portions of the energy landscape. Two are the main EM approaches:

Non-derivative methods

The SIMPLEX approach works by building a geometric shape with N + 1 linked vertices, where N is the dimension of the potential energy function. As an example, for a two-dimensional function, the figure will Form a triangle with each vertex representing a distinct coordinate set for which energy may be calculated. The geometric figure is given a set of allowed moves (ie. reflection, contraption, expansion) used to move on the PES, when a move leads to a lower energy point, the coordinates are updated. On the other hand, the sequential univariate method selects a coordinate from the system and generates two new structures each round altering the coordinate, a parabola is then fitted between the three structures (original and altered) and if the energy potential reached a minimum the coordinate is then changed to the position of the minimum. The random movements of non-derivative methods make them often inefficient and slow.

Derivative methods

Thanks to an understanding of the PES steepness this class of methods consistently outperforms the non-derivative ones. They can be divided into:

- First order methods: such as Steepest Descent and Conjugate gradient approaches, the structure moves towards decreasing energy potentials drastically reducing the simulation time.

- Second order methods: employ the inversion of the Hessian matrix, usually reserved for relatively small systems due to the high computational costs.

1.5.4 Molecular Dynamics

Molecular Dynamics (MD) relies on MM representations of a molecular system and provides an outlook on the dynamic evolution over time, letting the particles interact and solving numerically Newton's equations of motion instant by instant. In order to deal with particle dynamics methods that make use of generalized coordinates p and generalized momenta $q.$, the p and q together determine a point in the 6N-dimensional *phase space*. MD calculates the properties of a system sampling the microstates in a specific statistical ensemble, where an ensemble is a cluster of systems with different microscopic states but an identical macrostate. The main ensemble are:

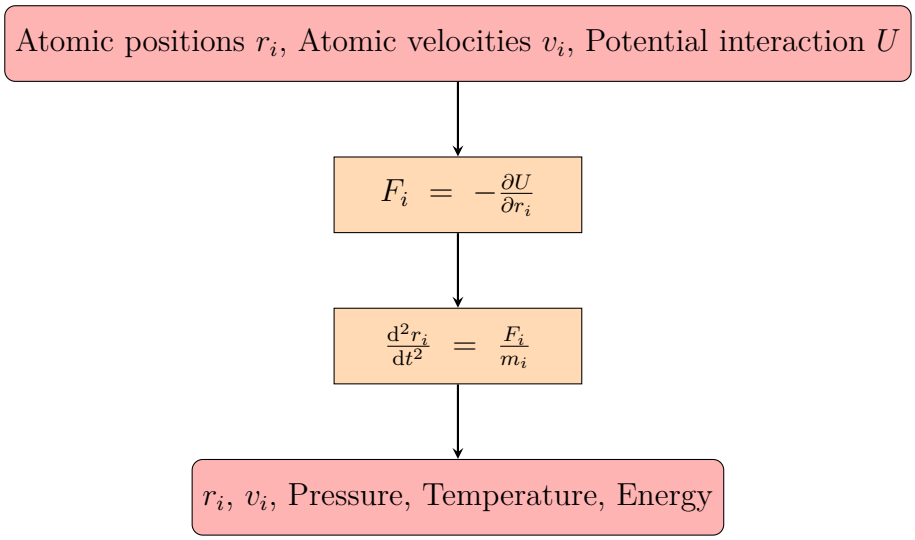
1. Thermally isolated equilibrium (NVE): appropriate for isolated systems, all the macrostate occupation probabilities are equal.

2. Thermal equilibrium with a heat reservoir (NVT): for systems in thermal equilibrium with a reservoir at a specific temperature, the macrostates have Boltzmann occupancies.

3. Isobaric-isothermal (NpT): fixed number of particles, pressure and temperature.

4. Grand canonical (muVT): fixed volume and temperature.

In order to sample the phase space many algorithm can be applied, such as Verlet, Velocity Verlet or Leap-frog; they differ in what parameters they use to calculate the future coordinates of the system. Essentially every MD algorithm pipeline can be summarized as:



Chapter 2

An overview on therapeutic peptides

Peptide therapeutics have played a major role in medicine since the introduction of insulin therapy a century ago and after decades of feeble innovation driven by technological limitations, a new era for peptide drugs is drawing near [1, 2]. The current surging interest in this field could be traced back to novel discovery, production, modification and delivery methods. Advancing further beyond endogenous human peptides, we are witnessing an unprecedented amount of raw data coming from the sequencing of structures from the animals and plants kingdom, as well as artificially designed ones, which are rapidly becoming available. Furthermore new efficient *in silico* and *in vitro* technologies are emerging, providing an essential tool for addressing well-known drawbacks as the expensive manufacturing and unfavorable pharmacokinetic properties [3]. For example, some natural product peptides (dietary-derived bioactive peptides) have been found to be orally bioavailable and many peptide toxins have remarkable stability profiles. Furthermore, advances in carriers able to efficiently deliver naturally active drugs such as siRNA or peptides [**Ebenezer2023-af**] open the path for an easier administration. Such examples yield the evidence that peptides can be optimized into effective and broadly active pharmaceutical agents [4]. From macrocyclic and cysteine-rich peptides to nonstandard chemistries, vaccines, carriers and antimicrobial peptides; the large gap between small molecules (<500 Da) and biologics (>5000 Da) offers a huge opportunity for peptides to find a prominent place in the pharmaceutical landscape [5].

When compared to small molecules, peptide drugs show a large number of advantages [6]. To begin with they have high safety and target affinity being usually synthesized from endogenous peptides templates, are 2 times more likely to be approved for marketing through clinical trials than small molecule drugs,

and the average R&D cycle is 0.7 years less [4]. The biggest drawback of peptides is the proteolytic degradation and the difficulty in crossing the intestinal mucosa, hindering drug accumulation effects and affecting the choice of successful route of administration. Despite low stability in serum and plasma medium, recent studies recorded an increased stability in fresh blood, indicating how the worry and effort put into increasing peptides stability may often be only relevant in vitro, but of lower importance in vivo [7].

2.1 Properties

An ideal peptide drug should therefore possess certain characteristics [4]:

1. Agonist. Only low receptor occupancy (5–20%) is necessary for receptor activation, whereas antagonists generally must occupy more than 50% of receptors to be effective. Furthermore, antagonism can be achieved by allosteric receptor interactions, leaving ample room for competing orally available small-molecule drugs.

2. Antagonist/inhibitor for targets where a broad surface area provides an advantage over small-molecule drugs (for example, protein–protein interactions and ion channels).

3. Stable due to rigid three-dimensional structure with secondary structural motifs stabilized by disulfide bonds or cyclization.

4. Replacement of methionine residues (for example, with norleucine) to avoid oxidative shelf-life problems.

5. Pharmacokinetics and dynamics matched with biological action, dose regime and safety

6. Incorporation of fatty acids (C14/16), pegylation or protein conjugation to evade renal clearance, if required.

7. Compatible with a delivery route/formulation strategy that maximizes patient adherence.

Ideal peptide drug target:

1. Extracellular and peripheral to bypass delivery challenges.

2. Multiple receptor subtypes to exploit the selectivity advantages of peptides.

3. Targets that require a large surface area for a therapeutic response.

In addition to the high binding affinity and low metabolic toxicity shared with antibodies and the high stability and ease of manufacture shared with small molecules, cyclic peptides have increasingly been developed in the last two decades thanks to their unique characteristics. Cyclization improves not only the structural properties of peptide chains but also the pharmacokinetic properties for absorption and biological membrane permeability that is necessary for reaching protein targets. Cyclic peptides can be further improved by the introduction of non-canonical elements (improving both their pharmacokinetic and pharmacodynamic properties), lipophilic side chains (avoid renal clearance and improves their pharmacokinetic properties) or conjugated with albumin and immunoglobulin to extend the half-life.[8]

2.2 Market trends

There are currently more than 80 peptide drugs in the global market, 170 at an advanced stage of clinical testing. [9] Peptide therapeutics covers a market size of USD 39 billion in 2021 and is expected to grow at a compound annual growth rate of 6.4% in the next ten years, reaching a market size of almost USD 50 billion in the next five years.[10] Among the >60 FDA- and EMA-approved peptides, two-thirds are in the cyclic form and have an important role in the modern pharmaceutical industry. [8] The choice of these commercially available compounds is also motivated by their relevance, the top three selling peptide drugs are in fact GLP-1 analogues. An analysis of the literature suggests peptide therapeutic agents are still strictly linked to the treatment of metabolic disorders (conidition worsened by unhealthy lifetsyles) occupying 35% of the market share in 2021, with a rise of cancer drugs emerging as second best seller, the reasons could be found in the patients concerns over the side effects of traditional therapies.[10] A global increasing demand for efficient and safe therapeutics leads to the need of a robust discovery pipeline, essential to further propel the progress.

Chapter 3

Peptides Selection

After an analysis of the current trends in peptide drugs [9], 26 entries were chosen from the Protein Data Bank filtering the available peptide-receptor complexes, the criteria was a resolution < 2.5 Å (16 structures), of which a subset where the resolution < 2 Å was extracted (10 structures). 2 Å is commonly accepted as a reasonable threshold between good and poor quality models. Seven entries were selected from the initial set because of their prominent role in the peptide drugs market and their positive response to the tested docking algorithms. The selected structures are derived from all of the three major research techniques in structural biology (NMR, X-ray and EM). Essentially there is no optimal choice since all of the aforementioned have their own unique advantages and disadvantages, nonetheless some methods achieve a better resolution.

Peptide	**Method**	**Resolution [Å]**
Sincalide [11]	Solution NMR	high resolution
GLP-1 [12]	X-ray diffraction	2.10
Semaglutide [13]	X-ray diffraction	1.80
Salmon Calcitonin [14]	X-ray diffraction	1.78
Secretin [15]	EM	2.30
Arginine Vasopressin [16]	EM	2.80
Oxytcin [17]	EM	2.90

Table 3.1: Peptide structures methods and resolutions

Along with a brief description of the peptide drugs analyzed, three figures are attached to each section: the 3D structure of the peptide, the peptide-receptor complex and a last one emphasizing the interface showing the *Interaction Surface*. This feature is obtained thanks to the Surfaces and Maps function implemented in

MOE 2020.09, representing the van der Waals potential of a probe atom with the receptor molecule, where on the interior of the surface vdW energy is positive and on the exterior is negative. A surface highlighting the vdW radii closely resembles the SASA of a structure. The distance cutoff between peptide and receptor atoms has been set at 5 Å as it is implemented for the i-RMS calculations. Additionally a color scheme has been generated where hydrophilic = purple and lipophilic = green, providing a valuable insight on the different regions binding affinity.

3.1 Sincalide

Sincalide is an injection-based medicine used to aid in the diagnosis of gallbladder and pancreatic problems. [18] It is the 8-amino acid C-terminal portion of cholecystokinin, often known as CCK-8. Cholecystokinin is a naturally occurring gastrointestinal peptide hormone that is generally required for accelerating protein and fat digestion in the body. Sincalide, when given intravenously, causes the gallbladder to constrict, resulting in a significant decrease in gallbladder size. The resulting bile evacuation is analogous to what occurs physiologically in response to endogenous cholecystokinin. Moreover, sincalide promotes pancreatic bicarbonate and enzyme secretion. Sincalide, when administered intravenously, causes the gallbladder to constrict, resulting in a significant decrease in gallbladder size. The resulting bile evacuation is analogous to what occurs physiologically in response to endogenous cholecystokinin. Sincalide, like cholecystokinin, promotes pancreatic secretion; when combined with secretin, it enhances both the amount of pancreatic secretion and the output of bicarbonate and protein (enzymes) by the gland. The combined impact of secretin and sincalide allows for the assessment of particular pancreatic function by duodenal aspirate measurement and analysis.

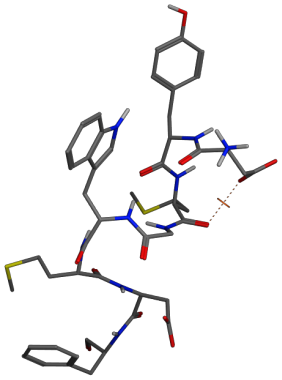


Figure 3.1: Sincalide

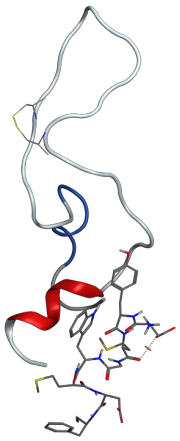


Figure 3.2: Sincalide complex with CCK(1) receptor

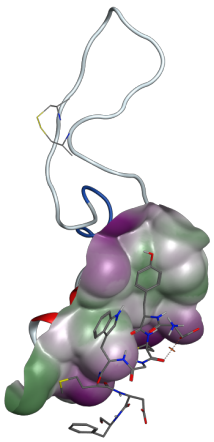


Figure 3.3: Sincalide complex, the *Interaction Surface* is highlighted on the receptor

3.2 Semaglutide and Liraglutide

An introduction is required for these peptides, GLP-1 analogues have recently been in the spotlight thanks to their weight control effect, promising to drastically change the fight against obesity.[19]

Liraglutide[20, 21, 22] is 97% similar to native human GLP-1 analyzed in this work, with the primary difference being the substitution of arginine for lysine at position 34. Liraglutide is synthesized by attaching a C-16 fatty acid (palmitic acid) with a glutamic acid spacer to the remaining lysine residue at position 26 of the peptide precursor [23]. Previous research has shown that GLP-1 has anti-diabetic effects on pancreatic beta cells, hence it was investigated further.[24] Liraglutide is a once-daily GLP-1 derivative utilized to treat type 2 diabetes. Liraglutide's protracted effect is achieved by binding a fatty acid molecule to position 26 of the GLP-1 molecule, allowing it to bind reversibly to albumin in the subcutaneous tissue and circulation and be released slowly over time. As compared to GLP-1, binding with albumin induces slower degradation and less elimination of liraglutide from the bloodstream by the kidneys. Liraglutide causes increased insulin secretion and reduced glucagon production in response to glucose, as well as delayed stomach emptying. Liraglutide had no effect on glucagon release in response to low blood sugar. Liraglutide is a GLP-1 agonist linked to adenylate cyclase. To improve blood sugar regulation, an increase in cyclic AMP promotes the glucose-dependent release of insulin, inhibits the glucose-dependent release of glucagon, and delays stomach emptying.

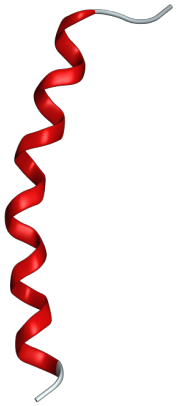


Figure 3.4: GLP-1

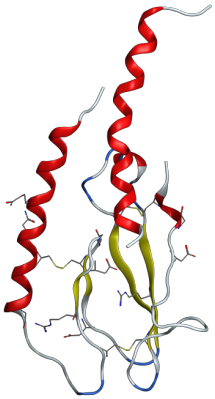


Figure 3.5: GLP-1 complex with GLP-1 receptor

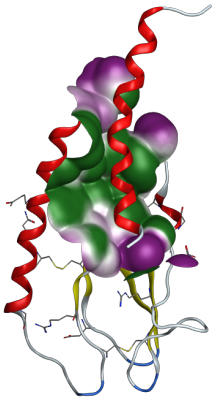


Figure 3.6: GLP-1 complex, the *Interaction Surface* is highlighted on the receptor

Semaglutide [25] is a glucagon-like peptide 1 (GLP-1) analog used to treat type 2 diabetes in conjunction with lifestyle modifications such as food restrictions and increased physical activity.[26] Exenatide and Liraglutide are other elements of this therapeutic class. Novo Nordisk developed semaglutide, which was approved by the FDA for subcutaneous injection in December 2017. In September 2019, the tablet formulation was authorized for oral use. Semaglutide stimulates insulin production and lowers blood glucose levels by attaching to and activating the GLP-1 receptor.[27] The subcutaneous injection is given once a week, and the pill is taken once a day. Semaglutide has a strategic advantage over other diabetic medications, which may require several daily doses. Clinical trials have shown that this medicine decreases glycosylated hemoglobin (HbA1c) levels and body weight, proving to be helpful for type 2 diabetic patients. The FDA authorized semaglutide in June 2021 for chronic weight control in individuals with general obesity or overweight who have at least one weight-related disease, making it the first licensed medicine for such usage since 2014. Health Canada and the EMA have both authorized the use of semaglutide in weight control as well.

