TÉCNICO LISBOA

Design de novos péptidos dirigidos ao recetor de androgénio no cancro da mama triplo negativo: uma abordagem *in silico*

Ana Catarina Ramos Pires Pedrosa

Dissertação para obtenção do Grau de Mestre em

Engenharia Química

Orientador(es): Doutora Maria Cristina das Neves Oliveira
Doutor João Domingos Galamba Correia

Júri

Presidente: Prof.Luísa Margarida Dias Ribeiro de Sousa Martins
Orientador: Doutora Maria Cristina das Neves Oliveira
Vogal: Doutora Rita Lourenço Paiva de Melo
Doutora Irina Sousa Moreira

Junho 2022

O trabalho apresentado nesta tese foi realizado no Grupo de Ciências Radiofarmacêuticas no Centro de Ciências Tecnologias e Nucleares (C2TN) do Campus Tecnológico e Nuclear (CTN), Instituto Superior Técnico-Polo de Loures, sob orientação da Doutora Cristina Oliveira e o Doutor João Galamba Correia.

Agradecimentos

Em primeiro lugar, gostaria de agradecer à Doutora Cristina Oliveira e ao Doutor João Galamba Correia pela orientação científica deste trabalho. Pelo apoio prestado, pela constante disponibilidade em esclarecer as minhas dúvidas, pela motivação e, ainda, por todas as sugestões que foram essenciais para o desenvolvimento e conclusão da presente dissertação.

Os meus sinceros agradecimentos à Doutora Salete Baptista pelo apoio incansável e fundamental para a concretização deste trabalho. Por toda a paciência e disponibilidade que sempre demonstrou, desde o primeiro dia, para esclarecer todas as minhas dúvidas. Por todas as sugestões e revisão científica deste documento. Pela simpatia, amizade, motivação e, acima de tudo, por ter tornado toda a realização deste trabalho muito mais fácil com a sua ajuda e dedicação.

Um agradecimento especial à Doutora Rita Melo, cujo apoio foi indispensável para o desenvolvimento deste trabalho. Agradeço a amizade, a boa disposição e a permanente disponibilidade para me ajudar sempre nas minhas dúvidas.

Um agradecimento à Doutora Irina Moreira que me possibilitou a realização dos estudos de dinâmica molecular.

Quero deixar um agradecimento especial às minhas melhores amigas Ana Franco e Cátia Alcobia por estarem sempre comigo e por serem um pilar fundamental na minha vida.

Agradeço, ainda, a todos os colegas que conheci no C^2TN por terem tornado esta etapa da minha vida muito mais fácil e marcada por bons momentos. Agradeço-lhes pela boa disposição, por todos os bons e divertidos momentos que partilhámos juntos e que espero que continuem daqui para a frente. São pessoas das quais não me irei esquecer.

Por último, um profundo agradecimento à minha família a quem devo tudo. Aos meu pais, por todo o amor, pela eterna ajuda ao longo do meu percurso académico e ao longo de toda a minha vida, por sempre me incentivarem a crescer enquanto pessoa e por sempre das primeiras pessoas a acreditarem nas minhas capacidades. Às minhas irmãs, por tudo aquilo que acrescentam à minha vida, por tornarem todos os momentos melhores e mais divertidos na sua presença, por serem das pessoas mais especiais da minha vida e com quem eu posso sempre contar. Aos meus tios e primos, que apesar de não estarmos sempre juntos, sei que posso sempre contar e estarão sempre lá para mim.

Resumo

O cancro da mama (CM) constitui, a nível mundial, a principal causa de morte por cancro entre as mulheres. Dos diferentes subtipos de CM, o cancro da mama triplo negativo (CMTN) é o que apresenta um comportamento mais agressivo e um pior prognóstico. Cerca de 25%-35% das células tumorais deste subtipo expressam recetores de androgénio (RA). Nestes casos, a existência de péptidos dirigidos a este recetor representa um enorme potencial para fins de diagnóstico e/ou terapêuticos. Neste contexto, foram desenvolvidos novos péptidos dirigidos ao RA com uma boa capacidade e eficiência de internalização ao nível celular.

Os novos péptidos foram desenvolvidos através do *web-server* CellPPD e avaliou-se a sua eficiência de internalização celular através do *web-server* MLCPP. As estruturas 3D dos novos péptidos foram previstas utilizando os *web-servers* PEP-FOLD e o I-TASSER. Para avaliar a afinidade dos novos péptidos para o RA, realizaram-se estudos de *docking* molecular no *web-server* HADDOCK e selecionaram-se os três péptidos mais promissores para estudos de dinâmica molecular (DM). Considerando os contactos polares, a percentagem de resíduos da interface que interagem, o HADDOCK-*score*, o desvio médio quadrático (RMSD), o CPP-*score* e a eficiência de internalização, os três péptidos mais promissores selecionados foram: o Pep19, o Pep23 e o Pep58. As simulações de DM demonstraram que o péptido que apresenta maior potencial é o pep23 tendo em conta a conjugação das diferentes análises: o RMSD, a flutuação quadrática média (RMSF), as ligações de hidrogénio, a energia livre de ligação e a área acessível ao solvente (SASA).

Palavras-chave: Cancro da mama triplo negativo, recetor de androgénio, *docking* molecular, dinâmica molecular, previsão da estrutura 3D

Abstract

Breast cancer (BC) is still the leading cause of cancer death in women worldwide. A specific subtype of BC, known as triple negative breast cancer (TNBC) is frequently associated with a worse prognosis, due mainly to the lack of targeted treatments. Androgen receptor (AR) has been detected in 25%–35% of TNBC carcinoma cells. In these cases, the existence of peptides targeting this receptor represents a huge potential for diagnostic and/or therapeutic purposes. In this context, novel cellular penetrating peptides (CPPs) targeting AR were developed.

The novel CPPs were developed through the CellPPD web-server and their uptake efficiency was predicted using MLCPP. The 3D structures of the peptides were predicted using the PEP-FOLD and I-TASSER web servers. Docking studies of the predicted AR-peptides complexes were performed in the HADDOCK web server to determine the peptides with the highest affinity to AR. The three most promising AR-binding peptides from docking studies were selected for molecular dynamics (MD) simulations. Based on the polar interactions, the percentage of the interface residues that interact, the HADDOCK-score, the root mean square deviation (RMSD), the CPP-score, and the uptake efficiency, the three most promising peptides selected were: Pep-19, Pep-23 and Pep-58. The MD simulations revealed that, of these three peptides, the one that shows the most potential is pep23 considering the conjugation of the different analyses: the RMSD, the root mean square fluctuation (RMSF), the hydrogen bonds, the binding free energy, and the solvent accessible area (SASA).

Keywords: Triple negative breast cancer, androgen receptor, molecular docking, molecular dynamics, 3D structure prediction

Índice

Agradecimentos v
Resumo vii
Abstract ix
Lista de Tabelas xiii
Lista de Figuras xv
Abreviaturas xvii
Nomenclatura de aminoácidos xx

1 Introdução **1**
1.1 Motivação 1
1.2 Enquadramento do trabalho 2
1.3 Objetivos 3
1.4 Organização da dissertação 3

2 Estado da Arte **5**
2.1 Cancro da mama 5
2.1.1 Cancro da mama triplo negativo 6
2.2 Recetor de Androgénio 7
2.2.1 Características estruturais 9
2.2.2 Motivo estrutural dos co-reguladores 11
2.3 Péptidos: principais aplicações biomédicas 13
2.4 Métodos *in silico* 15
2.4.1 Dinâmica molecular 16
2.4.2 *Docking* molecular 17
2.4.3 Previsão da estrutura de proteínas 18
2.4.4 *De novo design* 19

3 Métodos computacionais **21**
3.1 Estrutura primária dos péptidos iniciais 21
3.2 Previsão e *design* dos novos péptidos 22
3.3 Previsão da estrutura 3D 23
3.4 *Docking* molecular 24

3.5 Dinâmica molecular 25

4 Resultados e discussão **29**

4.1 Estrutura primária dos novos péptidos 29
4.2 Estrutura 3D dos péptidos 31
4.3 *Docking* dos péptidos iniciais 32
4.4 *Docking* dos novos péptidos 33
4.4.1 Péptidos com motivo FxxFY 34
4.4.2 Péptidos com motivo FxxLY 37
4.4.3 Péptidos com motivo FxxLF 40
4.4.4 Resultados finais 43
4.5 Dinâmica molecular 47
4.5.1 RMSD e trajetória 47
4.5.2 RMSF 49
4.5.3 Área da interface acessível ao solvente 52
4.5.4 Ligações de hidrogénio 53
4.5.5 Energia livre de ligação 55

5 Conclusões e perspetivas futuras **57**

Bibliografia **59**

A Estrutura 3D dos péptidos **67**

A.1 Validação da estrutura dos péptidos (ProSa-*web*) 67
A.2 Estruturas 3D finais dos péptidos (I-TASSER e PEP-FOLD) 71

B Dinâmica Molecular **74**

B.1 RMSD dos complexos e dos péptidos 74
B.2 SASA 75
B.3 Energia livre de ligação 75

Lista de Tabelas

2.1 Classificação dos principais subtipos de cancro da mama. *Adaptado de* [30]. 6

3.1 Sequência dos péptidos Bud31, C320, 3-18 e dos péptidos iniciais Pep-A, Pep-B e Pep-C, motivo estrutural e código PDB. 21

3.2 Parâmetros de *docking* utilizados no *web-server* HADDOCK. 25

4.1 Sequências dos péptidos iniciais e dos novos péptidos previstos pelo CellPPD, probabilidade associada à previsão de serem bons CPPs e à sua *uptake efficiency* (eficiência de internalização celular). Os aminoácidos que sofreram mutações são representados a vermelho. 30

4.2 Resultados obtidos do *docking* molecular dos péptidos iniciais Pep-A, Pep-B e Pep-C. . . 33

4.3 Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do *docking* dos péptidos com motivo FxxFY pertencentes à família 1. 35

4.4 Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do *docking* dos péptidos com motivo FxxLY pertencentes à família 2. 38

4.5 Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do HADDOCK dos péptidos com motivo FxxLF pertencentes à família 3. 41

4.6 Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 1: % de resíduos do motivo estrutural do péptido que interagem com o domínio LBD do RA, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos. 44

4.7 Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 2: % de resíduos do motivo estrutural do péptido que interagem, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos. 45

4.8 Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 3: % de resíduos do motivo estrutural do péptido que interagem, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos. 46
4.9 Ligações de hidrogénio estabelecidas entre o domínio LBD do RA e o pep19 (CRAFRRLFYPS). 53
4.10 Ligações de hidrogénio estabelecidas entre o o domínio LBD do RA e o pep23 (DRRFRRLYTRS). 53
4.11 Ligações de hidrogénio estabelecidas entre o domínio LBD do RA e o pep58 (RRKFRLFYKRK). 53

Lista de Figuras

1.1 Conjunto de etapas desenvolvidas no presente trabalho. 3

2.1 Agonistas e antagonistas (esteróides e não esteróides) do RA. 8
2.2 Representação do mecanismo de transcrição do RA. Abreviaturas: Hsp-Proteínas *heat-shock*; RA-Recetor de androgénio; DHT- Dihidrotestosterona. *Adaptado de* [5]. 9
2.3 Estrutura esquemática dos três principais domínios funcionais do RA. *Adaptado de* [17]. 10
2.4 Domínio de ligação do ligando (LBD) do RA. (A) Estrutura cristalina do domínio LBD do RA (PDB ID: 4OEZ). (B) Região AF-2 representada a azul (H12), rosa (H4/5) e a vermelho (H3). A cor de laranja está representado o péptido co-regulador. 11
2.5 Estrutura cristalina do domínio LBD do RA a interagir com o péptido (azul claro) com motivo FxxLF (PDB ID: 1T7R). (A) Região AF-2 (local de ligação hidrofóbico) formada pelas hélices-α H3, H4, H5 e H12. (B) Interface de interação entre o péptido com motivo FxxLF (azul claro) e os resíduos-chave para a interação K720 e E897 12
2.6 Representação dos contactos polares estabelecidos entre o motivo dos péptidos (branco) e os resíduos-chave K720, E897 e Q733 (azul) do RA. A) Interações polares entre a K720 e o E897 e o motivo FxxLF do péptido (PDB ID: 1T7R). B) Interações polares entre a K720, o E897 e a Q733 e o motivo FxxLY do péptido (PDB ID: 4OEZ). 13
2.7 Exemplos das aplicações de péptidos na área da biomedicina. (1) Utilização de péptidos em terapias-alvo. (2) Exploração de péptidos marcados com radionuclídeos, enquanto potenciais agentes de diagnóstico e/ou terapia. 14

3.1 Conjunto de etapas desenvolvidas no *web-server* CellPPD. 22
3.2 Exemplo ilustrativo do Z-*score* (A) e da visualização molecular dos resíduos de maior/menor energia da cadeia peptídica (B) no *web-server* ProSA-*web*. 24
3.3 Configuração da caixa dodecaédrica (d=1nm) incluindo um dos sistemas estudados solvatado em água (representada a vermelho) e com iões cloreto (representados a roxo). . 26

4.1 RMSD dos três complexos calculado em relação aos átomos Cα das estruturas. (a) RMSD do complexo domínio LBD-pep19; (b) RMSD do complexo domínio LBD-pep23; (c) RMSD do complexo domínio LBD do RA e pep58. 48
4.2 RMSD dos péptidos calculado em relação aos átomos Cα das estruturas. (a) RMSD do pep19; (b) RMSD do pep23; (c) RMSD do pep58. 48

4.3 RMSF dos complexos (recetor-péptido) calculado em relação aos átomos Cα. Resíduos do recetor: 670-918; Resíduos do péptido: 919-929 (a) RMSF do recetor-pep19; (b) RMSF do recetor-pep23; (c) RMSF do recetor-pep58. 50
4.4 Média das três réplicas do RMSF dos péptidos. (a) Média das três réplicas do RMSF pep19; (b) Média das três réplicas RMSF do pep23; (c) Média das três réplicas RMSF do pep58. 51
4.5 Área da interface dos péptidos. (a) Área da interface do pep19; (b) Área da interface do pep23; (c) Área da interface do pep58. 52
4.6 Energia livre de ligação (kJ/mol) do sistema do pep19, do pep23 e do pep58. 56

5.1 Conjunto de etapas desenvolvidas no presente trabalho e perspetivas futuras. 58

A.1 Z-*score* da estrutura 3D e visualização molecular da energia dos resíduos dos péptidos iniciais (Pep-A, Pep-B e Pep-C). 67
A.2 Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 1. 68
A.3 Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 2. 69
A.4 Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 3. 70
A.5 Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após a previsão do estado de protonação pelo PDB2PQR dos péptidos da família 1. 71
A.6 Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após a previsão do estado de protonação pelo PDB2PQR dos péptidos da família 2. 72
A.7 Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após previsão do estado de protonação pelo PDB2PQR dos péptidos da família 3. 73

B.1 Representação do RMSD dos complexos, calculado em relação aos átomos Cα. (a) RMSD do complexo recetor-pep19; (b) RMSD do complexo recetor-pep23; (c) RMSD do complexo recetor-pep58. 74
B.2 Representação do RMSD dos péptidos, calculado em relação aos átomos Cα.(a) RMSD do pep19; (b) RMSD do pep23; (c) RMSD do pep58. 74
B.3 Área da interface dos péptidos. (a) Área da interface do pep19; (b) Área da interface do pep23; (c) Área da interface do pep58. 75
B.4 Energia livre de ligação (kJ/mol) do sistemas. (a) Energia livre de ligação do sistema do pep19; (b) Energia livre de ligação do sistema do pep23; (b) Energia livre de ligação do sistema do pep58. 75

Abreviaturas

AF-1 Função de ativação 1, do Inglês *Activation function 1*

AF-2 Função de ativação 2, do Inglês *Activation function 2*

AIR Restrições de interações ambíguas, do inglês *Ambiguous Interaction Restraints*

CMTN Cancro da mama triplo negativo

CM Cancro da mama

DBD Domínio de ligação ao DNA, do Inglês *DNA-binding domain*

DHT Dihidrotestosterona

DM Dinâmica molecular

DNA Ácido desoxirribonucleico, do inglês *Deoxyribonucleic acid*

FDA Agência norte-americana do medicamento, do Inglês *Food and Drug Administration*

HER2 Fator do crescimento epidérmico humano do tipo 2, do Inglês *Human epidermal growth factor receptor-type 2*

HTS *Screening* de alto rendimento, do Inglês *High throughput screening*

I&D Investigação e desenvolvimento

IHQ Imunohistoquímica

Ki67 Índice de proliferação celular

LBBD *Design* de fármacos baseado no ligando, do Inglês Ligand-based drug design

LBD Domínio de ligação do ligando, do Inglês *Ligand-binding domain*

NTD Domínio N-terminal, do Inglês *N-terminal domain*

OS Sobrevivência global, do inglês *Overall survival*

PDB Banco de dados de proteínas, do Inglês *Protein Data Bank*

QSAR Relação quantitativa estrutura/atividade, do Inglês *Quantitative structure-activity relationship*

RA Recetor de Androgénio

RE Recetor de Estrogénio

RFS Sobrevivência livre de recorrência, do inglês *Recurrence free survival*

RMN Ressonância magnética nuclear

RMSD Desvio médio quadrático, do inglês *Root mean square deviation*

RMSF Flutuação quadrática média, do inglês *Root mean square fluctuation*

RP Recetor de Progesterona

SASA Área da interface acessível ao solvente, do inglês *solvent accessible surface area*

SBBD *Design* de fármacos baseado no recetor, do Inglês *Structure-based drug design*

Nomenclatura de aminoácidos

Aminoácido	Uma letra
Alanina	A
Arginina	R
Asparagina	N
Ácido Aspártico	D
Cisteína	C
Glutamina	Q
Ácido Glutâmico	E
Glicina	G
Histidina	H
Isoleucina	I
Leucina	L
Lisina	K
Metionina	M
Fenilalanina	F
Prolina	P
Serina	S
Treonina	T
Triptofano	W
Tirosina	Y
Valina	V

Capítulo 1

Introdução

No presente capítulo será abordada a motivação, o enquadramento e os objetivos do trabalho desenvolvido. Para além disto, será, ainda, apresentada uma descrição sumária do conteúdo dos capítulos que constituem o corpo principal desta dissertação.

1.1 Motivação

O cancro da mama triplo negativo (CMTN) é um subtipo de cancro da mama caracterizado pela ausência de expressão dos recetores hormonais (recetor de estrogénio e recetor de progesterona) e da proteína HER2 [1–3]. Estes recetores são biomarcadores fundamentais não só para fins de diagnóstico e prognóstico, como também para o desenvolvimento de uma estratégia terapêutica mais eficaz [1, 4, 5]. De facto, a ausência de expressão destes biomarcadores reflete-se em opções de tratamento muito limitadas e pouco eficazes, uma vez que não permite que os doentes possam beneficiar de uma terapia dirigida a um alvo molecular. Por conseguinte, esta condição intrínseca do tumor triplo negativo contribui para que este subtipo esteja, frequentemente, associado a um pior prognóstico [1, 6, 7]. Logo, é de primordial importância a descoberta de um potencial marcador clínico para o carcinoma triplo negativo.

Cerca de 25%-35% das células tumorais do CMTN expressam o recetor de androgénio (RA) [1, 8]. Nestes casos (CMTN/RA$^+$), este recetor constitui um potencial alvo molecular para diagnosticar precocemente a doença e desenvolver terapias dirigidas, que permitam a implementação de tratamentos mais específicos e eficazes [1, 5]. Atualmente, o tratamento aplicado nos casos de CMTN é a quimioterapia convencional neoadjuvante que acaba por ser uma abordagem limitada e pouco específica, o que intensifica a necessidade de desenvolver tratamentos mais promissores [1, 4, 9].

Assim, o âmbito de estudo do presente trabalho contempla o *design* de novas famílias de biomoléculas, mais especificamente de péptidos, com afinidade para o RA e com uma boa capacidade de internalização celular, recorrendo-se a um conjunto de ferramentas computacionais. Estes péptidos, quando marcados com radionuclídeos com características adequadas, poderão desempenhar uma função importante no diagnóstico e/ou terapia de tumores triplo-negativos com sobre-expressão do RA.

1.2 Enquadramento do trabalho

Face à pouca eficácia dos tratamentos existentes, associada à elevada toxicidade e, em muitos casos, ao desenvolvimento dos mecanismos de resistência às terapêuticas convencionais, as estratégias mais inovadoras adotadas atualmente no tratamento do cancro incluem agentes terapêuticos dirigidos a alvos moleculares [10]. Esta abordagem permite tratamentos mais individualizados e eficazes [10, 11]. De facto, a exploração de novas moléculas dirigidas a alvos moleculares específicos, sobre-expressos nas células tumorais, é cada vez mais relevante e, na realidade, o principal desafio passa pela descoberta de novos fármacos menos tóxicos, mais específicos e eficazes do que os atuais agentes de quimioterapia. Nesta perspetiva, a abordagem computacional no processo de investigação e desenvolvimento (I&D) de novos fármacos assume um papel muito importante no combate às doenças oncológicas [12, 13]. O elevado custo e tempo associados a este processo tem conduzido à crescente utilização de ferramentas computacionais, que permitem uma redução dos recursos monetários, dos recursos humanos e do tempo despendido [12, 14]. Logo, as metodologias computacionais constituem um valioso complemento aos métodos experimentais, sendo estes últimos indispensáveis para a validação dos resultados teóricos advindos dos estudos *in silico*.

Uma das estratégias terapêuticas que tem sido proposta para os carcinomas triplo negativos, que expressam o RA, inclui o desenvolvimento de novos agentes dirigidos a esse recetor [1, 15]. A existência de biomoléculas, tais como péptidos, com seletividade e afinidade para o RA constitui um enorme potencial do ponto de vista clínico. Este tipo de biomoléculas podem comportar-se como um vetor biológico, podendo ser utilizados para o transporte de agentes citotóxicos ou radionuclídeos para o alvo molecular. Neste último caso, à semelhança do que tem acontecido para outras neoplasias, a marcação radioativa de péptidos representa uma abordagem promissora para o desenvolvimento de agentes de diagnóstico e/ou terapêuticos para cancros de mama triplo negativos com sobre-expressão do RA.

O RA é uma proteína que pertence à família dos recetores nucleares e que no seu estado inativo se localiza no citoplasma da célula. Este recetor é composto por 3 domínios funcionais principais, nomeadamente: o domínio N-terminal (NTD), o domínio de ligação do DNA (DBD) e o domínio de ligação do ligando (LBD) [15–17]. Este último é o domínio de interesse primário no presente trabalho, uma vez que contempla um local de ligação hidrofóbico onde se ligam os péptidos, que desempenham uma função reguladora da atividade de transcrição (co-reguladores). Tendo em consideração que o alvo é intracelular, os péptidos dirigidos ao RA devem apresentar uma boa capacidade de atravessamento da membrana celular, de forma a aumentar o seu potencial terapêutico. Os péptidos com habilidade para translocarem através das membranas celulares são genericamente designados, em língua inglesa, por *Cell-Penetrating Peptides* (CPPs) [18].

Para o *design* dos novos péptidos foram consideradas sequências peptídicas que demonstraram afinidade para o RA. Como tal, alguns estudos reportados na literatura permitem destacar péptidos com motivos estruturais específicos, tais como FxxFY e FxxL(Y/F) (x=qualquer aminoácido), capazes de conferir uma maior seletividade e afinidade para o RA, particularmente para o domínio de ligação LBD [19–22]. Estes péptidos apresentam um motivo estrutural conhecido por revelar experimentalmente

afinidade para o recetor em estudo, e, como tal, foram o ponto de partida para a criação das novas famílias de péptidos. Neste trabalho, aplicando diferentes ferramentas computacionais, foram previstas novas sequências de CPPs com uma afinidade de ligação para o RA superior às apresentadas pelos péptidos já conhecidos e descritos na literatura.

