



In Silico design of novel peptides for treating cancer
through inhibition of the RANK-TRAF6 Interaction

Sara Batalha Galhofo Celestino Perdigão

Thesis to obtain the Master Degree in

Pharmaceutical Engineering

Supervisor(s): Doctor Rita Lourenço Paiva de Melo

Doctor. João Domingos Galamba Correia

Examination Committee

Chairperson: PhD. Ana Margarida Nunes da Mata Pires de Azevedo

Supervisor: PhD. Rita Lourenço Paiva de Melo

Members of the Committee: PhD. Maria de La Salette de Jesus Baptista

PhD. Vera Luísa Santos Neves

October 2021

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the University of Lisbon.

Preface

This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Pharmaceutical Engineering of the author at Instituto Superior Técnico. The work described here was performed at the Center for Nuclear Sciences and Technologies (C²TN), Instituto Superior Técnico, University of Lisbon (Lisbon, Portugal), during the period February-September 2021, under the supervision of Doctor Rita Lourenço Paiva de Melo and Doctor João Domingos Galamba Correia.

Acknowledgements

Os meus agradecimentos são destinados aqueles que, de uma forma de outra, tornaram possível a concretização desta tese.

Eu gostaria de agradecer aos meus orientadores, Doutor João Domingos Galamba Correia e Doutora Rita Lourenço Paiva de Melo, pela oportunidade de fazer a minha tese no Campus Tecnológico e Nuclear (CTN), por todo o apoio e à vontade que me foi transmitido ao longo destes 7 meses, por todos os conhecimentos que me passaram e por tudo aquilo que aprendi no campo da Biochemistry and Computational Chemistry.

Obrigada aos meus colegas de laboratório, Kyle Gonçalves e Ruben Silva, por toda a ajuda ao longo de todo o trabalho experimental, por toda a aprendizagem das bases essenciais dos laboratórios, por toda a paciência e disponibilidade, por todas as dicas e por todo o ambiente bom que nos rodeava.

Tenho de agradecer à Cátia Rosa, pela colega incansável que foi desde o início, por todas as entreajudas que existiram entre nós, por todo o apoio, companhia e a cima de tudo pela nossa amizade.

Um agradecimento especial ao meu namorado, por todo o amor, ajuda, paciência, motivação e sobretudo por todo o apoio emocional que me deu, sem dúvida que foi muito importante para a minha dissertação de tese.

Um obrigada enorme à minha família e amigos, por todo o carinho, diversão e apoio que foram essenciais nesta fase final.

O maior agradecimento vai para os meus pais, sem eles não teria a oportunidade de estudar e de chegar onde cheguei. Obrigada do fundo do coração por tudo aquilo que fazem por nós, por mim e pela irmã, por todos os ensinamentos, por todas as repreensões e por todo o amor. Espero um dia ser 1/3 para os meus filhos, do que aquilo que vocês foram para nós. Amo-vos muito.

Por fim, agradeço-me a mim mesma por todo o esforço, tempo e dedicação investido na realização da minha tese de mestrado.

Abstract

The RANK-TRAF6 metabolic pathway is commonly associated with osteoporosis and the development of breast and prostate cancer. As such, inhibition of binding between RANK and TRAF6 has been studied. One of the hypotheses is through decoy peptides that bind to one of the targets blocking the binding, decoy peptides based on the part of RANK that binds to TRAF6 were selected. Thus, potentially the peptide will bind to TRAF6 by inhibiting its binding to RANK. The aim of this work is the prediction of the 3D structure, selection and synthesis of peptides that potentially inhibit the RANK-TRAF6 interaction.

The 3D structures of the sequences based on the selected peptides were predicted using three web services, PEP-FOLD, PEPstrMOD and I-TASSER. Subsequently, a molecular docking study was performed to analyse and validate their stability with TRAF6. The molecular docking study was performed using the HADDOCK online server. Here, two approaches were performed: i) the active residues chosen for peptides 1, 2, 3 and 4 are those that are experimentally described as important in the interaction with the TRAF6; and ii) all residues belonging to the peptides were considered as active. Then, an analysis was performed to the generated complexes through a balance between three parameters, HADDOCK score, percentage of residues at the interface and root mean square deviation (RMSD), which allowed the selection of the most predicted stable complex of each peptide. For peptides 1, 2, 3 and 4 the complexes predicted to be the most stable were 154, 9, 109 and 50, respectively.

The solid phase synthesis of the peptides was performed using ultrasonic agitation and their analysis and purification was performed by Reverse-Phase High Performance Liquid Chromatography (RP-HPLC), where a purity of >95% was obtained for all the peptides. Their presence in the samples was confirmed by Mass Spectrometry using Electrospray Ionization method (ESI-MS).

At this point, the peptides are ready to be used in in vivo studies, and thus evaluate their efficiency in inhibiting RANK-TRAF6 binding.

Keywords: TRAF6, Protein-Protein Interaction, Computational Chemistry, HADDOCK score, Solid-Phase Peptide Synthesis, Reverse-Phase High Performance Liquid Chromatography

Resumo

A via metabólica RANK-TRAF6 está geralmente associada à osteoporose e ao desenvolvimento do cancro da mama e da próstata. Como tal, a inibição da ligação entre RANK e TRAF6 tem sido estudada. Uma das hipóteses é através de peptídeos de engodo que se ligam a um dos alvos para bloquear a ligação, os péptidos de engodo foram selecionados baseando-se na parte do RANK que se liga ao TRAF6. Assim, potencialmente o peptídeo ligar-se-á ao TRAF6, inibindo a sua ligação ao RANK. Assim, o objetivo desta tese é a previsão da estrutura 3D destes péptidos, a sua seleção e a síntese de peptídeos que potencialmente inibem a interação RANK-TRAF6.

As estruturas 3D das sequências baseadas nos péptidos selecionado foram previstas através de três servidores web, PEP-FOLD, PEPstrMOD e I-TASSER. Posteriormente foi realizado um estudo de docking molecular para analisar e validar a sua estabilidade com o TRAF6. O estudo de acoplagem molecular foi realizado através do servidor online HADDOCK. Aqui, foram realizadas duas abordagens: i) os resíduos ativos escolhidos para os péptidos são os que estão experimentalmente descritos como importantes na interação com o TRAF6; e ii) todos os resíduos pertencentes aos péptidos foram considerados ativos. Em seguida, efetua-se uma análise aos complexos gerados através de um balanço entre três parâmetros, score HADDOCK, percentagem de resíduos na interface e desvio de raiz média quadrada (RMSD), o que permitiu selecionar o complexo mais estável de cada peptídeo. Para os péptidos 1, 2, 3 e 4 os complexos previstos de serem os mais estáveis foram o 154, 9, 109 e 50, respetivamente.

A síntese em fase sólida dos péptidos foi realizada através de agitação ultrassónica e a sua análise e purificação foi realizada por Cromatografia Líquida de Alto Desempenho em Fase Reversa (RP-HPLC), onde foi obtida uma pureza >95% para todos os péptidos. A presença dos mesmos foi confirmada por Espectrometria de Massa, utilizando o método de Ionização por Eletrospray (ESI-MS).

Neste ponto, os peptídeos estão prontos para serem utilizados em estudos in vivo, e assim avaliar a sua eficiência na inibição da ligação RANK-TRAF6.

Palavras-chave: TRAF6, Interação Proteína-Proteína, Química Computacional, Pontuação HADDOCK, Síntese de Péptidos em Fase Sólida, Cromatografia Líquida de Alto Desempenho em Fase Reversa

Contents

Declaration	ii
Preface	iii
Acknowledgements.....	iv
Abstract	v
Resumo	vi
List of figures	ix
List of tables	x
List of equations.....	xi
Abbreviations.....	xii
1. Introduction	1
1.1. Cancer overview	1
1.2. RANK-TRAF6 pathway	1
1.2.1. TRAF6 and decoy peptides.....	3
1.2.2. Protein-Protein Interactions.....	4
1.3. Computational Chemistry	6
1.3.1. Three-Dimensional Structure Prediction	6
1.3.2. Molecular Docking	11
1.4. Solid phase peptide synthesis (SPPS).....	14
1.5. High Performance Liquid Chromatography (HPLC)	16
1.6. Mass Spectrometry	17
2. Motivation and Aim of the Thesis.....	19
3. Methodologies.....	20
3.1. Decoy peptides	20
3.2. Computational Chemistry.....	20
3.2.1. Three-Dimensional structure prediction	20
3.2.2. Molecular Docking	21
3.3. Solid-Phase Peptide Synthesis.....	22
3.3.1. Aluminum foil test	22
3.3.2. Solid-Phase Peptide Synthesis	23
3.4. Peptides chemical characterization and purification.....	29

3.4.1.	Reverse-Phase High Performance Liquid Chromatography (Analytical RP-HPLC)	29
3.4.2.	Reverse-Phase High Performance Liquid Chromatography (Preparative RP-HPLC)	30
3.4.3.	Mass Spectrometry (MS)	30
4.	Results and Discussion	31
4.1.	Computational Chemistry	31
4.1.1.	Three-dimensional structure prediction	31
4.1.2.	Molecular Docking	32
4.2.	Synthesis, Characterization and Purification of Peptides	38
5.	Conclusion and future perspectives	43
6.	Bibliography	44

List of figures

Figure 1: Schematic diagram of RANK-TRAF6 pathway[17].	2
Figure 2: Non-scale illustration of human TRAF6 domain structure [14].	3
Figure 3: TRAF6 binding motif Pro-X-Glu-X-X-(Ar/Ac), in RANK (TRANCE-R)[22].	4
Figure 4: Complex RANK-TRAF6[22].	4
Figure 5: Four levels of protein structure [35].	7
Figure 6: Computational Methods of Protein Structure Prediction vs Experimental Protein Structure Determination Methods [41].	8
Figure 7: Bond length, bond angles and torsion angles [41].	8
Figure 8: Dihedral Angles: Omega (Ω), Psi (ψ) and Phi (ϕ) [41].	9
Figure 9: Comparative Modeling Process [41].	10
Figure 10: Example of an amino acid sequence where the amino acid residues at the N- and C-termini (blue), are both at the α -carbon [14].	14
Figure 11: Solid phase peptide synthesis (SPPS)[14].	15
Figure 12: Ionization by ESI-MS [82].	18
Figure 13: Aluminum foil test at the bottom of the bath without the stainless-steel basket.	22
Figure 14: Aluminum foil test on the surface with the stainless-steel basket.	23
Figure 15: Ultrasounds.	27
Figure 16: An example of the Kaiser test where the amine group is unprotected and the colour is blue(a) and where the amine group is protected and it is colourless (b).	28
Figure 17: Residues Arg-392, Phe-410, Phe-471 and Tyr-473 are part of the interaction in complex_109w from TRAF6_Seq_3_1.	38
Figure 18: HPLC chromatogram: (a) Peptide 1 and (b) Peptide 2.	38
Figure 19: RP-HPLC chromatogram: (c) Peptide 3 and (d) Peptide 4.	39
Figure 20: Peptide 1 by ESI-MS after purification.	40
Figure 21: Peptide 2 by ESI-MS after purification.	41
Figure 22: Peptide 3 by ESI-MS after purification.	41
Figure 23: Peptide 4 by ESI-MS after purification.	42

List of tables

Table 1: Peptides to be produced.....	20
Table 2: Conditions for each docking.....	21
Table 3: Quantity of each reagent to be used during SPPS, for peptide 1.....	24
Table 4: Quantity of each reagent to be used during SPPS, for peptide 2.....	24
Table 5: Quantity of each reagent to be used during SPPS, for peptide 3.....	25
Table 6: Quantity of each reagent to be used during SPPS, for peptide 4.....	26
Table 7: Optimised method for the smallest peptides (peptide 1 and 2).	29
Table 8: Optimised method for the largest peptides (peptide 3 and 4).....	29
Table 9: Molecular weight and expected molecular ions, of each peptide.	30
Table 10: Z-score valeus.....	31
Table 11: Final 3D structure of each peptide.	32
Table 12: Top 10 complexes according to HADDOCK score of TRAF6_Seq_1 and TRAF6_Seq_1_1.....	33
Table 13: Top 10 complexes according to HADDOCK score of TRAF6_Seq_2 and TRAF6_Seq_2_1.....	34
Table 14: Top 10 complexes according to HADDOCK score of TRAF6_Seq_3 and TRAF6_Seq_3_1.....	35
Table 15: Top 10 complexes according to the HADDOCK score of TRAF6_Seq_4 and TRAF6_Seq_4_1.....	36
Table 16: Final complex of each peptide.	37
Table 17: Purity of each amino acid after purification.....	40
Table 18: m/z ratio of each peptide obtained by ESI-MS.....	40

List of equations

Equation 1: Buried Surface Area [29].	6
Equation 2: Binding free energy [26].	6
Equation 3: HADDOCK scoring function in the it0 stage [68].	13
Equation 4: HADDOCK scoring function in the it1 stage [68].	13
Equation 5: HADDOCK scoring function in the water (final) stage [68].	13

Abbreviations

ACN – Acetonitrile

AIRs - Ambiguous Interaction Constraints

Boc - t-butyloxycarbonyl group

BSA – Buried Surface Area

CPP - Cell Penetrating Peptides

DCM - Dichloromethane

DMF - N, N-Dimethylformamide

ESF - Energy Score Function

ESI-MS – Electrospray Ionisation Mass Spectrometry

Fmoc - fluorenylmethoxycarbonyl group

HADDOCK - High Ambiguity Driven DOCKing

HPLC - High Performance Liquid Chromatography

iK β – Inhibitor Kappa-Beta

NF-K β - Nuclear Factor Kappa-Beta

PDB - Protein Data Bank

PPIs - Protein-protein Interaction

PS – Polystyrene

RANK - receptor activator of nuclear factor-kappa β

RANKL - receptor activator of nuclear factor-kappa β ligand

RMSD – Root Mean Square Deviation

RP-HPLC - Reversed-Phase High-Performance Liquid Chromatography

SF - Scoring Function

SPPS - Solid Phase Peptide Synthesis

TD – TRAF Domain

TFA - Trifluoroacetic Acid

TNF – Tumor Necrosis Factor

TNFR - Tumor Necrosis Factor Receptor

TRAF - Tumor Necrosis Factor Receptor-Associated Factor

WHO - World Health Organization

3D – Three Dimensional

1. Introduction

1.1. Cancer overview

According with the World Health Organization (WHO) in 2020, 10 million deaths were counted due cancer, where 2.26 million were from breast cancer, 2.21 million from lung cancer, 1.93 million from colon and rectum cancer, 1.41 million from prostate cancer, 1.20 million from skin cancer and 1.09 million from stomach cancer [1].

Cancer is a group of diseases in which some of the body's cells grow uncontrollably and spread to other parts of the body [2]. There are two main categories of cancer, hematologic (blood) cancers (including leukemia, lymphoma and multiple myeloma) and solid tumor cancers (cancer in other organs or tissues of the body). When damaged cells grows and multiply abruptly, they may form tumors, called lumps. These tumors can also be divided into two, not cancerous (benign) or cancerous (malignant). The difference between them, is that the benign tumors can not spread to other parts of the body, when they are removed usually do not grow back, while the malignant can [3][2]. Those that spread into nearby tissues or can travel to distant places in the body, in order to form new cancerous tumors, this process is called metastasis [2].

Studies show relationship between some solid tumor cancers, such as breast and prostate cancer with the propensity to metastasize to bone and cause osteolysis and abnormal new bone formation [4].

The structural and metabolic integrity of bone is maintained through the dynamic process of bone remodeling, by two main kinds of bone cells, osteoblasts and osteoclasts, osteoblasts are responsible for the formation of new bone and osteoclasts for bone resorption [5]. When these cells are both working the way they should, new bone is being formed while old bone is being destroyed, like a maintenance, this way the bone remains strong. However, when cancer cells block this process, they speed up the action, meaning there is an increase in the activity of regulators of bone [4][6].

Currently, no efficient therapy or treatment has been found yet, so several therapeutical targets have been studied, namely those involving the receptor activator of nuclear factor-kappa β (RANK) signaling pathway [7][8]. This pathway is important because it is responsible for osteoclast activation.

Others diseases are also commonly associated with the RANK signaling pathway, such as postmenopausal osteoporosis (a condition of reduced bone mass causing abnormalities in the cytoskeletal structure), Parget's disease, rheumatoid arthritis, ankylosing spondylosis and chronic periodontitis [9][10][11].

1.2. RANK-TRAF6 pathway

As mentioned in the section above, the cytoskeletal structure is maintained through a balance between bone resorption by osteoclasts and new bone formation by osteoblasts, leading to bone remodeling. Therefore, the activation of the nuclear factor KB (NF- κ B) is essential since is responsible for bone resorption (osteoclast formation)[9][5].