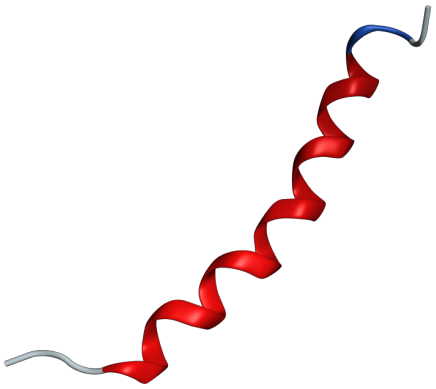


Figure 3.7: Semaglutide

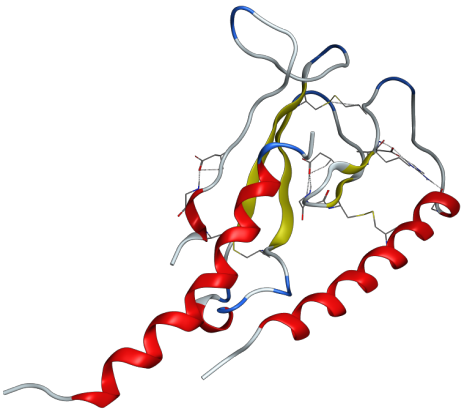


Figure 3.8: Semaglutide complex with GLP-1 receptor

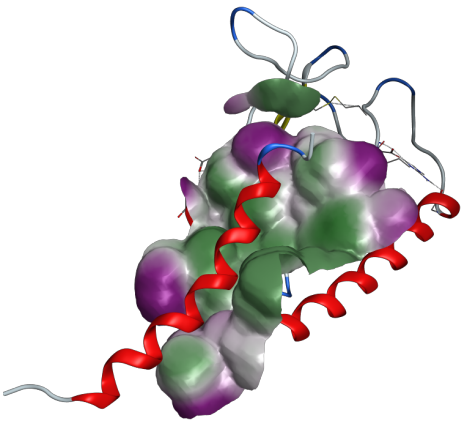


Figure 3.9: Semaglutide complex, the *Interaction Surface* is highlighted on the receptor

Mechanism of glycemic control GLP-1 is a physiological hormone that enhances glycemic control through a variety of mechanisms, including insulin secretion, slower stomach emptying, and decreased postprandial glucagon secretion. Glucose homeostasis is dependent on hormones such as insulin and amylin, which are released by pancreatic beta cells. Semaglutide is 94% identical to GLP-1 in humans. Analogs of this hormone, such as semaglutide, enhance insulin 3 production by activating pancreatic islet cells while decreasing glucagon release. They bind to the GLP-1 receptor with high selectivity, resulting in a variety of positive downstream actions that lower blood glucose in a glucose-dependent manner.[28, 29]

Mechanism of cardiovascular benefit and weight loss Semaglutide is thought to slow the development of atherosclerosis in hypercholesterolemia through decreasing intestinal permeability and inflammation. Weight loss is thought to occur as a result of reduced appetite and food cravings following semaglutide therapy.[30]

3.3 Salmon calcitonin

A synthetic peptide version of calcitonin used to treat hypercalcemia, osteoporosis, and Paget's disease by inhibiting bone resorption [31]. Calcitonin reduces osteoclast-mediated bone resorption by regulating osteoclast quantity and activity [32]. The impact of human calcitonin on osteoclasts is caused by a breakdown in cytoskeletal architecture, the distraction of actin rings, and the removal of osteoclast cellular polarity. Calcitonin is thought to exert its function at the subcellular level by regulating the cAMP-PKA signaling pathway.

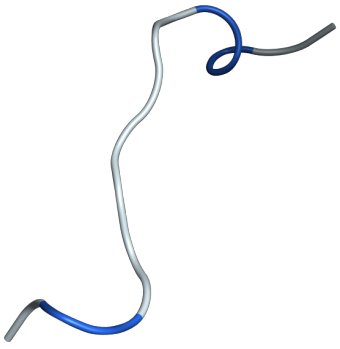


Figure 3.10: Salmon calcitonin

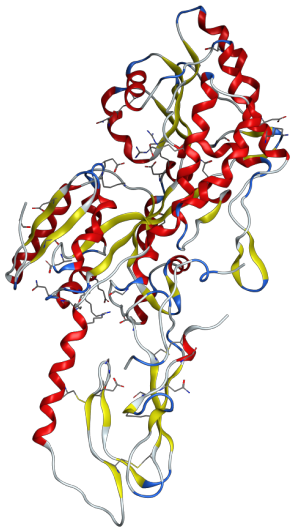


Figure 3.11: Salmon calcitonin complex with human calcitonin receptor

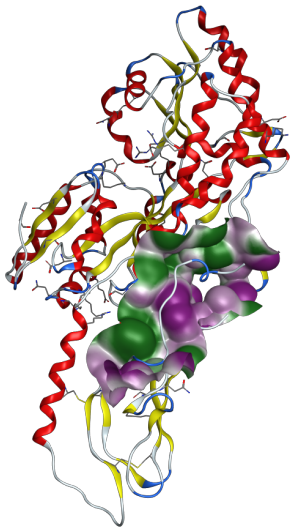


Figure 3.12: Salmon calcitonin complex, the *Interaction Surface* is highlighted on the receptor

3.4 Secretin

Human secretin is a secretin hormone that is used to induce pancreatic or gastric secretions in to detect exocrine pancreas dysfunction, gastrinoma, and bile and pancreatic duct abnormalities [33]. Human secretin is a peptide hormone found in the gastrointestinal tract that controls secretions in the stomach, pancreas, and liver. In reaction to duodenal content with a pH less than 4.5, enterochromaffin cells in the duodenum generate the hormone. [34]. Secretin's primary activity is to stimulate the pancreas to release pancreatic juice to regulate pH in the small intestines. Secretin is also involved in fluid homeostasis and bile formation. While secretin is a gastrointestinal hormone, it is also classified as a neuropeptide hormone since it is expressed in the central nervous system. [35, 36, 37, 37].

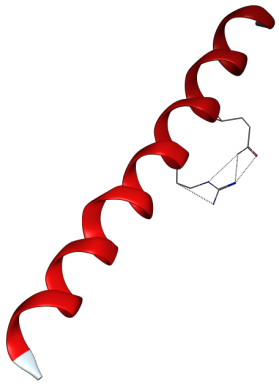


Figure 3.13: Secretin

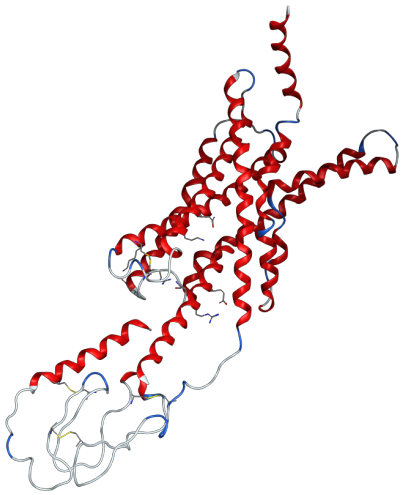


Figure 3.14: Secretin complex with human secretin receptor

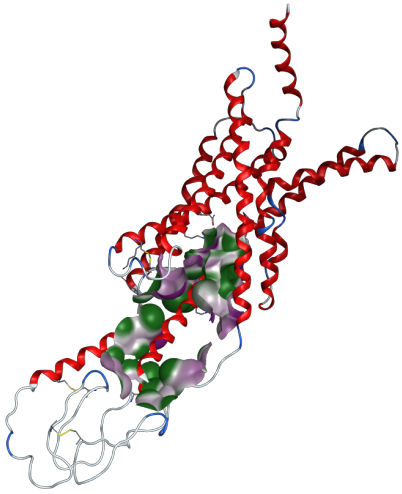


Figure 3.15: Secretin complex, the *Interaction Surface* is highlighted on the receptor

3.5 Arginine vasopressin

Vasopressin (AVP) is a peptide hormone that is used to increase blood pressure in patients with vasodilatory shock who are resistant to fluid and catecholamine treatments.[38, 39, 40, 41, 42] Vasopressin, Cyclo (1-6) L-Cysteinyl-L-Tyrosyl-L-PhenylalanylL-Glutaminyl-L-Asparaginyl-L-Cysteinyl-L-Prolyl-L-Arginyl-L-Glycinamide¦ is a cyclic nonapeptide hormone primarily produced by the supraoptic and periventricular nuclei of the hypothalamus.[43] Vasopressin release is mediated by sensory pathways, in which either a 2% increase in plasma osmolarity or a 10% decrease in blood pressure causes the release of endogenous vasopressin [44].

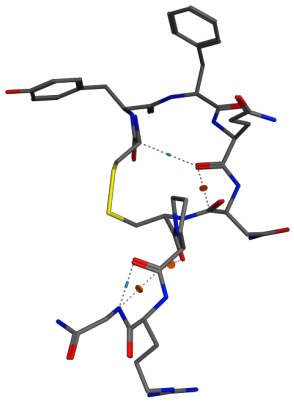


Figure 3.16: AVP

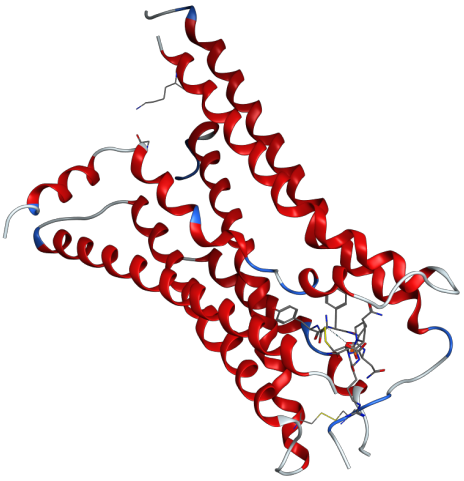


Figure 3.17: AVP complex with vasopressin receptor

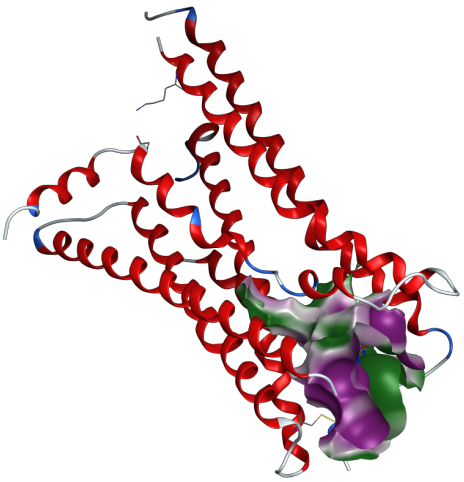


Figure 3.18: AVP complex, the *Interaction Surface* is highlighted on the receptor

3.6 Oxytocin

Oxytocin [45] is a recombinant hormone used to induce or strengthen uterine contractions in pregnant women to aid in labor and delivery or to control postpartum bleeding [46, 47]. Oxytocin is a pleiotropic nonapeptide hormone with significant physiological effects. It is most recognized for its ability to induce parturition and breastfeeding, but it also has significant physiological effects on metabolic and cardiovascular systems, sexual and maternal behavior, pair bonding, social cognition, and fear conditioning [48, 49, 50, 51, 52]. It is worth noting that oxytocin receptors are not limited to the reproductive system but can be found in many peripheral tissues and in central nervous system structures including the brain stem and amygdala [53].

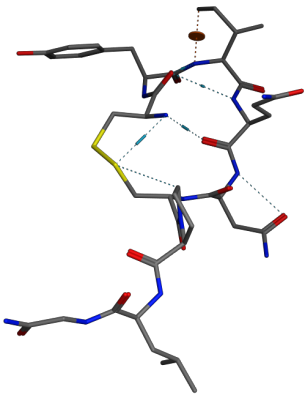


Figure 3.19: Oxytocin

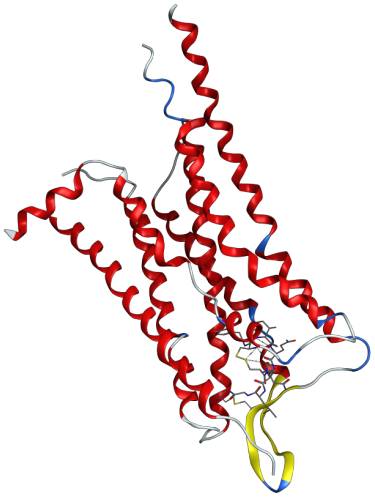


Figure 3.20: Oxytocin complex with oxytocin receptor

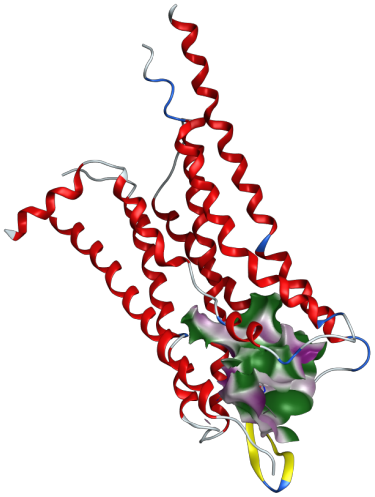


Figure 3.21: Oxytocin complex, the *Interaction Surface* is highlighted on the receptor

PDB entry	Name	Residues number	Cyclic	Cystein number	Commercial name
1d6g	CCK-8	8	no	0	Kinevac
3iol	GLP-1 Liraglutide*	30	no	0	Victoza*
4zgm	Semaglutide	30		0	Wegovy
6pfo	Salmon Calcitonin	27	yes	2	Fortical
6wzg	Secretin	27	no	0	SecreFlo
7kh0	Arginine Vasopressin	9	yes	2	VASOSTRICT
7ryc	Oxytocin	9	yes	2	Pitocin

Table 3.2: Selected peptides

Chapter 4

Computational methods

4.1 Peptide-protein docking

Occupying a middle ground in the biologic landscape, between the fairly simple modeling of small molecules and the large protein-protein interactions, peptides docking remains an arduous task. Peptides conformation predictions are indeed particularly challenging because of their high flexibility and large size [54] and the number of degrees of freedom drastically increases when combined with a protein binding pocket [55, 56]

The quality of docking algorithms is of crucial importance to the development of novel therapeutics; therefore, the spotlight on peptide drugs is at risk of fading if computational predictions fail to meet market expectations. When working with peptides, molecular docking studies often struggle to provide positive results; in fact, peptides are generally more flexible than proteins and tend to adopt a broad variety of conformations. At the same time, they are structurally much more complex than small molecules, and the method of predicting their affinity with receptors should be up to par. Novel docking methods are therefore focusing on the unique properties of peptides and how to model their flexibility [57, 58].

Molecular docking methods can be classified into three main categories: i) protein-peptide, ii) protein-protein, and iii) protein-small-molecule docking. Although peptide therapeutics databases are not lacking in number, only a few algorithms are currently specifically designed for peptide-protein docking. Previous studies have observed how some methods, such as ZDOCK, despite being developed for protein-protein docking, can also be used to dock peptides on a protein. Similarly, some software developed for docking small molecules on a protein, such as AutoDock and AutoDock Vina, can be adapted for peptide docking.

Protein-peptide interactions are modeled into the following three categories: Template docking, Ensemble docking and *de novo* methods. Briefly summarized in

the following table [59].

Category	Flexibility	Description
Template docking	None or little	Use sequence-based homology model to predict docking
Ensemble docking	Conformation Ensemble	Prepare a conformation ensemble to describe peptide, flexibility and then dock the conformations back into receptor
de novo methods	Fully flexible	Fully flexible & Model peptide flexibility with the respect to the receptor

Table 4.1: Protein-protein docking categories

Each of these methods has its own advantages and disadvantages and is not inherently better in every situation, Template docking algorithms (eg. GalaxyPepDock, FRODOCK) require homologue structures for both the receptor and the peptide, thus limiting the range of their applicability. Ensemble docking methods sample peptide conformations as a pre-processing step without knowledge of the receptor. Next, these conformations are docked rigidly or semi-rigidly into the receptors (eg. HPepDock, pepATTRACT). Despite these methods good accuracy for small and medium sized peptides (typically *le* 9 amino acids), their success rates tend to drop rapidly with longer peptides. Lastly, *de novo* methods sample the peptide's conformation during the docking process (eg. AutoDock CrankPep, FlexPepDock, HADDOCK) While de novo methods yield high accuracy and are less affected by the length of the peptides, these methods tend to be computationally expensive and often rely on lengthy molecular dynamics simulations to refine solutions.

According to previous studies [60, 61] results; FRODOCK [62], AutoDock CrankPep[63, 59], and HPepDock [64, 65, 66, 67, 68, 69, 70] web server were chosen as the best-performing methods, each with a different approach in order to cover a broader range of solutions. Seeking the highest customization level of the algorithms parameters FRODOCK and ADCP were run through command line. Due to the scalable properties this study is looking for, external Python scripts were used whenever the methods would have required time-consuming operations or the user intervention in order to obtain a smooth docking pipeline.

4.2 FRODOCK

The first algorithm tested focuses on the first stage of docking, which consists on rigid-body orientational sampling of a ligand molecule with respect to a fixed receptor molecule while a docking scoring function is maximized. The 6D sampling

space of the relative orientations between ligand and receptor is huge, and therefore computationally demanding [62]. Here, the molecules are represented by 3D grids that carry information of the shape, the ligand and receptor grids are then correlated using FFT to efficiently scan the translational space. After the Fourier-based evaluation has been complemented by an implicit orientational search, a large number of docked conformations with favorable surface complementarity can be obtained. This initial shape-based scoring function has been further enhanced by including additive correlation terms to consider electrostatics, solvation or even statistical interaction potentials. This approach permitted a superior efficiency and a more exhaustive search by speeding up the three rotational degrees of freedom using SH and a convenient formulation of the 3D rotation group. This approach permits a superior efficiency and a more exhaustive search by speeding up the three rotational degrees of freedom using SH and a convenient formulation of the 3D rotation group. The application of Fast Rotational Method to protein–protein docking has derived in new mathematical expressions, and hence in a new docking method called FRODOCK (Fast ROtational DOCKing).

In contrast to other approaches, FRODOCK has the advantage of combining the capability to express the interaction terms into 3D grid-based potentials with the efficiency of a SH-based rotational search. The binding energy upon complex formation is approximated by a sum of three types of potentials: van der Waals, electrostatics and desolvation, each of which can be written as a correlation function. These potentials are conveniently pre-calculated on a 3D grid, using appropriate energy thresholds. The interaction energy minima, and hence the potential docking solutions, are identified by a new fast and exhaustive rotational docking SH-based search combined with a simple translational scanning. A parallel version of FRODOCK can perform the docking search in just a few minutes, and the competitive docking accuracy achieved on standard protein–protein benchmarks demonstrates its applicability and robustness.

This method was run in parallel on ComputeCanada Graham cluster using a single script split in three sections. First the ligand and receptor structures are prepared using **pdb2pqr**, an helpful tool for reconstructing missing atoms, adding hydrogens, assigning atomic charges and radii from specified force fields, and generating PQR files, CHARMM27 was chosen as force field, indicating which charge and radius parameters to use. Second a Python script (making use of the BioPython library [71]) parsing the ligand's PDB file and extracting the interface coordinates. At last the job containing the FRODOCK algorithm is submitted with ligand, receptor and interface coordinates as inputs. The main code is divided in four steps:

1. frodockgrid: three grid potential maps are generated (van der Waals, electrostatic and desolvation) for the receptor and desolvation only for the ligand. Atomic properties such as van der Waals radius, charges etc. are taken from

CHARMM 19 force field and the SASA calculations were performed using analytical methods. [72]

2. frodock: the docking 6D search is performed on the pre-calculated grid maps, where the interface coordinates are given as inputs using the –**around** flag. The Message Passing Interface (MPI) binary files allow the process parallelization, thus a sensible time saving.