1.3 Objetivos

O objetivo principal do trabalho proposto é o desenvolvimento de novas famílias de péptidos que apresentem uma maior afinidade de ligação para o RA e simultaneamente uma melhor capacidade de internalização ao nível celular (CPPs), através de uma abordagem computacional multifatorial que inclui: i) modelos de *Machine Learning* para a previsão e *design* dos novos CPPs; ii) previsão da sua estrutura tridimensional (3D); ii) estudos de *docking* para identificação dos péptidos com maior afinidade de ligação para o RA; iii) simulações de dinâmica molecular dos melhores complexos selecionados através de estudos de *docking*, de forma a eleger o péptido mais promissor, isto é, com maior afinidade de ligação para o domínio LBD do recetor. Na figura 1.1 encontra-se um esquema representativo das principais etapas desenvolvidas neste trabalho.

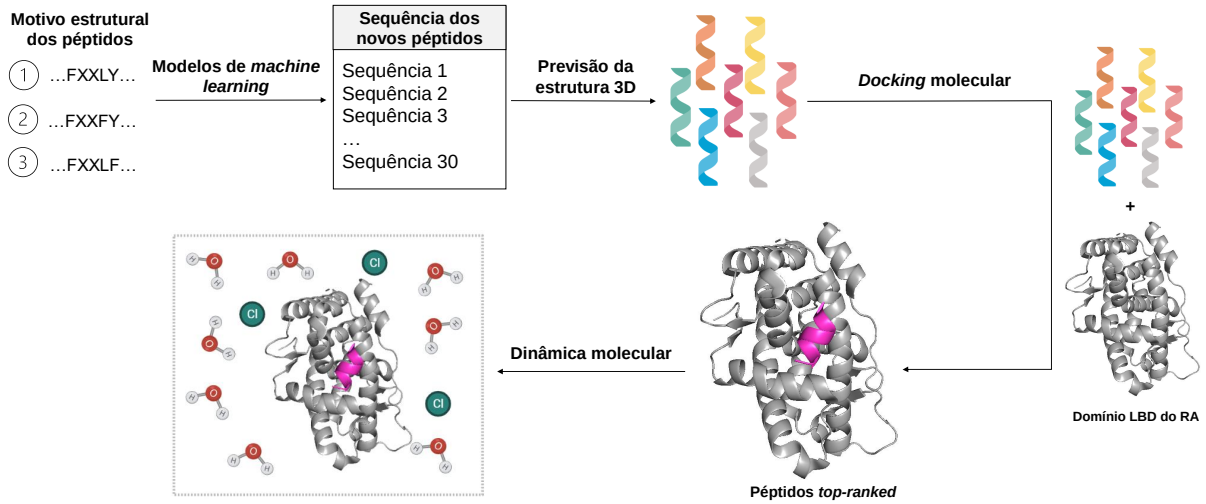


Figura 1.1: Conjunto de etapas desenvolvidas no presente trabalho.

1.4 Organização da dissertação

Esta dissertação encontra-se dividida em 5 capítulos principais. O primeiro capítulo fornece uma breve introdução e contempla os objetivos do trabalho proposto. O segundo capítulo contém uma breve revisão da literatura inserida na temática abordada no presente trabalho. O terceiro capítulo integra a descrição detalhada das metodologias aplicadas. O quarto capítulo inlcui os resultados e a discussão do trabalho realizado. Finalmente, o quinto capítulo apresenta as conclusões gerais e perspetivas futuras deste trabalho.

Capítulo 2

Estado da Arte

2.1 Cancro da mama

Apesar dos avanços clínicos verificados no desenvolvimento de tratamentos mais eficazes para o cancro da mama (CM), este carcinoma apresenta ainda elevados índices de mortalidade [23]. Esta neoplasia é uma das doenças oncológicas mais diagnosticada e a principal causa de morte por cancro entre as mulheres. No continente europeu é o tipo de cancro mais diagnosticado nas mulheres, com cerca de 531 mil novos casos estimados em 2020. À escala mundial, nesse mesmo ano, foram diagnosticados, aproximadamente, 2,3 milhões de novos casos de CM e registaram-se mais de 680 mil vítimas mortais do sexo feminino [23].

O CM é uma neoplasia maligna caracterizada pelo crescimento anómalo e descontrolado das células do tecido mamário, o que conduz à formação de uma massa designada por tumor primário [4, 24, 25]. Esta doença oncológica tem origem nas células do tecido mamário iniciando-se, vulgarmente, nos lóbulos (glândulas produtoras de leite) ou nos ductos (canais que transportam o leite dos lóbulos) [26]. Esta neoplasia pode, ainda, ser acompanhada pela disseminação das células tumorais para outros órgãos (metástases), conduzindo ao desenvolvimento de tumores secundários [4]. A metastização do CM ocorre, preferencialmente, nos seguintes órgãos do corpo humano: nos ossos (30%-60%), no pulmão (21%-32%), no fígado (15%-32%) ou no cérebro (4%-10%) [27].

O carcinoma da mama caracteriza-se sobretudo por uma acentuada heterogeneidade clínica, morfológica e molecular. Neste sentido, a sua classificação é crucial não só para o seu estadiamento e prognóstico da doença, mas também para o desenvolvimento de uma terapia direcionada e, dessa forma, mais específica e capaz de promover resultados clínicos mais promissores [4, 28, 29]. Na prática clínica atual, a classificação mais utilizada tem por base a combinação de duas abordagens principais: a classificação histológica e a classificação molecular [30].

A classificação de tumores mamários de acordo com parâmetros histológicos é, ainda, o método mais utilizado atualmente. A classificação histológica tem por base a observação microscópica e a avaliação do padrão histológico das células tumorais [31]. A diversidade de padrões histológicos observada reflete a heterogeneidade morfológica intrínseca do CM. Os carcinomas da mama são geralmente

tumores epiteliais e podem ser divididos em diversos tipos de acordo com vários fatores relacionados com as células envolvidas, sendo os dois mais comuns o carcinoma ductal e o carcinoma lobular, que se podem apresentar na forma *in situ* (não invasiva) ou na forma invasiva (infiltrante) [30].

A classificação molecular recorre à imunohistoquímica para determinar a presença ou ausência de marcadores tumorais, nomeadamente o recetor de estrogénio (RE), o recetor de progesterona (RP), o recetor do fator de crescimento epidérmico humano do tipo 2 (HER2) e a proteína Ki67 (marcador de proliferação celular) [30]. A avaliação do *status* destes biomarcadores é de primordial importância e tem vindo a tornar-se uma prática clínica comum dada a sua relevante utilidade no diagnóstico, na seleção de terapêuticas seletivas e no prognóstico da doença [32]. Tendo em consideração características moleculares e histológicas, o cancro da mama pode ser dividido em 5 subtipos distintos: o Luminal A, o Luminal B (HER2+ e HER2-), o HER2+ e o triplo negativo [30, 33]. A tabela 2.1 reúne esta classificação e a sua relação com o valor prognóstico.

Tabela 2.1: Classificação dos principais subtipos de cancro da mama. *Adaptado de* [30].

Subtipo	**Luminal A**	**Luminal B**		**HER2+**	**Triplo negativo**
		HER2-	HER2+		
Biomarcadores	RE+ RP+ HER2- Ki67 baixo	RE+ RP- HER2- Ki67 elevado	RE+ RP-/+ HER2+ Ki67 elevado	RE- RP- HER2+ Ki67 elevado	RE- RP- HER2- Ki67 elevado
Frequência (%)	60-70	10-20	13-15	13-15	10-15
Prognóstico	Bom	Intermédio		Mau	Mau
Tipo histológico	Ductal Invasivo Cribriforme Lobular invasivo	Ductal invasivo Lobular invasivo (pleiomórfico) Micropapilar	Ductal invasivo Lobular invasivo (pleomórfico)	Ductal invasivo	Ductal invasivo
Terapia	Endócrina (Tamoxifeno ou Anastrazole)	Endócrina Quimioterapia	Endócrina Quimioterapia Terapia dirigida	Quimioterapia Terapia dirigida (Trastuzumab e Lapatinib)	Quimioterapia

2.1.1 Cancro da mama triplo negativo

De entre os diferentes subtipos referidos de CM, o cancro da mama triplo negativo (CMTN) é considerado o mais agressivo e está, frequentemente, associado a um pior prognóstico [6, 7, 34]. O aparecimento de metástases, particularmente de metástases cerebrais e pulmonares, o elevado risco de recorrência e o crescimento acelerado, tornam este cancro biologicamente mais agressivo e mais resistente à terapia sistémica convencional [5, 6, 34]. Tendo consideração a natureza agressiva inerente a este subtipo, o seu tratamento e deteção precoces são fundamentais para se alcançarem resultados clínicos mais promissores.

O CMTN apresenta opções de tratamento muito limitadas [35, 36]. Uma vez que as células tumorais deste subtipo não expressam nenhuma proteína especifica que possa atuar como alvo terapêutico, não existem presentemente abordagens terapêuticas dirigidas eficazes [1, 5, 7, 36]. Os tratamentos mais usuais, como a terapia hormonal ou terapias direcionadas a um alvo molecular, não são indicadas nestes casos clínicos [1, 5, 36]. Logo, torna-se imperativo a necessidade de investigar novas abordagens

terapêuticas capazes de contribuir para um desfecho clínico mais promissor.

Face à natureza heterogénea deste tumor, o conhecimento dos diferentes tipos de CMTN é fundamental para o desenvolvimento de terapias personalizadas e mais eficazes [5, 6]. Em 2011, *Lehmann et al.* propuseram uma classificação molecular baseada na expressão genética de 587 pacientes diagnosticados com CMTN. Esta classificação assenta em 6 subtipos principais de acordo com o seu padrão molecular: o Luminal com sobre-expressão do recetor de androgénio (LAR), o imunomodulador (IM), o tipo-basal 1 (BL1), o tipo-basal 2 (BL2), o mesenquimal (M) e o mesenquimal tipo-estaminal (MSL) [37]. Cada um destes subtipos apresenta características moleculares distintas e diferentes respostas à quimioterapia convencional [37]. Em 2016, os mesmos autores redefiniram a sua anterior classificação molecular, propondo a passagem de seis subtipos de CMTN (LAR, M, MSL, IM, BL1 e BL2) para quatro subtipos: LAR, M, BL1 e BL2 [38]. Esta classificação molecular revelou-se a divisão mais consensual no meio científico até à data [4].

O subtipo LAR é caracterizado por apresentar elevados níveis de expressão do recetor de androgénio (cerca de 10 vezes superior aos restantes) [6, 37, 39]. O subtipo M expressa genes que envolvem a motilidade e diferenciação celular [4, 6, 37]. O tumor do tipo BL1 está associado a um elevado nível de proliferação celular, confirmado pela observação da expressão genómica e imunohistoquímica da proteína Ki67. Para além disto, neste subtipo, verificam-se, ainda, mutações nos genes envolvidos no mecanismo de reparação do DNA (BRCA1 e BRCA2) [6, 37, 39]. Relativamente ao BL2, verifica-se a elevada presença de recetores responsáveis por auxiliar os processos de crescimento e divisão celular, tais como os recetores EGFR, MET e EPAH2 [37, 39].

Utilizando a classificação assente nos 4 subtipos de CMTN (LAR, M, BL1 e BL2) descrita anteriormente, *Lehman et al.* avaliaram igualmente o risco de desenvolvimento de metástases e os indicadores de sobrevivência com maior relevância clínica, nomeadamente a sobrevivência global (OS) e a sobrevivência livre de doença (RSF) [38]. Este último indicador define o intervalo de tempo em que se verifica a ausência de sinais ou sintomas da doença, enquanto que o indicador OS compreende o período de tempo desde o diagnóstico até à morte do paciente [40, 41]. Os resultados deste estudo demonstraram que a sobrevivência global é significativamente melhor no subtipo BL1, enquanto que nos subtipos LAR e BL2 este indicador apresenta piores resultados. No que diz respeito ao indicador RFS, o subtipo BL1 revela melhores resultados. O subtipo M apresentou um maior aparecimento de metástases no pulmão quando comparado com os restantes subtipos. No subtipo LAR a incidência de metástases nos ossos foi significativamente superior em relação aos restantes [38].

2.2 Recetor de Androgénio

O recetor de androgénio (RA) é uma proteína que se encontra expressa em mais de 70% dos casos de cancro de mama e, no caso particular do CMTN, é detetado em cerca de 25%-35% das células tumorais deste subtipo [1, 42, 43]. Os tumores triplo negativos, que sobre-expressam o RA, são geralmente denominados pela terminologia CMTN/RA^{+} [1].

O RA pertence à família dos recetores nucleares constituída também por outro tipo de recetores,

onde se incluem o recetor de estrogénio, o recetor de progesterona, o recetor de glicocorticóides e o recetor de mineralocorticóides [15, 17, 44]. A expressão deste recetor é observada em diferentes tecidos do corpo humano, nomeadamente nos ossos, cérebro, próstata e no tecido mamário [45]. Quando ativado pelo seu ligando, o RA atua como um fator de transcrição responsável pela regulação da expressão de genes específicos (genes alvo). Dependendo das vias de sinalização envolvidas, este recetor controla a expressão de genes envolvidos em diferentes processos celulares, tais como a proliferação, diferenciação e apoptose celular [15, 46].

Existe um conjunto de pequenas moléculas (ligandos) que interagem com o RA e influenciam a sua atividade de transcrição. Esses ligandos podem ser classificados em agonistas (androgénios) ou antagonistas (anti-androgénios) de acordo com a sua capacidade para ativar ou inibir a transcrição do gene alvo, respetivamente (Figura 2.1) [17, 47]. Os principais agonistas são a testosterona e a dihidrotestosterona (DHT). Esta última resulta da conversão da testosterona, por ação da enzima 5α-*redutase*, na sua forma biologicamente mais ativa [17, 44]. O ligando DHT apresenta o dobro da afinidade para o RA e uma taxa de dissociação cerca de 5 vezes menor comparativamente à testosterona [17]. Os antagonistas do RA inibem a ação dos androgénios (inibição competitiva) e podem ser esteróides, incluindo a oxandrolona e a espironolactona, ou não esteróides, tais como a flutamida e a bicalutamida [17].

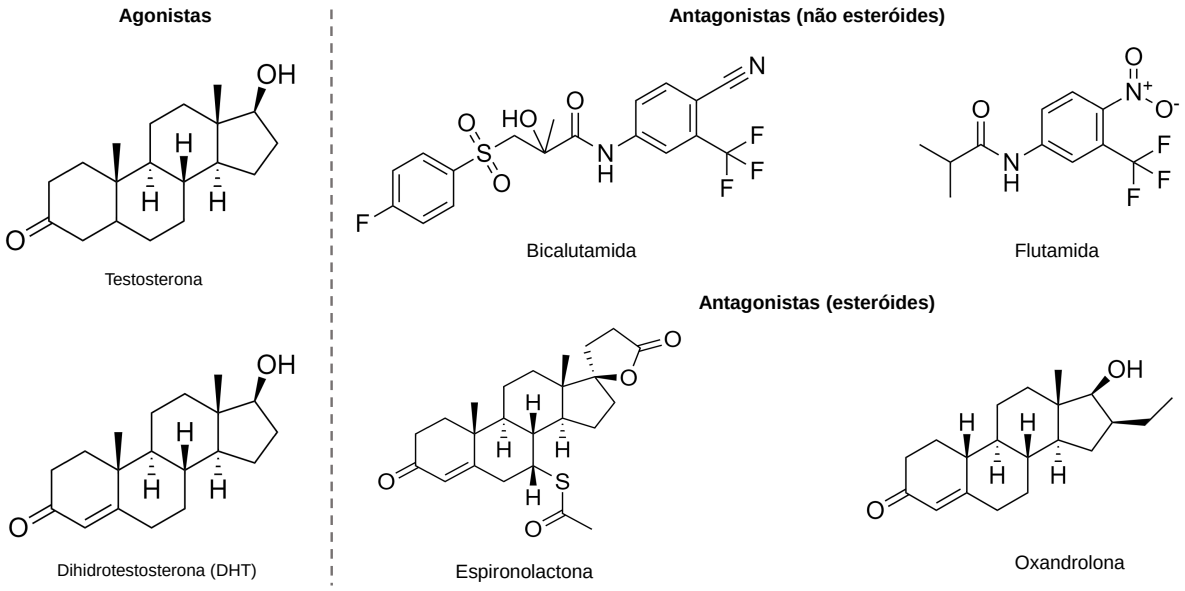


Figura 2.1: Agonistas e antagonistas (esteróides e não esteróides) do RA.

Na ausência do ligando, o RA localiza-se no citoplasma numa conformação inativa ligado a proteínas *heat-shock* (proteínas de choque térmico) responsáveis por estabilizar o recetor na sua forma *apo* (sem ligando) [15–17]. Após a ligação do ligando, o recetor adquire uma conformação ativa e dissocia-se das proteínas *heat-shock* [16, 44]. O complexo ativado dimeriza e, posteriormente, desloca-se para o núcleo celular, onde se dá início à atividade de transcrição dos genes alvo [5, 48]. No núcleo, o complexo de transcrição liga-se a sequências especificas do DNA e induz o recrutamento de co-reguladores, que auxiliam no processo de transcrição [5, 48]. Na figura 2.2 está representado o mecanismo celular

subjacente à atividade de transcrição do RA.

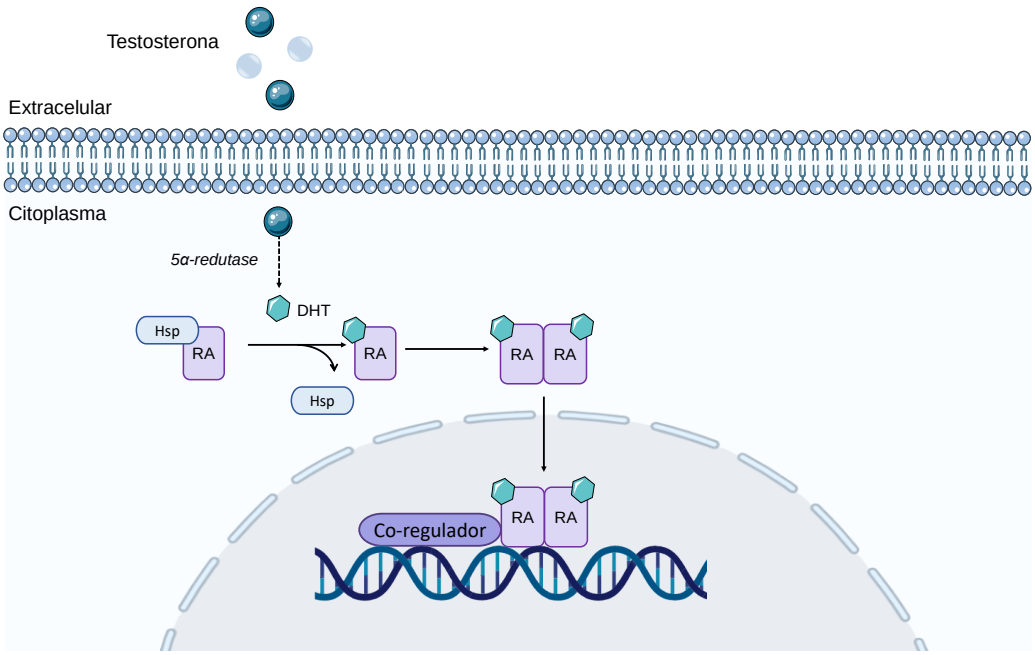


Figura 2.2: Representação do mecanismo de transcrição do RA. Abreviaturas: Hsp-Proteínas *heat-shock*; RA-Recetor de androgénio; DHT- Dihidrotestosterona. *Adaptado de* [5].

2.2.1 Características estruturais

O RA é uma proteína composta por 919 aminoácidos com uma massa molecular de, aproximadamente, 110 kDa [15–17]. O gene que codifica este recetor encontra-se localizado no cromossoma X (Xq11-12) e é constituído por 2757 nucleótidos, incluindo 8 exões intercalados por intrões [15, 17, 49].

Do ponto de vista estrutural, o RA, à semelhança de outros recetores nucleares, é composto por 3 domínios funcionais principais (Figura 2.3), nomeadamente: o domínio N-terminal (NTD, resíduos 1-555), o domínio de ligação do DNA (DBD, resíduos 555-623) e o domínio de ligação do ligando (LBD, resíduos 665-919), presente na região C-terminal [15–17]. Os domínios DBD e LBD estão interligados através de uma região flexível (*hinge*, resíduos 623-665) [17, 44].

O **domínio NTD** destaca-se, dentro da família dos recetores nucleares, por ser o domínio menos conservado (mais variável) quer no tamanho, quer na sua sequência [44, 50, 51]. É constituído pela região AF-1 (resíduos 142-485), responsável pela regulação da atividade de transcrição genética e indispensável para a atividade do recetor [44]. Uma vez que o domínio N-terminal apresenta uma estrutura terciária altamente desordenada, a obtenção da sua estrutura cristalina, com recurso a técnicas de cristalografia de raios-X ou espectroscopia de ressonância magnética nuclear (RMN), tem sido desafiante [17, 50, 52].

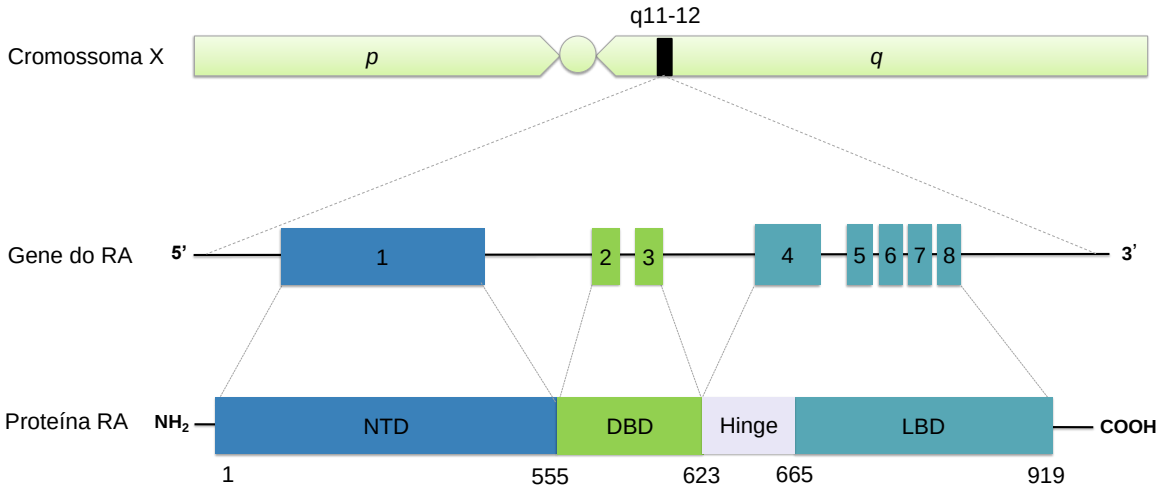


Figura 2.3: Estrutura esquemática dos três principais domínios funcionais do RA. *Adaptado de* [17].

O **domínio DBD** do RA apresenta uma natureza altamente conservada relativamente aos restantes recetores nucleares [17, 44]. Este domínio é constituído por dois "*zinc fingers*" compostos por quatro resíduos de cisteína ligados a um ião zinco central [17]. No primeiro "*zinc finger*" existe uma zona responsável pelo reconhecimento da sequência específica do DNA, designada por *P-box*, enquanto que no segundo "*zinc finger*" encontra-se uma região envolvida no processo de dimerização do recetor, denominada por *D-box* [17, 50].

No local que interliga os domínios DBD e LBD (*hinge*) encontra-se presente um sinal de localização nuclear responsável por direcionar o RA do citoplasma para o núcleo celular [16, 17]. Neste contexto, estudos realizados demonstraram que a ligação do ligando (i.e., do androgénio) ao recetor induz a ativação do sinal de localização nuclear, promovendo o transporte do complexo ligando-recetor para o núcleo da célula [17].

Por fim, o **domínio LBD** é constituído por 3 regiões principais: o local de ligação do ligando (LBP), a região de ativação 2 (AF-2) e a função de ligação 3 (BF-3). A estrutura cristalina deste domínio foi pela primeira vez obtida no ano 2000, tendo sido identificadas 11 hélices-α (H1, H3-H12) e 2 folhas-β antiparalelas (Figura 2.4A) [17, 53]. Ao contrário dos restantes recetores nucleares, a hélice-α H2 não faz parte da estrutura do RA [17]. Assim, no lugar ocupado por esta hélice, existe um ligamento flexível que interliga as hélices H1 e H3 [17].