NF- κ B is a heterodimer composed of two subunits, p65 and p50, and it is inactivated in the cytoplasm bound to the inhibitor Kappa-beta ($i\kappa$ B) protein, forming the NF- κ B/ $i\kappa$ B complex [12][13]. To activate NF- κ B an extracellular stimulus is needed. Afterward an intracellular pathway is activated leading to the activation of NF- κ B. After its activation, NF- κ B will be transported to the nucleus, binding to a specific DNA and activate the expression of the target gene responsible for osteoclast formation, and carry out its transcription. When the expression of the target gene is activated can trigger a diversity of responses depending on the stimulus.

The stimulus can be done by neurotransmitters, neurotoxin proteins, cytokines, among others. Cytokines have 5 groups: interleukins (IL); colony-stimulating factors (CSFs), tumor necrosis factors (TNFs), interferons (IFN) and growth factors (TGF), thus being able to signal through different surface receptors [14][12][15].

Within TNFs receptors there is a transmembrane protein receptor who plays an important role on the NF- κ B activation, known as RANK or TRANCE-R (TNF related activation-induced cytokine receptor). This receptor activator has a ligand, RANKL (receptor activator of nuclear factor-kappa β ligand), which binds to the extracellular domain of RANK, it is a member of TNF cytokine family. The protein-protein interaction between RANK and RANKL (extracellular stimulus) is responsible for the activation of NF- κ B, however, the same interaction needs to be mediated by TNF receptor-associated factors (TRAFs), adaptor proteins that bind to the cytoplasmic tail of TNFRs (tumor necrosis factors receptors). Among all TRAFs, TRAF6 is the only one that can activate NF- κ B [16].

When TRAF6 binds to the activating receptor RANK (intracellular stimulus), the intracellular signaling pathway is activated and recruits NF- κ B, subsequently the kinase IKK is induced by intermediate factors, which causes NF- κ B to be free in the cytoplasm. In this way, NF- κ B enters the nucleus and carry out the target gene transcription, leading to the activation of osteoclasts [16][14]. A scheme of RANK-TRAF6 pathway is described in figure 1.

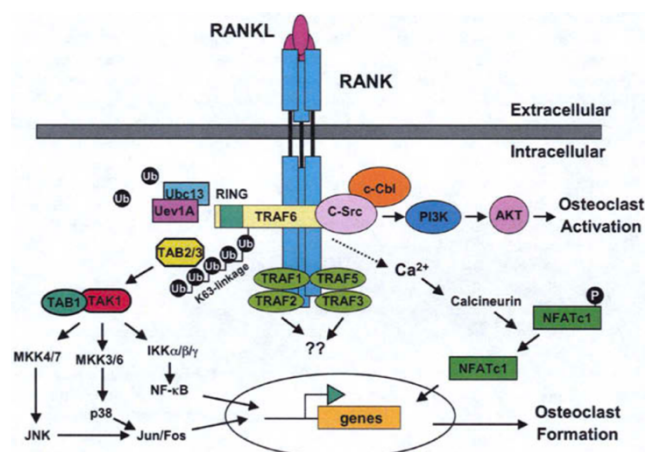


Figure 1: Schematic diagram of RANK-TRAF6 pathway[17].

The mechanism described above is crucial for bone resorption by osteoclasts to occurs, however, as mentioned earlier, increased osteoclast activation leads to a loss of bone mass, leading to an imbalance in bone remodeling and giving rise to all disease mentioned previously[14][18].

1.2.1. TRAF6 and decoy peptides

Tumor necrosis factor receptor-associated factor (TRAFs) were first discovered as adaptor proteins that couple the tumor necrosis factor receptor family to signaling pathways [19], therefore, they are the key signaling molecules that function in various cellular signaling events including immune response, development, thrombosis and cell death and survival. TRAFs also can transduce signals in various types of receptor-mediated cellular signaling, including tumor necrosis factor receptor (TNF-R), interleukin 1 receptor/Toll-like receptor (TLR), nucleotide-binding oligomerization domain-like receptor (NLR), RIG-1 like receptor (RLR), and even cytokine receptor family signaling pathways, and also plays a critical role in the regulation of the immune system and apoptosis [20].

TRAFs family include seven members, from TRAF1 to TRAF7. These proteins share a common structural domain at their C-terminal, called TRAF domain (TD), also (except for TRAF1) containing an N-terminal, a RING finger domain and a zinc finger domain, that is crucial for their functions. The C-terminal of the TRAF family, containing 180 residues, is a conserved region that allows them to be characterized and are responsible for the interaction between TRAFs and TNFR members, other kinases and proteins [20][21][19][14]. TRAF6 is illustrated in figure 2.

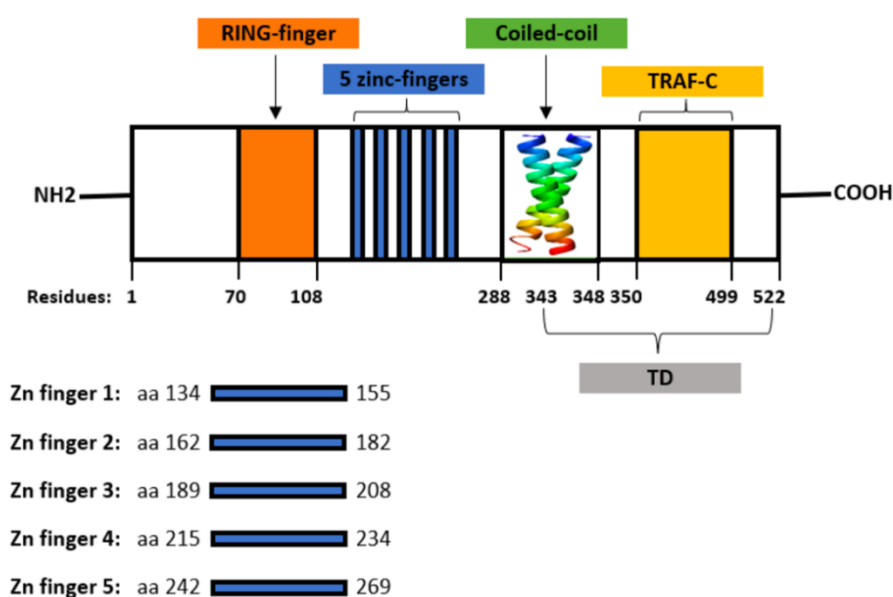


Figure 2: Non-scale illustration of human TRAF6 domain structure [14].

Studies have revealed, that although TRAF6 has the same structural characteristics of the other members of the family, it is the less conserved in its TD, more specifically in the TRAF-C domain, since he only shares 30 % of the sequence identity with the others TRAFs members (1-5). So he does not have the same binding site as the others, meaning that TRAF6 binds in a different region of the RANK than TRAFs 1, 2, 3, 5 and 7. Also, TRAF-C is what determines the function of the protein, the biological function of TRAF6 is related with the osteoclast formation, which makes its biological function unique. There are three important residues that are all involved in TRAF6 interaction and signal transduction, playing an essential role in NF- κ B activation mediated by RANKL [14][20][5].

In order to reach the TRAF6 binding motif, several studies and hypotheses were conducted. A sequence alignment based on the structure of TRAF6 binding sites in mouse and human, was performed. Based on the sequence alignment, figure 3, allowed us to conclude that TRAF6 binding motif, in RANK (TRANCE-R), and other TNFR family members, is a generalized amino acid sequence pattern, designated Pro-X-Glu-X-X-(Ar/Ac), where Ar is an aromatic and Ac an acid residue. The residue Glu has been designated in position P₀, Pro in position P₋₂ and Ar/Ac in position P₃. These residues are the most crucial for TRAF6 interaction and essential for maintaining the integrity of the binding interface [22][20].

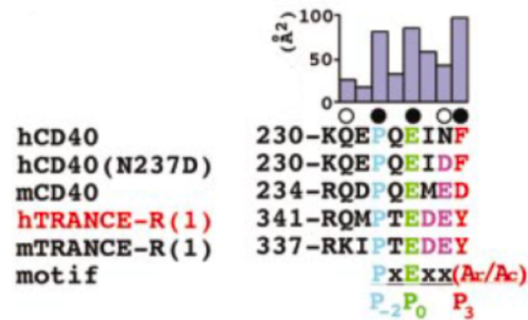


Figure 3: TRAF6 binding motif Pro-X-Glu-X-X-(Ar/Ac), in RANK (TRANCE-R)[22].

The RANK-TRAF6 complex (PDB ID: 1LB5) is represented in figure 4.

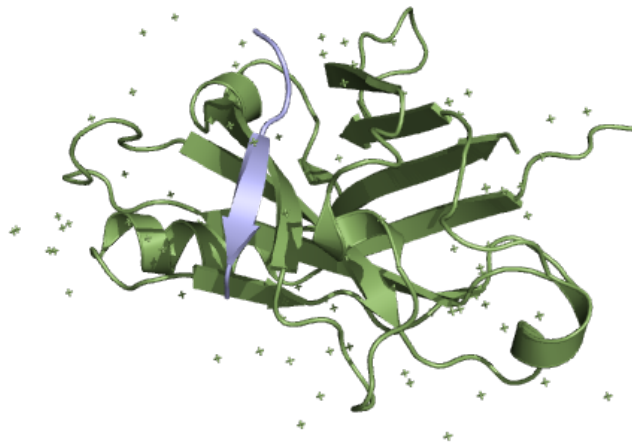


Figure 4: Complex RANK-TRAF6[22].

According to, *Hong Ye and Hyun (2018)*, the residues Arg-392, Phe-471, Tyr-473, Phe-410, Leu-432 and Phe-459 are key residues in the interaction TRAF6-RANK [22][20].

1.2.2. Protein-Protein Interactions

Protein-protein interactions (PPIs) plays an important role in numerous physiological and biological processes, like differentiation, growth, cell proliferation, signal transduction, metabolic pathways, and apoptosis. Moreover, it has been shown that changes or disturbances in PPIs are commonly associated with many diseases, including cancer [9][23][24].

PPIs involve the formation of complexes, between two or more proteins [25]. Specific interactions, the PPIs, are required for the formation of these complexes, and it is due to their complexity that they have a great diversity of functions, they mediating protein folding and they are the basis of practically all biological processes, in particular signal transduction [26][27].

PPIs can be classified based on their interaction surface, where they can be homomeric or heteromeric; based on their stability, where they can be obligatory or not, and they can be classified as transient or permanent based on their persistence[14]. PPI interfaces, depending on their type of obligation, are different. An area between 1150-1200Å² (non-obligate) is considered small, standard size interfaces are approximately 1600 Å² (+/- 400 Å²) and large interfaces bury 2000 to 4660Å². There are also important factors that influence PPIs such as flexibility, hydrophobicity and electrostatic interactions [26].

The structure of PPIs can be divided by some regions, such as the surface, the interior and the interface, the last of which consists of a set of amino acids that presents a region connecting two chains by non-covalent interactions [28].

During the protein folding process, the polypeptide undergoes a physical process in which its tertiary structure folds, causing the hydrophobic residues to protect themselves from the solvent, giving rise to a hydrophobic protein interior and a hydrophilic protein surface. The latter is where the exposed residues, those that are not protected, are found. The fact that there is a distinction between residues on the inside and on the surface brings certain profound implications for stability and evolution, because residues located on the surface of proteins can easily mutate unlike those that are protected on the inside. However if residues located on the inside were to mutate it would further destabilize the protein either in structure or function [28].

Regarding the interface region, this is where the important residues are found, called hot spots according to Clackson and Wells, where the interior of the interface is called the core and the surface of the interface is called the rim [28]. Thus the energy distribution is not uniform in a given protein-protein interaction, because at the core interface hot spots correspond to a small subset of residues that contribute especially to binding affinity. Hot spots are structurally conserved, this is evident because the mutation rate is slower compared to the other surface residues [26]. In general, the core residues are better preserved than those of the rim [14].

In this way Buried Surface Area (BSA), in PPI, is defined as the surface buried away from the solvent when a complex is formed, between two or more proteins. The most widely used method to calculate BSA is the solvent-accessible surface, where the solvent accessible surface is traced with a probe sphere as it rolls over the protein. The protein atoms are assigned their corresponding van der waals radii. However, there are other methods to calculate BSA so the calculated area depends on the method used. Another concerning limitation for the calculation of BSA of protein complexes is that proteins do not associate as rigid entities, but can undergo small to large conformational changes upon binding. It is therefore necessary to have a good knowledge of the 3D structures of the interacting proteins. This is calculated using the following equation[29]:

$$BSA = \sum_{n=1}^{NN_{complexes}} ASA_{free}^n - ASA_{complex}$$

Equation 1: Buried Surface Area [29].

Where:

ASA_{free}^n specifies the accessible surface area of the unbound molecules

$ASA_{complex}$ the bound complex accessible surface area

As mentioned earlier, the energy distribution is not uniform across the interface due to hot spots, which by definition are like a residue whose mutation to alanine results in a decrease of at least 2.0 kcal/mol in binding free energy ($\Delta\Delta G_{binding}$). The free energy of binding, or $\Delta\Delta G$, is defined as [26]:

$$\Delta\Delta G = \Delta G^{mut} - \Delta G^{wt}$$

Equation 2: Binding free energy [26].

Where:

ΔG^{mut} is the free energy of binding to the substituted alanine

ΔG^{wt} are the wild type residues in the protein complex

1.3. Computational Chemistry

Computational chemistry is a powerful approach dedicated to solving chemical, biochemical, technological and industrial problems, also help to study the structures and properties of molecules and materials. It combines computational and theoretical chemistry, using principles and equations of physical, mathematical algorithms, statistics and requires a massive amount of data [30][31][32].

Computational chemistry is used in the creation of models of synthesis reactions to demonstrate the effects of kinetics and thermodynamics, and also to simulate and identify protein sites that are most expected bind to a new drug molecule or a specific target. With this approach is possible to improve significantly the efficiency of a new drug project [31][33].

One of the focuses of computational chemistry has been the protein-peptide interaction, to know how its mechanism works, how they will relate, predict drug effects and mutations, know which protein or peptide has more chances to bind, modeling studies, among others. Generally, the modeling of peptide-protein complexes is approached in two phases, the initial phase where the binding site of the peptide on a protein is identified and the second phase where the native position of the peptide is determined [14].

1.3.1. Three-Dimensional Structure Prediction

To understand the function of a protein and its life at the molecular level [34], knowledge of tertiary structure is necessary. Therefore, there are four levels of protein structure. The primary structure (amino acid sequences) is used to predict secondary and tertiary structures; the secondary structure is

folded within the polypeptide chain, stabilized by hydrogen bonds, taking the form of alpha helices or beta sheets; the tertiary structure is the three-dimensional (3D) arrangement of the secondary structure in a polypeptide chain, finally the quaternary structure involves more than one amino acid chain (figure 5) [35].

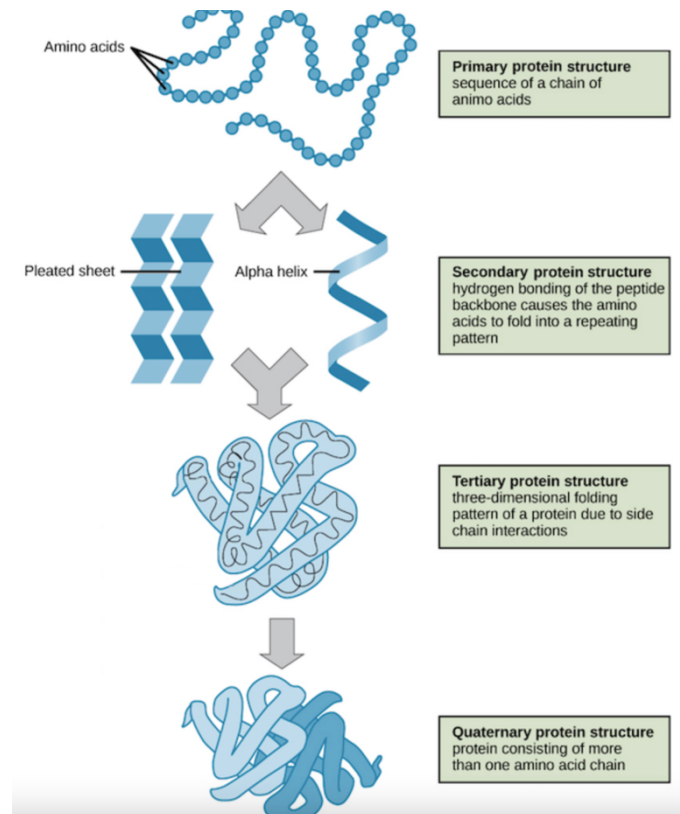


Figure 5: Four levels of protein structure [35].

The interest in determining the 3D structure of a protein has been growing, since it increases knowledge and understanding of the protein, making it possible to know how to affect, control or modify it. For example, through it it is possible to generate structural models of many hypothetical protein-protein complexes or create new protein-protein interactions [36][37].

Currently, there are four main computational alternatives to predict tertiary structures (figure 6), *Ab initio* methods (first principle methods without database information), Homology or Comparative modeling, Fold recognition or Threading [38][39]. All these methods fit in two main categories, *Ab initio* methods (or Template-free methods) and heuristic methods (or template-based methods), including the other two methods mentioned [40].