3. frodockcluster: the solutions generated by the docking algorithm are clustered, a maximum of 100 clusters is allowed with an RMSD distance of 2 Åbetween clusters.

4. frodockview: the best 100 models are saved, using the native peptide conformation as reference for RMSD calculations.

For this experiment, a receptor center was considered more appropriate for blind docking, and therefore no previous knowledge of the native peptide-protein interaction. Nonetheless an additional script calculating the interface center has been created as well and can be found in Appendix A.

Listing 4.1: Retrieve the receptor center coordinates

```
from Bio import PDB
import sys
import os
import math

code = sys.argv[1]
code1 = []
code1.append(code)
code1.append('-rec.pdb')
rec = ''.join(code1)
center = []
parser = PDB.PDBParser(QUIET=True)

models = os.listdir('/home/pietro5/env/frodock/')

# receptor center coordinates
receptor = parser.get_structure('rec', rec)
total = []
for model in receptor:
    for chain in model:
        for residue in chain:
            for atom in residue:
                center = atom.get_coord()
                total.append(center)
```

```
25
26 #rec_list = PDB.Selection.unfold_entities(receptor, 'A')
27 #t = []
28 #for i in range(len(rec_list)):
29 #    c = rec_list[i].get_coord()
30 #    t.append(c)
31
32 rec_center = sum(total)/len(total)
33 foo = []
34 foo.append(str(round(rec_center[0],1)))
35 foo.append(',')
36 foo.append(str(round(rec_center[1],1)))
37 foo.append(',')
38 foo.append(str(round(rec_center[2],1)))
39 coords = ''.join(foo)
40 print(coords)
```

4.3 AutoDock CrankPep

ADCP is a novel *de novo* method that folds the peptide in the energy landscape generated by the receptor and based off CRANKITE, an efficient software package originally developed for protein and peptide conformation sampling and folding [73, 74]. CRANKITE samples the conformational space of proteins or peptides using a Metropolis Monte Carlo (MC) search and a Gō-Type representation of amino acid side-chains [75, 76]. ADCP incorporates CRANKITE's conformation sampling ability with the grid-based AutoDock representation of a rigid receptor [77] to optimize the peptide conformation and its interactions with the receptor, thus yielding docking poses. The noteworthy modifications and additions are as follows: (i) the addition of new MC moves to boost the exploration of peptide position and orientation relative to the receptor; (ii) the addition of an energy term based on the AutoDock affinity grids to describe the peptide-receptor interactions; (iii) the use of a rotamer library [78] to interactively construct side-chain atoms; and (iv) the addition of a pose cache swapping mechanism to enhance the search. The overall workflow of the MC procedure implemented by ADCP could be condensed in two phases. First, a randomly selected MC move is applied to alter the current pose. The altered pose is then scored, and the move is either accepted or rejected based on a metropolis-like MC criterion. If the move is rejected, the pose before the move is restored and another move is attempted. If it is accepted, the altered pose becomes the current one and is used to update the cache of docking poses. This procedure repeats until one of the termination criteria is met. More details about the various elements of this workflow are provided below. This method requires both the peptide and the receptor to be processed by AutoGridFR [79]

and AutoGrid4[74] respectively, generating the target files containing the affinity maps, subsequently, the docking is performed using an extended set of MC moves which further increment the range of conformations explored by the peptide. In addition to the local alteration of the peptide structure, a new function called *translational jump* allows the peptide to translate to another position if deemed accurate by *AutoGridFR* [80]. The scoring function is composed by two terms, referred correspondingly to the conformation of the peptide, derived from an enhanced version of CRANKITE Go-Type potential [81, 77] and the interaction between the peptide and the receptor, based off the AutoDock affinity grids. Like in the previous test, the algorithm was incorporated in a script divided in three sections in order to provide helpful data to the main code: i) a python script parsing the PDB file of the ligand and output the amino acid sequence ii) a second python script counting the number of residues of the ligand iii) ADCP pre-processing and docking. The last section follows the guidelines recommended by the Center for Computational Structural Biology (CCSB) and can be better described as:

1. reduce: uses a script found in the ADFRSuite to protonate both ligand and receptor.

2. prepare_ligand and prepare_receptor: convert the PDB files into PDBQT format (two columns for the partial charge and AutoDock atom type are added), both scripts are included in the ADFRSuite)

3. agfr: generates the target files receiving as inputs the ligand and receptor PDBQT files. Using a BioPython script the ligand size in residues is specified through the **-ls** flag. The smallest box encompassing the receptor was used as docking box and the binding pocket was identified as the one with the best Autosite score.

4. adcp: docks the peptide from the sequence, specified in the **-s** flag, and receptor coordinates file. The **-cyc** and **-cys** flags allow further customization of the docking process, supporting tailored solutions for cyclic peptides through backbone and CYS-S-S-CYS respectively. A native contacts cutoff of 0.8 was used for clustering the solutions, generating 20 replicas performing a maximum of 1000000 steps per replica.

Outlined below are two original scripts implemented in the docking pipeline.

Listing 4.2: Extract the FASTA sequence from a PDB file

```
1 from Bio import PDB
2 import sys
```

```
3
4 d3to1 = {'CYS': 'C', 'ASP': 'D', 'SER': 'S', 'GLN': 'Q', 'GLU': 'E',
      'LYS': 'K', 'MET': 'M', 'ILE': 'I', 'PRO': 'P', 'PYL': 'O', 'THR':
       'T', 'PHE': 'F', 'ASN': 'N', 'GLY': 'G', 'HIS': 'H', 'LEU': 'L',
      'ARG': 'R', 'TRP': 'W', 'ALA': 'A', 'TYR': 'Y', 'VAL': 'V', 'SEC':
       'U', 'ASX': 'B', 'GLX': 'Z', 'XAA': 'X', 'XLE': 'J'}
5
6 code = sys.argv[1]
7 code1 = []
8 code1.append(code)
9 code1.append('-lig.pdb')
10 lig = ''.join(code1)
11
12 parser = PDB.PDBParser(QUIET=True, PERMISSIVE=True)
13 structure = parser.get_structure('struct', lig)
14 seq = []
15 for model in structure:
16     for chain in model:
17         for residue in chain:
18             if residue.resname in d3to1:
19                 lig = seq.append(d3to1[residue.resname])
20 sequence = (''.join(seq))
21 print(sequence.lower())
```

Listing 4.3: Count the number of AA in the peptide

```
1 from Bio import PDB
2 import sys
3
4 parser = PDB.PDBParser(QUIET=True, PERMISSIVE=True)
5 code = sys.argv[1]
6 code1 = []
7 code1.append(code)
8 code1.append('-lig.pdb')
9 lig = ''.join(code1)
10
11 d3to1 = {'CYS': 'C', 'ASP': 'D', 'SER': 'S', 'GLN': 'Q', 'GLU': 'E',
      'LYS': 'K', 'MET': 'M', 'ILE': 'I', 'PRO': 'P', 'PYL': 'O', 'THR':
       'T', 'PHE': 'F', 'ASN': 'N', 'GLY': 'G', 'HIS': 'H', 'LEU': 'L',
      'ARG': 'R', 'TRP': 'W', 'ALA': 'A', 'TYR': 'Y', 'VAL': 'V', 'SEC':
       'U', 'ASX': 'B', 'GLX': 'Z', 'XAA': 'X', 'XLE': 'J'}
12
13 # clean C and N termini ions before running
14 ligand = parser.get_structure('lig', lig)
15 num = 0
16 for model in ligand:
17     for chain in model:
```

```
18          seq = []
19          for residue in chain:
20              if residue.resname in d3to1:
21                  seq.append(d3to1[residue.resname])
22                  num = num + 1
23          sequence = ''.join(seq)
24  print(num)
```

4.4 HPepDock

This protocol can be divided into three branches, combining MODPEP sampling the space of conformations of the peptide, a revised version of MDOCK for the docking procedure and a number of programs for structure modeling. In the first place inputs are submitted to the server. If the receptor structure is provided it will be used, otherwise, if only the sequence is available, a structure modeling server will handle it using a homology modeling algorithm. After selecting a template a combination of programs build the 3D structure. On the other hand, only sequences are accepted as peptide input. Additional information about binding sites or binding modes is optional. Subsequently 1000 peptide conformations are generated using MODPEP, forming the ensemble that will be later submitted to the docking program. The docking process is performed by a modified version of MDock, employing a rigid docking protocol, where peptide flexibility is achieved by running the simulation on the previously generated template. Whether information regarding the binding site is provided or not, local spheres (around the binding sites) or global spheres (covering the entirety of the receptor) are generated. In the first case local docking will be performed, otherwise the spheres will be clustered around the most probable binding sites. The modifications applied to MDock in order to better portray peptide-protein interaction are two: p

1. Peptides are first docked as reduced models where each residue is represented by the Ca atom and a center of mass atom for the rest of the residue, and after the orientations are generated the atoms are replaced, highly facilitating the calculations.

2. The original (ligand-protein) scoring function is replaced with a knowledge-based (protein-protein) function

Finally, a SIMPLEX EM algorithm is performed on the system considering both the peptide and the protein rigid bodies, aiming to minimize their binding energies.

Chapter 5

Molecular Dynamics

5.1 Introduction

After testing out different algorithms a general quality assessment can already be presumed, although in order to further improve the understanding an EM and MD run were performed on the peptides, relaxing the structure in order to mimic a physiological state. This way the goal becomes finding a robust algorithm able to predict the correct docking pose starting from the bound drug, but to extend the peptides pool to relaxed and energy minimized structures, more likely to be found in existing databases. After performing an EM the peptides did not go through any significant change of conformation, therefore a short MD simulation was required to allow the structures to escape from the energy minima. The AMBER package offers two major MD engines, in this study PMEMD was employed; it is an extensively revised version of the other package (SANDER) highly optimized to improve both single-processor performance and parallel scaling. The process required two distinct steps: first ions and water are added until the system reaches an equilibrium, the peptide structures tested in the previous chapter were indeed only composed of residues atoms (downloaded from PDB with no further modification), subsequently EM and MD simulations are performed and the last frame extracted. MD simulations require a specific set of input files to give reliable results and for complex structures with a considerable amount of missing residues the best choice is usually to split the procedure into different steps.

5.2 Molecular Dynamics simulation

5.2.1 Structure preparation

Because of the many naming conventions used in pdb files the peptides were processed through the *pdb4amber* program, which makes the input files readable by the AMBER packages renaming atom and residue names. [82, 83] The output structure files were then given as inputs to *tleap*, generating a topology (.parm7) and a coordinates (.rst7) file using a text-based interface. The protein force field ff14SB is selected along with the the water model tip3p, the choice of a not so recent force field is guided by its well-tested performances and compatibility with the popular tip3p water model. Afterward, an octagonal box is created around the structure, following solvation and charges equilibration with ions addiction. The following sections have been run on the Beluga cluster of the Alliance Canada which provided the necessary computational power, significantly shortening the time needed. For each peptide a short protocol inspired by Ross Walker (2013) AMBER advanced tutorial number 22, was executed, the steps as follows.

5.2.2 Minimization, heating, and equilibration

Before proceeding to the main simulations the system needs to be prepared, this will help stabilize the native conformation in the solvent box. While starting a simulation it is crucial to minimize the energy to avoid the generation of large forces which can lead to a crash.

1. Water minimization: the water potential in the box is firstly minimized, the peptide is restrained.

2. Water relaxation: the water is let free to move (NpT, 300K) while the protein is restrained.

3. System minimization: for 2000 cycles the both the peptide and water are minimized.

4. Heating: the system is heated up restraining the peptide position (NVT, 0K to 300K).

5. Equilibration: system relaxation restraining the protein heavy atoms (NpT, 300K, 0.5ns).

This first part was compiled using the MPI executables of PMEMD called *pmemd.mpi*, allocating 4 tasks, 1 cpu-per-task and 10G of memory size required per cpu. The system is now ready for the EM and MD and simulations.

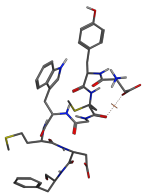
5.2.3 Energy Minimization and Molecular Dynamics run

In this second phase the energy is first minimized relaxing the system and afterwards an MD simulation provides an agitated state of the peptide, finally extracted as a pdb file using *ccptraj*. PMEMD is run in its single-GPU accelerated version called *pmemd.cuda* because of the significant higher required computational power.

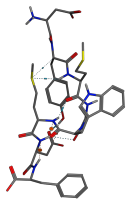
1. Relaxation: relative to the whole system ((NpT, 300K, 5ns)
2. MD: after obtaining an initial relaxed structure a simulation is now performed (NVT, 500ns), resulting in a perturbed system at the temperature of 300K
3. pdb file extraction: the last frame of the simulation is extracted with the *ccptraj* command.

5.3 Comparison with the native structures

The relaxed peptides were finally analyzed to ensure the positive outcome of the simulation. Graphic comparison and RMSD outlined below.



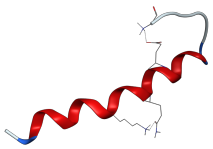
(a) Native



(b) Altered

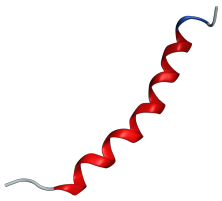
Figure 5.1: Sincalide, RMSD = 3.07 Å

(a) Native

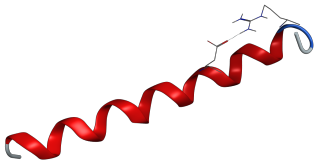


(b) Altered

Figure 5.2: GLP-1, RMSD = 5.35 Å

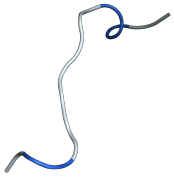


(a) Native

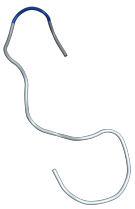


(b) Altered

Figure 5.3: Semaglutide, RMSD = 2.04 Å

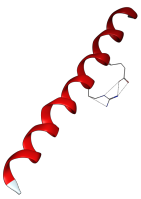


(a) Native

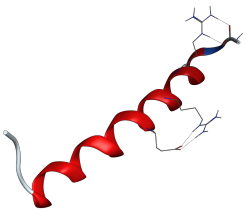


(b) Altered

Figure 5.4: Salmon calcitonin, RMSD = 4.38 Å

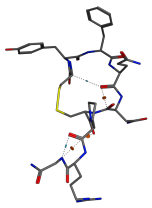


(a) Native

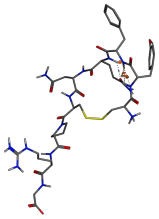


(b) Altered

Figure 5.5: Secretin, RMSD = 3.44 Å

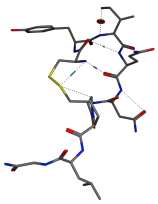


(a) Native

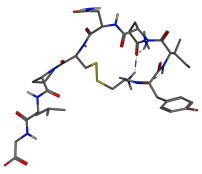


(b) Altered

Figure 5.6: AVP, RMSD = 1.57 Å



(a) Native



(b) Altered

Figure 5.7: Oxytocin, RMSD = 1.64 Å

Chapter 6

Results

6.1 Native peptides

The peptides native structures were first docked to their receptors using ADCP, FRODOCK and HpepDock using the parameters described in Chapter 3. The analyzed parameters are i-RMSD, L-RMSD. A *de novo* (ADCP) and a rigid docking (FRODOCK) method were additionally chosen to compare their RMSF. In order to obtain accurate results ProFit Version 3.3 (Martin, A.C.R. and Porter, C.T., http://www.bioinf.org.uk/software/profit/), a dedicated protein least squares fitting program, was employed. Fitting was performed by implementing the McLachlan algorithm (McLachlan, A.D., 1982 "Rapid Comparison of Protein Structures", Acta Cryst A38, 871-873). Thanks to a user-friendly interface it is possible to isolate specific zones of the analyzed molecules. Additionally, the **-f** flag allows to skip the interactive mode making the analysis of large data sets smoother.

Listing 6.1: An example of the script submitted to ProFit to calculate L-RMSD and RMSF for Sincalide

```
1 atoms Ca,C,N,O
2 fit
3 zone A10-A47
4 fit
5 rzone B1-B8:A1-A8
6 residue
7 nfitted
```

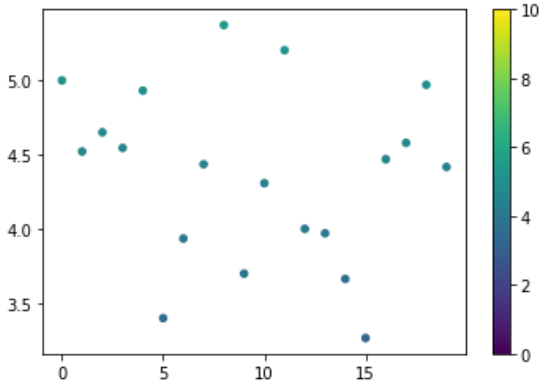
The calculations are referred to the backbone heavy atoms, first the receptors are superposed (this RMSD value is close to zero) and subsequently the RMSD is evaluated on the ligand residues. Finally the **residue** command plots the RMSD value of each residue.