A região LBP é formada pelas hélices H3, H5 e H11 e é nesta região do recetor que ocorre a ligação do ligando (agonista ou antagonista) [17]. O domínio AF-2 é a região do RA responsável pela interação com co-reguladores, que desempenham uma função importante na regulação do processo de transcrição genética [17, 50]. Estruturalmente, a região AF-2 é formada pelas hélices-α H3, H4, H5 e H12 (Figura 2.4B), sendo esta última fundamental para a atividade de transcrição do RA [17, 54]. A hélice H12 sofre alterações conformacionais desencadeadas pela ligação do agonista ou antagonista ao recetor [54]. Por um lado, após a ligação do agonista, a hélice H12 reposiciona-se de forma a criar um local de ligação hidrofóbico, juntamente com as hélices H4, H5 e H3, onde se irão ligar pequenos

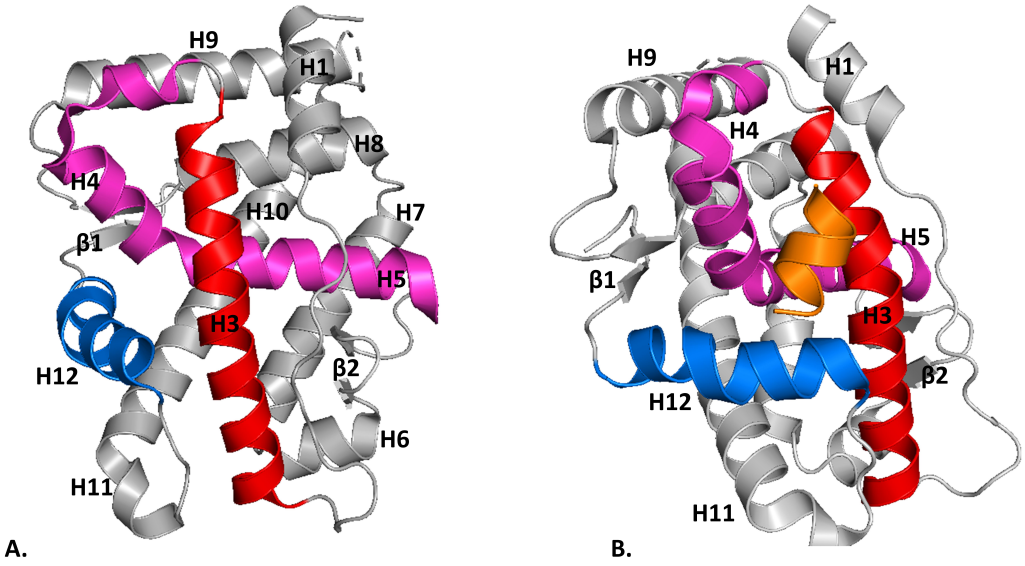


Figura 2.4: Domínio de ligação do ligando (LBD) do RA. (A) Estrutura cristalina do domínio LBD do RA (PDB ID: 4OEZ). (B) Região AF-2 representada a azul (H12), rosa (H4/5) e a vermelho (H3). A cor de laranja está representado o péptido co-regulador.

péptidos co-reguladores (Figura 2.4B). Por outro lado, uma vez que os antagonistas apresentam cadeias laterais mais volumosas (Figura 2.1), o espaço ocupado pelo ligando no LBP é superior. Assim, a expansão do local ativo LBP induz um rearranjo da H12 que impede a ligação de co-reguladores no local de ligação AF-2, inibindo, por conseguinte, a atividade de transcrição do RA [54]. Finalmente, no que concerne ao domínio BF-3, este local desempenha um papel importante na regulação da atividade do domínio AF-2 [17, 50]. A região BF-3 é igualmente formada por um local de ligação hidrofóbico e a ligação de pequenas moléculas a este região do recetor pode comprometer a interação do AF-2 com os co-reguladores [17, 54, 55].

Atualmente, ainda não existem informações robustas sobre a estrutura cristalina completa do RA [53]. No entanto, no presente trabalho, apenas o **domínio LBD** é o domínio de interesse primário, particularmente a região AF-2, uma vez que é neste local do recetor que ocorre a interação com péptidos co-reguladores.

2.2.2 Motivo estrutural dos co-reguladores

De um modo geral, muitos recetores nucleares para exerceram a sua atividade de transcrição ligam-se, preferencialmente, a péptidos co-ativadores que contenham na sua sequência peptídica um motivo estrutural composto por 5 aminoácidos, nomeadamente o motivo ^{1}LxxLL^{5} (x= qualquer outro aminoácido) [17, 50]. No caso particular do RA, estudos de *phage display* demonstraram que péptidos que contenham no seu motivo estrutural (formado por 5 aminoácidos) a combinação dos seguintes aminoácidos, revelam uma elevada capacidade de ligação ao RA, nomeadamente: i) aminoácidos hidrofóbicos de caráter aromático como a fenilalanina (F) na posição +1 do motivo e como a fenilalanina

ou a tirosina (Y) na posição +5 do motivo; ii) aminoácidos hidrofóbicos como a leucina (L) ou a fenilalanina na posição +4 do motivo estrutural [17, 20, 22].

Péptidos co-reguladores baseados no motivo estrutural FxxLF apresentam uma elevada afinidade e especificidade para o RA [17, 56]. Estes péptidos formam uma pequena hélice-α com uma interface hidrofóbica composta pelos grupos hidrofóbicos das posições +1, +4 e +5 do motivo estrutural (^{1}FxxLF^{5}), possibilitando a ligação ao local hidrofóbico da região AF-2 [17, 20, 22]. Para além do motivo estrutural mencionado, existem igualmente outros motivos com uma elevada afinidade para o RA, particularmente os motivos FxxFY e FxxLY [22]. Dos 5 aminoácidos que constituem estes motivos estruturais (^{1}FxxLF^{5}, ^{1}FxxFY^{5}, ^{1}FxxLY^{5}), os aminoácidos presentes nas posições +1 e +5 são fundamentais para a interação com o RA, enquanto que a posição +4 é menos envolvida na interação do complexo (Figura 2.5B) [20, 22]. Além disto, os dois aminoácidos representados pela letra X, por se encontrarem mais expostos ao solvente, não desempenham um papel significativo para a interação com o RA [20].

No local de ligação hidrofóbico da região AF-2 do RA existem dois resíduos de aminoácidos fundamentais (resíduos-chave) para a interação com péptidos co-reguladores, nomeadamente o E897 e a K720, presentes nas hélices H12 e H3, respetivamente [17, 20, 50]. Para péptidos compostos pelo motivo FxxLF, o E897 estabelece interações eletroestáticas com o átomo de azoto do *backbone* da F+1 e com o aminoácido presente na posição -1 do motivo, enquanto que a K720 interage com o grupo carbonilo do *backbone* da F+5 do motivo (Figura 2.5B e Figura 2.6A) [17, 22, 50]. O estabelecimento destes contactos polares permitem estabilizar a interação entre o RA e o péptido co-regulador [22].

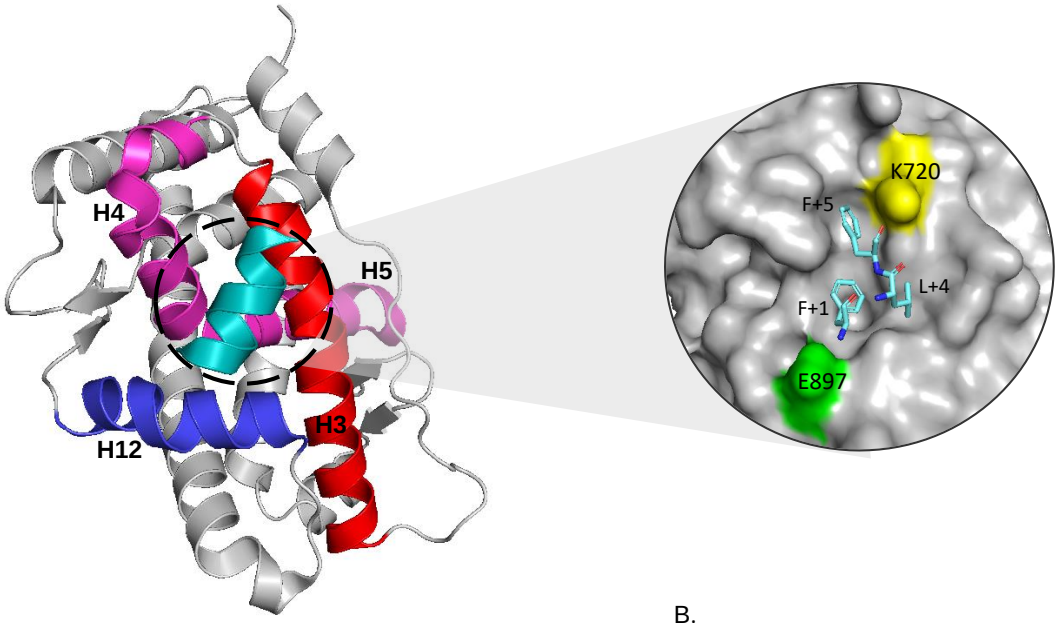


Figura 2.5: Estrutura cristalina do domínio LBD do RA a interagir com o péptido (azul claro) com motivo FxxLF (PDB ID: 1T7R). (A) Região AF-2 (local de ligação hidrofóbico) formada pelas hélices-α H3, H4, H5 e H12. (B) Interface de interação entre o péptido com motivo FxxLF (azul claro) e os resíduos-chave para a interação K720 e E897 .

Os motivos mencionados (FxxLF, FxxFY e FxxLY) apresentam características hidrofóbicas que favorecem e promovem a interação com o RA [17]. No entanto, apesar deste recetor apresentar uma elevada afinidade para péptidos que contenham na sua sequência peptídica estes motivos estruturais, a presença da tirosina na posição +5 do motivo, ao invés da fenilalanina, contribui para o acréscimo da afinidade destes péptidos para o RA [22]. Neste sentido, o resíduo Q733 do recetor assume um papel fundamental para a interação de péptidos com motivo FxxxY com o domínio LBD do RA. A cadeia lateral da Q733 reorganiza-se de forma a expor o átomo de azoto ao grupo hidroxilo, promovendo a formação de uma ligação de hidrogénio adicional entre o grupo OH da tirosina e o átomo de azoto da Q733 do recetor (Figura 2.6B) [22]. Esta característica confere ao RA uma maior afinidade para péptidos constituídos por uma Y na posição +5 do motivo estrutural, como é o caso dos motivos estruturais FxxFY e FxxLY [22, 57].

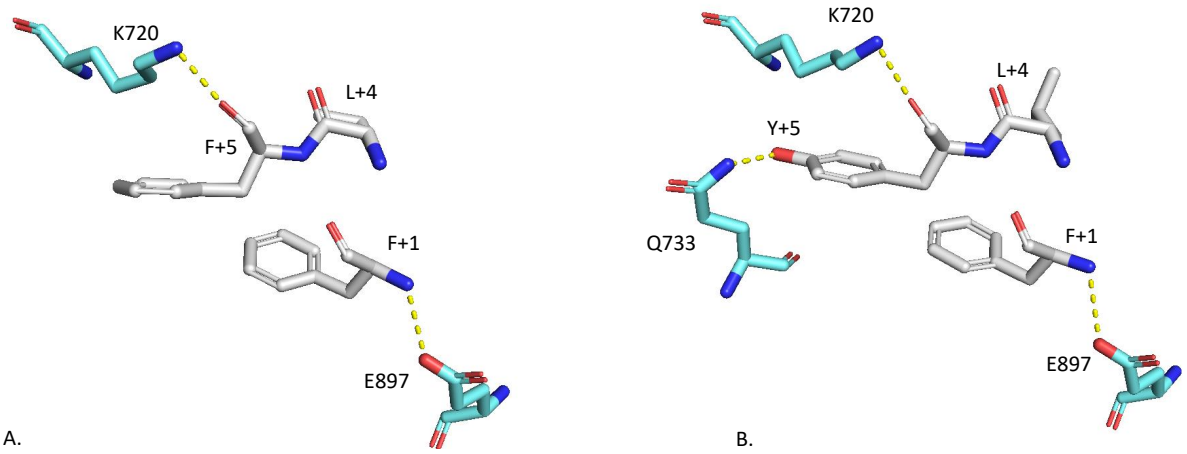


Figura 2.6: Representação dos contactos polares estabelecidos entre o motivo dos péptidos (branco) e os resíduos-chave K720, E897 e Q733 (azul) do RA. A) Interações polares entre a K720 e o E897 e o motivo FxxLF do péptido (PDB ID: 1T7R). B) Interações polares entre a K720, o E897 e a Q733 e o motivo FxxLY do péptido (PDB ID: 4OEZ).

No que diz respeito aos péptidos com o motivo estrutural LXXLL, não se estabelece interação com o resíduo E897, mas apenas com a K720. Este facto, conduz à diminuição dos contactos polares estabelecidos entre o péptido e o local de ligação hidrofóbico da região AF-2. Deste modo, este decréscimo das interações estabelecidas permite justificar a reduzida preferência que o RA apresenta para péptidos com motivo LXXLL [17, 20].

2.3 Péptidos: principais aplicações biomédicas

Ao longo dos anos tem se verificado uma crescente utilização de péptidos na área da biomedicina, enquanto potenciais agentes terapêuticos e de diagnóstico [58]. A utilização de péptidos oferece inúmeras vantagens, incluindo: elevada seletividade e afinidade, biocompatibilidade, baixo custo e reduzida toxicidade [58, 59]. Estas biomoléculas estão envolvidas num alargado espetro de aplicações terapêuticas, destacando-se a sua exploração como: i) péptidos CPPs, podendo ser utilizados como

vetores para o transporte de fármacos e/ou nanopartículas para o meio intracelular; ii) fármacos aplicados em terapias-alvo; iii) promissores agentes de imagem e/ou terapia, quando marcados com radionuclídeos, revelando uma enorme utilidade na área da medicina nuclear [58, 60, 61]. Eis alguns exemplos representativos de péptidos utilizados como fármacos aprovados pela FDA (*Food and Drug Administration*): o Ozempic® (semaglutida) indicado para o tratamento da diabetes, o OctreoScan® utilizado para o diagnóstico de tumores neuroendócrinos e, por último, o Lutathera ®, um radiofármaco indicado para o tratamento de tumores neuroendócrinos gastro-enteropancreáticos não operáveis ou metastáticos [62, 63].

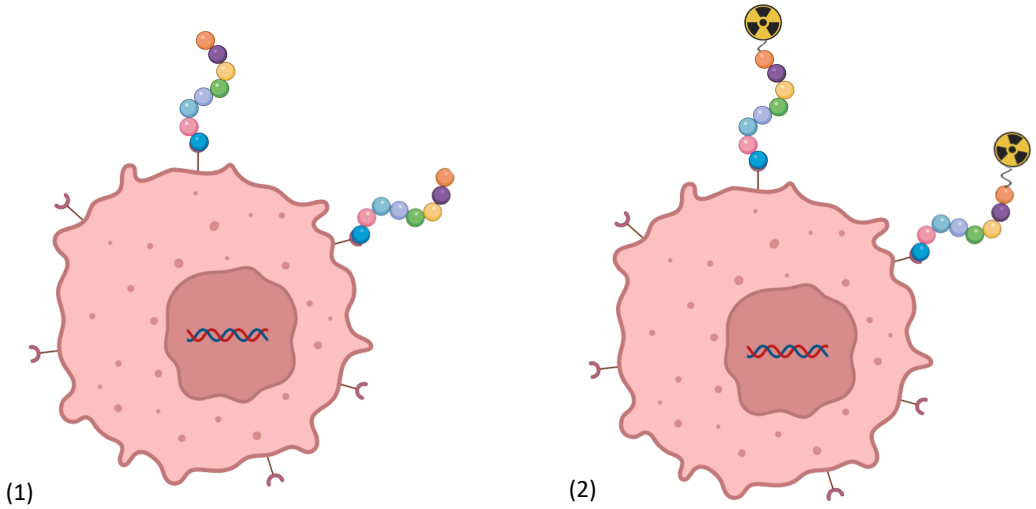


Figura 2.7: Exemplos das aplicações de péptidos na área da biomedicina. (1) Utilização de péptidos em terapias-alvo. (2) Exploração de péptidos marcados com radionuclídeos, enquanto potenciais agentes de diagnóstico e/ou terapia.

A exploração de péptidos CPPs constitui uma promissora abordagem enquanto potenciais agentes de diagnóstico e/ou terapêuticos [64]. Estes péptidos, devido às sua características físico-químicas, conseguem penetrar a membrana das células, apresentando uma boa capacidade de internalização celular. Quando conjugados com fármacos, nanopartículas ou radionuclídeos, os CPPs permitem a entrada destas moléculas para o meio intracelular. A presença de determinadas propriedades físico-químicas contribui para uma melhor capacidade dos péptidos atravessarem a membrana celular. De facto, péptidos anfipáticos e péptidos ricos em aminoácidos carregados positivamente, tais como a arginina, revelam uma boa capacidade de internalização celular devido às fortes interações eletrostáticas estabelecidas com a membrana das células, carregada negativamente [18, 64].

Existem alguns servidores disponíveis *on-line*, que permitem prever e desenhar novos péptidos CPPs. Neste sentido, destaca-se o *web-server* CellPPD (*Designing of Cell Penetrating Peptides*) [65]. Este *web-server* permite não só prever a capacidade de internalização celular dos péptidos, como também criar novos e melhores CPPs do que os péptidos originais [65, 66]. Ao submeter a estrutura primária do péptido, o *web-server* CellPPD origina todas as suas possíveis sequências mutantes, de forma a torná-lo um bom CPP. Além disto, esta ferramenta computacional permite, ainda, prever algumas das propriedades físico-químicas mais relevantes, nomeadamente a hidrofobicidade, a carga e a massa molecular dos péptidos [65].

2.4 Métodos *in silico*

O avanço tecnológico verificado ao longo dos últimos anos tem sido acompanhado pela utilização crescente dos métodos computacionais, destacando-se a sua enorme aplicabilidade na descoberta e desenvolvimento de novos fármacos (CADD, *Computer Aided Drug Design*) [13, 14, 67]. Esta nova abordagem de investigação e desenvolvimento (I&D) constitui uma ferramenta bastante promissora e um valioso complemento aos métodos experimentais, sendo estes últimos indispensáveis para a validação das abordagens teóricas computacionais [14, 67]. Através de simulações computacionais é possível prever as interações que ocorrem entre moléculas, simular e compreender a dinâmica e a natureza de inúmeros processos bioquímicos [13].

A introdução de um novo fármaco no mercado é um processo bastante moroso, envolve uma equipa multidisciplinar e implica um elevado investimento [12]. De facto, este processo demora, em média, cerca de 10-15 anos, com gastos que rondam os 700 milhões-1,6 biliões de euros, sendo que, aproximadamente, 75% dos custos envolvidos correspondem a erros ocorridos aquando deste processo [12, 13]. Neste sentido, a integração de metodologias mais eficientes e economicamente mais rentáveis têm despertado o interesse na indústria farmacêutica, nomeadamente os métodos CADD [12]. Estes métodos computacionais assumem um papel de primordial importância e a sua utilização oferece inúmeras vantagens, entre elas: a diminuição dos erros ocorridos e uma redução dos custos e do tempo despendido associados à I&D de novos fármacos [12–14].

As metodologias CADD são um valioso complemento aos ensaios de *high throughput screening* (HTS) [12]. O HTS é uma abordagem experimental que permite, através de um equipamento automatizado, avaliar um elevado número de compostos e aferir o seu efeito na atividade biológica de um determinado alvo molecular [12, 67, 68]. Dessa forma, é possível identificar os compostos mais promissores (*hits*), de entre um elevado número de possíveis candidatos [68]. Apesar disto, este *screening* experimental envolve custos elevados e requer uma enorme quantidade de alvos e ligandos que, frequentemente, podem não estar disponíveis [68]. Logo, com recurso aos métodos *in silico* é possível otimizar este processo e reduzir o número de compostos testados experimentalmente, conduzindo a um processo de descoberta de novas moléculas terapêuticas mais rápido e eficiente [12, 67].

Os métodos CADD podem ser classificados em dois tipos de abordagens computacionais, nomeadamente: o *design* de fármacos baseado na estrutura do alvo molecular (SBDD, *structure-based drug design*) e o *design* de fármacos baseado no ligando (LBDD, *ligand-based drug design*) [12, 14]. O que diferencia estas duas abordagens é o conhecimento da estrutura 3D do alvo molecular [12].

Por um lado, o **método LBDD** é aplicado quando a estrutura 3D da molécula alvo não é conhecida e a sua previsão não é concretizável [12, 14]. Este método *in silico* parte de um conjunto de ligandos, já conhecidos do alvo molecular de interesse, para identificar as propriedades físico-químicas mais relevantes e necessárias para a interação, permitindo o *design* de novas moléculas dirigidas ao alvo molecular [13, 14]. O método LBDD recorre a duas abordagens principais: modelos farmacofóricos e relações quantitativas entre estrutura e atividade (QSAR).

Os modelos farmacofóricos utilizam as características estruturais dos ligandos, conhecidos de um

determinado alvo e que partilham o mesmo local de ligação, para definir os requisitos estruturais necessários que uma determinada molécula deverá apresentar para ligar-se ao alvo molecular [14, 69]. As relações quantitativas entre estrutura e atividade (QSAR) são utilizadas a fim de encontrar uma correlação entre as características estruturais de um conjunto de ligandos e a sua atividade biológica [14, 67]. Neste sentido, segundo esta abordagem, moléculas com características estruturais idênticas apresentam atividades biológicas semelhantes [12].

Por outro lado, a aplicação do **método SBDD** pressupõe o conhecimento da estrutura 3D do alvo molecular de interesse, providenciada por meio de diferentes técnicas experimentais, nomeadamente cristalografia de raios-X ou espectroscopia de RMN [14]. Caso a estrutura não seja conhecida por nenhuma destas técnicas, é possível através de estudos de homologia e outros métodos explicados seguidamente, prever a estrutura [69]. As estruturas 3D são, frequentemente, encontradas numa base de dados de proteínas, designada por PDB (*Protein Data Bank*) [70]. Nesta base de dados encontram-se depositadas as estruturas 3D de diversas biomoléculas (na sua maioria proteínas) determinadas com recurso a técnicas experimentais como as anteriores mencionadas [70]. Assim, conhecendo a estrutura 3D do alvo é possível prever as interações de diferentes potenciais ligandos com o alvo em estudo, bem como os locais de ligação mais prováveis e, dessa forma, selecionar os ligandos mais promissores [13]. Os principais métodos computacionais utilizados pela abordagem SBDD são: simulações de dinâmica molecular, estudos de *docking* molecular e o *de novo design* [12, 13].

2.4.1 Dinâmica molecular

Com recurso a estudos de dinâmica molecular (DM) é possível simular o comportamento dinâmico de um conjunto de moléculas recorrendo à mecânica newtoniana clássica e à utilização de campos de força [14, 67]. Esta ferramenta computacional permite prever à escala atómica o comportamento de sistemas dinâmicos, tais como proteínas e outras biomoléculas, ao longo do tempo, mimetizando de uma forma simplificada o comportamento *in vivo* das mesmas [71].

A aplicação de um campo de forças permite descrever as interações entre os átomos, isto é, permite estimar as forças que atuam nos átomos do sistema e, dessa forma, é possível a previsão do movimento/trajetória das moléculas, ao longo do tempo, tendo por base a aplicação da equação de movimento de Newton (2ª Lei de Newton), descrita na equação 2.1 [14, 67, 71]. Entre os simuladores mais utilizados para a realização dos estudos de DM estão: CHARMM, AMBER e GROMACS [67, 69].

$$F_i = m_i \times a_i \tag{2.1}$$

Nesta equação, i representa cada átomo do sistema, sendo m_i a massa atómica, a_i a aceleração e F_i a força que atua sobre cada átomo.