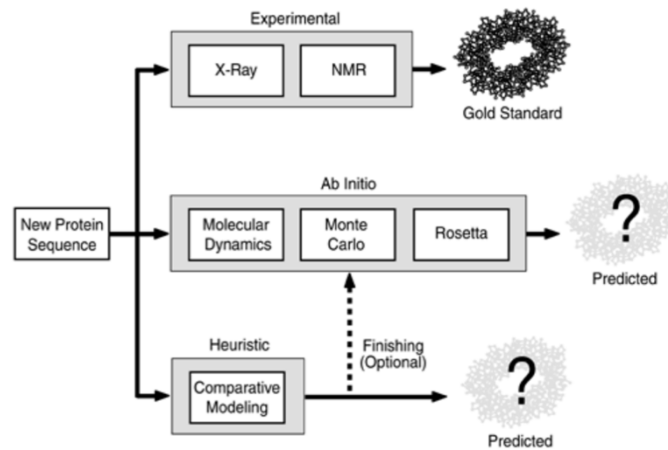


Figure 6: Computational Methods of Protein Structure Prediction vs Experimental Protein Structure Determination Methods [41].

Ab initio and Comparative modeling will be used to predict the tertiary structure.

1.3.1.1. *Ab initio* methods

Ab initio methods are based on the first principle laws, focusing on the chemical and physical properties of the amino acid sequence, without prior knowledge, in other words do not rely on any previously solved structure, and it is based on the fact that the native structure (primary structure) of protein is always at energy minimum [38][40][42]. Due to the computational requirements associated with *ab initio* methods, assumptions and simplifications for all proteins are necessary in order to facilitate the calculations involved. Therefore, there is the assumption that a protein of secondary structure, can be entirely defined as a function of bond length, bond angles, and torsion (dihedral) angles (figure 7) [39][41].

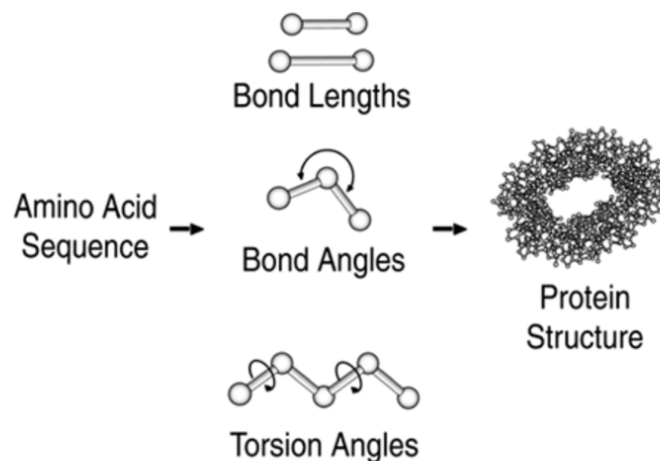


Figure 7: Bond length, bond angles and torsion angles [41].

The bond length and bond angle are assumed to be constant, however, the torsion angles are considered variable. Torsion (dihedral) angle, involves a torsion between two planes defined by two atoms that form the bond, so in total four atoms are involved. These atoms can vary between 0 to 360 degrees and there are three main angles, often referred to as omega (Ω), psi (ψ) and phi (ϕ) (figure 8).

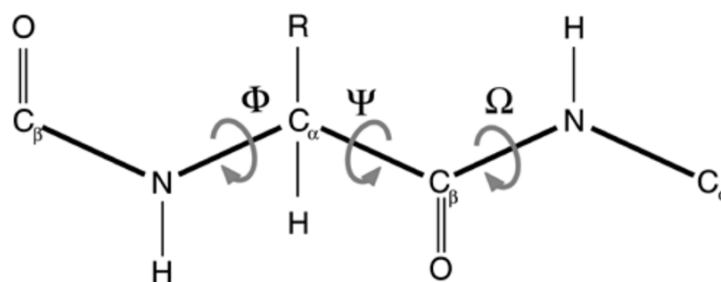


Figure 8: Dihedral Angles: Omega (Ω), Psi (ψ) and Phi (ϕ) [41].

Each amino acid in a polypeptide chain has a protein backbone defined by the set of the three angles mentioned above. At omega (Ω) 180 degrees C β and N are assumed constant; psi (ψ) C α and C β , and phi (ϕ) N and C α , due to their interaction with residues attached to the backbone, are restricted [41].

After generating the secondary structure from the amino acid sequence, using the bond length, bond angle and torsion angle, the next step is to generate the tertiary structure. This process involves methods like molecular dynamics, Monte Carlo and Rosetta. The first is based on Newtonian physics, where the force of each atom is calculated, the same atom is moved a certain distance in a short period of time. This process is repeated until the final structure has the most stable molecular conformation (lowest free energy) possible. Monte Carlo method, is used to identify structural conformation combinations with the lowest free energy. This process is repeated several times, through different simulations, the resulting structures are grouped into clusters and the structure that is located in the middle of these clusters is chosen and is considered the predicted structure [41]. Last but not least The Rosetta method, is based on generating a complete structure from three and nine residues chains of the unknown protein (target), since it is not known how the fragments will behave, several conformations are generated, however these conformations are limited and created with fragments libraries extracted from Protein Data Bank (PDB). The candidate structure chosen is the one whose local conformation has the lowest overall energy. [41][40][42].

1.3.1.2. Heuristic Methods

Heuristic methods, which are the most accurate and fastest, use a database of protein structure, namely amino acid sequence data, to predict the structure of tertiary proteins.

Homology and comparative modeling are used to predict protein structure from amino acid sequence data, however, they have differences. Homology modeling finds similarities between amino acid sequence considering ancestral relationships, because although time may have changed the exact composition of the protein, they can still be very similar in structure and function. It also assumes that proteins from the same families share folding motifs, conserved in certain regions, even they do not share the same sequences. If no homologous protein has been identified, then homologous modeling is impossible to use to predict the structure of a new protein [38][41].

In contrast, comparative modeling finds similarities between the target amino acid sequence and the amino acid sequence of another protein with known structure (template protein), independent

of the molecule's lineage. If there is similarity between the two sequences, the structural information obtained from the known structure is used to model the target protein, it also assumes that proteins with similar amino acid sequence share the same basic 3D structure [39][41].

There are four basic steps of comparative modeling procedure: template selection; template target alignment; model building and model evaluation [39][41]. All these steps are illustrated in figure 9.

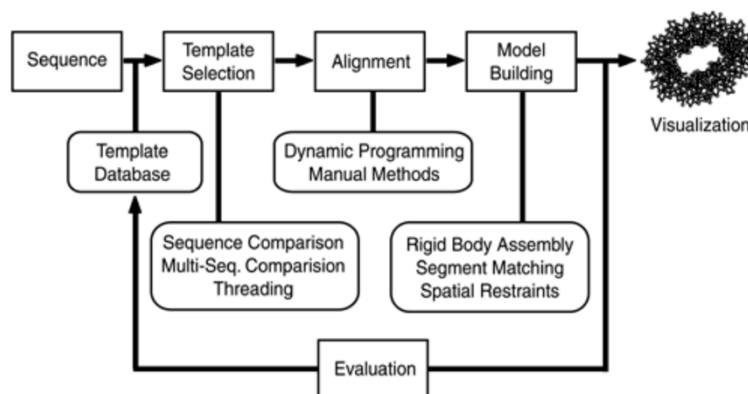


Figure 9: Comparative Modeling Process [41].

The template selection is based on the amino acid sequence of the target, where a database search is performed to find a template that is compatible with the target sequence. Usually, the protein structure template database used is PDB [39][41].

Template target alignment, consists of aligning the polypeptide sequence of the target sequence with the 3D structure of a template in order to position the target and template in the same 3D orientation, therefore it is possible to identify whether the amino acid sequence is spatially and chemically similar to the template. Only models that fit these parameters and best fit the target protein are potential models [39][41].

The construction of the template takes all previously chosen templates into consideration. Ideally, the structure of a template would fit exactly to the sequence of the target protein, however, due to numerous bends or curves in the backbone of the template it is likely that the template will not fully align to the target sequence. A change in a single angle is enough to change the conformation of the molecule. In order to avoid this, it is best to break the template into several parts and then fit these parts individually to the target. There are a few techniques, the most common being the rigid body, for this to be more precise their construction involves fragments of different templates, since in this way each area of the set of templates sequence has the conformation as similar as possible to the target sequence [39][41].

Finally, the model is evaluated according to its final accuracy, the greater the similarity between the model created and the template, through the Root Mean Square Deviation (RMSD), there is a higher probability that the model is an accurate prediction of the structure [41].

1.3.1.3. 3D Prediction and Evaluation

The webservers that will be used in this thesis to predict the three-dimensional structure are PEPstrMOD, I-TASSER and PEP-FOLD.

PEP-FOLD is a *de novo* approach that aims to predict peptide structures from amino acid sequences. This is based on the structural alphabet (SA) letters to describe the conformations of four consecutive residues, next it couples the predicted series of SA letters to a greedy algorithm and a coarse-grained force field. The *novo* approach incorporates *Ab initio* and knowledge-based methods [43][44][45][46].

PEPstrMOD is the second webserver used. This one predicts the tertiary structure of small peptides, whose strategy is based on the finding that β -turn is an important and consistent feature of small peptides in addition to regular structures. Therefore, this method uses both the regular secondary structure information predicted in PSIPRED as well as the β -turn information predicted in BetaTurns, making the structure more refined with energy minimization and molecular dynamical simulations [47][48][49].

I-TASSER automatically generates high-quality 3D structure prediction from their amino acid sequences. I-TASSER has several steps: in the first step it tries to identify similar fold model proteins (or super-secondary structures) from the PDB library through LOMETS, a locally installed meta-rotation approach (iterative model-based fragment assembly simulations). In the second, the excised continuous fragments from PDB templates are reassembled into full-length models through replica-exchange Monte Carlo simulations with the threading unaligned regions; the second step is also built by *Ab initio* modelling, then SPICKER through clustering the simulation decoys identify the low free energy states. Finally, in the third step the simulation of the fragment assembly from SPICKER clustering centroids occurs again, where spatial constraints collected from both LOMETS models as well as PDB structures by TM-align are used to guide simulations [50][51][52][53].

The models obtained by de webservers mentioned above need to be analysed, a very common and widely used webserver is the ProSA-Web. ProSA-Web (Protein Structure Analysis-web) is a tool used to validate 3D protein structures, where it identifies potential errors in their structure. This webserver calculates an overall quality score (z-score) for a specific input structure, this score is displayed in a graph that shows the scores of all experimentally determined protein chains currently available in the PDB (Protein Data Bank). Subsequently, both scores are related. In this way, the z-score indicates the overall quality of the model, where it checks whether the z-score of the input structure is within the range of scores typically found for native proteins of similar size [54][55][56].

1.3.2. Molecular Docking

Molecular docking has emerged over the past three decades, is a key tool in drug discovery and molecular modeling applications [57][58]. The goal of protein and peptide docking is to explore the predominant binding modes of a ligand when it binds to a protein with a known three-dimensional structure. In molecular docking, tens of thousands of possible ligand poses based on protein structures are generated through a search algorithm, where it searches the free energy landscape to find the best

ligand poses. Subsequently they are evaluated by an energy score function (ESF), and if this is correctly modeled it will correspond to the native binding mode, guiding and determining the ligand positions. Therefore, molecular docking consists in two main connected goals [59][57][58]:

- ◆ Determine the binding site and binding mode of a ligand to a protein (pose prediction);
- ◆ Estimate the binding affinity between the protein and the ligand.

Virtual screening, is the final stage of which can help provide a 3D hypothesis of how a ligand interacts with its protein target.

In general, molecular docking methods can be classified into three categories: i) protein-peptide docking, ii) protein-protein docking and iii) small molecule-protein docking [60]. In this case the category used is protein-peptide docking. In docking, there are two main strategies, *ab initio* docking (blind docking) and data-driven docking (Re-docking). The second is preferable since it not only takes into account the coordinates of the starting structures, but also knows the protein-peptide binding site, reducing the search space, making it faster, more accurate and increase the efficiency [61][62]. Among all docking methods, according to *Sjoerd J de Vries et. al.*, HADDOCK follows a data-driven strategy which enriches an overall vision of complexes formed [61].

1.3.2.1. High Ambiguity Driven DOCKing

High Ambiguity Driven DOCKing (HADDOCK) is a web server used to predict the 3D structure of protein-peptide complexes *in silico* using a variety of information sources to guide the docking process and score the predicted model. However, what makes it different from other software is its ability to incorporate experimental data as restraints and use it to guide the docking process along with traditional energy and shape complementarity [63][64][65][66].

The experimental data entered into HADDOCK is in the form of active and passive residues, which are converted into Ambiguous Interaction Restraints (AIRs), used to drive the docking [63][67][61][68].

- I. Passive residues, are those that contribute to the interaction, however, if no contacts are made they will not be scoring penalized, so they are considered less important, and they can only be in the interface.
- II. Active residues, described as the identified interface residues, are those of greatest importance for the interaction, such as residues whose elimination eliminates the interaction or those where the chemical disturbance is higher [63][67][61][68].

AIRs are created between each active residue from one partner and the combination of active and passive residue from the other partner. An important tool to conveniently define which residue are active and passive residue is penalty scoring. Because if the defined active residues are not at the interface it will directly influence the success of the molecular docking, so there must be a careful selection of the passive and active residues for the docking to be a success. This selection will generate a strict set of AIRs, leading to a very narrow sampling of the conformational space, giving rise to very similar poses and vice versa [14][63].

The HADDOCK protocol was designed so that molecules could experience different chemical environments and degrees of flexibility, consisting of three steps: Rigid-body energy minimization (it0); Semiflexible refinement in torsion angle space (it1); and a final refinement molecular dynamic in explicit solvent (water/itw) [61][63].

I. "it0":

$$\text{HADDOCKscore} - \text{it0} = 0.01E_{vdw} + 1.0E_{elec} + 1.0E_{desol} + 0.01E_{air} - 0.01BSA$$

Equation 3: HADDOCK scoring function in the it0 stage [68].

II. "it1":

$$\text{HADDOCKscore} - \text{it1} = 1.0E_{vdw} + 1.0E_{elec} + 1.0E_{desol} + 0.1E_{air} - 0.01BSA$$

Equation 4: HADDOCK scoring function in the it1 stage [68].

III. Itw "water":

$$\text{HADDOCKscore} - \text{water} = 1.0E_{vdw} + 0.2E_{elec} + 1.0E_{desol} + 0.1E_{air}$$

Equation 5: HADDOCK scoring function in the water (final) stage [68].

Where:

E_{vdw} is the non-bonded intermolecular van der Waals energy (adimensional)

E_{elec} is the non-bonded intermolecular electrostatic energy (adimensional)

E_{desol} is the empirical desolvation energy term (adimensional);

E_{air} is the AIR energy (adimensional);

BSA is the buried surface area (\AA^2).

The standard HADDOCK protocol generates 1000 models in the rigid body minimization stage, and then refines the best 200 in both it1 and water [63]. Afterwards the solutions are grouped (clusters) according to their pairwise RMSD values, next HADDOCK determines and ranks the clusters based on the average energy of the top four structures in each cluster [68]. The cluster numbering reflects the size of the cluster, with cluster 1 being the most populated cluster [14]. The final models are automatically clustered based on the positional interface ligand root mean square deviation (iL-RMSD), a specific similarity measure that captures conformational changes about the interface by fitting on the interface of the receptor and calculating the RMSDs on the interface of the smaller partner [63].

There is also statistics analysis of energetic terms and other structural measures for each cluster as the z-score, a quality measure parameter given by HADDOCK that indicates how many standard deviations from the average HADDOCK score a cluster is located, so the more negative the z-score, the better it is [14].

1.4. Solid phase peptide synthesis (SPPS)

Solid phase peptide synthesis (SPPS) is currently the method of choice for the chemical synthesis of peptides in the lab as well as in the industry [69][70]. The goal of SPPS is to sequentially add protected amino acids to an insoluble polymeric matrix (resin) to build a peptide chain. This is done through the carboxyl group of the first amino acid anchored to the resin in a covalent manner, through *linker* [71][14]. This strategy is used for small and medium peptides, up to 30 or 50 amino acids, depending on the literature source. To large peptides, a convergent strategy is used [69].

Peptide bonds result, in a successive way, from sequential reactions between the α -amino acid group of one amino acid with the α -carboxyl group of another amino acid, (figure 10) [72].

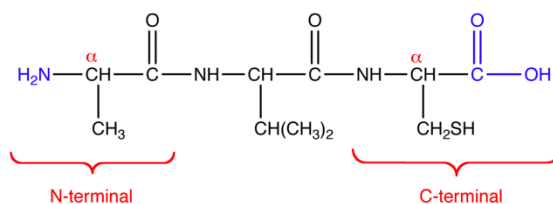


Figure 10: Example of an amino acid sequence where the amino acid residues at the N- and C-termini (blue), are both at the α -carbon [14].