Listing 6.2: An example of the script submitted to ProFit to calculate i-RMSD for Sincalide

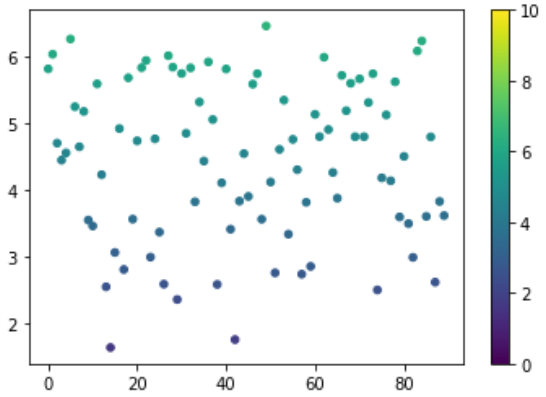
```
1 atoms Ca,C,N,O
2 zone A38–A47
3 zone B1–B8:A1–A8
4 fit
5 nfitted
```

Differently than the previous script, The residues corresponding to the peptide-protein interface are all superposed and RMSD is evaluated. Finally, the raw data

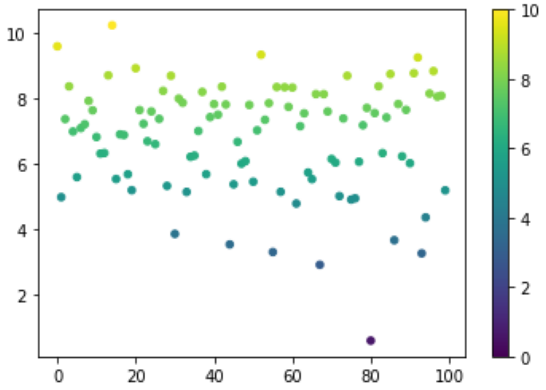
outputs were processed by an original Python script, portraying the results as scatter plots. In the L-RMSD and i-RMSD plots the x and y axis represent the model number and RMSD respectively; while in the RMSF results residues are on the x axis, RMSD on the y axis and each model is assigned a different color.