As simulações de dinâmica molecular apresentam um vasto conjunto de aplicações, destacando-se: o estudo da estabilidade e flexibilidade das biomoléculas, análise da resposta do sistema a uma determinada perturbação, identificação do local de ligação dos ligandos, compreensão das alterações conformacionais e dos mecanismos de ação de um determinado fármaco [14, 71].

2.4.2 *Docking* molecular

Os estudos de *docking* molecular apresentam como principal objetivo avaliar a afinidade e prever o modo de ligação que uma molécula (ligando) apresenta em relação a um alvo macromolecular (recetor ou enzima) [13, 14, 72]. A utilização desta ferramenta computacional possibilita o estudo das interações intermoleculares e a previsão da orientação e conformação mais promissoras de um ligando no local de ligação do recetor [12, 14]. Assim, entre um elevado conjunto de moléculas, é possível selecionar os ligandos que apresentam maior afinidade para um determinado alvo e que serão sujeitos a estudos *in vitro* [12].

Os métodos de *docking* podem ser divididos em três tipos: o rígido, o semi-flexível e o flexível [14]. O *docking* rígido pressupõe a rigidez das moléculas envolvidas na interação (ligando e o alvo molecular). O *docking* semi-flexível considera a estrutura macromolecular (alvo molecular) um corpo rígido, enquanto que a estrutura do ligando apresenta flexibilidade, sofrendo alterações conformacionais aquando do *docking*. Por último, no *docking* flexível, o ligando e o alvo comportam-se como estruturas flexíveis, permitindo uma previsão do comportamento molecular mais exata [14].

Existem diferentes programas que permitem realizar estudos de *docking* molecular, nomeadamente: o HADDOCK, o AutoDock Vina e o GOLD [69]. Os algoritmos de pesquisa (*search-algorithms*) destes programas analisam recursivamente diferentes conformações do ligando, até que seja alcançada a conformação de menor energia (mais estável) [72]. De forma a que a previsão da conformação seja a mais correta, antes de efetuar os estudos de *docking*, é necessário ter em consideração o estado de protonação dos resíduos de aminoácidos das cadeias laterais. Isto deve-se ao facto de diferentes estados de protonação promoverem diferentes interações entre o ligando e a proteína, o que irá refletir-se em diferentes poses do ligando [12]. A afinidade de ligação entre o alvo e o ligando, cuja pose foi determinada através dos algoritmos de pesquisa do *docking*, é prevista através de uma *scoring function*. As diferentes poses previstas ocupam um *ranking* de acordo com seu *score*. O modo de ligação energeticamente mais favorável será o da pose cujo *score* é o mais reduzido [12].

Em relação ao HADDOCK (*High Ambiguity Driven DOCKing*), este foi o servidor eleito para realizar os estudos de *docking* do presente trabalho. O HADDOCK é um *web-server*, frequentemente, utilizado para realizar *docking* de biomoléculas, mais concretamente de péptidos. Destaca-se, de entre os demais programas, por permitir, aquando da formação do complexo ligando-alvo, alterações conformacionais não só nas cadeias laterais, mas também no *backbone* das moléculas [73]. O funcionamento deste servidor envolve o conhecimento de dados experimentais que orientam o processo de docking [73]. Os dados experimentais são introduzidos sob a forma de resíduos ativos e passivos. Os resíduos ativos são aqueles que se encontram diretamente envolvidos na interação entre as moléculas do complexo, enquanto que os resíduos passivos são os resíduos vizinhos dos resíduos ativos e com uma participação menos importante para a interação [73]. Estes resíduos são os parâmetros de *input* utilizados no HADDOCK e são convertidos posteriormente em restrições de interações ambíguas (AIRs, *Ambiguous Interaction Restraints*) [73].

O procedimento envolvido nos cálculos de *docking*, recorrendo ao *web-server* HADDOCK, inclui 3 estágios no decorrer dos quais, ocorre, sucessivamente, um aumento da flexibilidade do complexo alvo-

ligando. No primeiro estágio (it0), ambas as moléculas envolvidas na interação são consideradas corpos rígidos e, nesse sentido, apenas movimentos de translação e rotação em torno do seu centro de massa são possíveis, de forma a otimizar as interações intermoleculares e promover o encaixe do ligando [74, 75]. No segundo estágio (it1), a cadeia lateral e o *backbone* dos resíduos pertencentes à interface de interação apresentam um comportamento flexível [75]. No último estágio de docking (itw), a adição de solvente (água) e curtas simulações de DM caracterizam este estágio [74, 75]. Após o docking, as estruturas são agrupadas em diferentes *clusters*, de acordo com o seu HADDOCK-*score* [73]. O HADDOCK-*score* resulta do somatório dos seguintes termos energéticos: energia intermolecular de Van der Walls (Evdw), energia eletroestática (Eelec), energia de dessolvatação (Edesolv) e a área não acessível ao solvente (BSA) [73]. As equações do HADDOCK que sustentam as diferentes fases (it0, it1 e itw) envolvidas no protocolo de *docking* são as seguintes:

$$Haddock - score(it0) = 0,01E_{vdw} + 1,0E_{elec} + 1,0E_{desol} + 0,01E_{air} - 0,01BSA \tag{2.2}$$

$$Haddock - score(it1) = 1,0E_{vdw} + 1,0E_{elec} + 1,0E_{desol} + 0,1E_{air} - 0,01BSA \tag{2.3}$$

$$Haddock - score(itw) = 1,0E_{vdw} + 0,2E_{elec} + 1,0E_{desol} + 0,1E_{air} \tag{2.4}$$

Como qualquer outro método, o *docking* molecular apresenta as suas limitações. Os principais desafios enfrentados por este método são os seguintes: os efeitos do solvente, os efeitos dos estados de protonação dos resíduos envolvidos na interação, os efeitos entrópicos e a flexibilidade do recetor [72].

2.4.3 Previsão da estrutura de proteínas

Como referido anteriormente, a estrutura 3D de proteínas e péptidos é determinada, frequentemente, com recurso a técnicas experimentais, tais como espectroscopia de RMN ou cristalografia de raios-X. No entanto, estas técnicas apresentam algumas limitações, entre elas: o elevado tempo associado e/ou os elevados custos envolvidos [76]. Desta forma, estas limitações incentivam a crescente utilização de metodologias computacionais capazes de prever a estrutura 3D de péptidos e proteínas a partir da sua estrutura primária (i.e., da sua sequência de aminoácidos) [77, 78]. Quando a estrutura 3D destas biomoléculas não é conhecida, existem métodos *in silico* (modelos teóricos) através dos quais é possível prever a sua estrutura, nomeadamente: i) modelação por homologia; ii) *threading*; iii) *ab initio folding* [12].

A modelação por homologia visa prever a estrutura de uma sequência alvo de interesse através do conhecimento da estrutura 3D de uma sequência de aminoácidos semelhante (sequência homóloga) à sequência cuja estrutura se pretende determinar [12]. A sequência com estrutura desconhecida é comparada com diversas sequências de proteínas com estruturas conhecidas, depositadas no PDB. Este método considera que, durante a evolução, a estrutura das proteínas mantém-se mais conservada

(menos variável) do que a sequência de aminoácidos. Assim, quando duas sequências apresentam um elevado grau de semelhança entre si, as suas estruturas são igualmente idênticas [79]. O método *threading* baseia-se no alinhamento da sequência de interesse (com estrutura desconhecida) com estruturas modelo existentes em bases de dados [80]. A sequência de interesse é colocada em diferentes posições estruturais da estrutura modelo. Os modelos com as estruturas previstas são avaliados tendo em conta um score [12]. Finalmente, o método *ab initio* é considerado o método de previsão da estrutura de proteínas mais desafiante [12]. Contrariamente aos restantes métodos mencionados, este método recorre apenas à informação da estrutura primária, não sendo utilizados estruturas modelo presentes em bases de dados. Através do conhecimento da sequência de interesse são exploradas diferentes conformações estruturais, até que seja alcançada a estrutura com o mínimo de energia (conformação da estrutura nativa da proteína) [12, 80, 81].

Existem ferramentas computacionais disponíveis *on-line* com o objetivo de determinar a estrutura 3D das sequências de aminoácidos, entre elas: o I-TASSER [78] e o PEP-FOLD [82].

O *web-server* I-TASSER é uma ferramenta computacional que permite prever não só a estrutura 3D de proteínas e péptidos, como também a sua função biológica [83]. A metodologia aplicada neste *web-server* assenta em 4 etapas principais. Na primeira etapa, denominada por *threading*, ocorre a identificação das estruturas modelo, presentes no PDB, através do meta-servidor LOMETS [80, 83]. A qualidade do alinhamento entre os diferentes modelos com a sequência de interesse é avaliada com base no Z-*score*, sendo que um bom alinhamento implica um Z-*score* maior do que 1 [77, 83]. Na segunda etapa, as secções da estrutura das proteínas modelo com melhor alinhamento são utilizadas para construir a estrutura 3D. Em relação às secções que não apresentam um bom alinhamento, a sua estrutura é construída com recurso ao método *ab initio* [83]. A terceira etapa, tem como objetivo refinar as estruturas construídas anteriormente e visa eliminar os efeitos estereoquímicos presentes [83] . Finalmente, na quarta e última etapa, com o objetivo de deduzir a função da proteína com a estrutura 3D prevista, os diferentes modelos obtidos são comparados com as estruturas existentes no PDB e cuja função é conhecida [83].

O PEP-FOLD é um *web-server* mais rápido do que o I-TASSER e pode ser utilizado com elevado grau de confiança para prever a estrutura 3D de sequências compostas por 9-25 aminoácidos. Este *web-server* recorre a uma abordagem *de novo* e utiliza um alfabeto estrutural composto por 27 letras para descrever a conformação de quatro resíduos consecutivos da cadeia peptídica [82].

2.4.4 *De novo design*

De novo design é uma metodologia que envolve a descoberta de novas moléculas através da estrutura 3D do alvo molecular (recetor) [13, 14]. Este método *in silico* tem como ponto de partida o conhecimento do local de ligação do recetor [84]. Assim, conhecendo a geometria e as propriedades físico-químicas do local de ligação é possível determinar as restrições da forma e as interações não covalentes estabelecidas com um determinado ligando. Através da aplicação de métodos de *docking* de diferentes grupos funcionais no local ativo do recetor é possível prever a conformação e orientação

energeticamente mais favoráveis. Os grupos funcionais, cuja ligação ao local ativo não é energeticamente favorável, são excluídos [84].

Capítulo 3

Métodos computacionais

3.1 Estrutura primária dos péptidos iniciais

Com a finalidade de criar famílias de novos péptidos dirigidos ao domínio LBD do RA, foram consideradas três sequências peptídicas constituídas por diferentes motivos estruturais com reconhecida afinidade experimental para este recetor. Estas sequências e a sua estrutura encontram-se reportadas na literatura, designadamente: o péptido Bud31 (PDB ID: 4OED), o péptido C320 (PDB ID: 4OEZ) e o péptido 3-18, cuja estrutura não está determinada [22]. Assim, através destes péptidos já conhecidos, foram desenvolvidos novos péptidos com uma melhor capacidade de internalização celular (CPPs) e simultaneamente com uma melhor afinidade de ligação para o recetor em estudo.

Com base nas sequências peptídicas Bud31, C320 e 3-18, selecionaram-se os cinco aminoácidos do motivo estrutural de interesse (reportado como determinante para a afinidade ao RA) juntamente com os três aminoácidos da cadeia peptídica imediatamente antes e após o motivo, de forma a criar péptidos compostos por onze aminoácidos: Pep-A, Pep-B e Pep-C. Na tabela 3.1 encontra-se a estrutura primária dos péptidos Bud31, C320 e 3-18, bem como dos péptidos iniciais (Pep-A, Pep-B e o Pep-C), que serviram de base para a criação dos novos péptidos, o respetivo motivo estrutural e código PDB.

Tabela 3.1: Sequência dos péptidos Bud31, C320, 3-18 e dos péptidos iniciais Pep-A, Pep-B e Pep-C, motivo estrutural e código PDB.

Péptido ID	**Sequência**	**Péptido ID**	**Sequência**	**Motivo estrutural**	**PDB ID**
Bud31	KTRYI**F**DL**FY**KRKAY	Pep-A	RYI**F**DL**FY**KRK	FxxFY	4OED
C320	SDSA**F**SR**LY**TRS	Pep-B	DSA**F**SR**LY**TRS	FxxLY	4OEZ
3-18	NTNA**F**SR**LF**YPS	Pep-C	TNA**F**SR**LF**YPS	FxxLF	-

Através da exploração de uma abordagem multifatorial envolvendo diferentes métodos computacionais, explicados detalhadamente nos capítulos seguintes, procedeu-se à previsão e *design* dos novos CPPs e à previsão da sua estrutura para a concretização de estudos de *docking* e simulações de dinâmica molecular.

3.2 Previsão e *design* dos novos péptidos

A previsão e o *design* dos novos CPPs com afinidade de ligação para o RA foram realizados através do *web-server* CellPPD [65]. Este *web-server* baseia-se em modelos de *Machine Learning* e tem por base o algoritmo SVM (*support vector machine*). Por cada submissão efetuada neste servidor são geradas todas as possíveis sequências mutantes do péptido, de forma a torná-lo um melhor CPP e, assim, aumentar a sua capacidade de penetrabilidade através da membrana celular. As etapas realizadas no *web-server* CellPPD encontram-se esquematizadas na figura 3.1.

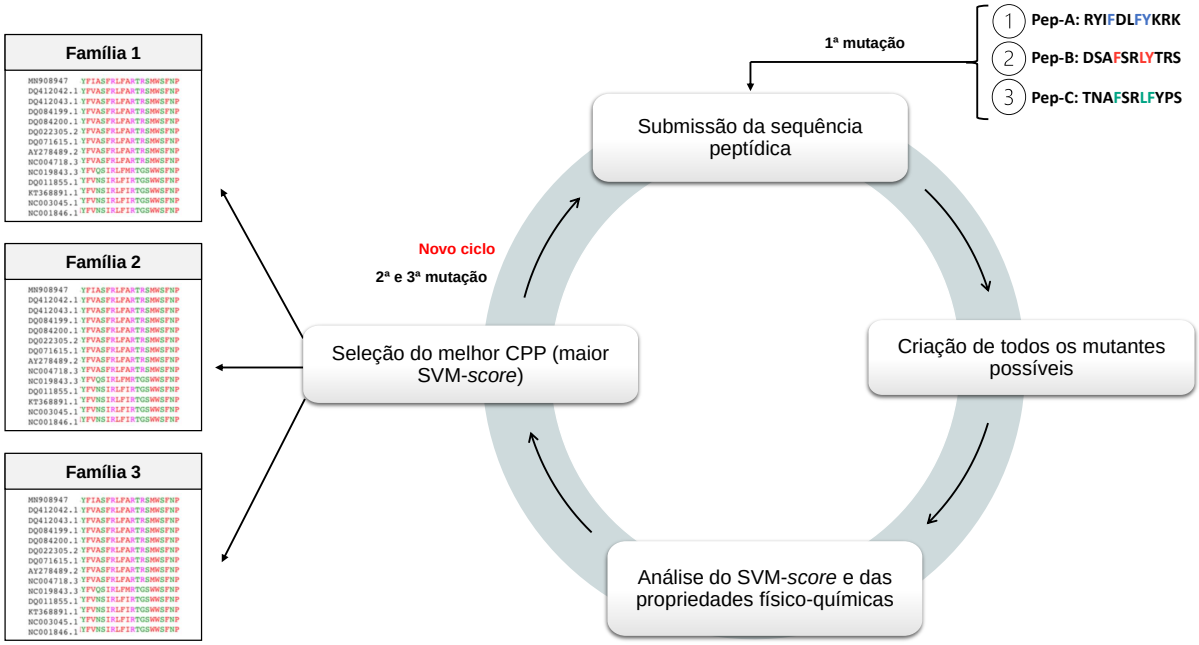


Figura 3.1: Conjunto de etapas desenvolvidas no *web-server* CellPPD.

Primeiramente, as sequências peptídicas do Pep-A, do Pep-B e do Pep-C (tabela 3.1) foram submetidas no CellPPD e sujeitas a três mutações dos seus aminoácidos. De cada uma destas sequências iniciais surge uma nova família de péptidos, sendo que do Pep-A, do Pep-B e do Pep-C resulta a família 1, 2 e 3, respetivamente. A estratégia empregue foi a mesma para cada sequência e inclui três etapas. A primeira etapa (1ªMutação) concretizada envolveu a submissão da sequência inicial no *web-server*. Na segunda etapa (2ªMutação) foi submetida a sequência com maior SVM-*score* resultante da primeira mutação. Por fim, na última etapa (3ªMutação) realizou-se a submissão da sequência peptídica com maior SVM-*score* da segunda mutação. A sequência com maior SVM-*score* da 1ª e 2ª mutação e as 18 sequências com um SVM-*score* mais elevado da 3ª mutação foram extraídas para as famílias 1, 2 e 3 dos novos péptidos. Aquando da escolha destes péptidos mutantes foi necessário ter em consideração que os aminoácidos das posições 4, 7 e 8 da cadeia peptídica (^{4}**F**DL**FY**^{8}, ^{4}**F**SR**LY**^{8}, ^{4}**F**SR**LF**^{8}) permanecem inalterados por serem fundamentais à afinidade com o RA. Assim, cada uma das famílias é constituída por 20 péptidos e, por conseguinte, a nova base de péptidos é composta por 60 novos péptidos.

Os parâmetros de *input* selecionados neste servidor foram o método de previsão e o SVM *threshold*.

Relativamente à previsão da capacidade do péptido ser CPP, existem dois modelos diferentes disponíveis: o modelo SVM e um modelo híbrido. Este último é mais robusto e conduz a uma previsão mais precisa e confiável, uma vez que contabiliza as informações do motivo dos péptidos e o SVM. Por este motivo, o modelo híbrido foi o modelo de eleição. Após a seleção do modelo de previsão, foi necessário escolher o SVM *threshold*. De notar que quanto mais elevado for este valor, mais rigorosa e confiável é a previsão da capacidade do péptido ser ou não um bom CPP. Neste sentido, para as três mutações efetuadas, os valores de SVM-*threshold* selecionados foram 0,3, 0,6 e 0,9 para a 1ª, 2ª e 3ª mutação, respetivamente. Estas condições foram iguais para as três sequências submetidas.

Em segundo lugar, uma vez que os péptidos dirigidos ao RA devem ser capazes de atravessar a membrana celular e apresentar propriedades físico-químicas e farmacocinéticas adequadas, utilizou-se, ainda, o *web-server* MLCPP (*Machine learning based prediction of cell penetrating peptides*) [85]. Este servidor permite prever a eficiência de internalização celular (*uptake efficiency*) dos melhores CPPs previstos através do CellPPD. Assim, os 20 novos péptidos de cada família foram submetidos no MLCPP e selecionaram-se os 10 melhores péptidos com maior probabilidade associada à *uptake efficiency*. Logo, as famílias 1, 2 e 3 passaram a ser constituídas por 10 péptidos e a nova base de péptidos formada por 30 péptidos.

Conhecendo a estrutura primária dos péptidos iniciais, bem como dos novos péptidos desenvolvidos procedeu-se à previsão da sua estrutura 3D.

3.3 Previsão da estrutura 3D

A previsão da estrutura 3D dos péptidos iniciais (Pep-A, Pep-B e Pep-C) e dos novos péptidos desenvolvidos foi realizada através de dois *web-servers*: o I-TASSER [78] e o PEP-FOLD [82]. Estes dois servidores oferecem diferentes modelos com as previsões das possíveis estruturas das sequências peptídicas alvo de estudo. A validação final das estruturas previstas foi realizada através do ProSA-*web* [86]. Este último permite analisar e avaliar a qualidade das diferentes estruturas obtidas, advindas do I-TASSER e do PEP-FOLD, permitindo selecionar a estrutura final mais promissora representativa de cada uma das sequências de péptidos. Nesse sentido, a escolha da melhor estrutura teve por base a seguinte estratégia: i) uma inspeção visual cuidada da estrutura 3D do péptido efetuada no PyMOL [87]; ii) análise do Z-*score*; iii) visualização molecular do péptido através da qual é possível observar os resíduos com maior ou menor energia.

O Z-*score* é um indicador que permite avaliar a qualidade dos modelos com as estruturas previstas dos péptidos. Este valor encontra-se disposto num gráfico, em função do número de resíduos, que reúne o Z-*score* do péptido avaliado, bem como o Z-*score* de todas as proteínas determinadas experimentalmente (RMN ou Raios-X) presentes no PDB (figura 3.2A). Através deste gráfico é possível verificar se o péptido apresenta um Z-*score* dentro do intervalo de valores de Z-*score* para péptidos nativos com número de resíduos próximo ao péptido alvo de estudo, ou seja, com tamanho semelhante. Para além disto, o ProSa-*web* permite, ainda, visualizar os resíduos da sequência peptídica com maior/menor energia. Os resíduos representados a vermelho apresentam maior energia, enquanto que

os representados a azul têm menor energia (figura 3.2B).

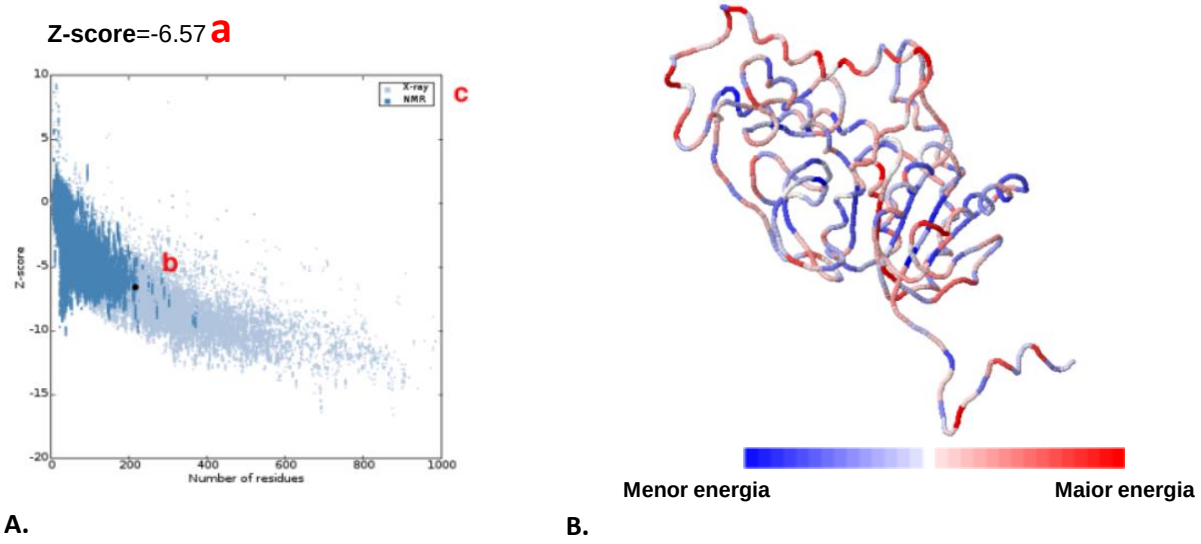


Figura 3.2: Exemplo ilustrativo do Z-*score* (A) e da visualização molecular dos resíduos de maior/menor energia da cadeia peptídica (B) no *web-server* ProSA-*web*.

Após a previsão das estruturas dos péptidos, encontram-se reunidas as condições para a realização dos estudos de *docking* molecular.

3.4 *Docking* molecular

Realizaram-se estudos de *docking* molecular com a finalidade de prever o modo de ligação dos novos péptidos, para os quais se determinou a estrutura, ao RA e investigar os determinantes estruturais e as principais interações estabelecidas entre os péptidos e a interface de ligação do domínio LBD. Os estudos de *docking* foram efetuados através do *web-server* HADDOCK (versão 2.4) [73].

O domínio LBD da estrutura cristalina com código PDB 4OEZ, determinada com recurso à técnica de cristalografia de Raios-X com uma resolução de 1,80 Å, foi o selecionado para perfazer os estudos de *docking* molecular. Esta estrutura 3D cristalizada contempla o péptido com motivo estrutural FxxLY e o ligando agonista DHT em interação com o domínio de interesse primário (domínio LBD) do RA.