There are two approaches to synthesize peptides, peptide synthesis in a Solid-Phase or peptide synthesis in solution. The difference between them lies, in solution synthesis the carboxyl group of the first amino acid is bound to a protecting group in solution, while in Solid-Phase synthesis the carboxyl group is protected through the resin. In this case, solid-phase will be used and the peptide grows in the C-terminal to N-terminal [72][70].

Within SPPS there are two different strategies that can be followed: Fmoc-SPPS and Boc-SPPS. The Fmoc-SPPS to protect the α -amino group of the amino acids, uses a fluorenylmethoxycarbonyl group (Fmoc), which is removed with a base and the Boc-SPPS to protect the α -amino group, uses a t-butyloxycarbonyl group (Boc) which is removed with an acid, usually hydrofluoric acid (HF). Nowadays, the most commonly used is Fmoc-SPPS. For this thesis we have used Fmoc-SPPS and piperidine for Fmoc deprotection [14].

The side chains of some amino acids are also protected; this protection is only removed when the peptide is cleaved from the solid support, unlike the protective α -amino groups that are removed along the synthesis [70].

The synthesis involves a set of cyclic steps: Cleavage of the α -amino protecting groups; Washing to remove the cleavage reagent; Coupling of the protected amino acid; Washing to remove non-reacting material and Cleavage the peptide from the resin [14]. This process is represented in the following figure 11.

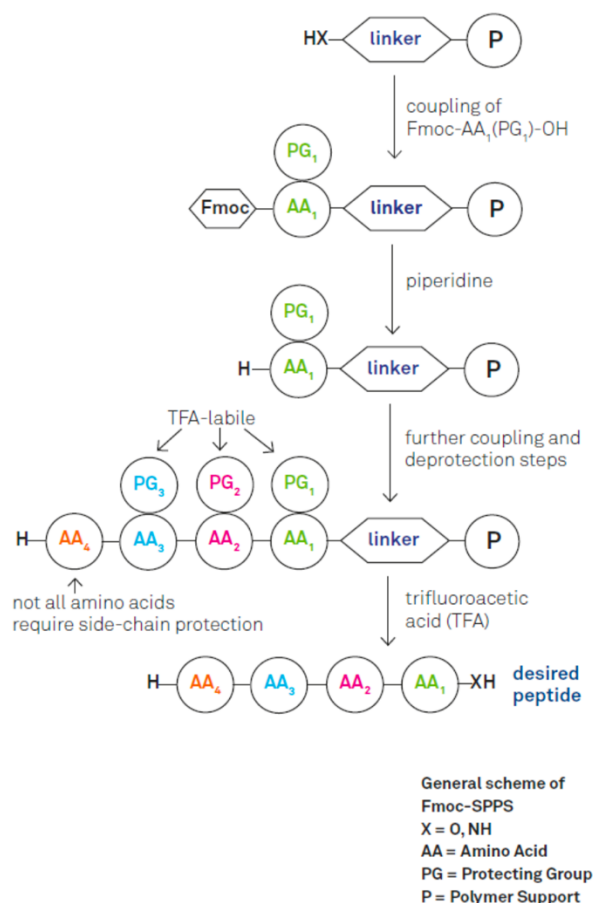


Figure 11: Solid phase peptide synthesis (SPPS)[14].

The resin is always washed between the steps and the primary solvents used for resin deprotection, coupling and washing are Dichloromethane (DCM) and N, N-dimethylformamide (DMF) [14][70]. In the final step, the cleavage is usually done with the proper cleavage solution, trifluoroacetic acid (TFA) and scavengers which trap highly reactive carbocations that are formed during the cleavage procedure and are responsible for the formation of unwanted byproducts [70]. The resulting peptide could have a carboxamide or carboxylic acid at its C-terminal, or other functional group, depending on the linker or the type of the resin used [14].

The choice of resin and linker is important for successful peptide synthesis. There are three types of resins: polystyrene (PS); polyethylene glycol (PEG) functionalized with PS and pure crosslinked PEG resins. PS resins are widely used and very successful in synthesis, especially for small to medium peptides. PS is usually crosslinked with 1% of divinylbenzene (DVB), swells well in non-polar solvents (DCM, DMF), is chemical inert; for these reasons it is the most commonly used resin in SPPS and is the one used in this thesis. Linkers play a dual role during peptide synthesis, offering protection against aggregation and providing reversible linkage between the peptide chain and the solid support. The linker that will be used is Rink Amide [70].

To control the manual SPPS, we performed a qualitative color test to detect free amine groups after deprotection as well as to monitor amino acid coupling. This control is referred to as, Kaiser test [73]. The Kaiser test uses ninhydrin (2,2-dihydroxyindane-1,3-dione) to observe the presence of primary

or secondary amines. If these functional groups (free primary amines) are present they will react with the ninhydrin, generating an intense blue color, indicating that deprotection has occurred successfully and the next amino acid can be attached. When the functional groups are not present, the ninhydrin does not react, a yellow color is observed, indicating that the coupling of the amino acid is complete, if you get a blue color, it means that the reaction is incomplete [74].

However, there are exceptions, the Kaiser test does not always generate a dark blue coloration, as is the case with serine, asparagine, aspartic acid and proline. In peptides that are to be synthesized there is only one exception, that is proline (secondary amines), where a red color is observed. [14].

1.5. High Performance Liquid Chromatography (HPLC)

In the last 25 years, High Performance Liquid Chromatography (HPLC) has proven to be extremely versatile and efficient in the separation and purification of molecules in general, but particularly peptides [75]. HPLC has some advantages over other conventional chromatographic techniques, such as: faster separation time; high recovery; better resolution and higher sensitivity. For all these reasons, the established method for the separation and purification of polypeptides is HPLC. This is a technique; based on highly resolute separation that can be achieved through uniform microparticle supports [76]. There are three major modes of HPLC utilized for peptides, like size-exclusion; ion-exchange and reversed-phase (RP) [75]. The mobile phase conditions, of each mode, can be handled in a way that maximizes the separation potential of a particular HPLC column [77].

The Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC) is the most widely used mode for separating peptides, as it is superior in both speed and efficiency when compared with the others. This mode is ideal for analytical and preparative separations because it offers the availability of volatile mobile phases. Its main components are: i) the pump, which moves the mobile phase and the sample through the column; ii) a column containing the stationary phase; and iii) a detector which shows the retention time of the molecules (this varies depending on the interaction of the sample with the mobile and stationary phases) [75].

HPLC columns that continue to be widely used for all major modes of HPLC (including RP-HPLC) have their stationary phase silica-based (solid particles) and their microparticulate rigidity allowing the use of high flow rates of mobile phases. It is important that the silica-based stationary phase (solid particles packed on the column) has a pore size up to 300 Å to be as stable as possible, and the particle size used should be between 5 and 10 µm. The conditions mentioned above are so that the solute has easy access to the pores present in the silica support, i.e. it should be as unrestricted as possible in order to perform fast analyses, since silica columns have a limited pH, between 2.0 to 8.0, and when in contact with basic eluents they are quickly dissolved. Also, the analytical dimensions of the RP-HPLC column typically include a length of 10 to 25 cm and a diameter of 4 to 4.6 mm. These conditions are applicable for separations of small peptides to medium molecular weight proteins [75][77].

The choice of column is based on microparticulated silica supports and organic polymers with a wide pH tolerance and its choice also depends on the nature of the substances to be analyzed and/or separated [77][75].

Generally, a pH below 3.0 is used in the mobile phase, because acidic pH values prevent undesirable ionic interactions from occurring with the stationary phase and protect the silica column core, which is soluble above pH 7.0 [75]. As such, the choice of solvent is also important, so that it is compatible with the pH range mentioned. The solvents used in mobile phase to separate peptides in HPLC, can be water-based or non-polar solvents, the last one are used as organic modifiers or to cancel non-ideal hydrophobic interactions [78]. The function of an organic solvent is to disaggregate the protein or peptide from the hydrophobic surface, where the concentration of the organic solvent is slowly raised until the peptides of interest disaggregate and elute [79]. The most used classical organic solvent is acetonitrile (ACN), a weaker solvent whose structure does not favor hydrogen bonding interactions and does not compete with water for the active sites of the stationary phase, thus decreasing the stability of the water layer used for the separation of compounds in HPLC. It is also very volatile, which favors the drying process of the collected fractions [80].

There are also undesirable interactions between the acidic silanols present in the stationary phase and the basic analytes, causing a broadening of the chromatographic peaks and loss of symmetry, so mobile phase additives are required [80]. Over the years various solvent systems have been developed, including trifluoroacetic acid, buffer components, salts and urea, but the most widely used in RP-HPLC is trifluoroacetic acid-based [78].

For optimal results, both solvents and mobile phase additives should be of the best possible HPLC grade. Otherwise, it will cause irreversible absorption of impurities on the column head, block adsorption sites, change the selectivity of the column and eventually break the peaks in the chromatogram and generate ghost peaks [81].

For the separation of peptides to occur successfully, there are still standard chromatographic conditions for RP-HPLC, and the best approach considered is to use aqueous trifluoroacetic acid (TFA) at linear TFA-acetonitrile gradients (pH=2.0), i.e., proportionally go inverting the percentage of solvents used, at a flow rate between 0.5 to 2.0 ml/min and also a gradient between 0.5 % to 2.0 % acetonitrile/min [78].

Before any run, the channels to be used should be purged and the system should be stabilized by means of a controlled flow unit, in order to prevent bubbles from forming in the system and to check for any leakage through pressure [81].

1.6. Mass Spectrometry

There is a high compatibility between HPLC and mass spectrometry due to the high percentage of organic solvent in the mobile phase, this can be demonstrated in numerous areas of science, one of them being pharmaceuticals [80].

Mass spectrometry is a technique for measuring/determining the mass-to-charge ratio (m/z) of one or more ionized species present in a sample, in gas phase, and is also often used to calculate the exact molecular weight of components present in the sample. In the study of peptides, this technique is very important because it can be used to identify unknown compounds by their molecular weight, determine the structure and chemical properties of the molecules, quantify known compounds and detect possible modifications [82][83].

The mass spectrometer consists of: sample input unit, ionization source, mass analyzer, ion detector and data system. The sample inlet unit is where the sample is injected; this can be a gas, liquid or solid. The sample is converted to vapor to obtain a stream of molecules, then the molecules flow to the ionization source (ionization chamber) where they are ionized. There are four ionization methods: electron ionization (EI); chemical ionization (CI); desorption ionization techniques (SIMS, FAB and MALDI) and electrospray ionization (ESI) [84].) However, because it is a technique that works with gas phase ions and because of the difficulty of producing stable gas phase ions with peptide molecules, for protein and peptide studies the most commonly used ionization sources are MALDI and ESI [82]. In this case, the technique used for mass spectrometry was ESI-MS.

In ESI ionization, the solution containing the analyte is pumped into a high-voltage capillary, causing the liquid to disperse (form a spray) and form droplets with multiple charges. These droplets are subjected to a drying gas, where they are desolvated and their volume decreases, but the charges remain the same. The decrease in volume causes the charges to come closer together, leading to an instability in the droplets that leads to a great deal of repulsion, to the point where the droplet is broken. As a consequence, the formation of the analyte with multiple charges occurs, which is very useful in MS analysis of proteins. A schematic for this technique is shown in figure 12 [82][84].

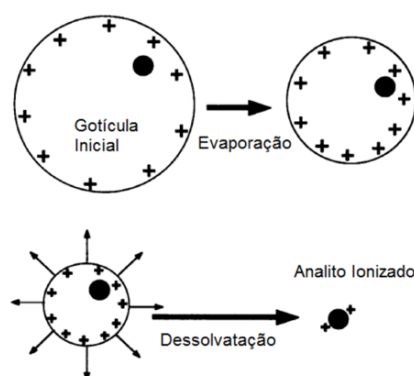


Figure 12: Ionization by ESI-MS [82].

Once ionized, the ion beam is accelerated by an electric field, where it enters the mass analyzer and is classified and separated according to the mass-to-charge (m/z) ratio[84]. The separated ions reach the detector, which is composed of a counter that produces a current proportional to the number of ions reaching it. Finally, the ions are inserted and sent to a data system where the m/z ratio is stored along with their relative abundance. In this way, each peak in a mass spectrum will show a unique m/z component in the sample, and the heights of the peaks show the relative abundance of the various components in the sample [83].

2. Motivation and Aim of the Thesis

According to the World Health Organization (WHO), cancer is one of the leading causes of death worldwide. In 2020, 10 million deaths were counted, the main one being breast cancer. Cancer is associated with modifications or disturbances in the PPIs responsible for physiological and biological processes. Studies have shown that both breast cancer and prostate cancer have a propensity for bone metastasis. Bone metastases lead to a deregulation of the structural and metabolic integrity of the bone, affecting the dynamic process of bone remodelling, which is maintained by two main types of bone cells, the osteoblasts (formation of new bone) and the osteoclasts (destruction of old bone). Currently, there is no efficient treatment, there are only therapeutic targets, so in this thesis we will study and explore one of them, the RANK signaling pathway. The interest in it is due to the fact that it is responsible for osteoclast activation. In-depth knowledge of PPIs has become very relevant to better understand the disease and thus develop new therapies. This concept is still not fully understood and so more studies become important to better understand the PPIs structure, function and dynamics. Only then it will be able to find more specific solutions to control a number of diseases.

There are already some studies reporting the production of peptides to inhibit the RANK-TRAF6 interaction. These peptides mimic the binding site of RANK to TRAF6, potentially binding to TRAF6 and thus inhibiting protein binding. However, the results are still preliminary and need further studies namely cellular studies to understand if there is indeed a binding inhibition. Therefore, this work has as main objective the prediction of 3D structures of these peptides, understand how occurs the interaction with TRAF6 and synthesize them.

3. Methodologies

3.1. Decoy peptides

Decoy peptides are peptides whose peptide sequence is based on the binding motif of the target protein, in this case, the binding motif of TRAF6 in RANK. Decoy peptides are being used like PPI inhibitors with the aim to improve productivity and reduce metabolism of peptides. Nowadays there are a large number of peptide based drugs being marketed. And this way help the development of novel therapeutics for various diseases, such as cancer [85][86].

Ann T. Poblentz conducted a study to identify the most effective inhibitor of the TRAF6 decoy peptides to prevent osteoclast differentiation. At the end of the study T6DP3 (RKIPTED EY) was found to be the most effective inhibitor [87].

Therefore, the inhibitor listed above will be produced, as well as three more inhibitors. In the second inhibitor to be produced the active residues that are experimentally described as important in the interaction RANK-TRAF6 are replaced by alanines to see if they are really important residues in the RANK-TRAF6 interaction. The third inhibitor is inhibitor one plus a Cell Penetrating Peptide (CPP) and the fourth is the second inhibitor plus CPP. The purpose of CPP is to facilitate the entry of the inhibitor into the membrane. The CPP used is, AAVALLPAVLLALLAP, a hydrophobic sequence of the Kaposi fibroblast growth factor signal peptide [22]. The four peptides to be produced are present in table 1.

Table 1: Peptides to be produced.

Peptide	Sequence
Peptide 1	RKIPTED EY
Peptide 2	RKIAT ADEA
Peptide 3	AAVALLPAVLLALLAP RKIPTED EY
Peptide 4	AAVALLPAVLLALLAP RKIAT ADEA

3.2. Computational Chemistry

3.2.1. Three-Dimensional structure prediction

Initially two different webservers were going to be used to predict the tertiary structures of the peptides, in order to decrease the error associated to its prediction, PEP-FOLD 3 and I-TASSER. However the I-TASSER only predicts structures from 10 amino acids and being that two of the sequences to be predicted have 9 amino acids. Therefore, it was necessary to find another webserver capable of predicting the tertiary structure of smaller peptides, so a third webserver was used, PEPstrMOD.

PEP-FOLD[44][45][46] and PEPstrMOD[47][48][49] webserver were used to predict the tertiary structure of sequences 1 and 2, and PEP-FOLD and I-TASSER[50][51][52][53] webserver will be used to predict the tertiary structure of sequences 3 and 4. Then the sequences are submitted on the different webserver, where the five best resulting models for each sequence on each webserver will be obtained. Finally, the models are analysed in the ProSA-Web[54][55][56] webserver, which allows us to select the best and most robust model for each sequence.

3.2.2. Molecular Docking

The molecular docking study was performed using the HADDOCK server. HADDOCK starts with a randomization of orientations and rigid body energy minimization (1000 solutions), then occurs semi-rigid simulated annealing in torsion angle space (200 solutions), and finally occurs a refinement in Cartesian space with an explicit solvent (200 solutions). HADDOCK also uses biological information to drive docking, introducing ARIs (Ambiguous Interaction Restraints)[65][66].

Two docking runs were performed for each peptide, where in the first docking the active residues chosen for peptides 1, 2, 3 and 4 were those that are experimentally described as important in the interaction with the TRAF6 and in the second docking all residues belonging to the peptides are considered active. The docking method used was protein or protein-ligand. These are illustrated in table 2.