(a) ADCP

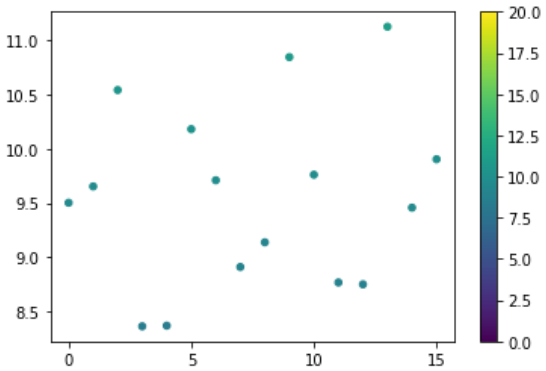


(b) FRODOCK

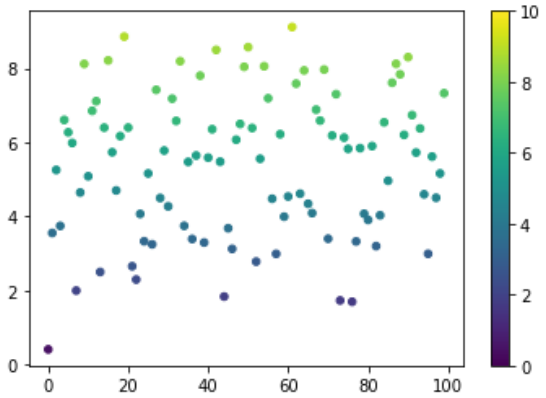


(c) HPepDock

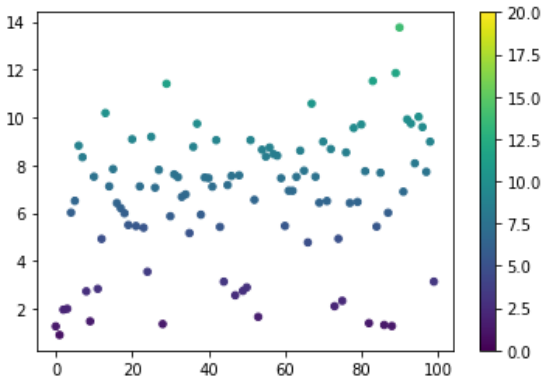
Figure 6.1: Sincalide i-RMSD



(a) ADCP

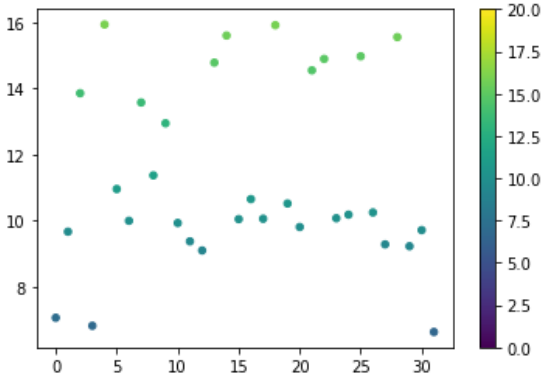


(b) FRODOCK

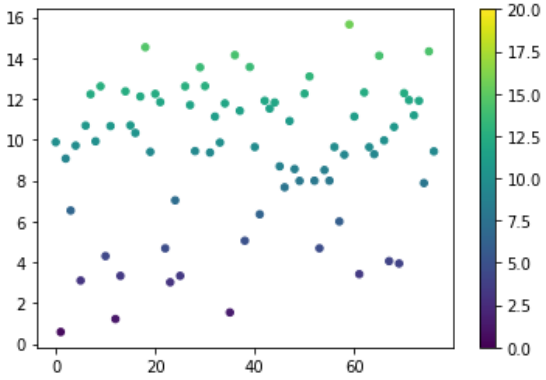


(c) HPepDock

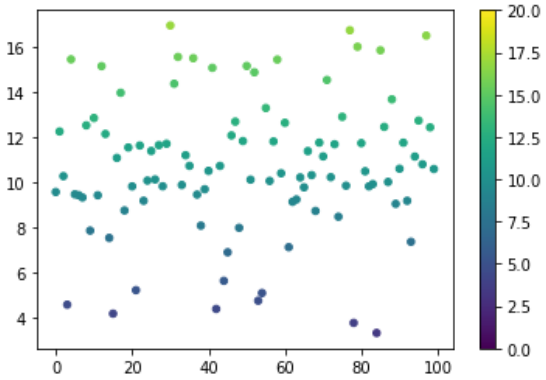
Figure 6.2: GLP-1 i-RMSD



(a) ADCP

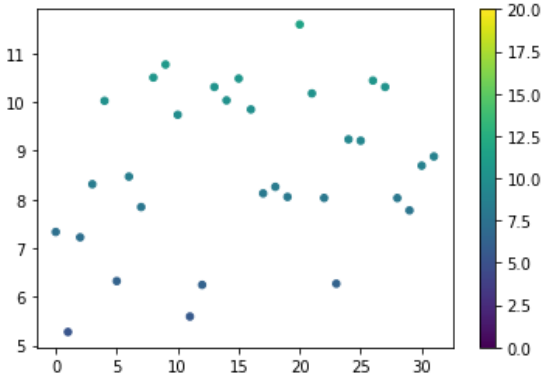


(b) FRODOCK

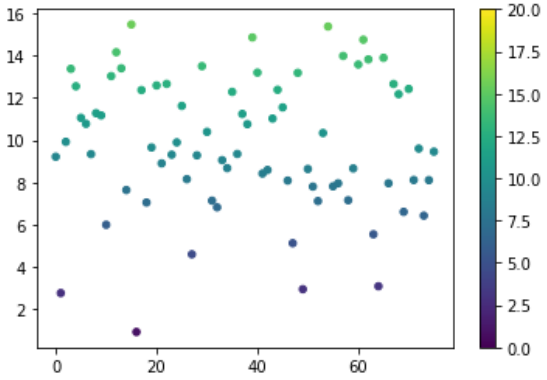


(c) HPepDock

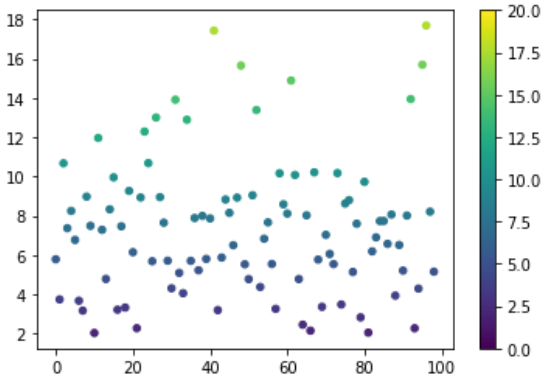
Figure 6.3: Semaglutide i-RMSD



(a) ADCP

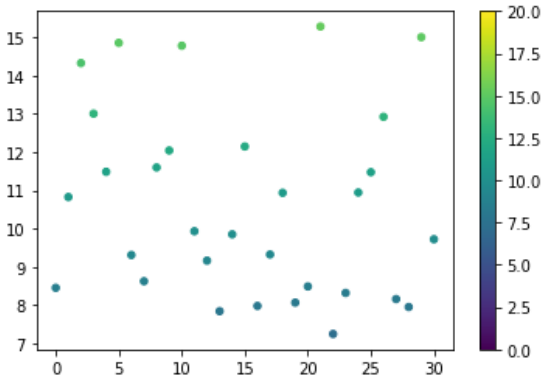


(b) FRODOCK

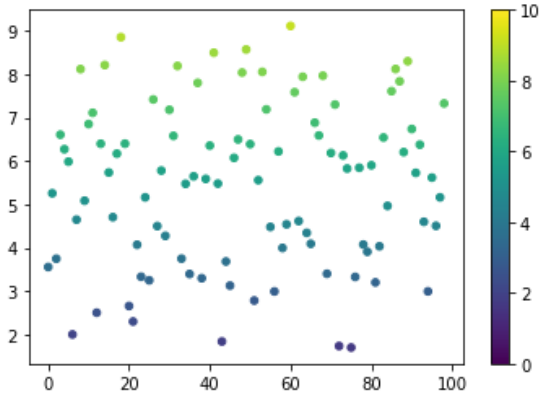


(c) HPepDock

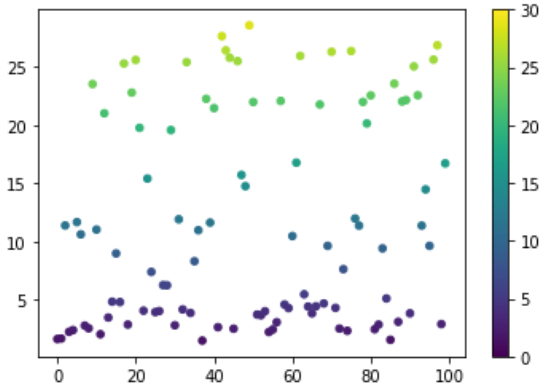
Figure 6.4: Salmon calcitonin i-RMSD



(a) ADCP

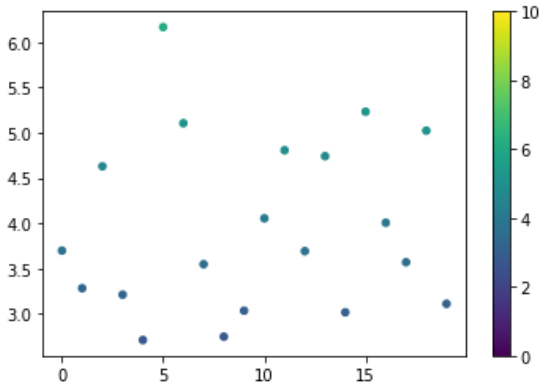


(b) FRODOCK

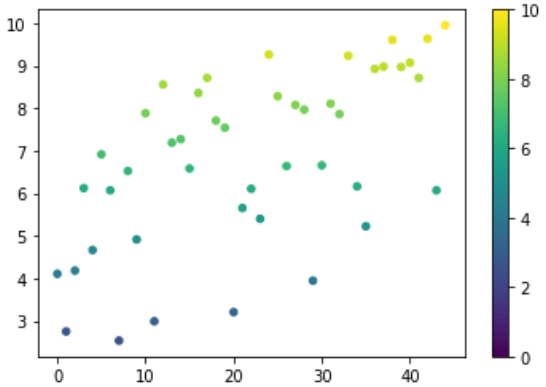


(c) HPepDock

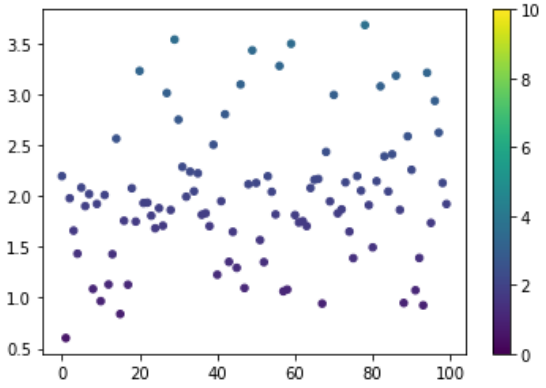
Figure 6.5: Secretin i-RMSD



(a) ADCP

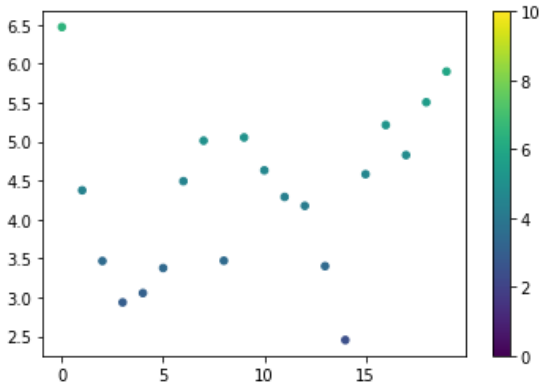


(b) FRODOCK

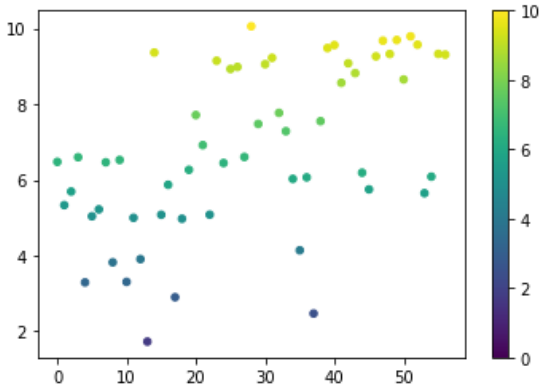


(c) HPepDock

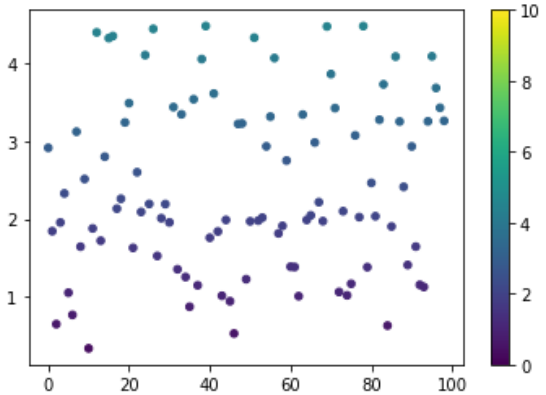
Figure 6.6: AVP i-RMSD



(a) ADCP

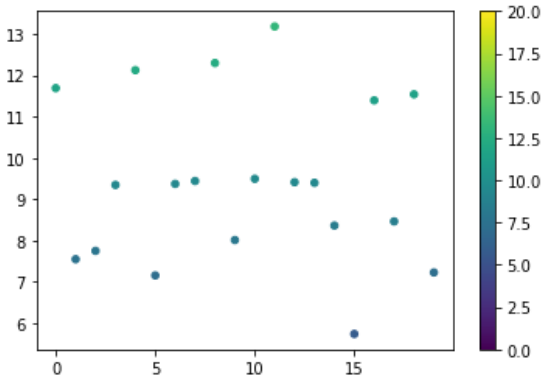


(b) FRODOCK

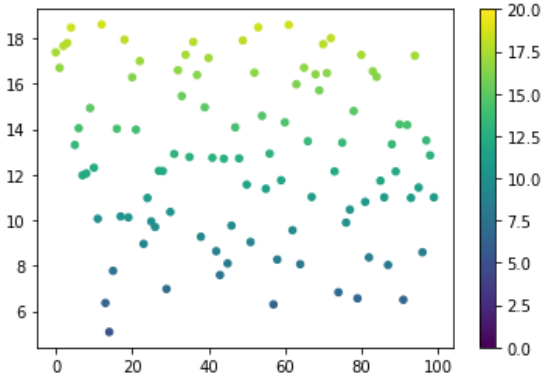


(c) HPepDock

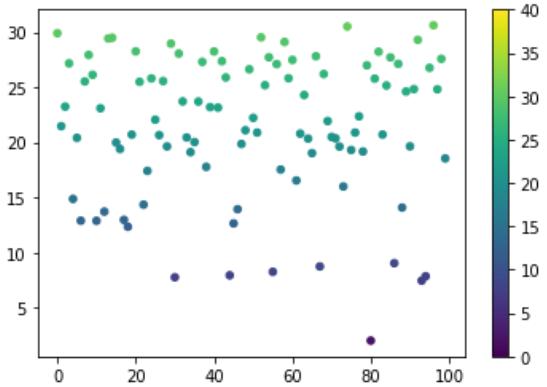
Figure 6.7: Oxytocin i-RMSD



(a) ADCP

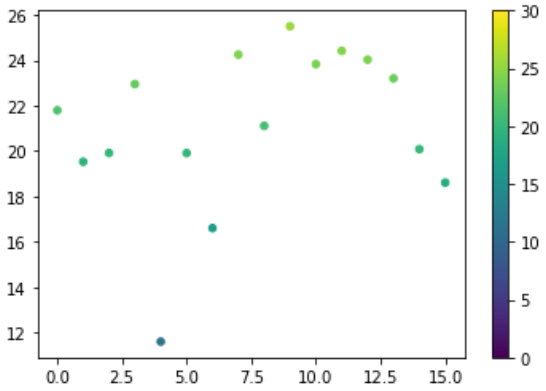


(b) FRODOCK

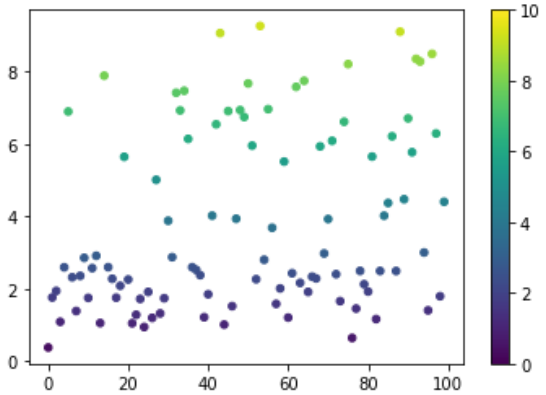


(c) HPepDock

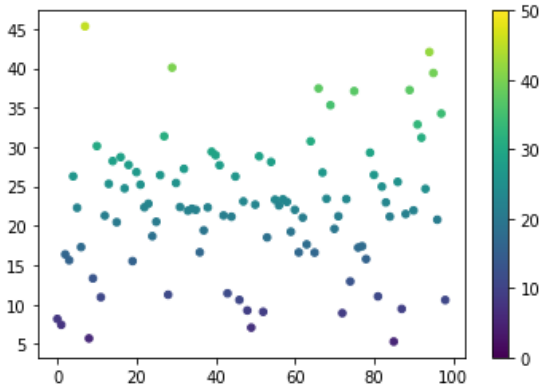
Figure 6.8: Sincalide L-RMSD



(a) ADCP

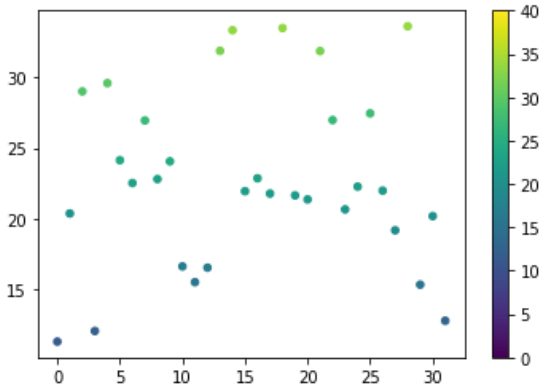


(b) FRODOCK

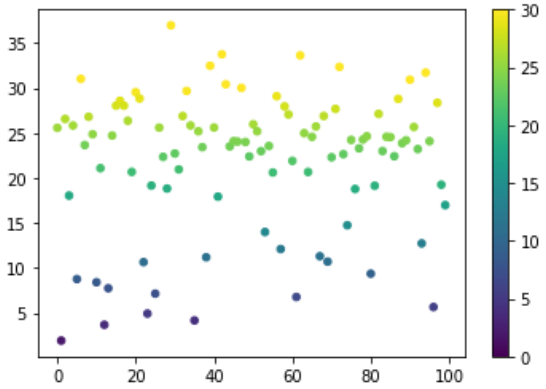


(c) HPepDock

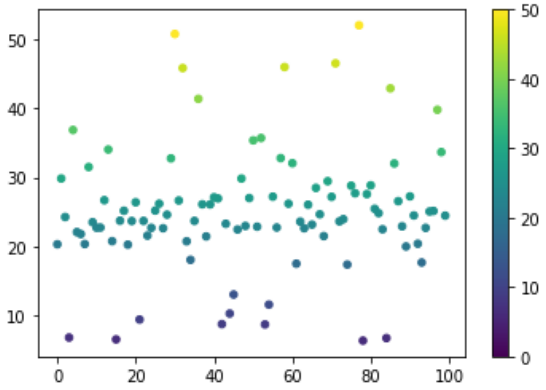
Figure 6.9: GLP-1 L-RMSD



(a) ADCP

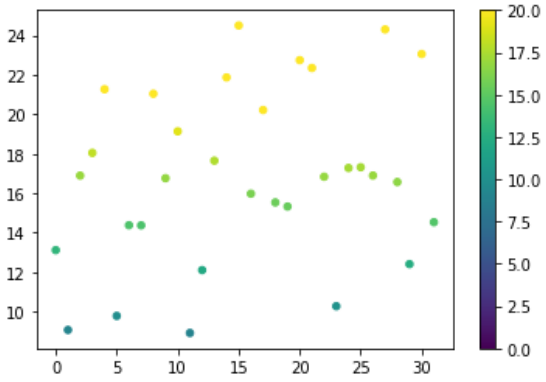


(b) FRODOCK

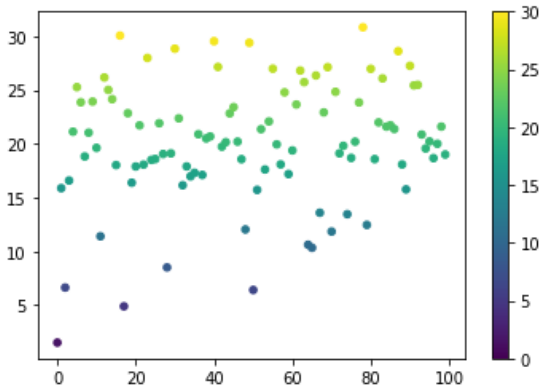


(c) HPepDock

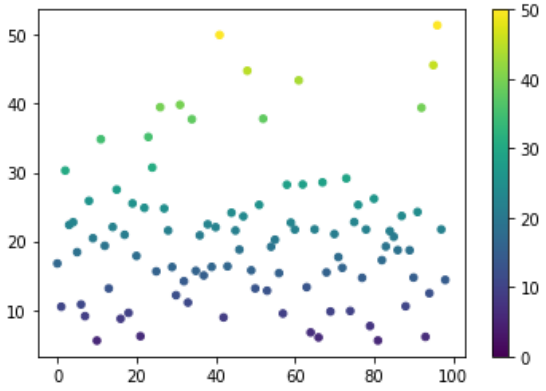
Figure 6.10: Semaglutide L-RMSD



(a) ADCP

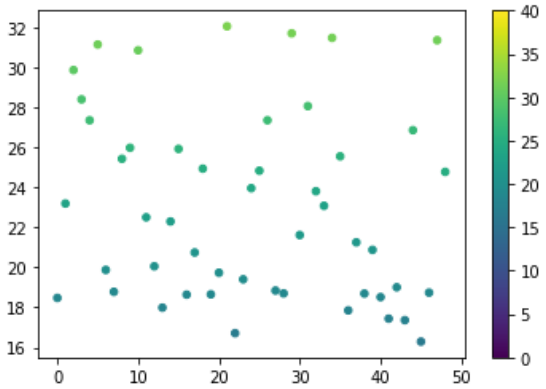