De forma a validar a estrutura selecionada e os parâmetros utilizados para os estudos de *docking*, o péptido co-cristalizado com o domínio LBD da estrutura 4OEZ foi submetido a *docking* molecular (controlo positivo). Através deste estudo inicial é possível otimizar as condições utilizadas no HADDOCK e, assim, selecionar os melhores parâmetros que permitem uma melhor reprodução do comportamento *in silico* do péptido co-cristalizado com o domínio LBD da estrutura experimental 4OEZ. Antes de efetuar o *docking* foi necessário preparar a estrutura cristalina 4OEZ a pH fisiológico (pH=7,4), utilizando o *web-server* PDB2PQR [88]. Este servidor define o estado de protonação dos aminoácidos da estrutura a pH fisiológico e adiciona os átomos de hidrogénio em falta. Assim, a estrutura 4OEZ preparada foi a primeira estrutura submetida ao HADDOCK. Como resíduos ativos consideraram-se todos os resíduos do domínio LBD do RA que se encontram a uma distância igual ou inferior a 5 Å do péptido co-cristalizado.

A seleção destes resíduos foi efetuada através do PyMOL [87]. Em relação aos resíduos ativos do péptido, selecionaram-se os cinco resíduos do motivo estrutural de interesse. Logo, os resíduos ativos do recetor e do péptido selecionados foram os seguintes:

1) Resíduos ativos do RA (Molécula 1): 712, 713, 716, 717, 719, 720, 725, 726, 730, 733, 734, 735, 737, 738, 893, 894, 897, 898.

2) Resíduos ativos do péptido (Molécula 2): 4, 5, 6, 7, 8.

Na tabela 3.2 encontram-se os parâmetros otimizados utilizados no HADDOCK e aplicados em todos os estudos de *docking* efetuados. De notar que os parâmetros que não foram alterados, mantiveram os parâmetros definidos por defeito utilizados neste *web-server*.

Tabela 3.2: Parâmetros de *docking* utilizados no *web-server* HADDOCK.

	Parâmetros
Cutoff for clustering	5,0
Clustering method	RMSD
Number of MD steps for rigid body high temperature TAD	2000
MD steps during first rigid body cooling stage	2000
MD steps during second cooling stage with flexible sidechains at interface	4000
MD steps during third cooling stage with folly flexible interface	4000

Após o *docking* do controlo positivo, realizou-se o *docking* dos três péptidos iniciais Pep-A,Pep-B,Pep-C e dos 30 novos péptidos desenvolvidos. Aplicou-se o mesmo protocolo descrito para o *docking* do controlo positivo. Como tal, primeiramente foi necessário prever o estado de protonação dos aminoácidos destas estruturas a pH 7,4 através do *web-server* PDB2PQR [88]. As estruturas finais, resultantes deste servidor, foram submetidas ao HADDOCK. Os resíduos ativos do recetor e do péptido, mencionados anteriormente, bem como os parâmetros presentes na tabela 3.2 foram os mesmos empregues para o *docking* dos péptidos iniciais e do novos péptidos.

Em relação aos resultados obtidos, analisaram-se os 10 melhores complexos tendo em conta o HADDOCK-*score*. Para a seleção dos péptidos mais promissores realizaram-se diferentes estudos efetuados com recurso ao PyMOL, nomeadamente: calculou-se o desvio médio quadrático (RMSD) entre os complexos e o controlo positivo (4OEZ), avaliaram-se as interações polares estabelecidas entre os péptidos e o domínio LBD do RA e determinaram-se os resíduos da interface que interagem, utilizando um *python script*.

3.5 Dinâmica molecular

As três estruturas mais promissoras, provenientes dos estudos de *docking*, foram sujeitas a simulações de dinâmica molecular, de forma a identificar o péptido com maior afinidade de ligação para o domínio LBD do RA. Estas simulações permitem obter uma visão mais aprofundada e detalhada do comportamento e da estabilidade dos péptidos em interação com o domínio LBD do recetor. As simulações

de dinâmica molecular foram efetuadas através do *software* GROMACS [89]. O protocolo efetuado inclui um conjunto de etapas sequenciais, nomeadamente: (1) criação da topologia do sistema; (2) configuração da caixa; (3) solvatação do sistema; (4) adição dos iões; (5) minimização de energia; (6) equilíbrio do sistema; (7) produção da simulação; (8) análise dos resultados.

Em primeiro lugar, a topologia do complexo (domínio LBD do RA+péptido) foi criada através do módulo *pdb2gmx* presente no GROMACS e o campo de forças utilizado foi o Amber99sb-ildn [90]. Posteriormente, foram adicionadas moléculas de água ao sistema numa caixa dodecaédrica com uma distância mínima entre o soluto e a caixa de 1 nm. O modelo teórico implementado para representar as moléculas de água foi o modelo TIP3. De forma a garantir a neutralidade do sistema, adicionaram-se iões cloreto. Na figura 3.3 é possível visualizar um dos sistemas estudados solvatado em água e com iões cloreto numa caixa dodecaédrica.

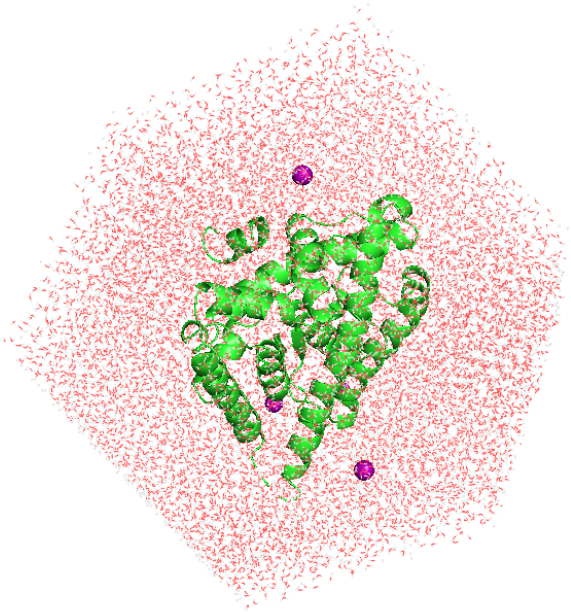


Figura 3.3: Configuração da caixa dodecaédrica (d=1nm) incluindo um dos sistemas estudados solvatado em água (representada a vermelho) e com iões cloreto (representados a roxo).

A adição do solvente e dos iões ao sistema pode promover interações desfavoráveis de maior energia que podem comprometer a simulação. Assim, de forma a alcançar o estado de menor energia do sistema, procedeu-se à minimização de energia utilizando o algoritmo *steepest descent*. Esta etapa inclui dois passos: no primeiro, o sistema encontra-se relaxado, enquanto que no segundo são aplicadas restrições das posições dos átomos através do algoritmo LINCS. Para cada sistema realizaram-se três réplicas, sendo que uma nova minimização de energia foi efetuada para cada réplica. Após o sistema estar na sua conformação de menor energia, este foi sujeito a uma fase de equilíbrio que compreende duas etapas. Na primeira etapa, o sistema foi aquecido até aos 300 K (termóstato *V-rescale*) em condições NVT (número de partículas, volume e temperatura constantes) durante um período de tempo de 100 ps com um passo de integração de 1 fs. A segunda etapa, incluiu o equilíbrio do sistema a 1 atm (Berendsen barostat) sob condições NPT (número de partículas, pressão e temperatura

constantes) durante, igualmente, 100 ps com um passo de integração de 1 fs. Posteriormente, uma vez que as simulações efetuadas não se realizaram com um passo de integração de 1 fs, mas sim de 2 fs, o passo de integração foi alterado para este último valor. Nestas três últimas etapas descritas, as posições dos átomos do soluto (recetor+ligando) encontram-se restringidas por uma força constante de 10 kcal/mol.Å^2, ocorrendo apenas o movimento das moléculas de solvente. As interações eletroestáticas foram avaliadas através do algoritmo PME (Particle-Mesh Ewald). Após a conclusão destas etapas, o sistema está pronto para a produção da dinâmica molecular. As simulações produzidas têm um tempo total de simulação de 100 ns com um passo de integração de 2 fs. A trajetória foi guardada a cada 20 ps, totalizando 5000 *frames*, ou seja, 5000 conformações.

Em segundo lugar, as análises das simulações produzidas integram a visualização da trajetória no PyMOL, o cálculo do RMSD, do RMSF (*root mean square fluctuation*), da área da interface acessível ao solvente (SASA), da energia livre de ligação e das interações de hidrogénio. O cálculo do RMSD, do RMSF e do SASA foi realizado através do GROMACS para as 5000 frames, enquanto que a energia livre de ligação foi calculada através do *package* gmx_MMPBSA [91] para 1000 frames (*skip_5*) e utilizando os parâmetros definidos por defeito. As ligações de hidrogénio foram calculadas tendo em consideração todas as frames e com recurso ao VMD [92]. Considerou-se a formação de uma ligação de hidrogénio até uma distância máxima de 3Å e com um ângulo entre o dador e aceitador de 20º.

Capítulo 4

Resultados e discussão

4.1 Estrutura primária dos novos péptidos

De forma a desenvolver novos péptidos constituídos por uma sequência peptídica que promova não só uma melhor internalização celular, como também uma maior capacidade de ligação ao domínio LBD do RA, utilizaram-se três sequências peptídicas que demonstram experimentalmente afinidade para o recetor em estudo, nomeadamente as sequências dos péptidos Bud31, C320 e 3-18 com motivos estruturais FxxFY, FxxLY e FxxLF, respetivamente [22]. Partindo destes péptidos já conhecidos na literatura, utilizou-se uma parte da sequência correspondente aos cinco aminoácidos do motivo estrutural de interesse mais os três aminoácidos imediatamente antes e após o motivo da cadeia peptídica, originando os péptidos iniciais Pep-A, Peb-B e Pep-C (tabela 2.1) compostos por onze aminoácidos. Através destes péptidos foram criadas, com recurso a um conjunto de ferramentas computacionais, três famílias de novos péptidos mais promissores ao nível celular e com uma elevada capacidade de ligação ao RA.

Primeiramente, com recurso ao *web-server* CellPPD foram previstas novas sequências peptídicas com uma boa capacidade de penetrabilidade celular, a partir das sequências dos péptidos iniciais Pep-A, Pep-B e Pep-C. Estas sequências foram sujeitas a três mutações, no sentido de aumentar a capacidade dos péptidos de atravessar a membrana das células. A abordagem aplicada foi a mesma para as três sequências. Da primeira mutação (i.e., da primeira submissão) realizada, selecionou-se o melhor CPP, ou seja, o péptido com o SVM-*score* mais elevado. Este péptido foi extraído para a família de novos péptidos e submetido no *web-server* para a 2ª mutação. Da segunda mutação concretizada, selecionou-se a sequência peptídica mutante com maior SVM-*score*. Aquando deste processo é necessário ter em consideração dois fatores: (1) o motivo estrutural de interesse (^{4}**F**xx**FY**$^{\mathbf{8}}$, ^{4}**F**xx**LY**$^{\mathbf{8}}$, ^{4}**F**xx**LF**$^{\mathbf{8}}$), à exceção dos aminoácidos que ocupam as posições XX, é imutável. Logo, não podem ser consideradas as sequências com mutações ocorridas nas posições 4, 7 e 8 da cadeia peptídica; (2) o aminoácido que sofreu a primeira mutação, não pode ser mutado novamente. Por este motivo, ter-se há que excluir, dentro das sequências mutantes do péptido, as sequências cujos aminoácidos sofreram mutações nas seguintes posições: 4, 7, 8, mais a posição do aminoácido que sofreu a primeira mutação. Por fim, realizou-se a submissão da sequência peptídica selecionada da segunda mutação.

Desta terceira mutação realizada, selecionaram-se as 18 sequências dos péptidos mutantes com maior SVM-*score*. Aplicando a mesma estratégia que na etapa anterior, os péptidos com mutações dos aminoácidos presentes nas posições 4, 7, 8, mais as posições dos aminoácidos que sofreram a primeira e a segunda mutações não foram tidos em consideração. Assim, do Pep-A, do Pep-B e do Pep-C resultam as famílias 1, 2 e 3, respetivamente. Cada família é composta por 20 novos péptidos.

Em segundo lugar, para prever a *uptake efficiency* (eficiência de internalização celular) dos novos CPPs previstos através do CellPPD, recorreu-se ao *web-server* MLCPP [85]. Dos 20 péptidos de cada família, selecionaram-se as 10 sequências peptídicas mais promissoras, isto é, com maior probabilidade associada à *uptake efficiency*.

Na tabela 4.1 encontram-se as sequências dos péptidos iniciais e do novos péptidos desenvolvidos, a probabilidade associada à previsão dos péptidos serem CPPs e à sua *uptake efficiency*.

Tabela 4.1: Sequências dos péptidos iniciais e dos novos péptidos previstos pelo CellPPD, probabilidade associada à previsão de serem bons CPPs e à sua *uptake efficiency* (eficiência de internalização celular). Os aminoácidos que sofreram mutações são representados a vermelho.

Família ID	**Péptido ID**	**Sequência inicial**	**CPP *score***	**Eficiência de internalização**	**Péptido ID**	**Sequência final**	**CPP *score***	**Eficiência de internalização**
Família 1	Pep-A	[1]RYIFDLFYKRK[11]	0,64	0,5	Pep43	RRIFRRFYKRK	0,98	0,59
					Pep44	RRRFRLFYKRK	0,99	0,54
					Pep48	RRIFRWFYKRK	0,97	0,62
					Pep51	RRIFRKFYKRK	0,97	0,61
					Pep52	RRWFRLFYKRK	0,98	0,58
					Pep53	RRIFRLFYPRK	0,94	0,56
					Pep54	CRIFRLFYKRK	0,91	0,56
					Pep57	WRIFRLFYKRK	0,95	0,54
					Pep58	RRKFRLFYKRK	0,98	0,57
					Pep59	RRIFRLFYKRC	0,91	0,58
Família 2	Pep-B	[1]DSAFSRLYTRS[11]	0,44	-	Pep23	DRRFRRLYTRS	0,89	0,49
					Pep24	DRAFRRLYRRS	0,88	0,46
					Pep25	DRAFRRLYTRR	0,85	0,44
					Pep26	DRAFRRLYKRS	0,84	0,45
					Pep28	DRIFRRLYTRS	0,81	0,52
					Pep29	DRAFRRLYPRS	0,79	0,44
					Pep31	DRAFRRLYTRC	0,76	0,48
					Pep32	RRAFRRLYTRS	0,97	0,43
					Pep33	DRAFRRLYTRP	0,80	0,45
					Pep35	DRWFRRLYTRS	0,83	0,47
Família 3	Pep-C	[1]TNAFSRLFYPS[11]	0,48	-	Pep3	TRRFRRLFYPS	0,92	0,48
					Pep5	TRAFRRLFYPR	0,93	0,48
					Pep9	TRAFRRLFYRS	0,91	0,48
					Pep10	KRAFRRLFYPS	0,86	0,45
					Pep11	TRIFRRLFYPS	0,81	0,59
					Pep13	TRAFRRLFYPC	0,80	0,50
					Pep15	RRAFRRLFYPS	0,88	0,48
					Pep17	TRWFRRLFYPS	0,86	0,59
					Pep19	CRAFRRLFYPS	0,73	0,52
					Pep20	WRAFRRLFYPS	0,81	0,56

Com base nos dados da tabela 4.1, verifica-se que os péptidos iniciais Pep-B e o Peb-C não são CPPs (CPP-*score* $<$ 0,5). O Pep-A pertence à classe de CPP, no entanto apresenta uma probabilidade reduzida. Os resultados obtidos sugerem, ainda, que os novos péptidos desenvolvidos são bons CPPs e, por isso, são capazes de atravessar a membrana das células. Em relação aos péptidos que lhes deram origem, os novos péptidos apresentam uma significativa melhor capacidade de internalização celular. As mutações ocorridas incluem a introdução nas sequências peptídicas dos seguintes aminoácidos: triptofano (W), lisina (K), isoleucina (I), prolina (P), cisteína (C) e arginina (R), destacando-se esta última por ser a mutação mais frequente.

Uma condição que é transversal a todos os péptidos da tabela 4.1 é a substituição dos seus aminoácidos por aminoácidos carregados positivamente, observando-se a incorporação de mais do que uma arginina na suas sequências. A elevada presença deste aminoácido revela ser essencial para aumentar a capacidade de internalização celular dos péptidos. De facto, sequências peptídicas ricas em aminoácidos carregados positivamente, tais como a arginina, apresentam maior facilidade em atravessar a barreira biológica celular [64, 93]. Para além disto, tendo em consideração os resultados *in silico* obtidos, as mutações que envolvem a substituição dos aminoácidos presentes nas posições 2 e 5 da cadeia peptídica por uma arginina, sugerem que esse aminoácido, nessas posições em particular, desempenha um papel importante para que os péptidos estudados se tornem melhores CPPs. Com menor frequência, observam-se mutações pontuais dos aminoácidos da cadeia peptídica por um triptofano, uma prolina ou por uma isoleucina. Estes aminoácidos apresentam características essencialmente hidrofóbicas que favorecem a translocação dos péptidos para o meio intracelular.

Em relação à *uptake efficiency*, os novos CPPs desenvolvidos apresentam uma maior eficiência ao atravessar a membrana das células do que os péptidos que lhes deram origem. De acordo com *web-server* MLCPP, os CPPs previstos podem apresentar uma maior eficiência de internalização celular ($\geq$ 0,5) ou, caso contrário, uma menor eficiência. Assim sendo, os péptidos pep48, pep51, o pep43, o pep11 e o pep17 destacam-se por apresentarem uma ligeira maior eficiência em penetrar a membrana celular em relação aos restantes péptidos.

4.2 Estrutura 3D dos péptidos

Uma vez conhecidas as sequências dos novos péptidos desenvolvidos e dos péptidos que lhes deram origem, procedeu-se à previsão da estrutura 3D dos mesmos, utilizando dois *web-servers*: o PEP-FOLD [82] e o I-TASSER [78]. Os diferentes modelos com as estruturas previstas, advindos destes dois servidores, foram submetidos no *web-server* ProSa-Web [86]. Através deste último é possível avaliar a qualidade global dos modelos obtidos com as diferentes estruturas previstas e selecionar a estrutura 3D final mais promissora de cada péptido. A escolha da melhor estrutura teve por base a análise do Z-*score*, uma inspeção visual cuidada no PyMOL das estruturas 3D previstas e a análise dos resíduos, que constituem a cadeia peptídica, de maior/menor energia. Relativamente ao Z-*score*, este valor deverá ser o mais reduzido e encontrar-se dentro do intervalo de valores de Z-*score* para estruturas determinadas experimentalmente com tamanho semelhante ao péptido em estudo. Assim sendo,

tendo por base esta estratégia, as estruturas mais promissoras do Pep-A, Pep-B, Pep-C, Pep3, Pep5, Pep13, Pep15, Pep19, Pep20, Pep23, Pep24, Pep25, Pep26, Pep29, Pep33, Pep35, Pep44, Pep52, Pep53, Pep54, Pep5 e o Pep58 foram selecionadas do PEP-FOLD, enquanto que as restantes estruturas foram selecionadas do I-TASSER. Todas estas estruturas previstas descrevem o comportamento de uma α-hélice. Após a seleção das estruturas finais mais promissoras de cada péptido, foi previsto o estado de protonação dos seus aminoácidos a pH fisiológico através do *web-server* PDB2PQR [88].

Nas figuras A.5, A.6 e A.7, presentes no anexo A, encontra-se representada a estrutura 3D prevista dos péptidos iniciais bem como dos novos péptidos. Os gráficos com o Z-*score* e os resíduos de maior/menor energia das estruturas finais escolhidas do Pep-A, do Pep-B, do Pep-C e dos novos péptidos das famílias 1, 2 e 3, podem ser visualizados nas tabelas A.1, A.2, A.3 e A.4 do anexo A.

Conhecendo a estrutura dos novos CPPs e dos péptidos iniciais, estas estruturas foram sujeitas a estudos de *docking*, cujos resultados serão apresentados e discutidos no seguinte capítulo.

4.3 *Docking* dos péptidos iniciais

Os estudos de *docking* molecular com os péptidos iniciais (Pep-A, Pep-B e Pep-C) foram realizados com as estruturas devidamente protonadas a pH fisiológico através do PDB2PQR [88]. O *docking* destes péptidos foi concretizado no sentido de avaliar o potencial dos novos CPPs desenvolvidos. Assim, é possível avaliar se os novos péptidos apresentam uma maior afinidade de ligação para o RA em relação aos péptidos que lhes deram origem.

O *docking* dos péptidos foi realizado com recurso ao *web-server* HADDOCK (versão 2.4) [73]. A abordagem aplicada na análise dos resultados obtidos teve por base a inspeção das estruturas presentes no top10, que inclui as dez poses do péptido com melhor HADDOCK-*score* (mais negativo). Para além disto, de forma a validar os resultados *in silico* obtidos, calculou-se o RMSD a partir do alinhamento dos complexos com a estrutura do cristal 4OEZ. Este cálculo permite avaliar a proximidade entre a conformação prevista dos complexos pelo HADDOCK e a conformação experimental da estrutura cristalina do controlo positivo 4OEZ. Na tabela 4.2, encontram-se os resultados obtidos do *docking* dos péptidos iniciais, tendo por base a análise das 10 melhores estruturas de acordo com o seu HADDOCK-*score*, e o RMSD calculado para os átomos Cα através do PyMOL.

O critério de seleção aplicado na escolha das estruturas mais promissoras, representativas da interação dos péptidos iniciais com o domínio LBD do RA, teve por base um compromisso entre o complexo com melhor HADDOCK-*score* e menor RMSD. Por um lado, um melhor HADDOCK-*score* reflete-se numa menor energia de ligação envolvida entre o recetor e o péptido, o que se traduz numa maior estabilidade do complexo. Por outro lado, relativamente ao RMSD, quanto mais próximo for este valor de zero, maior a proximidade conformacional entre a estrutura do complexo e a estrutura do cristal 4OEZ. Assim, os complexos *complex_14w*, *complex_105w* e *complex_122w* foram selecionados como sendo as estruturas mais promissoras representativas da interação do RA com os péptidos Pep-A, Pep-B e do Pep-C, respetivamente.

Tabela 4.2: Resultados obtidos do *docking* molecular dos péptidos iniciais Pep-A, Pep-B e Pep-C.

Péptido ID	Sequência	Complexo	HADDOCK-*score*	RMSD
Pep-A	RYIFDLFYKRK	*complex_160w*	-104,9	1,001
		complex_14w	-103,3	0,348
		complex_32w	-96,6	0,340
		complex_28w	-96,2	0,357
		complex_24w	-93,7	0,365
		complex_21w	-93,5	0,391
		complex_13w	-92,2	0,375
		complex_43w	-90,9	0,452
		complex_92w	-90,6	1,377
		complex_6w	-90,4	0,670
Pep-B	DSAFSRLYTRS	*complex_109w*	-98,4	0,725
		complex_105w	-98,1	0,691
		complex_117w	-92,6	1,057
		complex_74w	-91,6	1,042
		complex_159w	-89,1	1,041
		complex_32w	-83,7	1,152
		complex_187w	-83,6	0,995
		complex_167w	-83,5	1,480
		complex_169w	-83,2	1,061
		complex_178w	-82,4	1,082
Pep-C	TNAFSRLFYPS	*complex_122w*	-98,5	0,396
		complex_42w	-95,1	0,647
		complex_136w	-94,6	0,608
		complex_72w	-94,1	0,704
		complex_166w	-93,8	0,913
		complex_101w	-93,2	0,756
		complex_83w	-92,7	0,649
		complex_28w	-92,1	0,388
		complex_105w	-90,6	0,423
		complex_107w	-89,4	0,523

4.4 *Docking* dos novos péptidos

Os estudos de *docking* molecular dos novos CPPs das famílias 1, 2 e 3 foram realizados tendo como principal objetivo avaliar e prever o modo de ligação mais favorável de cada um destes péptidos ao RA, investigar os determinantes estruturais envolvidos e as principais interações estabelecidas na interface do domínio LBD-péptidos. Para concretizar este objetivo, primeiramente foi previsto o estado de protonação dos aminoácidos das estruturas dos novos péptidos a pH fisiológico através do *web-server* PDB2PQR, com a consequente adição dos átomos de hidrogénio em falta. As estruturas preparadas, resultantes deste servidor, foram sujeitas a estudos de *docking* no *web-server* HADDOCK [73]. A análise dos resultados obtidos teve por base a mesma estratégia aplicada no *docking* dos péptidos iniciais. Por isso, analisaram-se os complexos que se encontravam na lista das 10 melhores poses (top10), em termos de energia de ligação, classificadas pelo HADDOCK.