Table 2: Conditions for each docking.

	Run Name (ID)	Active Residue Chain A → TRAF6	Active Residue Chain B → Peptides
Active residues that are experimentally described as important in the interaction with the TRAF6	TRAF6-Seq_1	392, 410, 473, 471, 432, 459	4,6,9
	TRAF6-Seq_2		4,6,9
	TRAF6-Seq_3		20,22,25
	TRAF6-Seq_4		20,22,25
All residues belonging to the peptides are considered active	TRAF6-Seq_1_1		1,2,3,4,5,6,7,8,9
	TRAF6-Seq_2_1		1,2,3,4,5,6,7,8,9
	TRAF6-Seq_3_1		1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15, 16,17,18,19,20,21, 22,23,24,25
	TRAF6-Seq_4_1		1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15, 16,17,18,19,20,21, 22,23,24,25

For each run, 200 protein-peptide complexes were obtained, however only the 10 best complexes generated by HADDOCK and according with their HADDOCK score will be analysed with the aim of identifying the most stable complex. The analysis will be done by screening the 10 complexes generated by HADDOCK and by determining the interface residues and the hot spots present at the interface upon binding. For the determination of the interface residues and hot spots will be used the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC., a tool that allows the biomolecular visualization of the complexes, where from scripts *InterfaceResidues.py*, it is possible to obtain the interface residues, bond pose analysis and hot spots.

3.3. Solid-Phase Peptide Synthesis

3.3.1. Aluminum foil test

To start the synthesis, it is necessary to identify the best bath location in the ultrasonic to place the samples, in order to shorten the coupling time and consequently increase the yield. To determine the best location, the aluminum foil test was performed in the ultrasounds (Fisherbrand®)[88], where an aluminum foil at the bottom of the bath without the stainless-steel basket (figure 13) and another aluminum foil was placed on the surface with the stainless-steel basket (figure 14). The places where there is a hole in the aluminum foil are the indicated place to perform the synthesis, so all the synthesis was performed in the center.



Figure 13: Aluminum foil test at the bottom of the bath without the stainless-steel basket.



Figure 14: Aluminum foil test on the surface with the stainless-steel basket.

3.3.2. Solid-Phase Peptide Synthesis

To start the solid phase peptide synthesis, it was necessary to weigh the mass of resin to be used; the mass of each amino acid to be conjugated; the mass of HBTU (activator) and the volume of DIPEIA. These calculations depended on the reaction equivalents and the resin used.

Two Resins were used, Rink Amide MBHA resin (Novabiochem®) with a substitution coefficient of 0.78 mmole/g and Rink Amide MBHA resin (Novabiochem®) with a substitution coefficient of 0.38 mmol/g. The choice of the resin depends on the size of the peptide to synthesize. For sequence 1 the resin with substitution coefficient of 0.78 mmol/g was used and for the remaining sequences the Rink-amide resin with substitution coefficient of 0.38 mmol/g was used. For calculation purposes were considered 3.5 equivalents for amino acid and HBTU and for DIPEIA were considered 10 equivalents. The results of the calculations performed are present in table 3, 4, 5 and 6.

Table 3: Quantity of each reagent to be used during SPPS, for peptide 1.

Peptide 1	Amino acid	Amino acid mass (g)	Activator Mass (g) (HBTU)	Volume Dipeia (µl)	Coupling time (min)
<i>Tyrosine</i>	Fmoc-Tyr(tBu)-OH	0.161	0,133	174,180	15
<i>Glutamic Acid/Glutamate</i>	Fmoc-Glu(OtBu)-OH	0.104			15
<i>Aspartate</i>	Fmoc-Asp(OtBu)-OH	0.144			15
<i>Glutamic Acid</i>	Fmoc-Glu(OtBu)-OH	0.104			15
<i>Threonine</i>	Fmoc-Thr(tBu)-OH	0.139			20
<i>Proline</i>	Fmoc-Pro-OH	0.118			40
<i>Isoleucine</i>	Fmoc-Ile-OH	0.124			45
<i>Lysine</i>	Fmoc-Lys(Boc)-OH	0.158			50
<i>Arginine</i>	Fmoc-L-Arg(Pbf)-OH	0.227			60
Resin mass (g)	0.128				

Table 4: Quantity of each reagent to be used during SPPS, for peptide 2.

Peptide 2	Amino acid	Amino acid mass (g)	Activator Mass (g) (HBTU)	Volume Dipeia (µl)	Coupling time (min)
<i>Alanine</i>	Fmoc-Ala-OH	0.218	0,266	347,700	40
<i>Glutamic Acid/Glutamate</i>	Fmoc-Glu(OtBu)-OH	0.208			20
<i>Aspartate</i>	Fmoc-Asp(OtBu)-OH	0.288			30
<i>Alanine</i>	Fmoc-Ala-OH	0.218			30
<i>Threonine</i>	Fmoc-Thr(tBu)-OH	0.278			30
<i>Alanine</i>	Fmoc-Ala-OH	0.218			30
<i>Isoleucine</i>	Fmoc-Ile-OH	0.247			30
<i>Lysine</i>	Fmoc-Lys(Boc)-OH	0.317			30
<i>Arginine</i>	Fmoc-L-Arg(Pbf)-OH	0.227			70
Resin mass (g)	0.526				

Table 5: Quantity of each reagent to be used during SPPS, for peptide 3.

Peptide 3	Amino acid	Amino acid mass (g)	Activator Mass (g) (HBTU)	Volume Dipeia (µl)	Coupling time (min)
<i>Tyrosine</i>	Fmoc-Tyr(Tbu)-OH	0.161	0,133	174,180	15
<i>Glutamic Acid/Glutamate</i>	Fmoc-Glu(OTBu)-OH	0.104			15
<i>Aspartate</i>	Fmoc-Asp(OTBu)-OH	0.144			15
<i>Glutamic Acid/Glutamate</i>	Fmoc-Glu(OTBu)-OH	0.104			15
<i>Threonine</i>	Fmoc-Thr(tBu)-OH	0.139			15
<i>Proline</i>	Fmoc-Pro-OH	0.118			20
<i>Isoleucine</i>	Fmoc-Ile-OH	0.124			15
<i>Lysine</i>	Fmoc-Lys(Boc)-OH	0.158			15
<i>Arginine</i>	Fmoc-L-Arg(Pbf)-OH	0.227			40
<i>Proline</i>	Fmoc-Pro-OH	0.118			25
<i>Alanine</i>	Fmoc-Ala-OH	0.109			40
<i>Leucine</i>	Fmoc-Leu-OH	0.124			25
<i>Leucine</i>	Fmoc-Leu-OH	0.124			40
<i>Alanine</i>	Fmoc-Ala-OH	0.109			30
<i>Leucine</i>	Fmoc-Leu-OH	0.124			30
<i>Leucine</i>	Fmoc-Leu-OH	0.124			40
<i>Valine</i>	Fmoc-Val-OH	0.119			60
<i>Alanine</i>	Fmoc-Ala-OH	0.109			40
<i>Proline</i>	Fmoc-Pro-OH	0.118			40
<i>Leucine</i>	Fmoc-Leu-OH	0.124			50
<i>Leucine</i>	Fmoc-Leu-OH	0.124			55
<i>Alanine</i>	Fmoc-Ala-OH	0.109			70
<i>Valine</i>	Fmoc-Val-OH	0.119			60
<i>Alanine</i>	Fmoc-Ala-OH	0.109			70
<i>Alanine</i>	Fmoc-Ala-OH	0.109			80
Resin mass (g)	0.263				

Table 6: Quantity of each reagent to be used during SPPS, for peptide 4.

Peptide 4	Amino acid	Amino acid mass (g)	Activator Mass (g) (HBTU)	Volume Dipeia (µl)	Coupling time (min)
<i>Alanine</i>	Fmoc-Ala-OH	0.218	0.266	347.700	40
<i>Glutamic Acid/Glutamate</i>	Fmoc-Glu(OTBu)-OH	0.208			20
<i>Aspartate</i>	Fmoc-Asp(OTBu)-OH	0.288			30
<i>Alanine</i>	Fmoc-Ala-OH	0.218			30
<i>Threonine</i>	Fmoc-Thr(tBu)-OH	0.278			30
<i>Alanine</i>	Fmoc-Ala-OH	0.218			30
<i>Isoleucine</i>	Fmoc-Ile-OH	0.247			30
<i>Lysine</i>	Fmoc-Lys(Boc)-OH	0.317			30
<i>Arginine</i>	Fmoc-L-Arg(Pbf)-OH	0.227			70
<i>Proline</i>	Fmoc-Pro-OH	0.118	0.133	174.180	15
<i>Alanine</i>	Fmoc-Ala-OH	0.109			15
<i>Leucine</i>	Fmoc-Leu-OH	0.124			25
<i>Leucine</i>	Fmoc-Leu-OH	0.124			40
<i>Alanine</i>	Fmoc-Ala-OH	0.109			20
<i>Leucine</i>	Fmoc-Leu-OH	0.124			30
<i>Leucine</i>	Fmoc-Leu-OH	0.124			30
<i>Valine</i>	Fmoc-Val-OH	0.119			40
<i>Alanine</i>	Fmoc-Ala-OH	0.109			40
<i>Proline</i>	Fmoc-Pro-OH	0.118			40
<i>Leucine</i>	Fmoc-Leu-OH	0.124			50
<i>Leucine</i>	Fmoc-Leu-OH	0.124			50
<i>Alanine</i>	Fmoc-Ala-OH	0.109			60
<i>Valine</i>	Fmoc-Val-OH	0.119			50
<i>Alanine</i>	Fmoc-Ala-OH	0.109			60
<i>Alanine</i>	Fmoc-Ala-OH	0.109			75
Resin mass (g)	0.263				

After the calculation of the resin mass to be used, as well as the remaining calculations, the resin was weighed and transferred to the polymeric reactor with a incorporated frit (PP-Reactors 5 mL with PE frit, MultisynTech GmbH), where the synthesis was performed. Next, the resin was swollen where it was washed three times with DMF (CARLO ERBA reagents) and three times with DCM (CARLO ERBA reagents), respectively. After swelling, it was necessary to deprotect the resin before the first

amino acid was conjugated, this process is called deprotection or removal of the Fmoc group. Thus, 4 mL of 20 % piperidine (SIGMA-ALDRICH®) solution in DMF (deprotection solution) was added to the reactor, the reactor was sealed and taken to an ultrasonic (figure 15) bath for 5 to 10 min so that deprotection can occur. After the ultrasonic bath, the deprotection solution was discarded and the resin was washed with DMF (three times) and with DCM (three times). At the end of deprotection, a Kaiser test was performed (chapter 1.4).



Figure 15: Ultrasounds.

After deprotection, the first amino acid was conjugated where a solution was prepared in an eppendorf of 3.5 amino acid equivalents and 3.5 equivalents of HBTU (activator) (Iris Biotech GmbH) in 1mL of DMF, the solution was taken to an ultrasonic bath for 5 min to homogenize well and promote amino acid activation. Then, 10 equivalents of DIPEA (activating base)(SIGMA-ALDRICH®) were added to the solution and the eppendorf was again taken to the ultrasonic bath for 1 min to be well homogenized. Subsequently, the solution was transferred to the reactor where the first amino acid goes coupling for 10 min in the ultrasonic bath. The coupling time was adjusted according to the amino acid to be conjugated and the peptide chain length. After conjugation, the solution was removed and the resin washed with DMF (three times) and DCM (three times). At the end of each coupling, a Kaiser test was performed and then the deprotection of the amino acid is carried out in order to conjugate the next one. The coupling and deprotection process described above is repeated until the end of the peptide chain has been reached and the peptide has been synthesized. All amino acid used were from Novabiochem®.

The final step to complete the peptide synthesis was the cleavage process, where the resin with the chain was swollen, deprotected, followed by a Kaiser test. After these steps, a 4 mL cleavage solution (cocktail) was prepared containing 95 % trifluoroacetic acid (TFA)(3.8 mL) (SIGMA-ALDRICH®) 2.5 % deionized water (0.1 mL) and 2.5 % triisopropylene (TIS)(0.1 mL)(SIGMA-ALDRICH®). The prepared cleavage solution was added to the reactor and left for four hours under constant agitation. At the end of the four hours, the liquid was transferred to a falcon and was subjected to a constant nitrogen flow to evaporate the remaining amount of TFA, and allowed to evaporate to the 1mL mark. After the TFA was removed the peptide was precipitated. Ether was added to the falcon in a 1:10 ratio of TFA to

diethyl ether, peptide was seen to precipitate and the solution was homogenized. When correctly homogenized, the peptide was centrifuged (centrifuge HERMLE) for 5 minutes at 5000 rpm and an acceleration of 5. After centrifugation, the precipitated peptide was observed and the supernatant was transferred to a new falcon keeping the precipitate in the first falcon. The process of ether addition, centrifugation and removal of the supernatant was repeated three times. At the end of the third time, the excess of ether was removed under a constant nitrogen flow and finally the nitrogen dried peptide was obtained. This was stored in the freezer.

The monitoring of the synthesis was performed through the Kaiser test. For its realization it was necessary to prepare three different solutions. The solution A with 5 g of ninhydrin in 100 mL of ethanol; solution B with 80 g of phenol in 20 mL of ethanol and solution C with 2 mL of 0.001 M of aqueous potassium cyanide (KCN) in 98 mL of pyridine. Two drops of each solution A, B and C were added to a test tube containing a few dry resin spheres. The test tube was placed in a water bath for 5 to 10 minutes; after that time the resin spheres were thoroughly observed. During the coupling phase, if the spheres are colourless and the solution is yellowish we can proceed to the next conjugation, if the colour is blue, the conjugation has to be repeated. In case of deprotection, if the solution is blue or reddish (proline) it is deprotected; if it remains yellowish, the deprotection has to take place again. An example of the Kaiser test where the amine group is unprotected and where the amine group is protected is illustrated in figure 16.

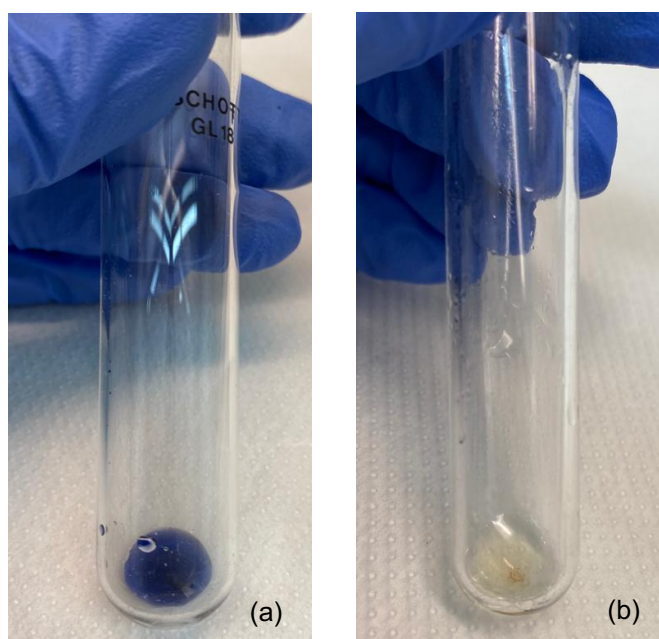


Figure 16: An example of the Kaiser test where the amine group is unprotected and the colour is blue(a) and where the amine group is protected and it is colourless (b).

3.4. Peptides chemical characterization and purification

The biological characterization and purification of the peptide was performed by Reverse Phase High Performance Liquid Chromatography (RP-HPLC). The solvents used were MilliQ® water and acetonitrile (ACN) (CARLO ERBA reagents).

3.4.1. Reverse-Phase High Performance Liquid Chromatography (Analytical RP-HPLC)

For the peptide isolation process a solution of 200 µL of water was prepared to which a small portion of cleaved and dried peptide was added, this was agitated to dissolve the all peptide. Sometimes is necessary to add one or two drops of ACN with 0.1% TFA to help dissolve, but it depends on the behaviour of the peptide. We took about 35 µL of the solution of each peptide prepared above, and injected it into the analytical RP-HPLC; the time between runs is about 35 min, depending on the method used. The system is composed with a pump (PerkinElmer® 200 Series Pump), a detector (PerkinElmer® 200 Series UV/Vis), a degasser (PerkinElmer® 200 Series vacuum degasser) and a column (Supelco Analytical, Discovery® BIO Wide Pore C18-5, 25 cm X 4.6 mm, 5 µm, SIGMAALDRICH®). The wavelength used for UV detection was between 210 nm to 220 nm, the eluents used in the system were H₂O with 0.1% TFA and ACN with 0.1% TFA, in channel A and B, respectively. Having peptides with different sizes, two HPLC methods were optimized, one for the smaller ones and the other for the larger ones, both methods are described in table 7 and 8.

Table 7: Optimised method for the smallest peptides (peptide 1 and 2).