(b) FRODOCK

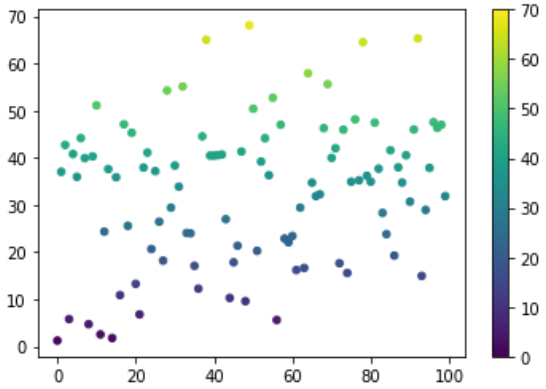


(c) HPepDock

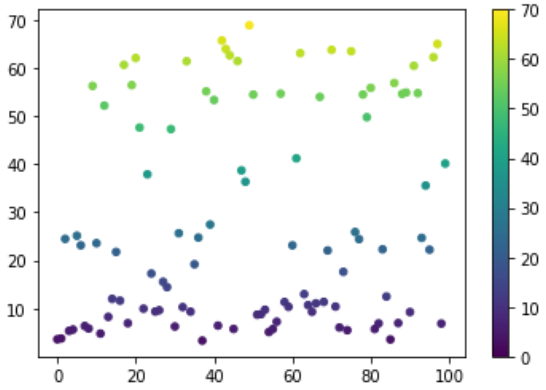
Figure 6.11: Salmon calcitonin L-RMSD



(a) ADCP

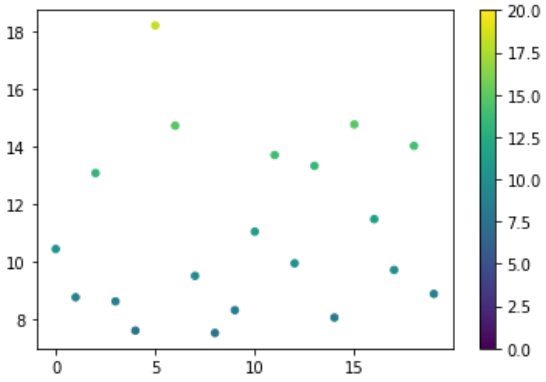


(b) FRODOCK

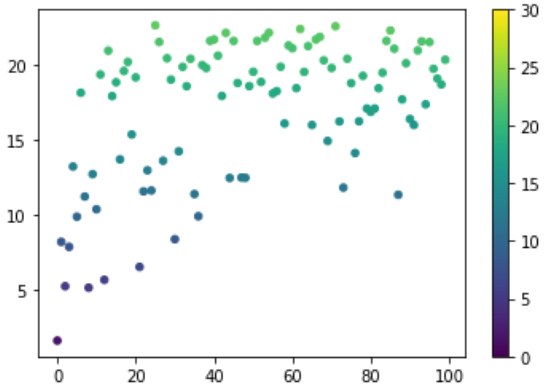


(c) HPepDock

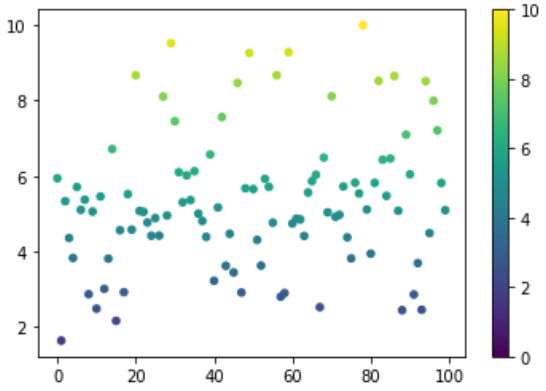
Figure 6.12: Secretin L-RMSD



(a) ADCP

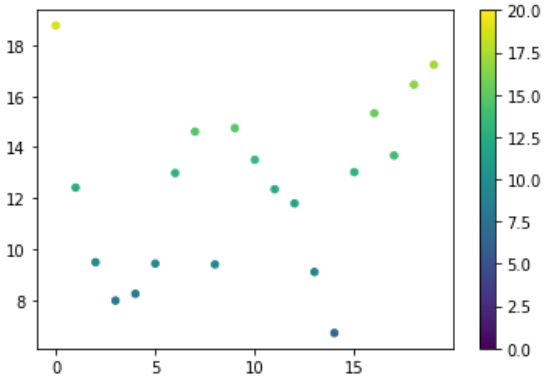


(b) FRODOCK

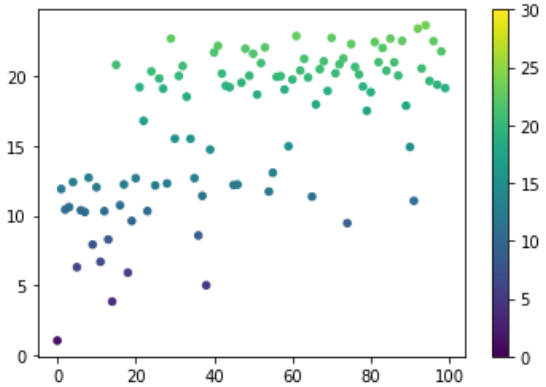


(c) HPepDock

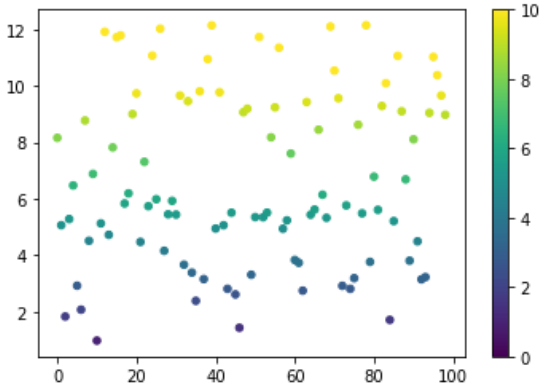
Figure 6.13: AVP L-RMSD



(a) ADCP

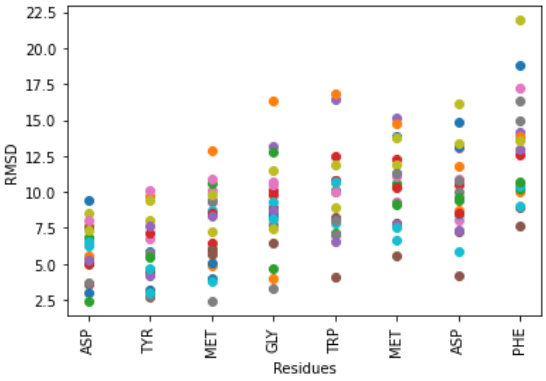


(b) FRODOCK

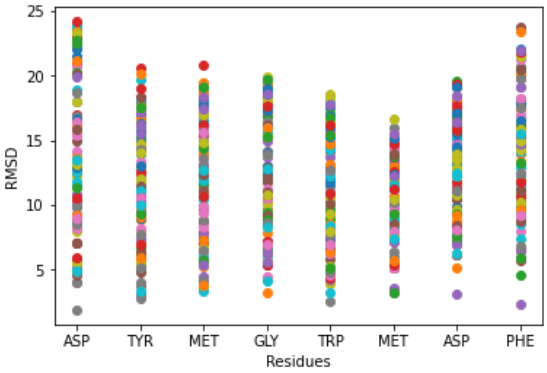


(c) HPepDock

Figure 6.14: Oxytocin L-RMSD

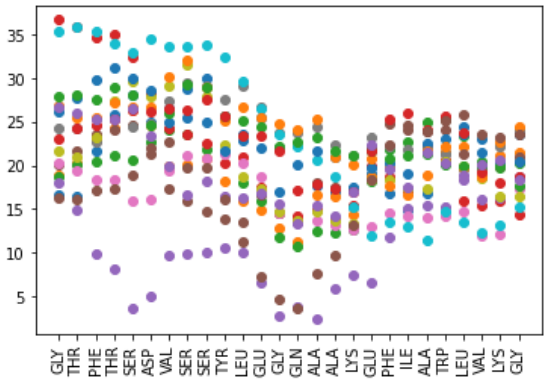


(a) ADCP

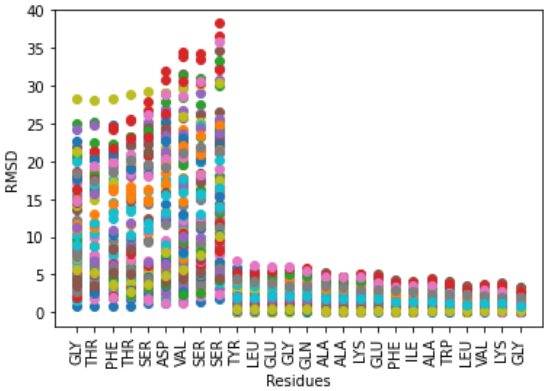


(b) FRODOCK

Figure 6.15: Sincalide RMSF

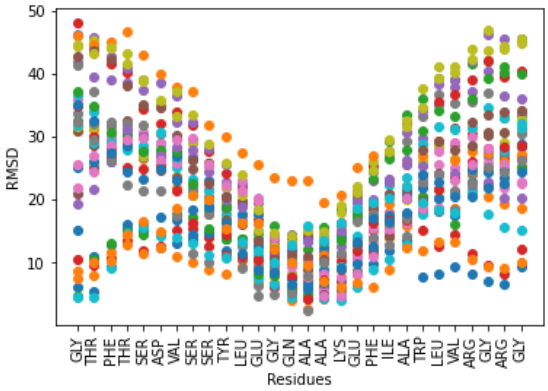


(a) ADCP

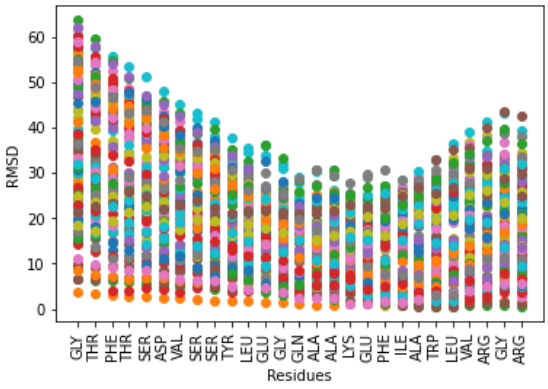


(b) FRODOCK

Figure 6.16: GLP-1 RMSF

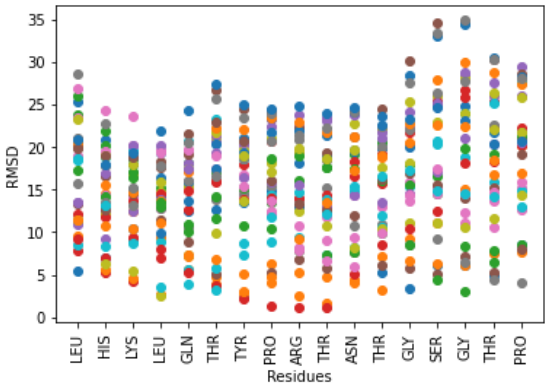


(a) ADCP

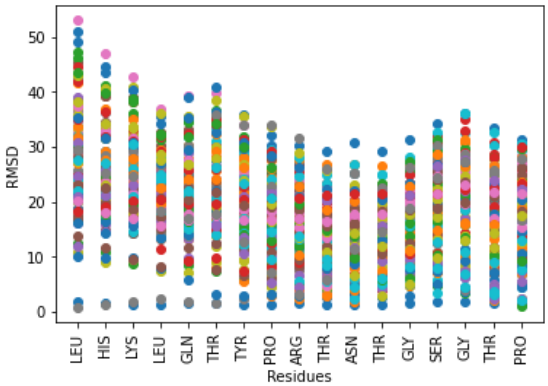


(b) FRODOCK

Figure 6.17: Semaglutide RMSF

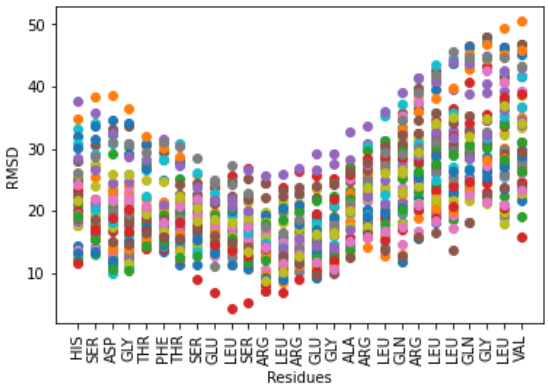


(a) ADCP

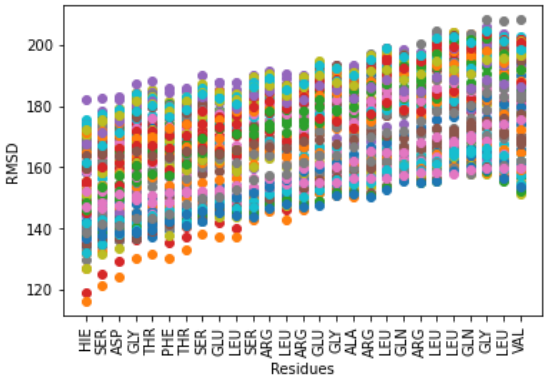


(b) FRODOCK

Figure 6.18: Salmon calcitonin RMSF

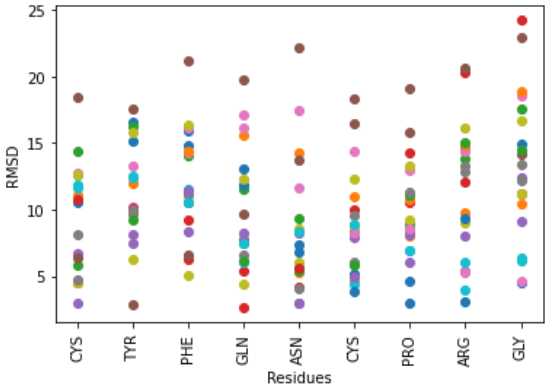


(a) ADCP

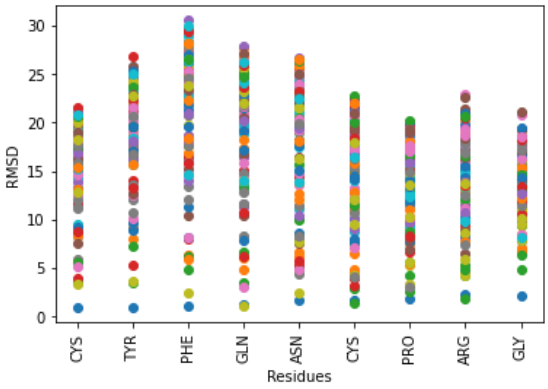


(b) FRODOCK

Figure 6.19: Secretin RMSF

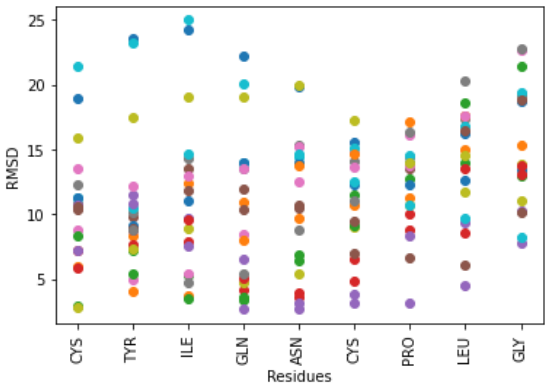


(a) ADCP

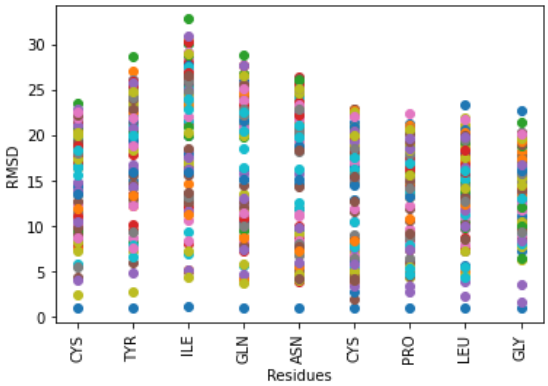


(b) FRODOCK

Figure 6.20: AVP RMSF



(a) ADCP

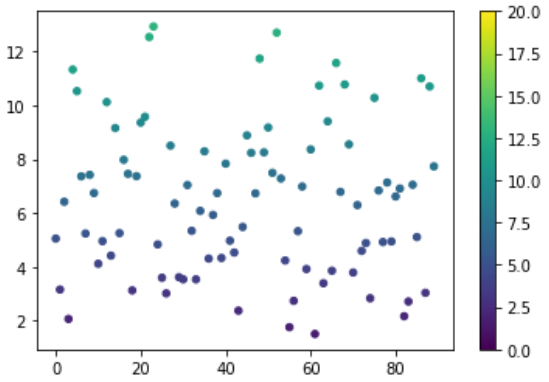


(b) FRODOCK

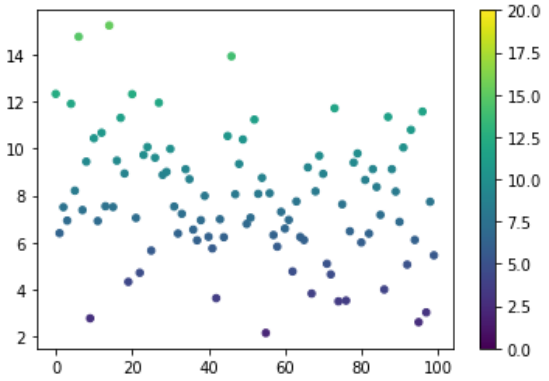
Figure 6.21: Oxytocin RMSF

6.2 Altered peptides

The following plots are the results from docking simulations on the relaxed peptides with their native receptors, seeking an impartial outlook on the docking performance of unbound peptides. The relaxed peptides were submitted only to the best-performing algorithm of each kind (FRODOCK and HPepDock).