Numa primeira abordagem, os complexos do top10 do HADDOCK foram sujeitos a uma inspeção visual cuidada no PyMOL, de forma a avaliar as diferenças existentes entre as diferentes conformações

previstas de cada complexo e averiguar a estabilidade do péptido. Paralelamente a esta inspeção visual, realizou-se uma pré-seleção das estruturas mais promissoras tendo em consideração o HADDOCK-*score* e o RMSD. Este último foi calculado através do PyMOL tendo por base o alinhamento da estrutura de cada complexo com a estrutura do controlo positivo (PDB ID: 4OEZ). Desta pré-seleção efetuada, selecionaram-se as estruturas com melhor HADDOCK-*score* e menor RMSD para uma análise mais robusta. Esta abordagem permite efetuar um *screening* virtual e excluir, de entre as diferentes conformações dos péptidos previstas pelo HADDOCK, as que revelam menor potencial. Posto isto, os melhores complexos, provenientes desta pré-seleção, encontram-se nas tabelas 4.3, 4.4 e 4.5 de acordo com os motivos estruturais FxxFY (família 1), FxxLY (família 2) e FxxLF (família 3), respetivamente.

Em segundo lugar, as estruturas provenientes da pré-seleção das 10 melhores poses foram sujeitas a uma análise mais robusta que inclui a investigação dos contactos polares estabelecidos entre o péptido e o recetor através do PyMOL. Assim, é possível identificar os complexos mais promissores e selecionar a estrutura que melhor representa a interação dos novos CPPs com o RA. Esta avaliação teve em consideração as interações-chave reportadas na literatura e consideradas essenciais para a interação entre domínio LBD do RA e péptidos co-reguladores com motivo estrutural FxxLF. Estas interações-chave são estabelecidas entre: (1) a K720 e o *backbone* da F+5 do motivo estrutural do péptido; (2) o E897 e o grupo amina da F+1 do motivo e o aminoácido presente nas posições -1 do motivo estrutural [17, 22, 94]. A Q733 revela um papel fundamental para interação do RA com péptidos constituídos por uma tirosina na posição 5 do seu motivo estrutural. Este resíduo do recetor estabelece uma ligação de hidrogénio adicional com a Y+5 do motivo do péptido, em relação à fenilalanina na mesma posição, tornando a interação do complexo mais forte e, por isso, mais estável [19].

Em suma, a estratégia aplicada na seleção do melhor complexo de cada péptido em interação com o domínio LBD do RA, teve por base um compromisso entre o complexo com melhor HADDOCK-*score*, menor RMSD e, ainda, a presença das interações-chave estabelecidas entre o domínio LBD do recetor e os péptidos, reportadas na literatura.

4.4.1 Péptidos com motivo FxxFY

As estruturas mais promissoras, provenientes da pré-seleção realizada tendo em consideração o HADDOCK-*score* e o RMSD, encontram-se na tabela 4.3. Estas estruturas são resultantes dos estudos de *docking* dos péptidos com motivo estrutural FxxFY da família 1.

No que diz respeito ao **Pep59**, de entre os quatro complexos previstos, é possível destacar o *complex_48w*. Esta estrutura encontra-se em primeiro lugar na lista do top10 dos melhores complexos e caracteriza-se por apresentar um menor RMSD e um melhor HADDOCK-*score*. Os resíduos do RA que estão envolvidos na interação são: R726, D731, Q733, Y739, E893, E897. A R726 e a Q733 estabelecem contactos polares com a Y8 do péptido, enquanto que o D731 e a Y739 estabelecem contactos polares com a R5 e com a R1, respetivamente. O E893 interage com a R2 e o E897 com os resíduos R1, R2 e I3 do péptido. A interação que envolve o resíduo-chave E897 do RA e o resíduo I3 está de

Tabela 4.3: Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do *docking* dos péptidos com motivo FxxFY pertencentes à família 1.

Péptido ID	Sequência	Complexo	HADDOCK-*score*	RMSD
Pep59	RRIFRLFYKRC	*complex_48w*	-114,84	0,417
		complex_106w	-101,94	0,576
		complex_28w	-99,87	0,470
		complex_111w	-99,78	0,489
Pep43	RRIFRRFYKRK	*complex_198w*	-111,3	0,671
		complex_43w	-109,3	0,762
		complex_169w	-108,7	0,530
Pep48	RRIFRWFYKRK	*complex_28w*	-105,6	0,593
		complex_27w	-105,1	0,878
Pep44	RRRFRLFYKRK	*complex_63w*	-121,0	0,540
		complex_45w	-106,8	0,493
		complex_97w	-105,2	0,551
		complex_22w	-104,3	0,535
Pep51	RRIFRKFYKRK	*complex_59w*	-98,2	0,479
		complex_6w	-97,3	0,533
Pep52	RRWFRLFYKRK	*complex_62w*	-113,2	0,598
		complex_56w	-104,9	0,577
		complex_94w	-104,4	0,611
Pep53	RRIFRLFYPRK	*complex_44w*	-119,6	0,702
		complex_162w	-118,6	0,648
		complex_111w	-102,1	0,652
		complex_181w	-101,4	0,568
Pep54	CRIFRLFYKRK	complex_147w	-115,0	0,416
		complex_6w	-114,2	0,400
		complex_33w	-112,9	0,346
		complex_10w	-109,0	0,383
		complex_36w	-109,0	0,366
		complex_64w	-108,2	0,403
Pep57	WRIFRLFYKRK	*complex_92w*	-118,0	0,509
		complex_191w	-117,5	0,504
Pep58	RRKFRLFYKRK	*complex_44w*	-121,8	0,356
		complex_143w	-121,4	0,380
		complex_20w	-118,3	0,451
		complex_98w	-115,1	0,408

acordo com a literatura [17, 22, 94].

Tendo em consideração a estratégia aplicada na seleção do complexo mais promissor (RMSD, HADDOCK-*score* e interações polares), a melhor estrutura que representa a interação do **Pep43** com o RA demonstra ser o *complex_169w*. Por apresentar um menor RMSD, a conformação desta estrutura é a que se encontra mais próxima à conformação da estrutura cristalina 4OEZ. Os resíduos do RA que participam na interação com o péptido são: D731, E893, E897. O D731 e o E893 estabelecem contactos polares com a R1 e com a R6 do péptido, respetivamente. No que diz respeito ao E897, este

interage com os resíduos R1, R2 e I3 do Pep43. A interação entre o E897 e a I3 está concordante com as interações-chave reportadas na literatura [17, 22, 94].

Em relação ao **Pep48**, dos 10 melhores complexos, apenas o *complex_28* e o *complex_27w* apresentam um RMSD inferior a 1. Por conseguinte, estes dois complexos apresentam um comportamento mais próximo ao do comportamento experimental do que as restantes estruturas do top10. Assim, estes dois complexos demonstram ser os mais promissores em relação ao RMSD. O *complex_28* e o *complex_27w* não revelam diferenças significativas ao nível do HADDOCK-*score*. De entre estes dois complexos, o potencialmente mais promissor parece ser o *complex_28* por apresentar não só um melhor HADDOCK-*score*, como também um menor RMSD. Neste complexo, os resíduos do recetor que estabelecem contactos polares com o péptido são os seguintes: K720, R726, Q733, E893 e o E897. A K720 estabelece contactos polares com a F7 e com a R10. Os resíduos R726 e Q733 interagem com a K11. O E893 interage com o W6 do péptido, enquanto que o E897 interage com a R1, R2 e I3 do péptido. Quer no *complex_28*, quer no *complex_27w* as interações-chave entre os resíduos essenciais K720 e a Q733 do recetor não são estabelecidas, enquanto que o resíduo E897 estabelece com a I3 a interação-chave descrita na literatura [17, 22, 94].

O *complex_63w* revela ser a melhor estrutura resultante do *docking* do **Pep44**. Este complexo encontra-se em primeiro lugar na lista do top10 dos melhores complexos de acordo com o HADDOCK-*score*. A L712, a K720, o E893 e o E897 são os resíduos do recetor que interagem com o Pep44. As interações polares estabelecidas são entre os seguintes resíduos: a L712 do recetor e a K11 do péptido, a K720 e a Y8, o E893 e o W6 e a R10 e, por último, entre o E897 e a R2 da cadeia peptídica. A interação estabelecida pela K720 está concordante com as interações-chave mencionadas na literatura [17, 22, 94].

Em relação ao **Pep51**, o *complex_59w* apresenta menor HADDOCK-*score* e menor RMSD. Os resíduos do RA que participam na interação são o D731 que interage com a R1, o E893 interage com a K6 e o E897 com os resíduos R1, R2 e I3 do péptido. Apenas as interações-chave que envolvem o resíduo E897 estão presentes. No *complex_6w* participam os mesmos resíduos do RA (D731, E893 e E897) na interação com este péptido. Assim, tendo por base a conjugação do RMSD, HADDOCK-*score* e as interações-chave, o *complex_59w* demonstra ser a estrutura mais promissora.

Os resultados obtidos do *docking* do **Pep52** sugerem que o complexo mais promissor parece ser o *complex_62w*. Este complexo diferencia-se dos restantes por apresentar um menor HADDOCK-*score*. Nesta estrutura, o Pep52 interage com os seguintes resíduos do recetor: K720, D731, E893 e o E897. A K720 interage com a F7, o D731 estabelece contactos polares com a R5 do péptido. O E893 interage com a R2 e o E897 do recetor com os resíduos R1 e R2.

Os resultados obtidos do *docking* do **Pep53** revelam que a estrutura do *complex_181w* demonstra ser a estrutura mais promissora. Este complexo apresenta uma conformação mais próxima à conformação da estrutura cristalina do controlo positivo 4OEZ, na medida em que apresenta um RMSD inferior aos restantes complexos. Os resíduos do recetor que participam na interação com o péptido são: V713, K720, D731, E893 e o E897. O resíduo V713 estabelece contactos polares com a R2 do péptido. A K720 e o D731 interagem com a L6 e com a K11, respetivamente. O E893 interage com a

R1 e com a R5 da cadeia peptídica. Por fim, o E897 interage com a R1 do péptido.

Relativamente ao **Pep54**, todos os complexos revelam valores de RMSD próximos entre si. No entanto, apenas o *complex_33w* parece ser a melhor estrutura face à estratégia aplicada. Este complexo experiencia um RMSD e um HADDOCK-score reduzidos e, ainda, apresenta a presença das interações com os resíduos-chave do recetor e do péptido. Os resíduos V713, K720, D731, Q733 e o E897 do recetor participam na interação com o pep54. A V713 interage com a K11, a K720 interage não só com a K11, mas também com a Y8. O D731 estabelece interações polares com a R5, enquanto que a Q733 interage com a Y8 do péptido. Finalmente, o E897 interage com os resíduos C1, R2, I3, F4 da sequência peptídica. Todas as interações dos resíduos K720, Q733 e do E897 do RA estão de acordo com as interações descritas na literatura [17, 22, 94]. Assim, este complexo foi selecionado como sendo o mais promissor de entre as restantes estruturas.

Os dois melhores complexos que envolvem o **Pep57**, provenientes da pré-seleção realizada, apresentam comportamentos próximos em termos de conformação, HADDOCK-*score* e RMSD. Apesar disto, o *complex_92w* destaca-se por apresentar as interações essenciais reportadas na literatura [17, 22, 94]. Os resíduos do recetor envolvidos na interação deste complexo são a R726, a Q733, o E897 e a Q902. Os resíduos R726 e Q733 interagem com a Y8 do péptido, o E897 com os resíduos W1, R2, I3 e, por fim, o Q902 com o W1 da cadeia peptídica. Como é possível verificar, as interações dos resíduos-chave do recetor (E897 e Q733) estabelecidas estão concordantes com as interações descritas na literatura [17, 22, 94]. No que diz respeito ao *complex_191w*, os resíduos que interagem são apenas o D731 e o E897. Assim, tendo em conta os resultados obtidos, o *complex_92* revela ser a estrutura mais promissora.

Finalmente, a pose do **Pep58** no *complex_44w* é a que promove a formação de um complexo recetor-péptido mais estável, estando, por isso, uma menor energia de ligação envolvida. Este complexo está presente em primeiro lugar na lista dos 10 melhores complexos em termos do HADDOCK-*score*. Além disto, apresenta um menor RMSD e os resíduos de aminoácidos do recetor que interagem são: V713, K720, D731, E893 e o E897. A V713 interage com a K11 do péptido, enquanto que a K720 estabelece contactos polares com a Y8 e com K11. O D731 interage com os resíduos R1 e R5. Por fim, o E893 interage com a K3 e o E897 com a R2, K3 e F4 do péptido. As interações dos resíduos-chave K720 e do E897 do RA estão de acordo com as interações reportadas na literatura [17, 22, 94].

4.4.2 Péptidos com motivo FxxLY

Os melhores complexos, advindos da pré-seleção realizada com base no HADDOCK-*score* e RMSD, encontram-se na tabela 4.4. Estas estruturas são provenientes dos estudos de *docking* dos péptidos com motivo FxxLY da família 2. A estratégica aplicada na seleção da melhor estrutura foi a mesma utilizada anteriormente.

Em relação ao **Pep23**, existem dois possíveis potenciais complexos ao nível do HADDOCK-*score* e do RMSD, particularmente: o *complex_119w* e o *complex_132*. Estas estruturas apresentam um RMSD inferior aos restantes complexos e um melhor HADDOCK-*score*. Tendo por base a estratégia aplicada

Tabela 4.4: Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do *docking* dos péptidos com motivo FxxLY pertencentes à família 2.

Péptido ID	Sequência	Complexo	HADDOCK-*score*	RMSD
Pep23	DRRFRRLYTRS	*complex_119w*	-113,9	0,403
		complex_132w	-107,1	0,324
		complex_140w	-102,1	0,492
		complex_71w	-101,7	0,415
Pep24	DRAFRRLYRRS	*complex_13w*	-108,2	0,421
		complex_68w	-103,7	0,353
		complex_103w	-102,8	0,386
Pep25	DRAFRRLYTRR	*complex_103w*	-109,3	0,387
		complex_35w	-103,5	0,388
		complex_183w	-101,2	0,377
		complex_78w	-88,9	0,373
		complex_37w	-87,5	0,374
Pep26	DRAFRRLYKRS	*complex_154w*	-107,0	0,337
		complex_109w	-100,9	0,713
		complex_180w	-97,6	0,516
		complex_198w	-96,3	0,443
		complex_146w	-95,6	0,401
Pep28	DRIFRRLYTRS	*complex_69w*	-100,3	0,717
		complex_119w	-99,7	0,735
		complex_63w	-96,2	0,657
Pep29	DRAFRRLYPRS	complex_127w	-90,8	0,689
		complex_65w	-87,9	0,517
		complex_89w	-83,0	0,594
		complex_104w	-81,3	0,530
		complex_19w	-81,0	0,531
		complex_33w	-80,6	0,560
Pep31	DRAFRRLYTRC	complex_179w	-98,4	0,771
		complex_78w	-90,2	0,387
		complex_51w	-87,8	0,543
		complex_8w	-87,5	0,354
		complex_19w	-85,3	0,353
		complex_14w	-83,5	0,399
Pep32	RRAFRRLYTRS	*complex_23w*	-96,0	0,468
		complex_85w	-94,9	0,438
		complex_2w	-91,1	0,450
Pep33	DRAFRRLYTRP	*complex_114w*	-108,2	0,449
		complex_69w	-107,7	0,434
		complex_14w	-105,6	0,374
		complex_72w	-104,0	0,476
Pep35	DRWFRRLYTRS	*complex_40w*	-103,9	0,640
		complex_44w	-102,7	0,599

na seleção do complexo mais promissor, a melhor estrutura demonstra ser a do *complex_119w*. Neste complexo, os resíduos de aminoácidos do RA que participam na interação são: K717, K720, R726, V730, D731, Q738, E893 e E897. O resíduo K717 interage com a S11 do péptido. A K720 e a R726 interagem ambos com a Y8 do Pep23, enquanto que os resíduos V730 e o D731 estabelecem contactos

polares com a R5 do péptido. A Q738 interage com o D1 da cadeia peptídica. Finalmente, o E893 interage com a R3 e com a R6, enquanto que o E897 estabelece contactos polares com a R3 e com a F4 do péptido.

Relativamente ao **Pep24**, verifica-se que o *complex_68w* é o que apresenta um RMSD inferior aos restantes e a pose do péptido nesta estrutura é a que permite uma interação com um maior número de resíduos do RA. As interações polares estabelecidas neste complexo são entre: a K720 e a Y8 e a S11, a R726 e a Y8, o D731 e a R5, o Q733 e a Y8, o Q738 e o D1, o E893 e a R6, o E897 e a R2, A2 e F4 e, por fim, o Q902 e a D1 da cadeia peptídica. As interações-chave reportadas na literatura são estabelecidas [17, 22, 94]. Assim, o *complex_68w* revela ser a estrutura que melhor representa a interação do Pep24 com o RA.

Os resultados obtidos do *docking* do **Pep25** sugerem que o melhor complexo parece ser o *complex_35w*. Observa-se uma proximidade dos valores de RMSD dos diferentes complexos previstos pelo HADDOCK. Por um lado, o *complex_103w* é o complexo que apresenta um melhor HADDOCK-*score*. No entanto, as interações essenciais envolvidas entre o domínio LBD do RA e o péptido não são estabelecidas. Por outro lado, o *complex_35w* destaca-se devido ao facto de apresentar um menor RMSD, melhor HADDOCK-*score* e, ainda, por apresentar a presença das interações essenciais. Os resíduos do recetor que participam na interação são: a K720, o D731, a Q733, a Q738, a E893, a E897 e a Q902. A K720 estabelece contactos polares com a Y8 e com a R11, o D731 interage com a R5 e a Q733 com a Y8 do péptido. Os resíduos Q738 e o Q902 interagem com o D1, enquanto que o E893 interage com a R2 da sequência peptídica. Por fim, o resíduo E897 do recetor estabelece contactos polares com a R2 e com a A3.

No que concerne ao **Pep26**, o *complex_154w* é a estrutura que apresenta um melhor HADDOCK-*score* e um menor RMSD. Este complexo encontra-se na primeira posição da lista do top10 dos melhores complexos advindos do HADDOCK. Neste complexo, o pep26 estabelece contactos polares com os resíduos K720, D731, Q733, E893, E897 e Q902 do RA. Os resíduos D731, Q902 e E893 interagem com a R5, o D1 e a R6, respetivamente. A Q733 interage com a Y8. O resíduo K720 estabelece contactos polares com a Y8 e com a S11 do péptido, enquanto que o resíduo E897 interage com a R2 e com a A3. Como é possível verificar as interações-chave reportadas na literatura são estabelecidas [17, 22, 94]. Posto isto, a melhor estrutura parece ser a estrutura do *complex_154w*.

O *complex_69w* que envolve o **Pep28** é o que satisfaz as condições utilizadas pela estratégia aplicada na escolha da melhor estrutura. Este complexo apresenta um menor RMSD, melhor HADDOCK-*score* e, ainda, estão envolvidas as interações mais relevantes no complexo. Os resíduos do recetor envolvidos na interação são V713, K720, D731, Q733, o E893 e o E897. A V713 interage com a S11, a K720 com a Y8 e com a S11, o D731 com a R5, a Q733 com a Y8, o E893 com a R6 e com a R10 e, por fim, a E897 interage com a R6.

A pose do **Pep29** no *complex_65w* é a que permite obter um HADDOCK-*score* e um RMSD inferior em relação aos restantes complexos. Os resíduos do RA que participam na interação com o péptido são os seguintes: E709, K720, E893, E897. O resíduo E709 interage com a R6, a K720 com a L7, o E893 com a R2 e, por fim, o E897 com o D1 e com a R2. Assim, a melhor estrutura revela ser a

estrutura do *complex_65w*.

No caso do **Pep31**, o complexo com maior potencial parece ser o *complex_78w* uma vez que apresenta um melhor HADDOCK-*score*, um valor de RMSD inferior aos restantes e, ainda, revela a presença das interações-chave. A K720, Q733 e o E897 são os resíduos do recetor envolvidos na interação com o péptido. Por um lado, os resíduos K720 e a Q733 estabelecem interações polares com o resíduo Y8 do péptido. Por outro lado, o resíduo E897 interage com os resíduos R2, A3 e F4 do péptido (contactos polares).

A melhor estrutura resultante do *docking* do **Pep32** demonstra ser o *complex_23w*. Este complexo foi eleito como o mais promissor por apresentar um melhor HADDOCK-*score*, menor RMSD e a presença das interações-chave. Os contactos polares estabelecidos entre o RA e o Pep32 são entre a K720 e a Y8 e a S11, o D731 e a R1, a Q733 e a Y8 e, finalmente, entre o E897 e a A3 e a F4 da cadeia peptídica.

Os resultados obtidos do *docking* do **Pep33** sugerem que o *complex_14w* revela ser a melhor estrutura. Esta escolha teve por base um compromisso entre o complexo com melhor HADDOCK-*score*, menor RMSD e a presença das interações consideradas essenciais entre os resíduos-chave do RA e o péptido. Os resíduos envolvidos na interação são: a K720, o R726, o D731, o E893 e o E897. A K720 interage com a L7 e com a Y8, a R726 interage com a Y8, o D731 com a R5, o E893 com a R2 e o E897 com a R2 e com A3 do péptido.

Por último, em relação ao **Pep35** destaca-se a estrutura do *complex_44w*. Este complexo apresenta um menor RMSD e um HADDOCK-*score* ligeiramente menos promissor do que o do *complex_40w*. Este último não revela a presença dos resíduos-chave do RA. Os resíduos K720, R726, E897 do RA estão envolvidos na interação do domínio LBD do RA com o Pep35 no *complex_44w*. A K720 e a R726 interagem com os resíduos L7 e S11 do péptido, enquanto que o E897 interage com o D1 e com a R2. Assim, considerou-se o *complex_44w* como a melhor estrutura.

4.4.3 Péptidos com motivo FxxLF

As melhores poses dos péptidos com motivo FxxLF da família 3, provenientes da pré-seleção realizada com base no HADDOCK-*score* e RMSD, encontram-se na tabela 4.5. A abordagem utilizada na seleção dos melhores complexos teve por base a mesma estratégia aplicada nos casos anteriores.

Relativamente ao **Pep3**, existem dois complexos que se destacam em termos de RMSD, nomeadamente: o *complex_64w* e o *complex_115w*. Por revelarem um menor RMSD, estas estruturas apresentam uma conformação mais próxima à conformação da estrutura do cristal do que os restantes complexos. Tendo por base a estratégia de seleção aplicada na escolha do complexo mais promissor, que envolve um compromisso entre o complexo com melhor HADDOCK-*score*, menor RMSD e as interações-chave descritas na literatura, o *complex_64w* foi eleito como a estrutura que melhor representa a interação do Pep3 com o domínio LBD do RA. Neste complexo, o péptido interage com os resíduos K720, R726, D731, E893 e E897 do recetor. A K720 interage com a F8 do péptido, enquanto que o E897 estabelece interações polares com os resíduos T1, R2, R3 e F4, a R726 interage com a Y9,

Tabela 4.5: Resultados obtidos dos melhores complexos, em termos de HADDOCK-*score* e RMSD, advindos do HADDOCK dos péptidos com motivo FxxLF pertencentes à família 3.