Time (minutes)	Eluent A (%)	Eluent B (%)	Flow (ml/min)
-	95	5	1
3	95	5	
28	75	25	
30	0	100	
33	0	100	
35	95	5	

Table 8: Optimised method for the largest peptides (peptide 3 and 4).

Time (minutes)	Eluent A (%)	Eluent B (%)	Flow (mL/min)
-	90	10	1
3	70	30	
28	40	60	
30	0	100	
33	0	100	
35	90	10	

3.4.2. Reverse-Phase High Performance Liquid Chromatography (Preparative RP-HPLC)

For the peptide purification process, a solution was prepared of approximately 10 mL of H₂O with 0.1% TFA to which a portion of cleaved and dried peptide was added; this was agitated to dissolve all the peptide: add one or two drops of ACN if necessary. At first about 500 μ L was removed to test how much could be injected without losing resolution, so the volume to be injected in the next runs was 1000 μ L; the time between runs is about 45-50 min depending on the method. The system is composed with a pump (Waters 2535 Quaternary Gradient Module), a detector (Waters 2998 Photodiode Array Detector), a degasser (Uniflows, DG- 3210) and a column semi-preparative (MACHEREY-NAGEL Nucleosil® 100-5 C18, 250 cm x 8 cm). The wavelength used for UV detection was between 210 nm and 220 nm, the eluents used in the system were H₂O with 0.1% TFA and ACN with 0.1% TFA, in channel A and B, respectively.

The HPLC methods used were the same as applied in the analytical part. During purification, three fractions of the pure peptide were removed and then lyophilized (CoolSafe 100-9 Pro).

3.4.3. Mass Spectrometry (MS)

Analysis by ESI-MS of compounds was performed with the aid of an electrospray ionization (ESI) mass spectrometer (Bruker HCT Esquire 3000 plus®). This analysis was performed according to the methodology developed by the research group of Radiopharmaceutical Sciences from Center for Nuclear Sciences and Technologies (C²TN).

To calculate the molecular weight of each peptide Expassy was used, where 1g was taken from the molecular weight obtained, because the peptides end with a carboxamide group instead of a carboxyl group, due to the resin used. The molecular ions of each peptide were then calculated by hand. The molecular weight and molecular ions are shown in table 9 [89][90][91].

Table 9: Molecular weight and expected molecular ions, of each peptide.

Peptide	MM (g/mol)	m/z		
		[M+1H ⁺] ⁺	[M+2H ⁺] ²⁺	[M+3H ⁺] ³⁺
Peptide 1	1149.25	1150.25	575.64	384.09
Peptide 2	973.08	974.08	487.55	325.37
Peptide 3	2647.18	2648.18	1324.60	883.40
Peptide 4	2471.01	2472.01	1236.52	824.68

4. Results and Discussion

4.1. Computational Chemistry

4.1.1. Three-dimensional structure prediction

Through the webservers mentioned in chapter 3.2.1, five models were generated for each peptide. They were further analysed in the ProSA-Web webserver according to their z-score and visualized in the PyMOL software to observe their structure in order to obtain the most stable and accurate model. The z-score values for each model of each peptide are presented in table 10 [55][56].

Table 10: Z-score values.

		PEP-FOLD	I-TASSER	PEPstrMOD
Peptide	Model	z-score		
Peptide 1	1	- 0.31	n/a*	- 0,6
	2	- 1.42	n/a*	n/a**
	3	- 0.24	n/a*	n/a**
	4	- 0.38	n/a*	n/a**
	5	- 0.93	n/a*	n/a**
Peptide 2	1	- 1.40	n/a*	0, 54
	2	- 1.34	n/a*	n/a**
	3	- 1.29	n/a*	n/a**
	4	- 0.25	n/a*	n/a**
	5	- 0.74	n/a*	n/a**
Peptide 3	1	- 2.53	- 2.96	n/a***
	2	- 3.91	- 1.95	n/a***
	3	- 3.16	- 3.21	n/a***
	4	- 4.86	- 2.41	n/a***
	5	- 2.05	- 0.42	n/a***
Peptide 4	1	- 2.37	- 1.75	n/a***
	2	- 2.62	- 0.09	n/a***
	3	- 3.22	- 1.26	n/a***
	4	- 3.03	- 1.37	n/a***
	5	- 2.59	- 0.13	n/a***


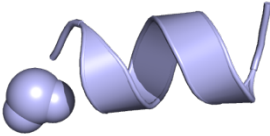
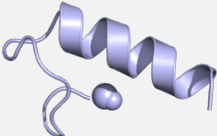
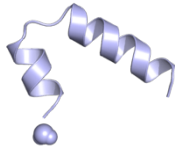
*peptide 1 and 2 with amino acid sequence < 10.

**PEPstrMOD create one model for each sequence.

***PEPstrMOD was not use for peptide 3 and 4

After the analysis according to the z-score, the best and most stable model for each peptide was chosen, shaded in blue in table 10; all those chosen were generated by the PEP-FOLD web server [44][45][46]. Thus, the chosen models are shown in table 11 and will be used in molecular docking.

Table 11: Final 3D structure of each peptide.

Peptideo	Final Structure
Peptide 1	
Peptide 2	
Peptide 3	
Peptide 4	

4.1.2. Molecular Docking

The 10 best complexes were selected according to the HADDOCK score considering that the more negative the better. Beyond the HADDOCK score, the top10 was analyzed according their percentage of residues at the interface and RMSD. The HADDOCK score is a data provided by HADDOCK when the complexes are downloaded and allows to perceive which is the best complex obtained according to an energy algorithm (see chapter 1.3.2.1.); the percentage of interface residues was calculated considering: i) the active residues that are experimentally described as important in the interaction with the TRAF6 protein (TRAF6_Seq_1; TRAF6_Seq_2; TRAF6_Seq_3 and TRAF6_Seq_4) and ii) all residues of the peptide sequence. (TRAF6_Seq_1_1, TRAF6_Seq_2_1, TRAF6_Seq_3_1 and TRAF6_Seq_4_1). The RMSD was calculated in PyMOL, where the complex with the lowest RMSD is the most similar to crystal (1LB5, chapter 1.2.1)[22]. The HADDOCK score, percentage of residues at the interface of the complex and RMSD values for each peptide are present in tables 12, 13, 14 and 15 [65][66].

Table 12: Top 10 complexes according to HADDOCK score of TRAF6_Seq_1 and TRAF6_Seq_1_1.

TRAF6_Seq_1				TRAF6_Seq_1_1			
Complexes	Haddock-Score	Interface Residues (%)	RMSD	Complexes	Haddock-Score	Interface Residues (%)	RMSD
Complex_182w	- 75.74	77.78	0.352	Complex_154w	- 80.83	73.33	0.335
Complex_161w	- 73.79	77.78	0.980	Complex_41w	- 75.63	73.33	0.990
Complex_62w	- 72.52	55.56	0.345	Complex_55w	- 74.41	80.00	0.357
Complex_159w	- 69.71	77.78	0.996	Complex_148w	- 74.38	73.33	0.340
Complex_51w	- 67.89	77.78	1.111	Complex_7w	- 74.18	80.00	1.124
Complex_6w	- 67.86	77.78	1.025	Complex_183w	- 72.57	73.33	1.336
Complex_110w	- 67.69	77.78	1.006	Complex_175w	- 72.49	86.67	0.359
Complex_1w	- 67.55	77.78	1.018	Complex_177w	- 71.04	80.00	0.401
Complex_123w	- 67.53	66.67	0.776	Complex_72w	- 70.89	80.00	0.836
Complex_28w	- 66.87	77.78	0.939	Complex_153w	- 69.56	73.33	0.357

Table 13: Top 10 complexes according to HADDOCK score of TRAF6_Seq_2 and TRAF6_Seq_2_1

TRAF6_Seq_2				TRAF6_Seq_2_1			
Complexes	Haddock-Score	Interface Residues (%)	RMSD	Complexes	Haddock-Score	Interface Residues (%)	RMSD
Complex_81w	- 63.03	55.56	0.908	Complex_9w	- 81.91	80.00	0.985
Complex_116w	- 62.92	66.67	0.335	Complex_195w	- 73.72	80.00	0.976
Complex_173w	- 62.61	66.67	0.323	Complex_106w	- 67.82	73.33	1.040
Complex_172w	- 62.24	66.67	0.338	Complex_33w	- 67.77	73.33	0.990
Complex_195w	- 60.99	66.67	0.353	Complex_160w	- 65.87	80.00	0.987
Complex_163w	- 60.15	66.67	0.350	Complex_74w	- 64.61	66.67	1.142
Complex_42w	- 60.13	55.56	0.931	Complex_199w	- 64.38	80.00	1.157
Complex_118w	- 60.09	77.78	0.363	Complex_3w	- 63.26	73.33	1.011
Complex_153w	- 59.94	66.67	1.011	Complex_107w	- 63.20	80.00	1.032
Complex_37w	- 59.79	55.56	0.979	Complex_6w	- 62.36	80.00	0.965

Table 14: Top 10 complexes according to HADDOCK score of TRAF6_Seq_3 and TRAF6_Seq_3_1.

TRAF6_Seq_3				TRAF6_Seq_3_1			
Complexes	Haddock-Score	Interface Residues (%)	RMSD	Complexes	Haddock-Score	Interface Residues (%)	RMSD
Complex_38w	- 89.76	77.78	0.384	Complex_109w	- 93.85	70.97	2.466
Complex_52w	- 83.22	77.78	1.624	Complex_184w	- 93.61	70.97	2.492
Complex_22w	- 80.85	66.67	0.341	Complex_50w	- 83.02	51.61	0.417
Complex_56w	- 76.03	77.78	0.331	Complex_171w	- 80.95	61.29	0.372
Complex_69w	- 74.10	77.78	0.977	Complex_10w	- 79.88	58.06	1.637
Complex_10w	- 72.83	77.78	0.371	Complex_14w	- 79.42	67.74	0.949
Complex_89w	- 72.44	77.78	1.568	Complex_190w	- 79.22	58.06	1.679
Complex_94w	- 70.72	77.78	2.118	Complex_35w	- 78.60	67.74	0.959
Complex_9w	- 70.20	77.78	0.394	Complex_7w	- 78.35	64.52	0.951
Complex_84w	- 70.11	77.78	0.343	Complex_148w	- 77.84	64.52	1.680

Table 15: Top 10 complexes according to the HADDOCK score of TRAF6_Seq_4 and TRAF6_Seq_4_1.

TRAF6_Seq_4				TRAF6_Seq_4_1			
Complexes	Haddock-Score	Interface Residues (%)	RMSD	Complexes	Haddock-Score	Interface Residues (%)	RMSD
Complex_6w	- 70.31	66.67	0.295	Complex_26w	- 80.17	67.74	1.511
Complex_11w	- 69.81	66.67	0.336	Complex_37w	- 79.87	67.74	1.917
Complex_139w	- 69.59	66.67	0.361	Complex_48w	- 79.69	67.74	1.895
Complex_84w	- 69.49	66.67	0.323	Complex_50w	- 77.87	70.97	0.433
Complex_95w	- 69.08	66.67	0.345	Complex_63w	- 77.25	67.74	1.487
Complex_76w	- 68.23	66.67	0.332	Complex_170w	- 74.69	74.19	0.418
Complex_83w	- 67.81	66.67	0.338	Complex_81w	- 74.47	74.19	1.925
Complex_30w	- 66.62	55.56	0.271	Complex_60w	- 74.33	70.97	1.351
Complex_162w	- 65.99	77.78	0.329	Complex_109w	- 73.86	67.74	0.431
Complex_119w	- 65.50	66.67	0.359	Complex_36w	- 73.00	67.74	1.335


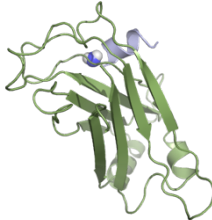
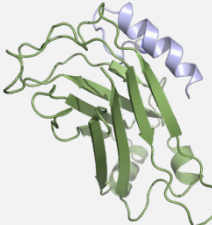
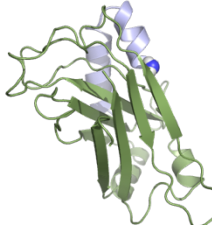
In table 12, for the sequence TRAF6_SEQ_1 the potential candidates are complex_182w, 161w and 62w. Despite having a high percentage of residues at the interface, the remaining complexes have a less negative HADDOCK score and a higher RMSD. In this way, between complex_182w and complex_161w, they have the same percentage of residues at the interface, however, complex_182w has a more negative HADDOCK score and a lower RMSD. Between complex_182w and complex_62w, although the first one has a slightly higher RMSD, it has a higher percentage of residues at the interface and its HADDOCK score is more negative. Thus, and according to what was described, the complex_182w is the chosen one.

For the sequence TRAF6_SEQ_4_1, the complex chosen was complex_50w. Complex_26w has a more negative HADDOCK score, however, its percentage of residues at the interface is lower and its RMSD is higher. There is another potential complex to be chosen, complex_170w, since it has a percentage of residues at the interface slightly higher than the percentage of residues at the interface

of complex_50w. However it presents a difference in its HADDOCK score when compared to complex_26w and its RMSD is similar to the RMSD of complex_50w.

In summary, the selected complexes were: complex154w of TRAF6_seq_1_1, complex_9w of TRAF6_seq_2_1, complex_38 of TRAF6_seq_3 and complex_50w of TRAF6_seq_4_1. However, since most of the selected complexes are relative to the sequence where all peptide residues were considered, for the sake of concordance, the second best complex in TRAF6_seq_3_1, complex_109w, was chosen. In table 16 the chosen complexes are described.

Table 16: Final complex of each peptide.

Peptide	Complex	Final Structure
TRAF6_Seq_1_1	Complex_154w	
TRAF6_Seq_2_1	Complex_9w	
TRAF6_Seq_3_1	Complex_109	
TRAF6_Seq_4_1	Complex_50w	

Through the percentage of residues at the interface, it was observed that residues Arg-392, Phe-410, Phe-471 and Tyr-473 described as experimentally important in the RANK-TRAF6 interaction are considered relevant in the interaction, thus reinforcing their relevance in the interaction. In figure 17 is an example of the complex 109w from TRAF6_Seq_3_1, that shows the four residues mentioned.

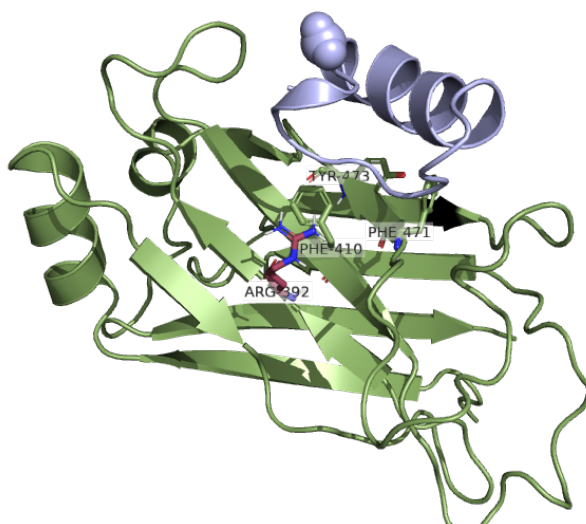


Figure 17: Residues Arg-392, Phe-410, Phe-471 and Tyr-473 are part of the interaction in complex_109w from TRAF6_Seq_3_1.

4.2. Synthesis, Characterization and Purification of Peptides

After synthesis, we proceeded to characterization, evaluation and purification of the peptides. The chromatograms of the synthesized peptides after purification are illustrated in figures 18 and 19.