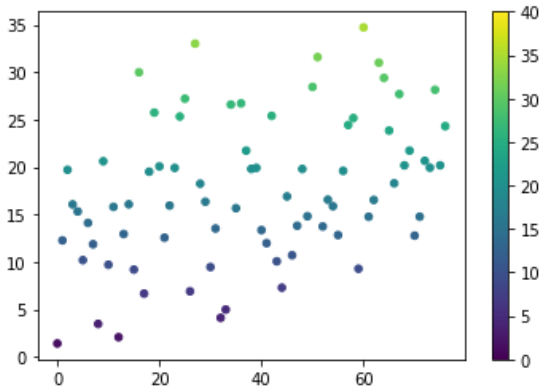


(a) FRODOCK

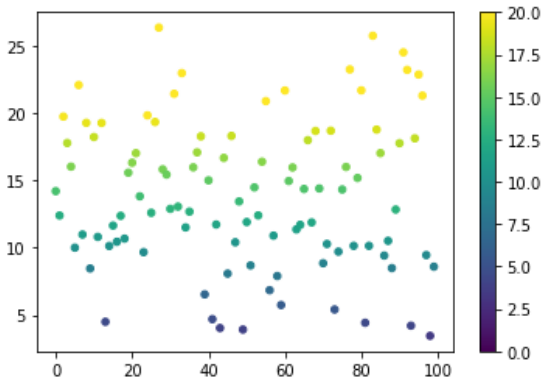


(b) HPepDock

Figure 6.22: Altered sincalide i-RMSD

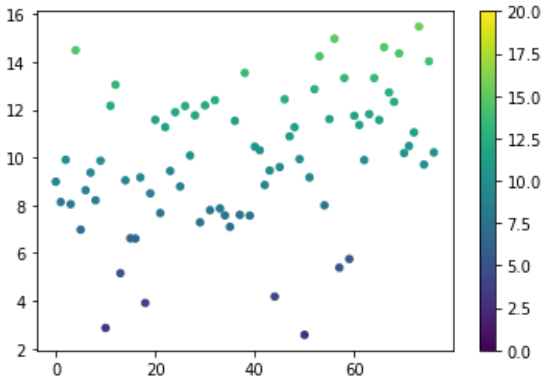


(a) FRODOCK

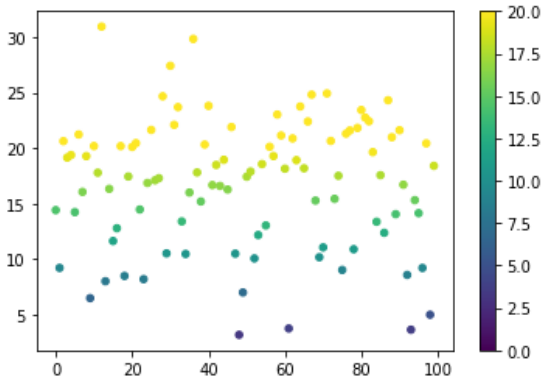


(b) HPepDock

Figure 6.23: Altered GLP-1 i-RMSD

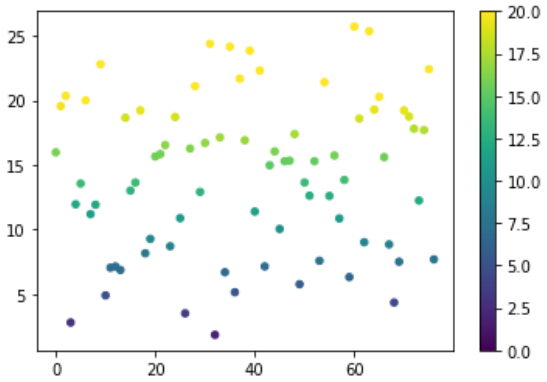


(a) FRODOCK

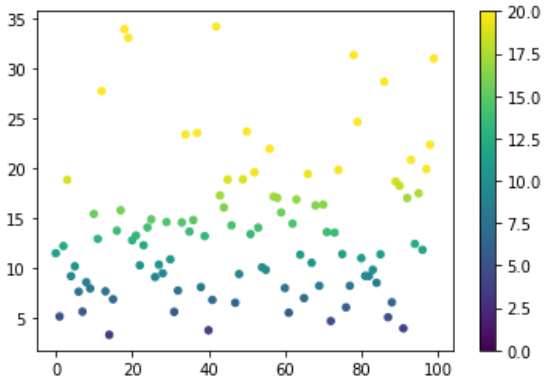


(b) HPepDock

Figure 6.24: Altered semaglutide i-RMSD

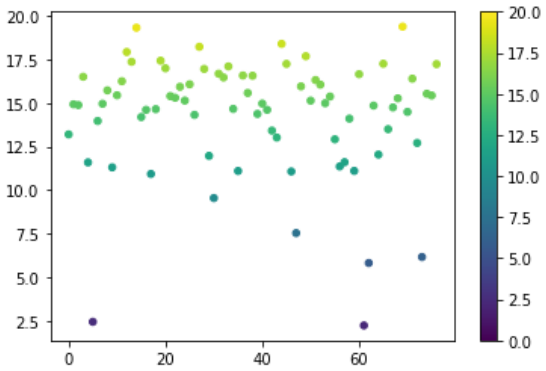


(a) FRODOCK

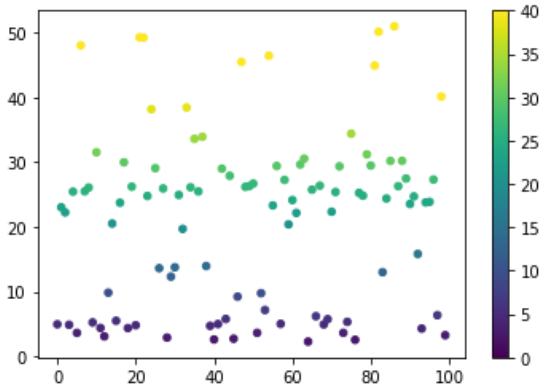


(b) HPepDock

Figure 6.25: Altered salmon calcitonin i-RMSD

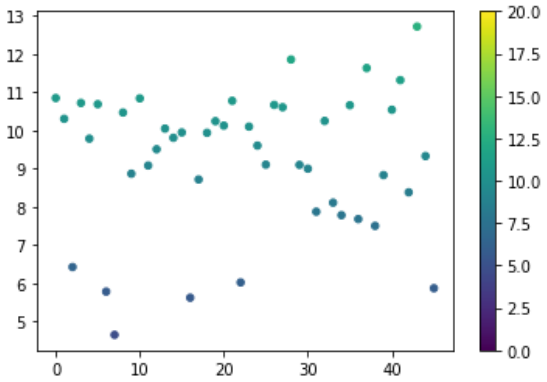


(a) FRODOCK

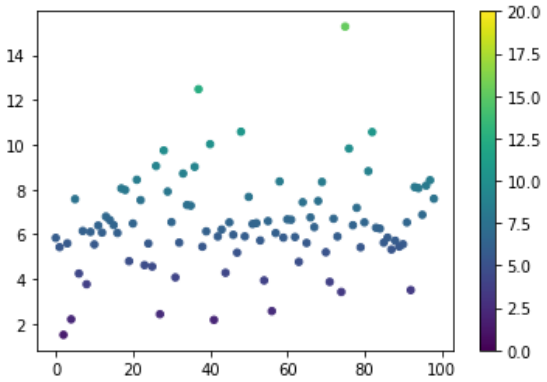


(b) HPepDock

Figure 6.26: Altered secretin i-RMSD

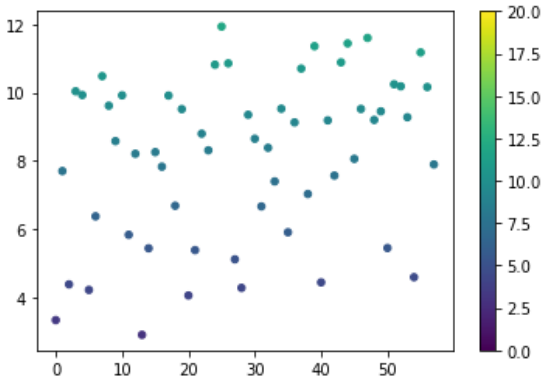


(a) FRODOCK

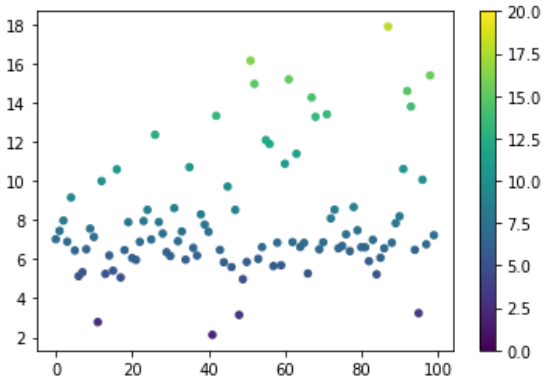


(b) HPepDock

Figure 6.27: Altered AVP i-RMSD

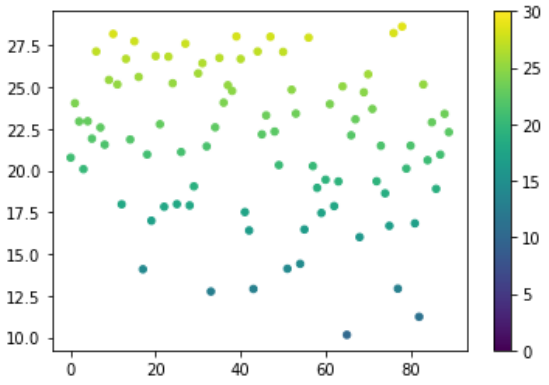


(a) FRODOCK

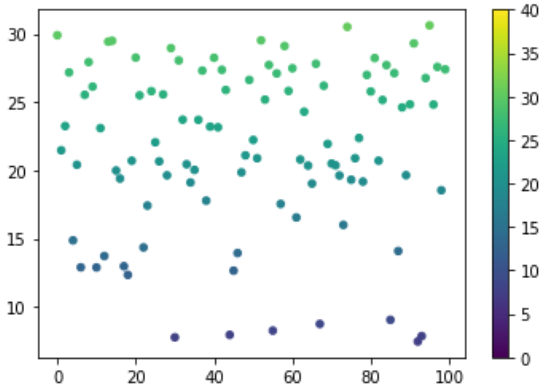


(b) HPepDock

Figure 6.28: Altered oxytocin i-RMSD

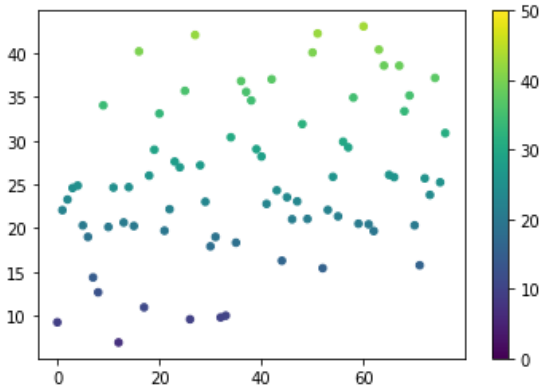


(a) FRODOCK

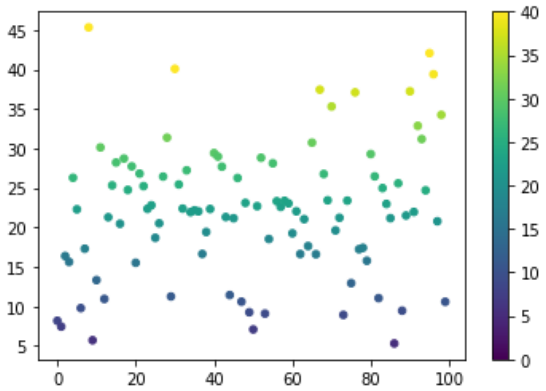


(b) HPepDock

Figure 6.29: Altered sincalide L-RMSD

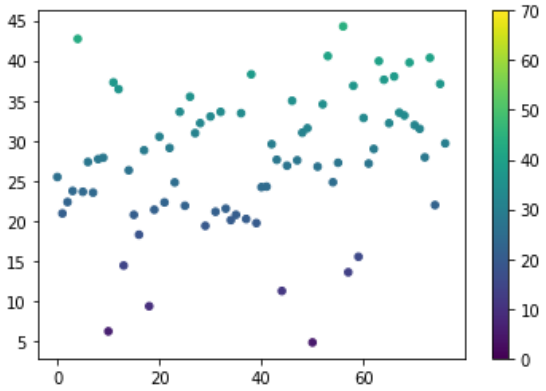


(a) FRODOCK

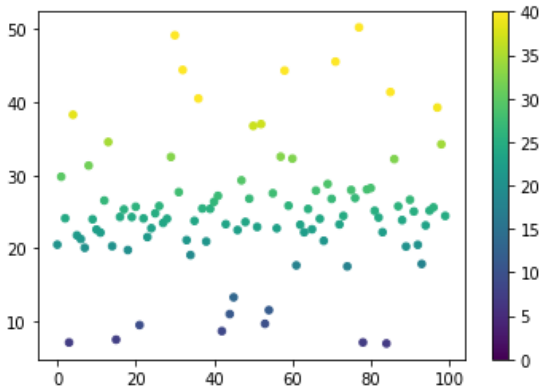


(b) HPepDock

Figure 6.30: Altered GLP-1 L-RMSD

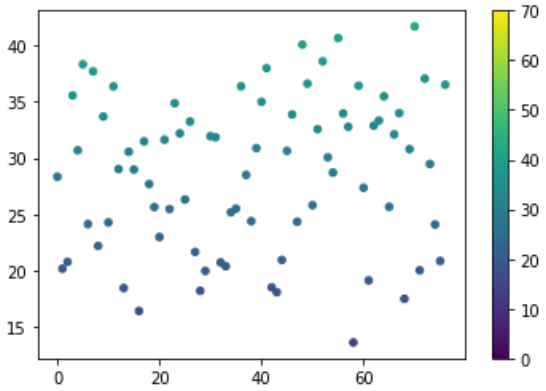


(a) FRODOCK

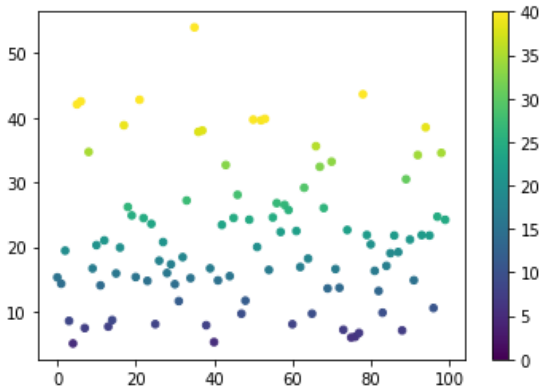


(b) HPepDock

Figure 6.31: Altered semaglutide L-RMSD

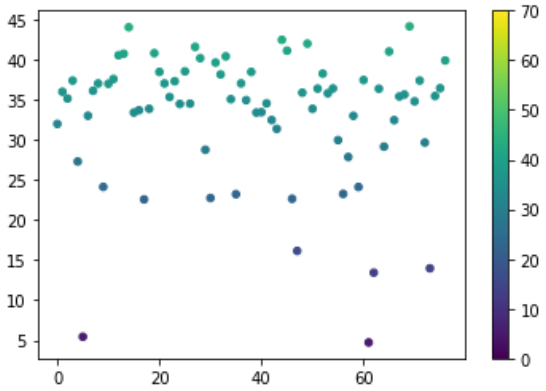


(a) FRODOCK

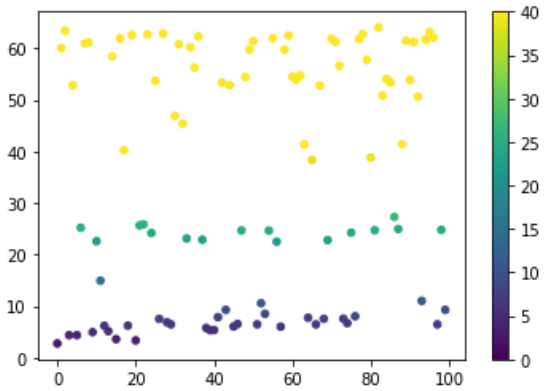


(b) HPepDock

Figure 6.32: Altered salmon calcitonin L-RMSD

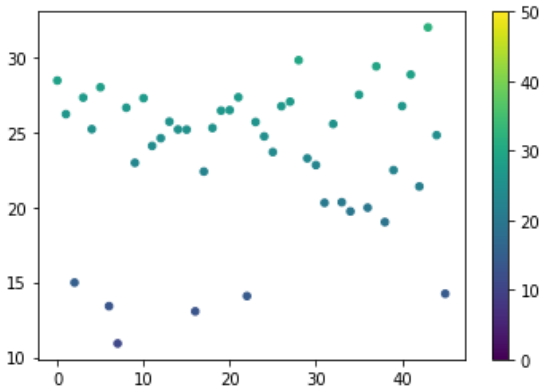


(a) FRODOCK

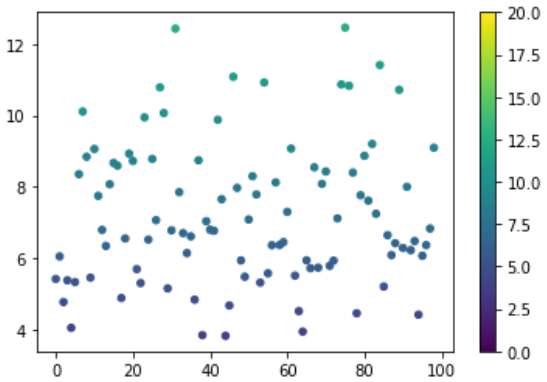


(b) HPepDock

Figure 6.33: Altered secretin L-RMSD

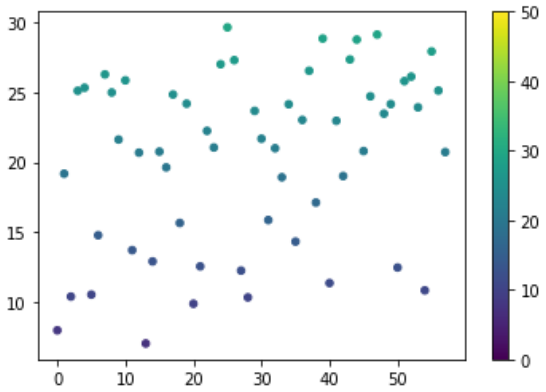


(a) FRODOCK

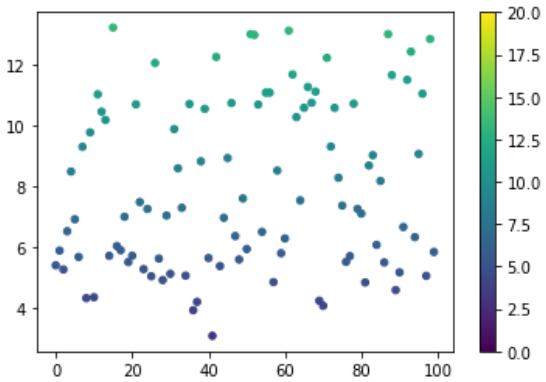


(b) HPepDock

Figure 6.34: Altered AVP L-RMSD



(a) FRODOCK



(b) HPepDock

Figure 6.35: Altered oxytocin L-RMSD

Chapter 7

Conclusions

The docking results can be summarized in the tables presented below, showing the best score for each of the two considered RMSD, relative to native and altered input peptides. ADCP showed inferior quality for both the examined parameters and further tuning of the docking parameters might be beneficial. Average RMSD increase is evaluated in order to highlight the behavior of a protocol over a set of ligands when changing any of the docking parameters. FRODOCK consistently outperformed the other two methods when provided the bound ligand structure, but the quality of predictions decreases with altered structures: moderately for what it concerns the i-RMSD (successfully reproducing the peptide-receptor interaction, ai-RMSD = 0.8 Å) and significantly relative to the L-RMSD (aL-RMSD = 5.9 Å). HPepDock managed to dock native and altered conformation, achieving an ai-RMSD of 1.4 Å and aL-RMSD of 0.8 Å relative to the change of input peptides. It is worthy of consideration how the cysteine-rich peptides (possessing a cyclic structure) salmon calcitonin, AVP and oxytocin averaged better RMSD scores (especially with HPepDock) compared to the rest, but did not perform better when altered (ai-RMSD = 1 Å, aL-RMSD = 9 Å with FRODOCK and ai-RMSD = 1.4 Å, iL-RMSD = 1.2 Å with HPepDock. Ultimately, an additional feature related to the predictive accuracy of docking methods is the type of peptide-protein interaction, where an increased binding surface leads to better results.

Peptide	ADCP	FRODOCK	HPepDock
Sincalide	3.2	1.8	0.6
GLP-1	8.4	0.5	0.9
Semaglutide	6.6	0.9	3.5
Salmon calcitonin	5.2	0.8	1.6
Secretin	7.2	1.8	1.4
AVP	2.6	2.1	0.6
Oxytocin	2.5	1.9	0.3

Table 7.1: i-RMSD [Å] of native peptides docked on their receptors

Peptide	ADCP	FRODOCK	HPepDock
Sincalide	5.9	4.9	2.0
GLP-1	11.8	0.8	5.7
Semaglutide	12.3	2.8	6.4
Salmon calcitonin	11.5	1.1	5.5
Secretin	16.1	3.4	5.1
AVP	7.5	1.0	1.6
Oxytocin	6.8	1.1	1.0

Table 7.2: L-RMSD [Å] of native peptides docked on their receptors

Peptide	FRODOCK	HPepDock
Sincalide	1.2	3.1
GLP-1	1.9	2.7
Semaglutide	2.3	2.0
Salmon calcitonin	1.7	3.4
Secretin	2.4	4.4
AVP	3.9	1.6
Oxytocin	2.2	1.7

Table 7.3: i-RMSD [Å] of altered peptides docked on their receptors

Peptide	FRODOCK	HPepDock
Sincalide	10.1	5.9
GLP-1	6.4	5.1
Semaglutide	5.0	7.4
Salmon calcitonin	13.8	4.7
Secretin	5.1	2.4
AVP	11.9	3.8
Oxytocin	4.7	2.4

Table 7.4: L-RMSD [Å] of altered peptides docked on their receptors

Appendix A

Additional resources

Listing A.1: Extract the coordinates of the center from two structures interface

```
    from Bio import PDB
import sys
import os
import math

def unique(list1):

    # insert the list to the set
    list_set = set(list1)
    # convert the set to the list
    unique_list = (list(list_set))
    return unique_list

code = sys.argv[1]
code1 = []
code1.append(code)
code1.append('-rec.pdb')
rec = ''.join(code1)
center = []
parser = PDB.PDBParser(QUIET=True)

models = os.listdir('/home/pietro5/env/frodock')

# interface center coordinates
code2 = []
code2.append(code)
code2.append('-lig.pdb')
lig = ''.join(code2)
ligand = parser.get_structure('rec', lig)
a_list = PDB.Selection.unfold_entities(ligand, 'A')
b_list = PDB.Selection.unfold_entities(receptor, 'A')
```

```
32
33 for i in range(0,len(a_list)):
34     center = a_list[i].get_coord()
35     ns = PDB.NeighborSearch(b_list)
36     neighbors = ns.search(center, 10.0)
37 close_to_a = unique(neighbors)
38
39 for i in range(0,len(b_list)):
40     center = b_list[i].get_coord()
41     ns = PDB.NeighborSearch(a_list)
42     neighbors = ns.search(center, 10.0)
43 close_to_b = unique(neighbors)
44
45 interface = close_to_a + close_to_b
46 int_coord = []
47 tot = []
48 for atom in interface:
49     int_coord = atom.get_coord()
50     tot.append(int_coord)
51
52 int_center = sum(tot)/len(tot)
53 foo = []
54 foo.append(str(round(int_center[0],1)))
55 foo.append(',')
56 foo.append(str(round(int_center[1],1)))
57 foo.append(',')
58 foo.append(str(round(int_center[2],1)))
59 coords = ''.join(foo)
60 print(coords)
```

Bibliography

[1] David J Diller, Mark Jarosinski, Tomi K Sawyer, and Joseph Audie. *Peptide drug discovery: innovative technologies and transformational medicines*. University of KwaZulu-Natal, South Africa & IRB-Barcelona, Spain, Nov. 2015 (cit. on p. 10).

[2] Keld Fosgerau and Torsten Hoffmann. «Peptide therapeutics: current status and future directions». en. In: *Drug Discov. Today* 20.1 (Jan. 2015), pp. 122–128 (cit. on p. 10).

[3] Jolene L Lau and Michael K Dunn. «Therapeutic peptides: Historical perspectives, current development trends, and future directions». en. In: *Bioorg. Med. Chem.* 26.10 (June 2018), pp. 2700–2707 (cit. on p. 10).

[4] Markus Muttenthaler, Glenn F King, David J Adams, and Paul F Alewood. «Trends in peptide drug discovery». en. In: *Nat. Rev. Drug Discov.* 20.4 (Apr. 2021), pp. 309–325 (cit. on pp. 10, 11).

[5] Evangelia Petsalaki and Robert B Russell. «Peptide-mediated interactions in biological systems: new discoveries and applications». en. In: *Curr. Opin. Biotechnol.* 19.4 (Aug. 2008), pp. 344–350 (cit. on p. 10).

[6] Lei Wang, Nanxi Wang, Wenping Zhang, Xurui Cheng, Zhibin Yan, Gang Shao, Xi Wang, Rui Wang, and Caiyun Fu. «Therapeutic peptides: current applications and future directions». en. In: *Signal Transduct Target Ther* 7.1 (Feb. 2022), p. 48 (cit. on p. 10).

[7] Roland Böttger, Ralf Hoffmann, and Daniel Knappe. «Differential stability of therapeutic peptides with different proteolytic cleavage sites in blood, plasma and serum». en. In: *PLoS One* 12.6 (June 2017), e0178943 (cit. on p. 11).

[8] Huiya Zhang and Shiyu Chen. «Cyclic peptide drugs approved in the last two decades (2001-2021)». en. In: *RSC Chem Biol* 3.1 (Jan. 2022), pp. 18–31 (cit. on p. 12).

[9] Danah Al Shaer, Othman Al Musaimi, Fernando Albericio, and Beatriz G de la Torre. «2021 FDA TIDES (Peptides and Oligonucleotides) Harvest». en. In: *Pharmaceuticals* 15.2 (Feb. 2022) (cit. on pp. 12, 13).

[10] *Global peptide therapeutics market size, share report, 2030*. en. https://www.grandviewresearch.com/industry-analysis/peptide-therapeutics-market. Accessed: 2023-3-5 (cit. on p. 12).

[11] M Pellegrini and D F Mierke. «Molecular complex of cholecystokinin-8 and N-terminus of the cholecystokinin A receptor by NMR spectroscopy». en. In: *Biochemistry* 38.45 (Nov. 1999), pp. 14775–14783 (cit. on p. 13).

[12] Christina Rye Underwood, Patrick Garibay, Lotte Bjerre Knudsen, Sven Hastrup, Günther H Peters, Rainer Rudolph, and Steffen Reedtz-Runge. «Crystal structure of glucagon-like peptide-1 in complex with the extracellular domain of the glucagon-like peptide-1 receptor». en. In: *J. Biol. Chem.* 285.1 (Jan. 2010), pp. 723–730 (cit. on p. 13).

[13] Jesper Lau et al. «Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide». en. In: *J. Med. Chem.* 58.18 (Sept. 2015), pp. 7370–7380 (cit. on p. 13).

[14] Sang-Min Lee, Yejin Jeong, John Simms, Margaret L Warner, David R Poyner, Ka Young Chung, and Augen A Pioszak. «Calcitonin Receptor N-Glycosylation Enhances Peptide Hormone Affinity by Controlling Receptor Dynamics». en. In: *J. Mol. Biol.* 432.7 (Mar. 2020), pp. 1996–2014 (cit. on p. 13).

[15] Maoqing Dong et al. «Structure and dynamics of the active Gs-coupled human secretin receptor». en. In: *Nat. Commun.* 11.1 (Aug. 2020), p. 4137 (cit. on p. 13).

[16] Lei Wang, Jun Xu, Sheng Cao, Dapeng Sun, Heng Liu, Qiuyuan Lu, Zheng Liu, Yang Du, and Cheng Zhang. «Cryo-EM structure of the AVP-vasopressin receptor 2-Gs signaling complex». en. In: *Cell Res.* 31.8 (Aug. 2021), pp. 932–934 (cit. on p. 13).

[17] Justin G Meyerowitz, Michael J Robertson, Ximena Barros-Álvarez, Ouliana Panova, Robert M Nwokonko, Yang Gao, and Georgios Skiniotis. «The oxytocin signaling complex reveals a molecular switch for cation dependence». en. In: *Nat. Struct. Mol. Biol.* 29.3 (Mar. 2022), pp. 274–281 (cit. on p. 13).

[18] «Sincalide». In: *Drugs and Lactation Database (LactMed®)*. Bethesda (MD): National Institute of Child Health and Human Development, Dec. 2018 (cit. on p. 15).

[19] Günaj Rakipovski et al. «The GLP-1 Analogs Liraglutide and Semaglutide Reduce Atherosclerosis in ApoE-/- and LDLr-/- Mice by a Mechanism That Includes Inflammatory Pathways». en. In: *JACC Basic Transl Sci* 3.6 (Dec. 2018), pp. 844–857 (cit. on p. 17).