Péptido ID	Sequência	Complexo	HADDOCK-*score*	RMSD
Pep3	TRRFRRLFYPS	*complex_98w*	-115,4	0,540
		complex_14w	-112,3	0,645
		complex_64w	-107,7	0,386
		complex_115w	-102,1	0,366
Pep5	TRAFRRLFYPR	*complex_111w*	-104,1	0,553
		complex_91w	-100,0	0,574
		complex_36w	-97,5	0,597
		complex_77w	-96,1	0,592
Pep9	TRAFRRLFYRS	*complex_33w*	-108,0	0,365
		complex_36w	-97,6	0,357
		complex_169w	-97,5	0,397
Pep10	KRAFRRLFYPS	*complex_20w*	-112,1	0,879
		complex_5w	-104,4	0,637
		complex_16w	-102,5	0,706
Pep11	TRIFRRLFYPS	*complex_193w*	-99,8	0,624
		complex_21w	-99,7	0,692
		complex_59w	-98,0	0,579
Pep13	TRAFRRLFYPC	*complex_26w*	-103,2	0,669
		complex_44w	-101,0	0,664
		complex_35w	-99,1	0,632
		complex_34w	-97,4	0,407
Pep15	RRAFRRLFYPS	*complex_48w*	-117,6	0,384
		complex_62w	-110,7	0,414
		complex_44w	-106,6	0,363
Pep17	TRWFRRLFYPS	*complex_9w*	-103,4	0,424
		complex_64w	-102,3	0,599
		complex_52w	-99,1	0,546
Pep19	CRAFRRLFYPS	*complex_2w*	-114,6	0,376
		complex_58w	-111,9	0,391
Pep20	WRAFRRLFYPS	*complex_11w*	-126,7	0,405
		complex_196w	-114,5	0,355
		complex_9w	-113,4	0,547

o D731 com a R5 e o E893 com a R2 e com a R3 da sequência peptídica. As interações estabelecidas pelos resíduos K720 e E897 do recetor estão concordantes com as interações-chave reportadas na literatura [17, 22, 94].

Em relação ao **Pep5**, verifica-se que todos os complexos apresentam valores de RMSD próximos entre si. No entanto, a interação dos resíduos-chave do RA (K720 e E897) com o péptido é verificada apenas no *complex_36w* e no *complex_77w*. Nos restantes complexos, estes resíduos não estabelecem contactos polares. No *complex_36w* o péptido interage com os resíduos K720, D731 e E897 do recetor, enquanto que no *complex_77w* interage com os resíduos K720, E709, E893 e E897. Por um lado, no *complex_36w*, a K720 do recetor estabelece interações polares com a K7 e com a F8 do péptido e o

E897 e o D731 interagem com a T1 e com a R5, respetivamente. Por outro lado, no *complex_77w* a K720 interage com a L7, o E897 com os resíduos T1 e R2, o E709 com a R6 e o E893 estabelece contactos polares com a R2 da sequência peptídica. Tendo isto em conta, apenas no *complex_36w* a K720 estabelece as interações-chave descritas na literatura [17, 22, 94]. Assim, tendo por base os resultados *in silico* obtidos, a estrutura do *complex_36w* parece ser a mais promissora.

Da pré-seleção realizada, entre as dez melhores poses do **pep9** previstas pelo HADDOCK, a pose deste péptido no *complex_33w*, no *complex_36w* e no *complex_169w* revelam maior potencial em termos de RMSD e HADDOCK-*score*. Estes três complexos apresentam valores de RMSD próximos entre si, sendo que o *complex_33w* é a estrutura que apresenta um melhor HADDOCK-*score* em relação aos restantes. Neste complexo, os resíduos do recetor que participam na interação são a K720, a V730, o D731, o E893, o E897 e a Q902. A K720 interage com os resíduos F8, R10 e S11 do péptido. Os resíduos V730 e o D731 do recetor interagem com a R5 da cadeia peptídica, enquanto que o E893 interage com a R6. O E897 estabelece contactos polares com os resíduos T1, R2 e A3 do pep9. Por fim, a Q902 interage com a T1 do péptido. Logo, as interações-chave são estabelecidas. A conjugação do valor do RMSD, HADDOCK-score e os contactos polares estabelecidos conduzem à seleção da estrutura do *complex_33w* como a mais promissora.

Relativamente aos resultados do *docking* do **Pep10**, a estrutura do *complex_5w* é a mais próxima em termos de conformação à estrutura do péptido co-cristalizado em 4OEZ. Este complexo revela igualmente ser promissor em relação ao HADDOCK-*score*. Neste complexo, os resíduos do recetor que participam na interação com o péptido são a K720, a R726, a Q738, o E893 e o E897. Os contactos polares estabelecidos neste complexo são entre: a K720 e a L7, a R726 e a S11, a Q738 e a K1, o E893 e a R2 e, por fim, o E897 e a K1 e a R2. Assim, a estrutura mais promissora parece ser a estrutura do *complex_5w*.

A estrutura que melhor representa a interação do **Pep11** com o RA demonstra ser a estrutura do *complex_193w*. Esta estrutura foi selecionada como a mais promissora em termos de RMSD, HADDOCK-*score* e interações polares. As interações polares estabelecidas são entre a K720 do recetor e os resíduos L7 e S11 do péptido, a R726 e a S11, o D731 e a R5 e, finalmente, o E893 e o E897 interagem com o resíduo R2 do péptido.

A estrutura mais promissora resultante do *docking* do **pep13** com o domínio LBD do RA selecionada foi a estrutura do *complex_34w*, tendo por base um compromisso entre o complexo com melhor HADDOCK-score, menor RMSD e a presença das interações-chave. Este complexo apresenta um RMSD inferior aos restantes e, por conseguinte, esta conformação é a que revela maior proximidade com a conformação experimental do cristal. No *complex_34w* são estabelecidos contactos polares com os resíduos do recetor E709, E893 e E897. Quer o resíduo E709, quer o E893 interagem com a R2, enquanto que o E897 estabelece interações polares com os aminoácidos R2, A3 e F1 da cadeia peptídica. As interações-chave que envolvem o E897 estão concordantes com a literatura [17, 22, 94].

A pose do **Pep15** do *complex_48w* revela ser a pose mais promissora não só por ser a que apresenta um menor RMSD, mas também uma menor energia de ligação envolvida (melhor HADDOCK-*score*). O péptido do *complex_48w* estabelece contactos polares com os resíduos K720, R726, D731, E893

e E897 do RA. A K720 estabelece ligações polares com a F8 e com a L7 da sequência peptídica, enquanto que o E897 interage com a R1, R2, A3 e F4. Além disto, os resíduos do recetor R726, D731 e E893 interagem com os resíduos do péptido Y9, R1 e R2, respetivamente.

Relativamente ao **Pep17**, o *complex_9w* é o que revela melhor HADDOCK-*score* e um menor RMSD. Os resíduos do RA que participam na interação são o Q717, D731 e o E897. Em relação aos resíduos Q717 e D731, estes estabelecem contactos polares com os resíduos S11 e R5, respetivamente. O resíduo E897 interage com os resíduos T1, R2, W3 e F4 do péptido. Estas últimas interações estabelecidas estão de acordo com as interações-chave que envolvem o resíduo E897 descritas na literatura. Logo, os resultados obtidos do *docking* deste péptido sugerem que o *complex_9w* parece ser a melhor estrutura.

O complexo mais promissor, representativo da interação do **Pep19** com o RA, revela ser a estrutura do *complex_2w*. Este complexo é o que apresenta maior potencial devido ao facto de apresentar um melhor HADDOCK-*score* e um menor RMSD. As interações-chave estão presentes. Os contactos polares estabelecidos envolvem o resíduo K720 com a F8 e com a L7, o E897 com a R2, A3 e com a F4 e, por fim, o E893 com o resíduo R2 do péptido.

Finalmente, a estrutura do *complex_11w* parece ser a melhor estrutura dos estudos de *docking* do **Pep20**. Este complexo revela um melhor HADDOCK-*score* e, por isso, em termos de energia de ligação a pose do pep20 nesta estrutura é mais favorável e estável. Relativamente às interações estabelecidas, os resíduos essenciais estão presente e observa-se as interações-chave descritas na literatura [17, 22, 94]. Os resíduos do RA envolvidos na interação são a K720, o D731, a Q733, o E893 e o E897. A K720 estabelece contactos polares com a F8, o D731 com a R5, a Q733 interage com a Y9 e o E897 com o W1, com a R2 e com a R3.

4.4.4 Resultados finais

Conhecendo as melhores complexos do *docking* dos novos CPPs desenvolvidos, realizaram-se estudos de dinâmica molecular com os três péptidos mais promissores provenientes cada um das famílias 1, 2 e 3. Este estudo permite um conhecimento mais aprofundado da estabilidade dos péptidos, das interações estabelecidas e dos principais determinantes estruturais envolvidos na interação e estabilidade dos péptidos no domínio LBD do RA.

O critério aplicado na seleção dos três péptidos com maior potencial (mais promissores) para estudos de dinâmica molecular teve por base o HADDOCK-*score*, a percentagem de resíduos do motivo estrutural do péptido que interagem com o domínio LBD do RA, a quantidade de resíduos da sequência peptídica que interagem, o RMSD, as interações polares estabelecidas entre o péptido e o domínio LBD do RA, a capacidade e a eficiência de internalização celular dos péptidos. Os resultados finais com os complexos mais promissores, resultantes do estudo de *docking* efetuado no HADDOCK, apresentam-se sumarizados nas tabelas 4.6, 4.7 e 4.8 de acordo com as famílias 1, 2 e 3, respetivamente.

Tabela 4.6: Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 1: % de resíduos do motivo estrutural do péptido que interagem com o domínio LBD do RA, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos.

Péptido ID	Complexo	% Interações	Resíduos	HADDOCK	RMSD	CPP	MLCPP	Interações Polares
Pep58	*complex_44w*	80	RRKFRLFYKRK	-121,8	0,356	0,982	0,568	V713-K11 K720-K11 K720-Y8 D731-R1 D731-R5 E893-K3 E897-R2 E897-K3 E897-F4
Pep44	*complex_63w*	80	RRRFRLFYKRK	-121,0	0,540	0,986	0,544	L712-K11 K720-Y8 E893-W6 E893-R10 E897-R2
Pep57	*complex_92w*	80	WRIFRLFYKRK	-118,0	0,509	0,952	0,544	R726-Y8 Q733-Y8 E897-W1 E897-R2 E897-I3 E902-W1
Pep59	*complex_48w*	80	RRIFRLFYKRC	-114,8	0,417	0,908	0,581	R726-Y8 Q733-Y8 D731-R5 Y739-R1 E893-R2 E897-R1 E897-R2 E897-I3
Pep52	*complex_62w*	80	RRWFRLFYKRK	-113,2	0,598	0,984	0,579	K720-F7 D731-R5 E893-R2 E897-R1 E897-R2
Pep54	*complex_33w*	80	CRIFRLFYKRK	-112,9	0,346	0,906	0,556	V713-K11 K720-Y8 K720-K11 D731-R5 Q733-Y8 E897-C1 E897-R2 E897-I3 E897-F4
Pep43	*complex_169w*	100	RRIFRRFYKRK	-108,7	0,530	0,979	0,595	D731-R1 E893-R6 E897-R1 E897-R2 E897-I3
Pep48	*complex_28w*	80	RRIFRWFYKRK	-105,6	0,593	0,974	0,617	K720-F7 K720-R10 R726-K11 Q733-K11 E893-W6 E897-R1 E897-R2 E897-I3
Pep53	*complex_181w*	100	RRIFRLFYPRK	-101,4	0,568	0,945	0,557	V713-R2 K720-L6 D731-K11 E893-R1 E893-R5 E897-R1
Pep51	*complex_59w*	100	RRIFRKFYKRK	-98,2	0,479	0,974	0,610	D731-R1 E893-K6 E897-R1 E897-R2 E897-I3

Tabela 4.7: Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 2: % de resíduos do motivo estrutural do péptido que interagem, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos.

Péptido ID	Complexo	% Interações	Resíduos	HADDOCK	RMSD	CPP	MLCPP	Interações Polares
Pep23	*complex_119w*	100	DRRFRRLYTRS	-113,9	0,403	0,887	0,490	K717-S11 K720-Y8 R726-Y8 V730-R5 D731-R5 Q738-D1 E893-R3 E893-R6 E897-R3 E897-F4
Pep26	*complex_154w*	100	DRAFRRLYKRS	-107,0	0,337	0,839	0,450	K720-Y8 K720-S11 D731-R5 Q733-Y8 E893-R6 E897-R2 E897-A3 Q902-D1
Pep33	*complex_14w*	100	DRAFRRLYTRP	-105,6	0,374	0,796	0,455	K720-L7 K720-Y8 R726-Y8 D731-R5 E893-R2 E897-R2 E897-A3
Pep35	*complex_44w*	80	DRWFRRLYTRS	-103,9	0,640	0,828	0,471	K720-L7 K720-S11 R726-Y8 R726-S11 E897-D1 E897-R2
Pep24	*complex_68w*	100	DRAFRRLYRRS	-103,7	0,353	0,878	0,458	K720-Y8 K720-S11 R726-Y8 D731-R5 Q733-Y8 Q738-D1 E893-R6 E897-R2 E897-A3 E897-F4 E902-D1
Pep25	*complex_35w*	80	DRAFRRLYTRR	-103,5	0,388	0,852	0,442	K720-Y8 K720-R11 D731-R5 Q733-Y8 Q738-D1 E893-R2 E897-R2 E897-A3 E902-D1
Pep28	*complex_69w*	100	DRIFRRLYTRS	-100,3	0,717	0,806	0,515	V713-S11 K720-Y8 K720-S11 D731-R5 Q733-Y8 E893-R6 E893-R10 E897-R6
Pep29	*complex_65w*	80	DRAFRRLYPRS	-87,9	0,517	0,786	0,437	K709-R6 K720-L7 E893-R2 E897-D1 E897-R2
Pep32	*complex_23w*	80	RRAFRRLYTRS	-96,0	0,468	0,969	0,432	K720-Y8 K720-S11 D731-R1 Q733-Y8 E897-A3 E897-F4
Pep31	*complex_78w*	80	DRAFRRLYTRC	-90,2	0,387	0,760	0,481	K720-Y8 Q733-Y8 E897-R2 E897-A3 E897-F4

Tabela 4.8: Resultados obtidos dos complexos mais promissores do HADDOCK, resultantes do *docking* dos péptidos da família 3: % de resíduos do motivo estrutural do péptido que interagem, resíduos da interface que interagem e os que não interagem (representados a vermelho), HADDOCK-*score*, RMSD, probabilidade do péptido ser um bom CPP e a sua eficiência de internalização celular e os contactos polares estabelecidos.

Péptido ID	Complexo	% Interações	Resíduos	HADDOCK	RMSD	CPP	MLCPP	Interações
Pep20	*complex_11w*	100	WRAFRRLFYPS	-126,7	0,405	0,814	0,562	K720-F8 E897-W1 D731-R5 E897-R3 Q733-Y9 E897-R2
Pep15	*complex_48w*	100	RRAFRRLFYPS	-117,6	0,384	0,884	0,481	K720-L7 K720-F8 R726-Y9 D731-R1 E893-R2 E897-R1 E897-A3 E897-F4 E897-R2
Pep19	*complex_2w*	80	CRAFRRLFYPS	-114,6	0,376	0,731	0,518	K720-F8 K720-L7 E893-R2 E897-R2 E897-A3 E897-F4
Pep9	*complex_33w*	100	TRAFRRLFYRS	-108,0	0,365	0,911	0,480	K720-F8 K720-R10 K720-S11 V730-R5 D731-R5 E893-R6 E897-T1 E897-R2 E897-A3 Q902-T1
Pep3	*complex_64w*	80	TRRFRRLFYPS	-107,7	0,386	0,918	0,479	K720-F8 R726-Y9 D731-R5 E893-R2 E893-R3 E897-T1 E897-R2 E897-R3 E897-F4
Pep10	*complex_5w*	80	KRAFRRLFYPS	-104,4	0,637	0,855	0,450	K720-L7 R726-S11 Q738-K1 E893-R2 E897-K1 E897-R2
Pep17	*complex_9w*	100	TRWFRRLFYPS	-103,4	0,424	0,859	0,594	D731-R5 E897-T1 E897-R2 E897-W3 E897-F4 K717-S11
Pep11	*complex_193w*	80	TRIFRRLFYPS	-99,8	0,624	0,809	0,594	K720-L7 K720-S11 R726-S11 D731-R5 E893-R2 E897-R2
Pep5	*complex_36w*	80	TRAFRRLFYPR	-97,5	0,597	0,935	0,478	K720-L7 K720-F8 D731-R5 E897-T1
Pep13	*complex_34w*	80	TRAFRRLFYPC	-97,4	0,407	0,796	0,500	K709-R2 E893-R2 E897-R2 E897-A3 E897-F1

Com base nos resultados obtidos, de forma a garantir que os novos péptidos são estruturas mais promissoras em termos de energia de ligação com o domínio LBD do RA do que os péptidos que lhes deram origem, estabeleceu-se um valor limite para o HADDOCK-*score* de -110. A estratégia aplicada na seleção dos três péptidos mais promissores de cada família engloba um compromisso entre o complexo com: i) melhor HADDOCK-*score*; ii) maior percentagem de resíduos do motivo estrutural do péptido que interagem com o domínio LBD do RA; iii) maior quantidade de resíduos da sequência peptídica que interagem com o RA; iv) menor RMSD; v) boa capacidade e eficiência de internalização celular.

Relativamente à percentagem de resíduos do motivo estrutural que interagem, verifica-se que, em

alguns péptidos, essa percentagem corresponde a 80%. Nesses casos, os resíduos da posição 5 e 6 da cadeia peptídica são os resíduos do motivo estrutural que não interagem com o domínio LBD do RA. De facto, estes resíduos não desempenham um papel importante para a interação do RA com péptidos co-reguladores por se encontrarem expostos ao solvente [20]. Avaliando no seu todo os resíduos da sequência peptídica que interagem, quanto maior o número de resíduos a interagir com o domínio LBD do RA, mais favorável será ligação (mais estável). Logo, tendo por base os resultados *in silico* obtidos, as estruturas submetidas aos estudos de dinâmica molecular foram o Pep58, o Pep23 e o Pep19 da família 1, 2 e 3, respetivamente. Para estes três péptidos, verifica-se um acréscimo de afinidade em relação aos péptidos iniciais e uma maior capacidade de internalização celular. Além disto, cada um contempla os três diferentes motivos estruturais estudados o que é relevante para avaliar a influência destes motivos e as diferentes interações estabelecidas.

4.5 Dinâmica molecular

Como referido no capítulo anterior, as simulações de dinâmica molecular foram realizadas com a finalidade de selecionar, de entre o pep19, o pep23 e o pep58, o péptido com maior afinidade de ligação para o RA. Estas simulações foram efetuadas utilizando o *software* GROMACS [89]. As análises das simulações produzidas incluem a visualização da trajetória, o cálculo do RMSD do complexo e dos péptidos, do RMSF, do SASA, das ligações de hidrogénio e, ainda, da energia livre de ligação. De notar que, de forma a visualizar com mais clareza a tendência do comportamento do RMSD (do complexo e do péptido), do SASA e da energia livre de ligação ao longo do tempo, calculou-se a média móvel para eliminar o ruído dos gráficos e suavizar o movimento das curvas obtidas. No caso do RMSD, do RMSF e do SASA, a média móvel foi calculada para uma janela com 100 pontos, o que corresponde à média móvel a cada 2 ns de simulação. Para a energia livre de ligação calculou-se a média móvel para uma janela com 25 pontos, o que corresponde à média móvel a cada 2,5 ns de simulação. Os gráficos originais (com ruído), sem a aplicação da média móvel, estão disponíveis no apêndice B.

Posto isto, os resultados *in silico* obtidos das simulações produzidas apresentam-se descritos nos capítulos seguintes.

4.5.1 RMSD e trajetória

A análise do RMSD permite avaliar a estabilidade e a convergência global dos sistemas estudados durante a simulação. Através desta análise é possível monitorizar as simulações produzidas. O cálculo do RMSD foi efetuado para os átomos Cα das estruturas e foi realizada através do GROMACS [89]. Tendo por base a aplicação da média móvel de 2 ns, os resultados do RMSD do complexo e do péptido ao longo dos 100 ns de simulação apresentam-se nas figuras 4.1 e 4.2, respetivamente.

Com base nos gráficos da figura 4.1, verifica-se que o RMSD dos três complexos encontra-se no mesmo intervalo de valores (0$<$RMSD$\leq$0,2 nm). Os resultados obtidos sugerem que os complexos são estáveis, uma vez que o valor do RMSD não apresenta oscilações acentuadas durante a simulação. Dos

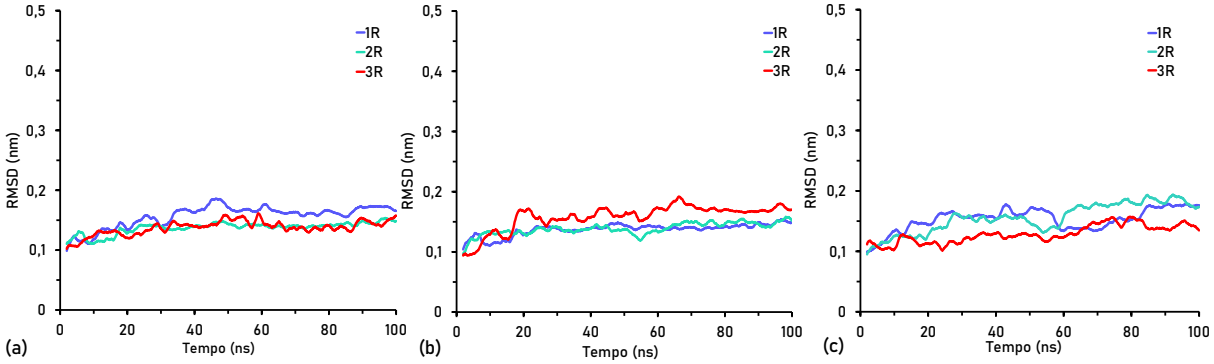


Figura 4.1: RMSD dos três complexos calculado em relação aos átomos Cα das estruturas. (a) RMSD do complexo domínio LBD-pep19; (b) RMSD do complexo domínio LBD-pep23; (c) RMSD do complexo domínio LBD do RA e pep58.

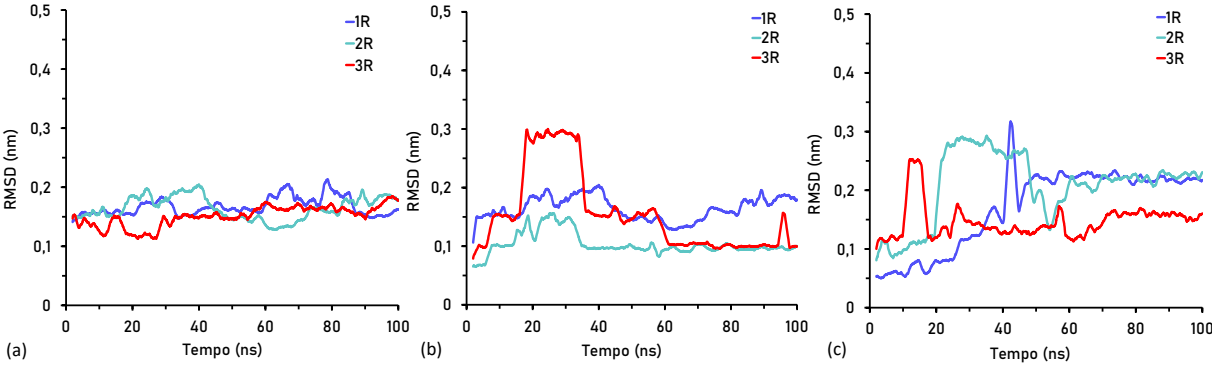


Figura 4.2: RMSD dos péptidos calculado em relação aos átomos Cα das estruturas. (a) RMSD do pep19; (b) RMSD do pep23; (c) RMSD do pep58.

três sistemas analisados, o sistema do pep58 demonstra um maior afastamento entre as três réplicas e experiencia maiores variações ao longo da dinâmica observadas principalmente na 1ª e 2ª réplica (Figura 4.1c). O sistema do pep19 revela uma maior proximidade entre a 2ª e 3ª réplica (Figura 4.1a), enquanto que no sistema do pep23 o comportamento da 1ª e 2ª réplica são mais semelhantes entre si (Figura 4.1b). Assim, tendo em consideração os gráficos da figura 4.1, os sistemas do pep19 e do pep23 demonstram ser tendencialmente mais estáveis do que o sistema do pep58. Além disto, verifica-se, ainda, uma tendência do sistema do pep19 e do pep23 para convergir a partir, aproximadamente, dos 65 ns. Esta tendência é reprodutível em todas as réplicas. No sistema do pep58 o sistema converge depois dos 80 ns.