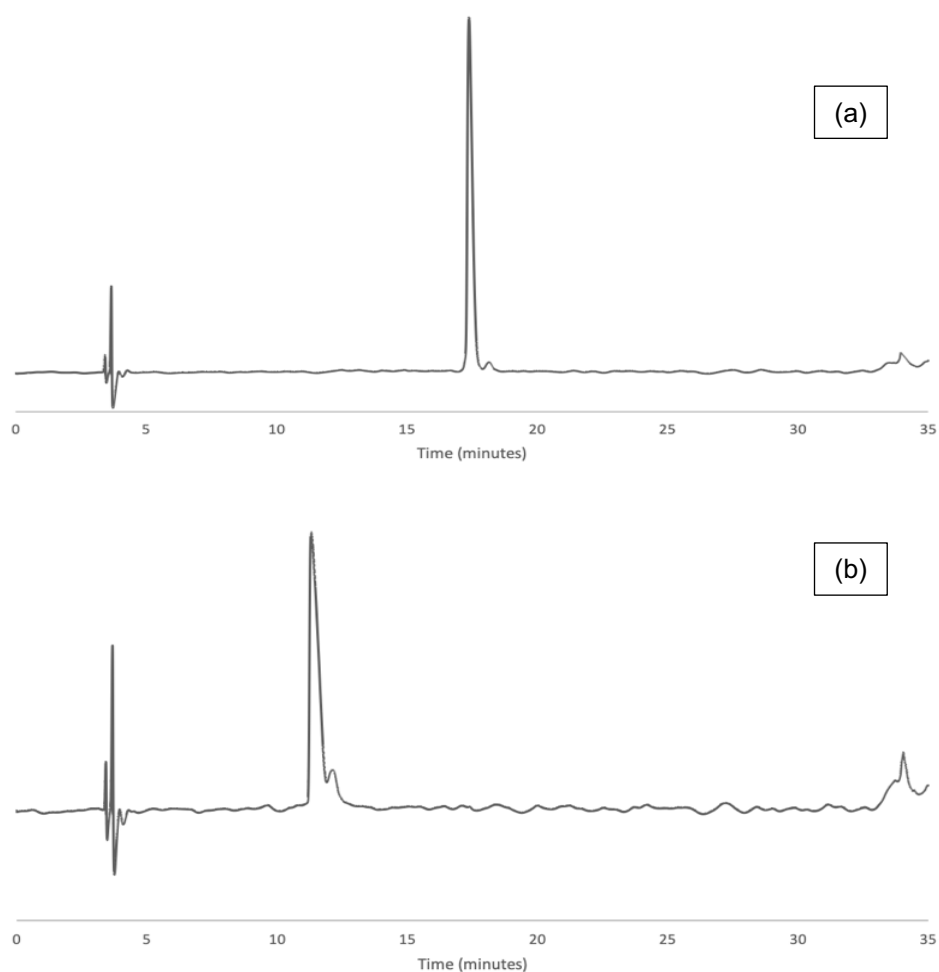


Figure 18: HPLC chromatogram: (a) Peptide 1 and (b) Peptide 2.

Peptides were compared according to the method used. In figure 18, the retention time of peptide 1 is longer than the retention time of peptide 2, 17.40 minutes and 11.35 minutes, respectively. This difference may be related to the fact that the peptide sequences are not totally equal, differing only in three amino acids.

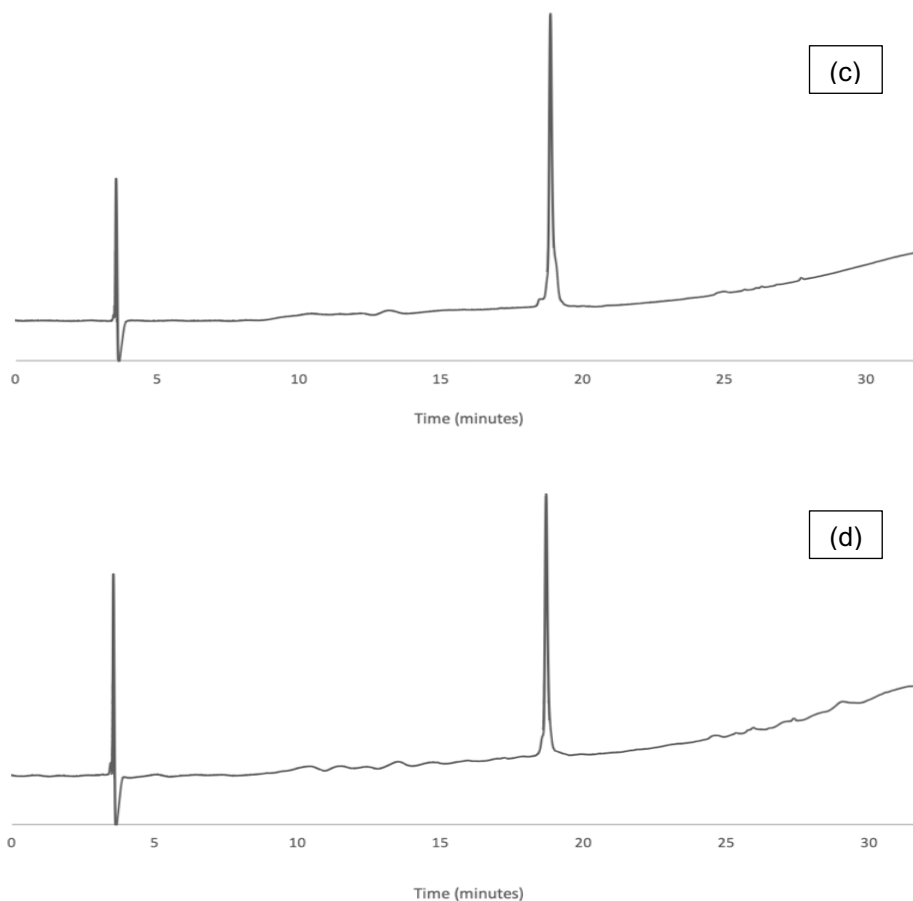


Figure 19: RP-HPLC chromatogram: (c) Peptide 3 and (d) Peptide 4.

Regarding figure 19, the retention times of peptide 3 and 4 are very close, 18.86 minutes and 18.71 minutes, respectively. The proximity of the retention times are not only due to the similarity between the two peptide chains, where the mutated active residues are the same as in peptides 1 and 2, but also possibly due to their peptide sequence being longer (25 amino acids) and these mutated amino acids don't interfere as much.

Table 17 shows the purities of each peptide.

Table 17: Purity of each amino acid after purification.

Peptide	Sequence	Purity (after purification) (%)
Peptide 1	RKIPTED EY-NH ₂	96
Peptide 2	RKIATADEA-NH ₂	99
Peptide 3	AAVALLPAVLLALLAP RKIPTED EY- NH ₂	97
Peptide 4	AAVALLPAVLLALLAP RKIATADEA- NH ₂	>99

After purification, a sample of each peptide was analyzed by mass spectrometry to validate the presence of the peptide. The m/z ratio of each peptide is shown in table 18, and their respective mass spectra are illustrated in figures 20, 21, 22 and 23.

Table 18: m/z ratio of each peptide obtained by ESI-MS.

Peptide	m/z		
	[M+1H ⁺] ⁺	[M+2H ⁺] ²⁺	[M+3H ⁺] ³⁺
Peptide 1	1149,6	575,4	-
Peptide 2	973,6	487,3	-
Peptide 3	-	1324,5	883,5
Peptide 4	-	1236,4	824,6

Through table 18, it is possible to conclude that all peptides were present in the sample, since their molecular ions were found. The obtained mass values can be compared with the expected mass values, through table 9 located in chapter 3.4.3.

The ESI-MS spectrum of peptide 1 after purification is illustrated in figure 20.

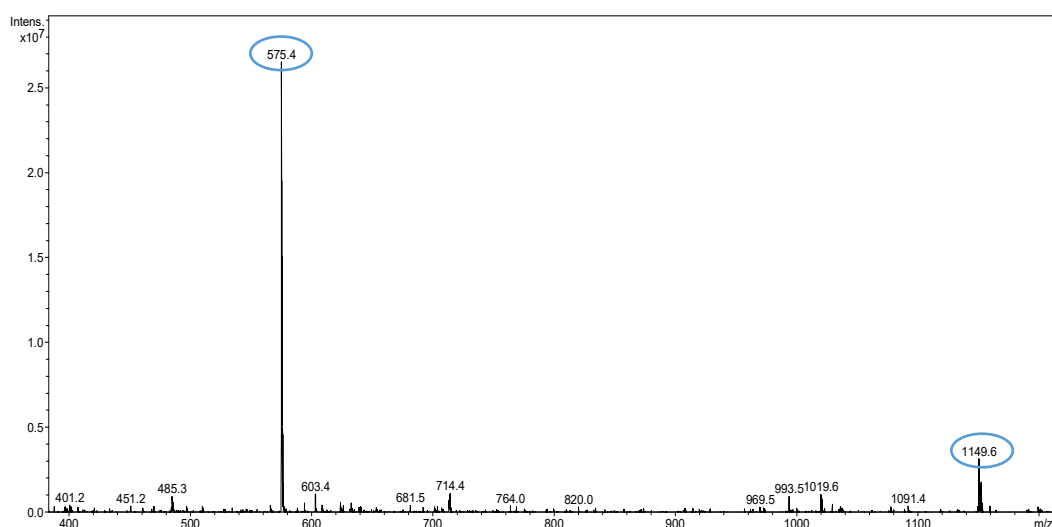


Figure 20: Peptide 1 by ESI-MS after purification.

The ESI-MS spectrum of peptide 2 after purification is illustrated in figure 21.

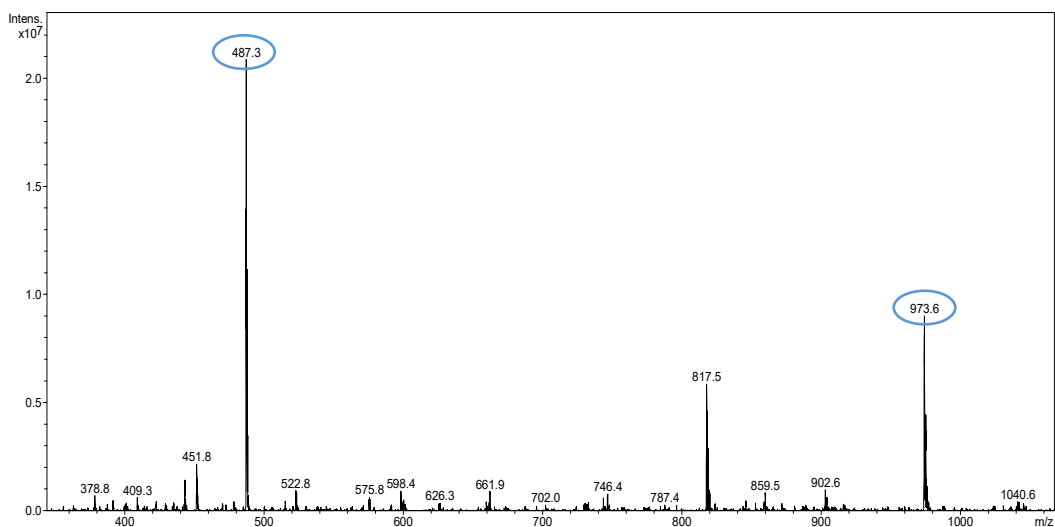


Figure 21: Peptide 2 by ESI-MS after purification.

The ESI-MS spectrum of peptide 3 after purification is illustrated in figure 22.

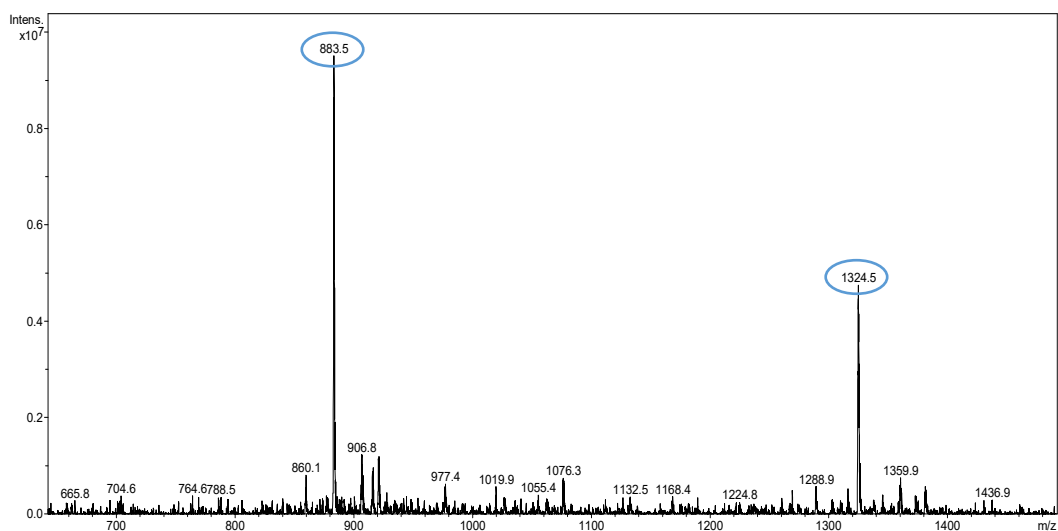


Figure 22: Peptide 3 by ESI-MS after purification.

The ESI-MS spectrum of peptide 4 after purification is illustrated in figure 23.

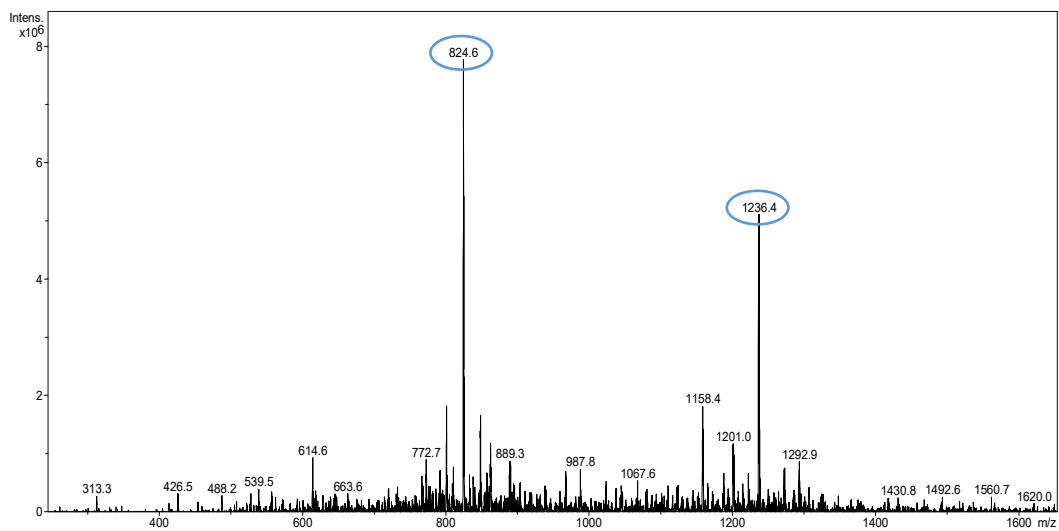


Figure 23: Peptide 4 by ESI-MS after purification.

The mass spectra were obtained in positive mode, positively charged, since in negative mode they were not found.

5. Conclusion and future perspectives

The aim of this work was the Prediction of the 3D structure, selection and synthesis of peptides that potentially inhibit the RANK-TRAF6 interaction.

3D structures of each peptide were predicted, and were subjected to molecular docking in order to predict the best position and orientation of the peptides compared to the target protein (TRAF6). Through the simulations performed, a balance between the parameters HADDOCK score, percentage of residues at the interface and RMSD, allowed the choice of the most stable complex for peptides 1, 2, 3 and 4, being them, complex 154, 9, 109 and 50, respectively. The use of computational techniques and methods will reduce the associated costs when performing *in vitro* and *in vivo* studies in several cell lines.

The synthesis of all peptides was successfully with >95% of purity, making them ready to proceed with *in vitro* studies. Meaning that all steps during synthesis were well executed accompanied by strict control.

This work served as a kickoff for the molecular dynamics' simulation studies of the chosen complexes. Thus, through these computational tools, we can understand at the atomic level how the peptides interact with the target protein. In a future perspective, we intend to evaluate the biological ability of the synthesized peptides to interact with the TRAF6 protein. For this it is necessary to conduct *in vitro* studies assessing their ability to inhibit NF- κ B activation in several cell lines. Finally, and if the results are promising, *in vivo* studies will allow a better understanding of the behaviour of these peptides in the biological system.

6. Bibliography

1. World Health Organization Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on Aug 5, 2021).
2. National Cancer Institute Available online: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer#definition> (accessed on Aug 5, 2021).
3. American Cancer Society Available online: <https://www.cancer.org/treatment/understanding-your-diagnosis/advanced-cancer/bone-metastases.html> (accessed on Aug 5, 2021).
4. Guise, T.A.; Mohammad, K.S.; Clines, G.; Stebbins, E.G.; Wong, D.H.; Higgins, L.S.; Vessella, R.; Corey, E.; Padalecki, S.; Suva, L.; et al. Basic Mechanisms Responsible for Osteolytic and Osteoblastic Bone Metastases. *Am. Assoc. Cancer Res.* **2006**, *12*, 6213–6217, doi:10.1158/1078-0432.CCR-06-1007.
5. Armstrong, A.P.; Tometsko, M.E.; Glaccum, M.; Sutherland, C.L.; Cosman, D.; Dougall, W.C. A RANK / TRAF6-dependent Signal Transduction Pathway Is Essential for Osteoclast Cytoskeletal Organization and Resorptive Function *. *J. Biol. Chem.* **2002**, *277*, 44347–44356, doi:10.1074/jbc.M202009200.
6. American Cancer Society Available online: <https://www.cancer.org/cancer/cancer-basics/what-is-cancer.html> (accessed on Aug 5, 2021).
7. Zhang, X. Interactions between cancer cells and bone microenvironment promote bone metastasis in prostate cancer. *Cancer Commun.* **2019**, *39*, 1–10, doi:10.1186/s40880-019-0425-1.
8. González-Suárez, E.; Sanz-Moreno, A. RANK as a therapeutic target in cancer. *FEBS (Federation Eur. Biochem. Soc.) Journal* **2016**, *283*, 2018–2033, doi:10.1111/febs.13645.
9. Darnay, B.G.; Besse, A.; Poblenz, A.T.; Lamothe, B.; Jacoby, J.J. TRAFs in RANK Signaling. In *TNF Receptor Associated Factors (TRAFs)*; Wu, H., Ed.; Landes Bioscience and Springer Science+Business Media, 2007.
10. Ferreira Dias, P. Instituto superior de ciências da saúde egas moniz, Egas Moniz, 2016.
11. Caparbo, V. de F. Avaliação do papel da osteoclastogênese e ativação dos osteoclastos em pacientes com espondilite anquilosante, Faculdade de Medicina da Universidade de São Paulo, 2018.
12. Glezer, I.; Marcourakis, T.; Avellar, M.; Gorenstein, C.; Scavone, C. O fator de transcrição NF- κ B nos mecanismos moleculares de ação de psicofármacos. *Rev. Bras. Psiquiatria* **2000**, *22*, 26–30.
13. Gohda, J.; Akiyama, T.; Koga, T.; Takayanagi, H.; Tanaka, S.; Inoue, J. RANK-mediated amplification of TRAF6 signaling leads to NFATc1 induction during osteoclastogenesis. *EMBO*

(*European Mol. Biol. Organ. J.* **2005**, 24, 790–799, doi:10.1038/sj.emboj.7600564).