[20] Mariana Cornelia Tilinca, Robert Aurelian Tiuca, Alexandru Burlacu, and Andreea Varga. «A 2021 Update on the Use of Liraglutide in the Modern Treatment of 'Diabesity': A Narrative Review». en. In: *Medicina* 57.7 (June 2021), p. 669 (cit. on p. 17).

[21] Bruce Bode. «Liraglutide: a review of the first once-daily GLP-1 receptor agonist». en. In: *Am. J. Manag. Care* 17.2 Suppl (Mar. 2011), S59–70 (cit. on p. 17).

[22] PubChem. *Liraglutide*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/16134956`. Accessed: 2023-3-11 (cit. on p. 17).

[23] Sam Lear, Zaid Amso, and Weijun Shen. «Chapter Eight - Engineering PEG-fatty acid stapled, long-acting peptide agonists for G protein-coupled receptors». In: *Methods in Enzymology*. Ed. by Arun K Shukla. Vol. 622. Academic Press, Jan. 2019, pp. 183–200 (cit. on p. 17).

[24] Young-Sun Lee and Hee-Sook Jun. «Anti-diabetic actions of glucagon-like peptide-1 on pancreatic beta-cells». en. In: *Metabolism* 63.1 (Jan. 2014), pp. 9–19 (cit. on p. 17).

[25] PubChem. *Semaglutide*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/56843331`. Accessed: 2023-3-11 (cit. on p. 19).

[26] Lene Jensen, Hans Helleberg, Ad Roffel, Jan Jaap van Lier, Inga Bjørnsdottir, Palle Jacob Pedersen, Everton Rowe, Julie Derving Karsbøl, and Mette Lund Pedersen. «Absorption, metabolism and excretion of the GLP-1 analogue semaglutide in humans and nonclinical species». en. In: *Eur. J. Pharm. Sci.* 104 (June 2017), pp. 31–41 (cit. on p. 19).

[27] Sylvie Hall, Diana Isaacs, and Jennifer N Clements. «Pharmacokinetics and Clinical Implications of Semaglutide: A New Glucagon-Like Peptide (GLP)-1 Receptor Agonist». en. In: *Clin. Pharmacokinet.* 57.12 (Dec. 2018), pp. 1529–1538 (cit. on p. 19).

[28] Panagiotis Andreadis, Thomas Karagiannis, Konstantinos Malandris, Ioannis Avgerinos, Aris Liakos, Apostolos Manolopoulos, Eleni Bekiari, David R Matthews, and Apostolos Tsapas. «Semaglutide for type 2 diabetes mellitus: A systematic review and meta-analysis». en. In: *Diabetes Obes. Metab.* 20.9 (Sept. 2018), pp. 2255–2263 (cit. on p. 21).

[29] John Blundell, Graham Finlayson, Mads Axelsen, Anne Flint, Catherine Gibbons, Trine Kvist, and Julie B Hjerpsted. «Effects of once-weekly semaglutide on appetite, energy intake, control of eating, food preference and body weight in subjects with obesity». en. In: *Diabetes Obes. Metab.* 19.9 (Sept. 2017), pp. 1242–1251 (cit. on p. 21).

[30] Carsten F Gotfredsen, Anne-Marie Mølck, Inger Thorup, Niels C Berg Nyborg, Zaki Salanti, Lotte Bjerre Knudsen, and Marianne O Larsen. «The human GLP-1 analogs liraglutide and semaglutide: absence of histopathological effects on the pancreas in nonhuman primates». en. In: *Diabetes* 63.7 (July 2014), pp. 2486–2497 (cit. on p. 21).

[31] PubChem. *Fortical*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/16129616`. Accessed: 2023-3-11 (cit. on p. 21).

[32] C H Chesnut 3rd, M Azria, S Silverman, M Engelhardt, M Olson, and L Mindeholm. «Salmon calcitonin: a review of current and future therapeutic indications». en. In: *Osteoporos. Int.* 19.4 (Apr. 2008), pp. 479–491 (cit. on p. 21).

[33] PubChem. *Secretin*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/16129665`. Accessed: 2023-3-11 (cit. on p. 23).

[34] Syeda Afroze, Fanyin Meng, Kendal Jensen, Kelly McDaniel, Kinan Rahal, Paolo Onori, Eugenio Gaudio, Gianfranco Alpini, and Shannon S Glaser. «The physiological roles of secretin and its receptor». en. In: *Ann Transl Med* 1.3 (Oct. 2013), p. 29 (cit. on p. 23).

[35] Jessica Y S Chu, Carrie Y Y Cheng, Vien H Y Lee, Y S Chan, and Billy K C Chow. «Secretin and body fluid homeostasis». en. In: *Kidney Int.* 79.3 (Feb. 2011), pp. 280–287 (cit. on p. 23).

[36] Stuart Sherman, Martin L Freeman, Paul R Tarnasky, C Mel Wilcox, Abhijit Kulkarni, Alex M Aisen, David Jacoby, and Richard A Kozarek. «Administration of secretin (RG1068) increases the sensitivity of detection of duct abnormalities by magnetic resonance cholangiopancreatography in patients with pancreatitis». en. In: *Gastroenterology* 147.3 (Sept. 2014), 646–654.e2 (cit. on p. 23).

[37] P G Lankisch and W Creutzfeldt. «Effect of synthetic and natural secretin on the function of the exocrine pancreas in man». en. In: *Digestion* 22.2 (1981), pp. 61–65 (cit. on p. 23).

[38] S R Wersinger, H K Caldwell, L Martinez, P Gold, S-B Hu, and W S Young 3rd. «Vasopressin 1a receptor knockout mice have a subtle olfactory deficit but normal aggression». en. In: *Genes Brain Behav.* 6.6 (Aug. 2007), pp. 540–551 (cit. on p. 25).

[39] Isadora F Bielsky, Shuang-Bao Hu, Kathleen L Szegda, Heiner Westphal, and Larry J Young. «Profound impairment in social recognition and reduction in anxiety-like behavior in vasopressin V1a receptor knockout mice». en. In: *Neuropsychopharmacology* 29.3 (Mar. 2004), pp. 483–493 (cit. on p. 25).

[40] Jean-Sébastien Pelletier, Bryan Dicken, David Bigam, and Po-Yin Cheung. «Cardiac effects of vasopressin». en. In: *J. Cardiovasc. Pharmacol.* 64.1 (July 2014), pp. 100–107 (cit. on p. 25).

[41] Maja Lozić, Olivera Šarenac, David Murphy, and Nina Japundžić-Žigon. «Vasopressin, Central Autonomic Control and Blood Pressure Regulation». en. In: *Curr. Hypertens. Rep.* 20.2 (Feb. 2018), p. 11 (cit. on p. 25).

[42] PubChem. *Argipressin*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/644077`. Accessed: 2023-3-11 (cit. on p. 25).

[43] Stephen J Lolait, Lesley Q Stewart, David S Jessop, W Scott Young 3rd, and Anne-Marie O'Carroll. «The hypothalamic-pituitary-adrenal axis response to stress in mice lacking functional vasopressin V1b receptors». en. In: *Endocrinology* 148.2 (Feb. 2007), pp. 849–856 (cit. on p. 25).

[44] Anna B Marcinkowska, Vinicia C Biancardi, and Pawel J Winklewski. «Arginine Vasopressin, Synaptic Plasticity, and Brain Networks». en. In: *Curr. Neuropharmacol.* 20.12 (Nov. 2022), pp. 2292–2302 (cit. on p. 25).

[45] PubChem. *Oxytocin*. en. `https://pubchem.ncbi.nlm.nih.gov/compound/439302`. Accessed: 2023-3-11 (cit. on p. 27).

[46] Pedro Hidalgo-Lopezosa, Marıa Hidalgo-Maestre, and Marıa Aurora Rodrıguez-Borrego. «Labor stimulation with oxytocin: effects on obstetrical and neonatal outcomes». en. In: *Rev. Lat. Am. Enfermagem* 24 (July 2016), e2744 (cit. on p. 27).

[47] G Gimpl and F Fahrenholz. «The oxytocin receptor system: structure, function, and regulation». en. In: *Physiol. Rev.* 81.2 (Apr. 2001), pp. 629–683 (cit. on p. 27).

[48] J Gutkowska, M Jankowski, and J Antunes-Rodrigues. «The role of oxytocin in cardiovascular regulation». en. In: *Braz. J. Med. Biol. Res.* 47.3 (Feb. 2014), pp. 206–214 (cit. on p. 27).

[49] R L Perry, A J Satin, W H Barth, S Valtier, J T Cody, and G D Hankins. «The pharmacokinetics of oxytocin as they apply to labor induction». en. In: *Am. J. Obstet. Gynecol.* 174.5 (May 1996), pp. 1590–1593 (cit. on p. 27).

[50] Mariela Mitre et al. «Sex-Specific Differences in Oxytocin Receptor Expression and Function for Parental Behavior». en. In: *Gend Genome* 1.4 (Dec. 2017), pp. 142–166 (cit. on p. 27).

[51] Joanne Paquin, Bogdan A Danalache, Marek Jankowski, Samuel M McCann, and Jolanta Gutkowska. «Oxytocin induces differentiation of P19 embryonic stem cells to cardiomyocytes». en. In: *Proc. Natl. Acad. Sci. U. S. A.* 99.14 (July 2002), pp. 9550–9555 (cit. on p. 27).

[52] Michael Kosfeld, Markus Heinrichs, Paul J Zak, Urs Fischbacher, and Ernst Fehr. «Oxytocin increases trust in humans». en. In: *Nature* 435.7042 (June 2005), pp. 673–676 (cit. on p. 27).

[53] S Arrowsmith and S Wray. «Oxytocin: its mechanism of action and receptor signalling in the myometrium». en. In: *J. Neuroendocrinol.* 26.6 (June 2014), pp. 356–369 (cit. on p. 27).

[54] Maciej Ciemny, Mateusz Kurcinski, Karol Kamel, Andrzej Kolinski, Nawsad Alam, Ora Schueler-Furman, and Sebastian Kmiecik. «Protein-peptide docking: opportunities and challenges». en. In: *Drug Discov. Today* 23.8 (Aug. 2018), pp. 1530–1537 (cit. on p. 30).

[55] Ilya A Vakser and Petras Kundrotas. «Predicting 3D structures of protein-protein complexes». en. In: *Curr. Pharm. Biotechnol.* 9.2 (Apr. 2008), pp. 57–66 (cit. on p. 30).

[56] David W Ritchie. «Recent progress and future directions in protein-protein docking». en. In: *Curr. Protein Pept. Sci.* 9.1 (Feb. 2008), pp. 1–15 (cit. on p. 30).

[57] Carlos J Camacho and Sandor Vajda. «Protein-protein association kinetics and protein docking». en. In: *Curr. Opin. Struct. Biol.* 12.1 (Feb. 2002), pp. 36–40 (cit. on p. 30).

[58] Alexandre M J J Bonvin. «Flexible protein-protein docking». en. In: *Curr. Opin. Struct. Biol.* 16.2 (Apr. 2006), pp. 194–200 (cit. on p. 30).

[59] Yuqi Zhang and Michel F Sanner. «Docking Flexible Cyclic Peptides with AutoDock CrankPep». In: *J. Chem. Theory Comput.* 15.10 (Oct. 2019), pp. 5161–5168 (cit. on p. 31).

[60] Piyush Agrawal, Harinder Singh, Hemant Kumar Srivastava, Sandeep Singh, Gaurav Kishore, and Gajendra P S Raghava. «Benchmarking of different molecular docking methods for protein-peptide docking». en. In: *BMC Bioinformatics* 19.Suppl 13 (Feb. 2019), p. 426 (cit. on p. 31).

[61] Raúl Méndez, Raphaël Leplae, Leonardo De Maria, and Shoshana J Wodak. «Assessment of blind predictions of protein-protein interactions: current status of docking methods». en. In: *Proteins* 52.1 (July 2003), pp. 51–67 (cit. on p. 31).

[62] José Ignacio Garzon, José Ramón Lopéz-Blanco, Carles Pons, Julio Kovacs, Ruben Abagyan, Juan Fernandez-Recio, and Pablo Chacon. «FRODOCK: a new approach for fast rotational protein-protein docking». en. In: *Bioinformatics* 25.19 (Oct. 2009), pp. 2544–2551 (cit. on pp. 31, 32).

[63] Yuqi Zhang and Michel F Sanner. «AutoDock CrankPep: combining folding and docking to predict protein-peptide complexes». en. In: *Bioinformatics* 35.24 (Dec. 2019), pp. 5121–5127 (cit. on p. 31).

[64] Sheng-You Huang and Xiaoqin Zou. «Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking». en. In: *Proteins* 66.2 (Feb. 2007), pp. 399–421 (cit. on p. 31).

[65] Sheng-You Huang and Xiaoqin Zou. «An iterative knowledge-based scoring function for protein-protein recognition». en. In: *Proteins* 72.2 (Aug. 2008), pp. 557–579 (cit. on p. 31).

[66] Pei Zhou, Bowen Jin, Hao Li, and Sheng-You Huang. «HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm». en. In: *Nucleic Acids Res.* 46.W1 (May 2018), W443–W450 (cit. on p. 31).

[67] Pei Zhou, Botong Li, Yumeng Yan, Bowen Jin, Libang Wang, and Sheng-You Huang. «Hierarchical Flexible Peptide Docking by Conformer Generation and Ensemble Docking of Peptides». en. In: *J. Chem. Inf. Model.* 58.6 (June 2018), pp. 1292–1302 (cit. on p. 31).

[68] Huanyu Tao, Yanjun Zhang, and Sheng-You Huang. «Improving Protein–Peptide Docking Results via Pose-Clustering and Rescoring with a Combined Knowledge-Based and MM–GBSA Scoring Function». In: *J. Chem. Inf. Model.* 60.4 (Apr. 2020), pp. 2377–2387 (cit. on p. 31).

[69] Huanyu Tao, Xuejun Zhao, Keqiong Zhang, Peicong Lin, and Sheng-You Huang. «Docking cyclic peptides formed by a disulfide bond through a hierarchical strategy». en. In: *Bioinformatics* 38.17 (Sept. 2022), pp. 4109–4116 (cit. on p. 31).

[70] Huanyu Tao, Qilong Wu, Xuejun Zhao, Peicong Lin, and Sheng-You Huang. «Efficient 3D conformer generation of cyclic peptides formed by a disulfide bond». en. In: *J. Cheminform.* 14.1 (May 2022), p. 26 (cit. on p. 31).

[71] Peter JA Cock et al. «Biopython: freely available Python tools for computational molecular biology and bioinformatics». In: *Bioinformatics* 25.11 (2009), pp. 1422–1423 (cit. on p. 32).

[72] Ján Buša, Jozef Džurina, Edik Hayryan, Shura Hayryan, Chin-Kun Hu, Ján Plavka, Imrich Pokorný, Jaroslav Skřivánek, and Ming-Chya Wu. «ARVO: A Fortran package for computing the solvent accessible surface area and the excluded volume of overlapping spheres via analytic equations». In: *Comput. Phys. Commun.* 165.1 (Jan. 2005), pp. 59–96 (cit. on p. 33).

[73] Ruth Huey, Garrett M Morris, Arthur J Olson, and David S Goodsell. «A semiempirical free energy force field with charge-based desolvation». en. In: *J. Comput. Chem.* 28.6 (Apr. 2007), pp. 1145–1152 (cit. on p. 34).

[74] Garrett M Morris, Ruth Huey, William Lindstrom, Michel F Sanner, Richard K Belew, David S Goodsell, and Arthur J Olson. «AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility». en. In: *J. Comput. Chem.* 30.16 (Dec. 2009), pp. 2785–2791 (cit. on pp. 34, 35).

[75] S Takada. «Go-ing for the prediction of protein folding mechanisms». en. In: *Proc. Natl. Acad. Sci. U. S. A.* 96.21 (Oct. 1999), pp. 11698–11700 (cit. on p. 34).

[76] H Taketomi, Y Ueda, and N Gō. «Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions». en. In: *Int. J. Pept. Protein Res.* 7.6 (1975), pp. 445–459 (cit. on p. 34).

[77] Alexei A Podtelezhnikov and David L Wild. «CRANKITE: A fast polypeptide backbone conformation sampler». en. In: *Source Code Biol. Med.* 3 (June 2008), p. 12 (cit. on pp. 34, 35).

[78] R L Dunbrack Jr and F E Cohen. «Bayesian statistical analysis of protein side-chain rotamer preferences». en. In: *Protein Sci.* 6.8 (Aug. 1997), pp. 1661–1681 (cit. on p. 34).

[79] Pradeep Anand Ravindranath, Stefano Forli, David S Goodsell, Arthur J Olson, and Michel F Sanner. «AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility». en. In: *PLoS Comput. Biol.* 11.12 (Dec. 2015), e1004586 (cit. on p. 34).

[80] Pradeep Anand Ravindranath and Michel F Sanner. «AutoSite: an automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms». en. In: *Bioinformatics* 32.20 (Oct. 2016), pp. 3142–3149 (cit. on p. 35).

[81] Alexei A Podtelezhnikov and David L Wild. «Exhaustive Metropolis Monte Carlo sampling and analysis of polyalanine conformations adopted under the influence of hydrogen bonds». en. In: *Proteins* 61.1 (Oct. 2005), pp. 94–104 (cit. on p. 35).

[82] Todd J Dolinsky, Jens E Nielsen, J Andrew McCammon, and Nathan A Baker. «PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations». en. In: *Nucleic Acids Res.* 32.Web Server issue (July 2004), W665–7 (cit. on p. 39).

[83] Todd J Dolinsky, Paul Czodrowski, Hui Li, Jens E Nielsen, Jan H Jensen, Gerhard Klebe, and Nathan A Baker. «PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations». en. In: *Nucleic Acids Res.* 35.Web Server issue (July 2007), W522–5 (cit. on p. 39).