14. Pereira, M. *List of Equations 1, Universidade Nova de Lisboa, 2018.
15. Info Escola Available online: <https://www.infoescola.com/sistema-imunologico/citocinas/> (accessed on Aug 6, 2021).
16. Xu, H.; Chen, F.; Liu, T.; Xu, J.; Li, J.; Jiang, L. Ellagic acid blocks RANKL – RANK interaction and suppresses RANKL-induced osteoclastogenesis by inhibiting RANK signaling pathways. *Chem. Biol. Interact.* **2020**, 331, 109235, doi:10.1016/j.cbi.2020.109235.
17. Mcmanus, S.; Roux, S. The adaptor protein p62 / SQSTM1 in osteoclast signaling pathways. *J. Mol. Signal.* **2012**, 1–8.
18. Yu, J.; Yun, H.; Shin, B.; Kim, Y.; Park, E.; Choi, S.; Yu, J.; Amarasekara, D.S.; Kim, S.; Inoue, J.; et al. Interaction of Tumor Necrosis Factor Receptor-associated Factor 6 (TRAF6) and Vav3 in the Receptor Activator of Nuclear Factor κ B (RANK) Signaling Complex Enhances. *J. Biol. Chem.* **2016**, 291, 20643–20660, doi:10.1074/jbc.M116.728303.
19. Bradley, J.R.; Pober, J.S. Tumor necrosis factor receptor-associated factors (TRAFs). *Nat. Publ. Gr.* **2001**, 1, 6482–6491.
20. Park, H.H. Structure of TRAF Family : Current Understanding of Receptor Recognition. *Front. Immunol.* **2018**, 9, 1–7, doi:10.3389/fimmu.2018.01999.
21. Xu, L.; Li, L.; Shu, H. TRAF7 Potentiates MEKK3-induced AP1 and CHOP Activation and Induces Apoptosis *. *J. Biol. Chem.* **2004**, 279, 17278–17282, doi:10.1074/jbc.C400063200.
22. Ye, H.; Arron, J.R.; Lamothe, B.; Cirilli, M.; Kobayashi, T.; Shevde, N.K.; Segal, D. Distinct molecular mechanism for initiating TRAF6 signalling. *Nat. Publishing Gr.* **2002**, 418, 443–447.
23. Thangudu, R.R.; Bryant, S.H.; Panchenko, A.R.; Madej, T. Modulating Protein – Protein Interactions with Small Molecules : The Importance of Binding Hotspots. *J. Mol. Biol.* **2012**, 415, 443–453, doi:10.1016/j.jmb.2011.12.026.
24. Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.* **2008**, 108, 1225–1244, doi:10.1021/cr040409x.
25. Jones, S.; Thornton, J.M. Review Principles of protein-protein interactions. *Proc. Natl. Acad. Sci.* **1996**, 93, 13–20.
26. Morrow, J.; Zhang, S. Computational Prediction of Hot Spot Residues. *Natl. Inst. Heal.* **2013**, 18, 1255–1265.
27. Stone, T.A.; Deber, C.M. Therapeutic design of peptide modulators of protein-protein interactions in membranes. *BBA (Biochimica Biophys. Acta) - Biomembr.* **2017**, 577–585, doi:10.1016/j.bbamem.2016.08.013.

28. Levy, E.D. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. *J. Mol. Biol.* **2010**, *403*, 660–670, doi:10.1016/j.jmb.2010.09.028.
29. Kastritis, P.L.; Bonvin, A.M.J.J. On the binding affinity of macromolecular interactions: Daring to ask why proteins interact. *J. R. Soc. Interface* **2013**, *10*, doi:10.1098/rsif.2012.0835.
30. Karas, L.J.; Wu, C.; Wu, J.I.; States, U. Computational Organic Chemistry. In *Chemistry Molecular Sciences and Chemical Engineering*; Elsevier Inc., 2019; pp. 1–4 ISBN 9780124095472.
31. Catapult Medicines Discovery Available online: <https://md.catapult.org.uk/the-value-of-computational-chemistry/>.
32. Nature Available online: <https://www.nature.com/subjects/computational-chemistry>.
33. Boyd, R.J. Theoretical and Computational Chemistry. In *Chemistry Molecular Sciences and Chemical Engineering*; Elsevier Inc., 2019; pp. 1–2 ISBN 9780124095472.
34. Márquez-Chamorro, A.E.; Asencio-Cortés, G.; Santiesteban-Toca, C.E.; Aguilar-Ruiz, J.S. Soft computing methods for the prediction of protein tertiary structures: A survey. *Appl. Soft Comput. J.* **2015**, *35*, 398–410, doi:10.1016/j.asoc.2015.06.024.
35. Helene Perrin Available online: <https://medium.com/@HeleneOMICtools/a-guide-for-protein-structure-prediction-methods-and-software-916a2f718cfe> (accessed on Aug 15, 2021).
36. Siebenmorgen, T.; Zacharias, M. Computational prediction of protein – protein binding affinities. *Comput. Mol. Sci.* **2019**, 1–18, doi:10.1002/wcms.1448.
37. DNASTAR Available online: <https://www.dnastar.com/blog/structural-biology/why-structure-prediction-matters/> (accessed on Aug 18, 2021).
38. Kicinski, M. AB INITIO PROTEIN STRUCTURE PREDICTION ALGORITHMS, San José State University, 2011.
39. Barbachan, M.; Lamb, S. Three-Dimensional Protein Structure Prediction: Methods and Computational Strategies. In *Computational Biology and Chemistry*; 2014; pp. 1–39.
40. Deng, H.; Jia, Y.; Zhang, Y. Protein structure prediction. *Int. J. Mod. Phys. B* **2018**, *32*, 1–17, doi:10.7498/aps.65.178701.
41. Bergeron, B. *Bioinformatics Computing*; 2002; ISBN 0131008250.
42. Kumar, P.; Halder, S.; Bansal, M. Biomolecular Structures : Prediction , Identification and Analyses. *Encycl. Bioinforma. Comput. Biol.* **2018**, doi:10.1016/B978-0-12-809633-8.20141-6.
43. PEP-FOLD Available online: <https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/#overview> (accessed on Sep 20, 2021).
44. Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuff, P. PEP-FOLD3 : faster

- denovo structure prediction for linear peptides in solution and in complex. **2016**, 1–6, doi:10.1093/nar/gkw329.
45. Shen, Y.; Maupetit, J.; Derreumaux, P. Improved PEP-FOLD Approach for Peptide and Mini-protein Structure Prediction. *JCTC - J. Chem. Theory Comput.* **2014**.
 46. Thénevet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. **2012**, *40*, 288–293, doi:10.1093/nar/gks419.
 47. PEPstrMOD Available online: <https://webs.iiitd.edu.in/raghava/pepstrmod/> (accessed on Sep 20, 2021).
 48. Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G.P.S. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct* **2015**, 1–19, doi:10.1186/s13062-015-0103-4.
 49. Kaur, H.; Garg, A.; Raghava, G.P.S. PEPstr: A de novo Method for Tertiary Structure Prediction of Small Bio- active Peptides. *Bentham Sci. Publ.* **2007**, 626–631.
 50. Zhang Lab - I-TASSER Available online: <https://zhanggroup.org/I-TASSER/about.html> (accessed on Sep 20, 2021).
 51. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Publ. Gr.* **2015**, *12*, 7–8, doi:10.1038/nmeth.3213.
 52. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Publ. Gr.* **2010**, *5*, 725–738, doi:10.1038/nprot.2010.5.
 53. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, BioMed Cent.* **2008**, *8*, 1–8, doi:10.1186/1471-2105-9-40.
 54. ProSA-Web Help Page Available online: https://prosa.services.came.sbg.ac.at/prosa_help.html.
 55. Sippl, M.J. Recognition of Errors in Three-Dimensional Structures of Proteins. *Wiley Online Libr.* **1993**, *362*, 355–362.
 56. Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Natl. Libr. Med. Natl. Cent. biotechnology Information, PubMed* **2007**, *35*, 407–410, doi:10.1093/nar/gkm290.
 57. Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein – Ligand Interactions in Molecular Docking. *Springer* **2019**, *0*, 1–9, doi:10.1007/s12539-019-00327-w.
 58. Morris, G.M.; Lim-wilby, M. Molecular Docking. In *Methods in Molecular Biology*; Vol. 443, pp. 365–382.
 59. Guedes, I.A.; Magalhães, C.S. De; Dardenne, L.E. Receptor – ligand molecular docking.

- Springer* **2013**, doi:10.1007/s12551-013-0130-2.
60. Agrawal, P.; Singh, H.; Srivastava, H.K.; Singh, S.; Kishore, G.; Raghava, G.P.S. Benchmarking of different molecular docking methods for protein-peptide docking. *BMC Bioinformatics* **2018**, *19*.
 61. Vries, S.J. De; Dijk, M. Van; Bonvin, A.M.J.J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Publ. Gr.* **2010**, *5*, 883–897, doi:10.1038/nprot.2010.32.
 62. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking : a review. *Springer* **2017**, doi:10.1007/s12551-016-0247-1.
 63. Education Available online: <https://www.bonvinlab.org/education/molmod/docking/#a-bite-of-theory> (accessed on Aug 18, 2021).
 64. Koukos, P.I.; Roel-Touris, J.; Ambrosetti, F.; Geng, C.; Schaarschmidt, J.; Trellet, M.E.; Melquiond, A.S.J.; Xue, L.C.; Honorato, R. V.; Moreira, I.; et al. An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. *Proteins Wiley* **2020**, *88*, 1029–1036, doi:10.1002/prot.25869.
 65. Van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.S.J.; Van Dijk, M.; De Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725, doi:10.1016/j.jmb.2015.09.014.
 66. Honorato, R. V.; Koukos, P.I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlató, M.; Giachetti, A.; Rosato, A.; Bonvin, A.M.J.J. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8*, 1–7, doi:10.3389/fmolb.2021.729513.
 67. Trellet, M.; Melquiond, A.S.J.; Bonvin, A.M.J.J. Chapter 10 Information-Driven Modeling of Protein-Peptide Complexes. In *Computational Peptidology, Methods in Molecular Biology*; 2015; Vol. 1268 ISBN 9781493922857.
 68. Karaca, E.; Melquiond, A.S.J.; Vries, S.J. De; Kastritis, P.L.; Bonvin, A.M.J.J. Building Macromolecular Assemblies by Information-driven Docking. *Mol. Cell. Proteomics* **9.8** 2010, 1784–1794.
 69. García-martín, F.; Bayó-Puxan, N.; Cruz, L.; Bohling, J.; Albericio, F. Chlorotriyl Chloride (CTC) Resin as a Reusable Carboxyl. *QSAR Comb. Sci.* **2007**, 1027–1035, doi:10.1002/qsar.200720015.
 70. Petrou, C.; Sarigiannis, Y. Peptide synthesis: Methods, trends, and challenges, chapter 1. In *Peptide Applications in Biomedicine, Biotechnology and Bioengineering*; Elsevier Ltd, 2018; pp. 1–22 ISBN 9780081007365.
 71. Kreutzer, A.G.; Salvesson, P.J.; Yang, H.; Guaglianone, G. *Standard practices for Fmoc-based solid-phase peptide synthesis in the Nowick laboratory*; version 1.;

72. Jaradat, D.M.M. Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation. *Springer* **2017**, doi:10.1007/s00726-017-2516-0.
73. Kaiser, E.; Colestcoot, R.L.; Bossinger, C.D.; Cook, P.I. Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. In *Armour Pharmaceutical Company*; 1970.
74. Suzuki, R.; Konno, H. Stain Protocol for the Detection of N - Terminal Amino Groups during Solid-Phase Peptide Synthesis. *ACS (American Chem. Soc. Publ.* **2020**, doi:10.1021/acs.orglett.0c00445.
75. Mant, C.T.; Chen, Y.; Yan, Z.; Popa, T. V; Kovacs, J.M.; Mills, J.B.; Tripet, B.P.; Hodges, R.S. HPLC Analysis and Purification of Peptides. In *Methods in Molecular Biology*; Vol. 386.
76. Smith, A.I. Peptide Characterization and Purification Using High-Performance Liquid Chromatography. In *Neuropeptide Analogs, Conjugates, and Fragments: Methods in Neurosciences*; ACADEMIC PRESS, INC., 1993; Vol. 13, pp. 91–106.
77. Mant, C.T.; Kondejewski, L.H.; Cachia, P.J.; Monera, O.D.; Hodges, R.S. Analysis of Synthetic Peptides by High-Performance, Liquid Chromatography. In *Methods in Enzymology*; 1997; Vol. 289, pp. 426–469.
78. Mant, C.T.; Hodges, R.S. Analysis of peptides by High-Performance liquid chromatography. In *Methods in Enzymology*; 1996; Vol. 271, pp. 3–50.
79. Hanberry, B.B.; Demarais, S.; Jones, J.C. A Guide to the Analysis and Purification of Proteins and Peptides by Reversed-Phase HPLC. *ISRN Ecol.* **2012**, 2012, 1–7, doi:10.5402/2012/359572.
80. Azevedo da Silva, C.; Bottoli, C.; Collins, C. CROMATOGRAFIA POR INTERAÇÕES HIDROFILICAS (HILIC): ESTADO DA ARTE E APLICAÇÕES. *Quim. Nova* **2016**, 39, 210–220, doi:http://dx.doi.org/10.5935/0100-4042.20160003.
81. Laboratory Technologies Available online: <https://www.dctech.com.br/cuidados-com-colunas-hplc-e-uso-de-colunas-base-de-silica/> (accessed on Aug 19, 2021).
82. Pires, D.A.T. Síntese, Purificação e Determinação Estrutural de Peptídeos como Ferramenta Exploratória de Atividades Biologicamente Relevantes, Universidade de Brasília, 2012.
83. Broad Institute Available online: <https://www.broadinstitute.org/technology-areas/what-mass-spectrometry> (accessed on Aug 19, 2021).
84. Pavia, D.L.; Lampman, G.M.; Kriz, G.S.; Vyvyan, J.R. Espectrometria de Massa. In *Introdução à Espectroscopia*; CENGAGE Learning.
85. Vidal, M.; Cusik, M.; Barabási, A.-L. Interactome Networks and Human Disease. *Cell* **2011**,

doi:10.1016/j.cell.2011.02.016.

86. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. Synthetic therapeutic peptides : science and market. *Drug Discov.* **2010**, *15*, doi:10.1016/j.drudis.2009.10.009.
87. Poblenz, A.T.; Jacoby, J.J.; Singh, S.; Darnay, B.G. Inhibition of RANKL-mediated osteoclast differentiation by selective TRAF6 decoy peptides. *Biochem. Biophys. Res. Commun.* **2007**, *359*, 510–515, doi:10.1016/j.bbrc.2007.05.151.
88. Wolczański, G.; Płóciennik, H.; Lisowski, M.; Stefanowicz, P. A faster solid phase peptide synthesis method using ultrasonic agitation. *Tetrahedron Lett.* **2019**, *60*, 1814–1818, doi:10.1016/j.tetlet.2019.05.069.
89. Bjellqvist, B.; Hughes, G.J.; Hochstrasser, D. The focusing positions of polypeptides in immobilized. *Electrophoresis* **1993**, *14*, 1023–1031.
90. Bjellqvist, B.; Basse, B.; Olsen, E.; Celis, J.E. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **1994**, *15*, 529–539, doi:10.1002/elps.1150150171.
91. Ning, J.; Yu, Z.; Xie, H.; Zhang, H.; Zhuang, G.; Bai, Z.; Yang, S.; Jiang, Y. *ExpASy: The proteomics server for in-depth protein knowledge and analysis.*; 2008; Vol. 24; ISBN 1588293432.