O cálculo do RMSD dos péptidos (Figura 4.2) fornece uma visão global da estabilidade de cada um individualmente. Face aos resultados obtidos, observa-se que tendencialmente o pep19 apresenta um comportamento mais estável durante a simulação (Figura 4.2a). Neste péptido não se verificam variações acentuadas ao longo das três réplicas. Como é possível verificar através dos gráficos da figura 4.2b e 4.2c, o pep23 e o pep58 sofrem maiores alterações conformacionais ao longo da dinâmica. Nestes dois péptidos todos os picos verificados apresentam um RMSD$<$0,4, observando-se que o pep58 atinge valores superiores de RMSD durante os 100 ns de simulação. No pep23 as oscilações mais acentuadas são experienciadas apenas na 3ª réplica, enquanto que no pep58 são reprodutíveis

em todas as réplicas.

Para uma melhor compreensão do comportamento dos péptidos estudados, complementou-se a análise do RMSD com a análise da trajetória através PyMOL. Pela análise da trajetória do pep19, observou-se que este péptido não sofre grandes alterações conformacionais ao longo da simulação de dinâmica molecular, tal como seria expectável tendo em consideração os valores obtidos do RMSD (Figura 4.2a). As variações mais significativas ocorridas correspondem a uma maior flexibilidade conformacional dos dois últimos resíduos de aminoácidos da cadeia peptídica. A partir dos 95 ns, o pep19 adota uma conformação mais estável e constante na 2ª e 3ª réplica e com menores flutuações dos últimos resíduos, enquanto que na 1ª réplica o péptido alcança uma conformação mais estável depois dos 90 ns de simulação. Em relação ao pep23, no primeiro pico presente entre os 15-35 ns (Figura 4.2b) e no segundo pico verificado entre os 92-96 ns (Figura 4.2b) da 3ª réplica, observa-se pela trajetória uma flutuação dos três últimos resíduos do péptido, marcada por uma variação mais acentuada dos dois últimos resíduos da sequência peptídica. Estes resíduos finais colocam-se numa outra posição momentaneamente. No intervalo de tempo entre os 40-60 ns (Figura 4.2b), o pep23 experiencia uma conformação muito semelhante entre a 1ª e 2ª réplica. A partir dos 60 ns até aos 70 ns, o pep23 adota uma conformação constante e muito próxima entre as três réplicas. Após os 70 ns, o comportamento do péptido na 1ª réplica afasta-se ligeiramente do das restantes réplicas devido essencialmente aos três resíduos finais que experienciam maiores oscilações conformacionais. Na 2ª e 3ª réplica a conformação do péptido mantêm-se próxima entre si até aos 92 ns. A pose do péptido mais estável e constante ao longo das três réplicas volta a ocorrer na fase final da simulação, mais particularmente depois dos 96 ns. Relativamente ao pep58, através da inspeção visual realizada, verificou-se que os picos ocorridos (Figura 4.2c) são igualmente devido a oscilações dos três últimos resíduos da cadeia peptídica, sendo estas oscilações mais acentuadas nos dois últimos resíduos. Neste péptido observa-se, ainda, uma alteração conformacional mais acentuada dos dois primeiros resíduos da sequência peptídica. Este péptido sofre alterações conformacionais reproduzidas em todas as réplicas até, aproximadamente, os 65 ns. A conformação adotada pelo péptido, após os 65 ns, é próxima entre a 1ª e 2ª réplica.

Em suma, pela análise combinada do RMSD e da trajetória, as oscilações verificadas no pep23 e no pep58 devem-se essencialmente a variações conformacionais dos resíduos das extremidades da sequência de aminoácidos, verificando-se que essas oscilações são mais acentuadas no pep58.

4.5.2 RMSF

O RMSF foi igualmente calculado através do GROMACS e tendo em consideração os átomos Cα das estruturas. O RMSF é uma análise individual por resíduo que permite avaliar a flutuação espacial dos mesmos ao longo da simulação. Assim, através desta análise é possível identificar a região estrutural da cadeia peptídica que sofreu maiores alterações conformacionais no decorrer dos 100 ns de simulação e aferir determinadas interações estabelecidas entre o péptido e o domínio LBD do RA. A figura 4.3 apresenta os valores do RMSF dos três sistemas estudados, sendo que os resíduos 670-918 correspondem ao RA (numeração concordante com a numeração do PDB), enquanto que os resíduos

919-929 são relativos ao péptido. Para uma melhor compreensão dos resultados obtidos do RMSF dos péptidos, realizou-se a média das três réplicas para cada péptido (Figura 4.4).

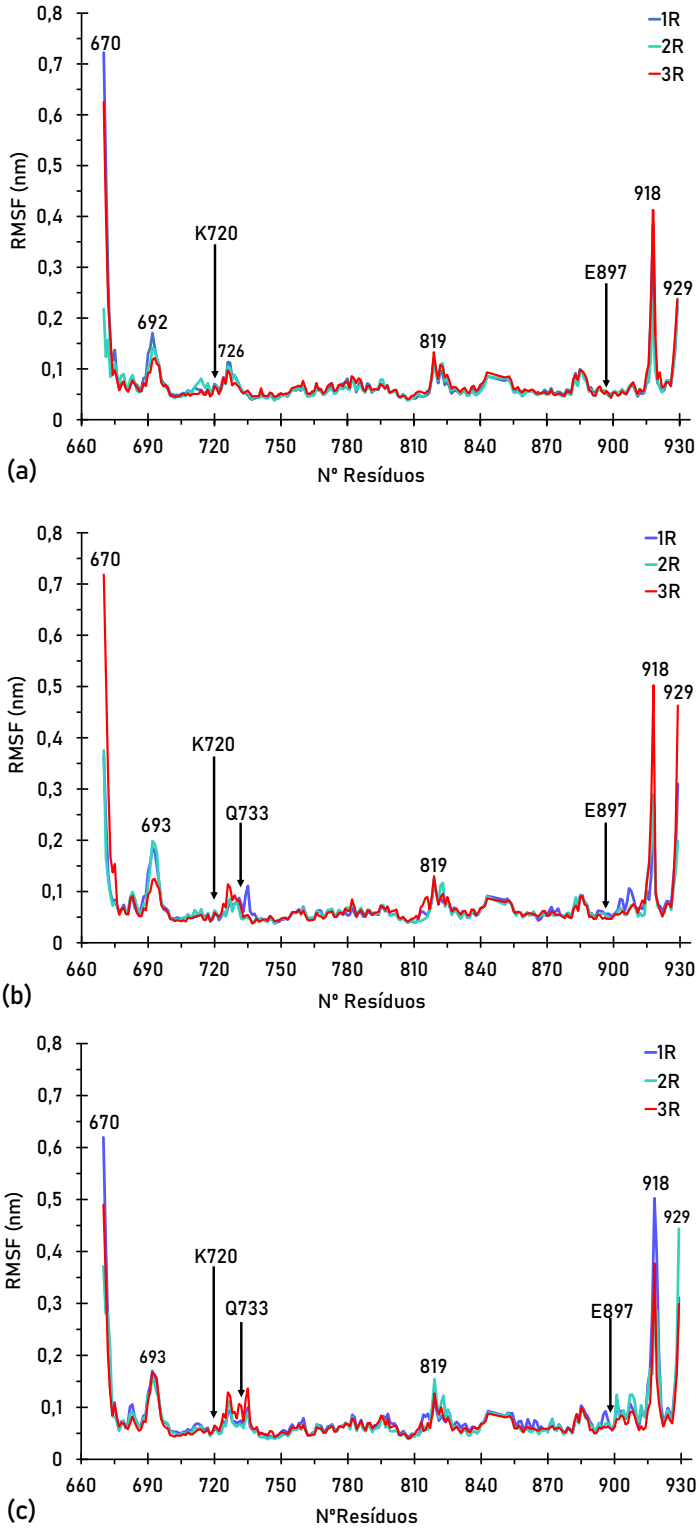


Figura 4.3: RMSF dos complexos (recetor-péptido) calculado em relação aos átomos Cα. Resíduos do recetor: 670-918; Resíduos do péptido: 919-929 (a) RMSF do recetor-pep19; (b) RMSF do recetor-pep23; (c) RMSF do recetor-pep58.

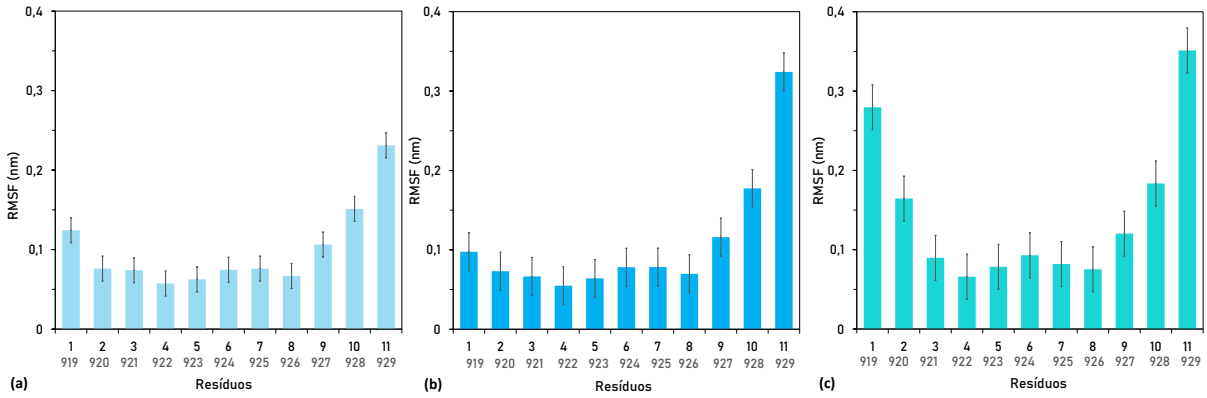


Figura 4.4: Média das três réplicas do RMSF dos péptidos. (a) Média das três réplicas do RMSF pep19; (b) Média das três réplicas RMSF do pep23; (c) Média das três réplicas RMSF do pep58.

Tendo em conta os resultados obtidos, verifica-se que os resíduos do recetor 670-672, 688-695, 723-734, 818-823, 916-918 experienciam maiores flutuações (Figura 4.3a, b e c). Este comportamento pode ser justificado devido a facto destes resíduos estarem presentes em regiões do recetor associadas a zonas de *loops*, caracterizadas por apresentarem maiores oscilações conformacionais e uma menor estabilidade associada. Contrariamente, os resíduos do RA que desempenham um papel importante para a interação (E897, K720 e a Q733) experienciam menores flutuações. Como é possível visualizar através dos gráficos da figura 4.3, os resíduos mencionados encontram-se presentes na base dos picos. Por isso, uma vez que estes resíduos apresentam um RMSF reduzido, não apresentam uma grande flexibilidade conformacional ao longo da dinâmica. Logo, estes resultados obtidos sugerem que os resíduos E897, a K720 e a Q733 são mais estáveis, o que pode ser um indicativo de que estabelecem ligações de hidrogénio no decorrer da simulação e, por isso, são estabilizados por essas mesmas interações.

Em relação aos péptidos, pela análise do RMSF observa-se que os cinco aminoácidos do motivo estrutural (resíduos 922-926) da cadeia peptídica são os que apresentam menores flutuações (Figura 4.3 e 4.4). Estes resíduos encontram-se no patamar dos picos dos resíduos 919 e 929 do péptido (Figura 4.3). Logo, a menor flutuação experienciada por estes resíduos é indicativo de uma maior estabilidade dos mesmos durante a simulação. Este comportamento é verificado para os três péptidos estudados. Estes resultados sugerem que a região do motivo estrutural do péptido é a região que estabelece mais interações com o RA, uma vez que é a parte estrutural que se encontra mais estável ao longo da dinâmica. De entre os cinco aminoácidos do motivo estrutural, os resíduos 922 e 926, correspondentes aos aminoácidos presentes nas posições 4 e 8 da sequência peptídica, revelam um RMSF inferior aos restantes aminoácidos que constituem o motivo (Figura 4.4). Além disto, em todos os péptidos verifica-se que os resíduos que apresentam maiores flutuações são essencialmente os resíduos iniciais e os resíduos finais da cadeia peptídica (Figura 4.4). Estas flutuações, experienciadas por estes resíduos, são mais acentuadas no pep23 e no pep58 (Figura 4.3b,c e 4.4b,c). Através da figura 4.4b, observa-se uma maior flexibilidade conformacional dos três últimos resíduos (resíduos 927-929) do pep23, enquanto que pela figura 4.4c verifica-se uma flutuação mais acentuada dos dois primeiros (resíduos 919-920) e dos três últimos resíduos (resíduos 927-929) do pep58. Este compor-

tamento sustenta as análises da trajetória e do RMSD dos péptidos descritas anteriormente. De facto, as oscilações acentuadas verificadas no gráfico do RMSD do pep23 e do pep58 (Figura 4.2b,c) são devidas a uma maior flexibilidade conformacional dos três últimos resíduos, no caso do pep23, e dos dois primeiros e três últimos resíduos, no caso do pep58. Assim, como seria expectável, estes resíduos assumem valores de RMSF superiores.

4.5.3 Área da interface acessível ao solvente

A área da interface acessível ao solvente (SASA) permite estimar a extensão de interação dos diferentes péptidos com o solvente durante os 100 ns. Este valor foi calculado através do GROMACS [89]. A abordagem efetuada teve por base a subtração da área acessível ao solvente do péptido com a proteína (complexo) e da área acessível ao solvente do péptido sem a proteína. O valor resultante desta diferença corresponde ao valor do SASA do péptido. Assim, esta análise permite identificar o péptido, por comparação com os restantes péptidos estudados, com uma maior e/ou menor área de exposição ao solvente.

Os resultados dos valores do SASA do péptido em interação com o recetor, ao longo dos 100 ns de simulação, apresentam-se no gráfico da figura 4.5. Este gráfico resulta da média das três réplicas realizadas para cada péptido.

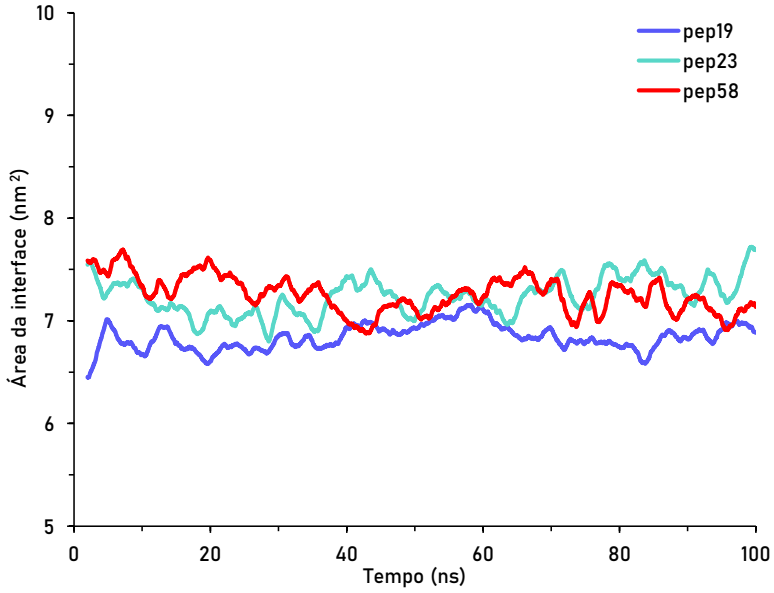


Figura 4.5: Área da interface dos péptidos. (a) Área da interface do pep19; (b) Área da interface do pep23; (c) Área da interface do pep58.

Os resultados obtidos sugerem que o pep19 apresenta uma menor área de exposição ao solvente em relação ao pep23 e ao pep58, ainda que as diferenças, especialmente entre o pep19 e pep23, sejam mínimas. Importa ainda realçar que o pep23, apresenta, a par do pep19 um valor de SASA relativamente constante ao longo da simulação, o que sugere a estabilidade dos péptidos ao longo da DM.

4.5.4 Ligações de hidrogénio

O cálculo das interações de hidrogénio estabelecidas entre o péptido e o RA é mais uma análise que permite averiguar a estabilidade dos sistemas estudados. Quanto maior o número de ligações de hidrogénio estabelecidas entre o domínio LBD do recetor e os péptidos, maior tenderá a ser a estabilidade do complexo. As ligações de hidrogénio foram calculadas utilizando do VMD [92]. O cálculo destas interações através deste programa permite conhecer os resíduos envolvidos na ligação de hidrogénio e a percentagem dessa interação ao longo da dinâmica. Analisaram-se as ligações de hidrogénio a partir de uma ocupância igual ou superior a 20% e reproduzidas em todas as réplicas. Posto isto, os resultados obtidos apresentam-se nas tabelas 4.9, 4.10 e 4.11 para o sistema do pep19, do pep23 e do pep58, respetivamente. Estes valores são resultado da média das três réplicas.

Tabela 4.9: Ligações de hidrogénio estabelecidas entre o domínio LBD do RA e o pep19 (CRAFRRLFYPS).

Dador	Aceitador	Ocupância
R2-Main	E897-Side	41,72% $\pm$ 0,10
A3-Main	E897-Side	37,16% $\pm$ 0,08
F8-Main	K720-Side	27,65% $\pm$ 0,04
S11-Side	K720-Side	32,81% $\pm$ 0,12
R6-Side	E893-Side	31,04% $\pm$ 0,02

Tabela 4.10: Ligações de hidrogénio estabelecidas entre o o domínio LBD do RA e o pep23 (DRRFRRLYTRS).

Dador	Aceitador	Ocupância
R2-Main	E897-Side	43,11% $\pm$ 0,03
R3-Main	E897-Side	44,00% $\pm$ 0,02
R5-Side	D731-Side	23,96% $\pm$ 0,03
S11-Side	K720-Side	35,54% $\pm$ 0,07
Y8-Main	K720-Side	24,53% $\pm$ 0,05
Y8-Side	Q733-Side	54,97% $\pm$ 0,11

Tabela 4.11: Ligações de hidrogénio estabelecidas entre o domínio LBD do RA e o pep58 (RRKFRLFYKRK).

Dador	Aceitador	Ocupância
K3-Main	E897-Side	34,77% $\pm$ 0,05
F4-Main	E897-Side	42,85% $\pm$ 0,03
Y8-Main	K720-Side	24,98% $\pm$ 0,03
Y8-Side	Q733-Side	63,03% $\pm$ 0,01

No pep19, os resíduos do RA que estabelecem interações de hidrogénio são a K720, o E893 e o E897. Assim, estes três resíduos estarão estabilizados por estas interações, não experienciando grandes oscilações conformacionais ao longo da dinâmica, o que sustenta os seus reduzidos valores

de RMSF. No caso da K720, a sua cadeia lateral interage com o *backbone* da F8 (como descrito na literatura) e com a cadeia lateral da S11. A interação entre as cadeias laterais dos resíduos S11 e K720 sugere que a presença da S na posição 11 da sequência peptídica é importante para a interação com este resíduo-chave do RA (K720). Observa-se, ainda, a interação da cadeia lateral do E897 do recetor com o *backbone* da R2 e da A3 do pep19. A interação entre o E897 e a A3 do péptido está concordante com as interações-chave reportadas na literatura e consideradas essenciais para a interação de péptidos com o RA [17, 22, 94]. A ligação estabelecida entre a R2 e o E897 é uma nova interação que não está reportada na literatura e que envolve a participação de um aminoácido presente numa posição da sequência peptídica onde ocorreu uma mutação. Isto demonstra que a presença da R na posição 2 permite uma maior capacidade deste péptido para interagir com o RA. Para além disto, verifica-se que a cadeia lateral do E893 interage com a cadeia lateral da R6, o que pode sugerir que a R nesta posição pode contribuir para a estabilidade do péptido e, consequentemente, para uma melhor interação com o recetor.

O pep23 estabelece um número superior de ligações de hidrogénio com o domínio LBD do RA em relação ao pep19 e ao pep58. No pep23 os resíduos do recetor envolvidos em interações de hidrogénio são a K720, o E897, a Q733 e o D731, enquanto que os do péptido são a R2, a R3, a R5, a Y8 e a S11. A cadeia lateral da S11 e o *backbone* da Y8 do péptido estabelecem uma ligação de hidrogénio com a cadeia lateral da K720 do recetor. A interação estabelecida entre a Y8 e a K720 está concordante com as interações-chave reportadas na literatura [17, 22, 94]. Quer no pep19, quer no pep23 a presença da serina na posição 11 da sequência peptídica parece desempenhar um papel importante, uma vez que quando a K720 não estabelece interação com a Y8, tendencialmente este resíduo do recetor é estabilizado por interação com a S11. Logo, no decorrer da dinâmica, a K720 consegue estabelecer interações de hidrogénio com a S11 ou com a Y8, o que reforça não só a estabilidade do péptido em interação com o local ativo do recetor, como também a importância desses dois resíduos do péptido para a interação. Para além disto, são ainda estabelecidas ligações de hidrogénio entre a cadeia lateral do E897 do recetor e o *backbone* dos resíduos R2 e R3 do péptido (ocupância $\geq$ 40%). A ligação da R3 com o E897 está concordante com a literatura [17, 22, 94]. Tal como no péptido anterior, destaca-se a interação entre a R2 e o E897. Esta interação apresenta uma % de ocupância relevante e é uma interação que envolve um aminoácido presente numa posição da sequência peptídica onde ocorreu uma mutação, o que pode contribuir para uma melhor interação deste péptido com o recetor. Com uma menor ocupância é estabelecida uma ligação de hidrogénio entre a cadeia lateral do D731 e a cadeia lateral da R5 do péptido (ocupância $>$ 20%). Esta interação não é verificada nos restantes péptidos estudados. Finalmente, a cadeia lateral da Q733 estabelece uma interação de hidrogénio com a cadeia lateral da Y8 do péptido, sendo esta ligação concordante com as interações-chave reportadas na literatura [17, 22, 94]. Esta última revela ser a ligação mais estável (ocupância $>$ 50%) ao longo dos 100 ns de simulação, o que permite evidenciar a importância da Y na posição 8 da sequência para uma melhor interação com o recetor. De facto, como descrito na literatura, a presença da Y8 da cadeia peptídica, ao invés da F8, promove uma maior afinidade para o RA [17, 22, 94].

Em relação ao pep58, de acordo com os resultados obtidos, os resíduos do recetor que participam

nas interações de hidrogénio são a K720, o E897 e a Q733. Os resíduos do péptido que estabelecem ligações de hidrogénio são a K3, a F4 e a Y8. A cadeia lateral da K720 do RA estabelece ligações de hidrogénio com o *backbone* da Y8, como reportado na literatura. Contrariamente aos restantes dois péptidos, no pep58 a interação do aminoácido da posição 11 da sequência peptídica com a K720 do RA ocorre em menos de 20% da simulação. Logo, a presença da serina na posição 11, ao invés da lisina, parece potenciar a interação do péptido com o recetor, particularmente com o resíduo-chave K720. Além disto, como verificado no pep23, no pep58 observa-se igualmente a interação entre a cadeia lateral da Y8 do péptido e a cadeia lateral da Q733 do recetor. Em ambos os péptidos (pep23 e pep58) esta interação revela ser a mais estável ao longo da simulação. No caso do pep58, verifica-se que esta interação ocorre em mais de 60% da simulação. Observa-se que a presença da Y na posição 8, ao invés da F, não revela um papel significativo para a interação com a K720, uma vez que este resíduo do recetor interage com o *backbone* dos aminoácidos da posição 8. No entanto, para a interação com a Q733, a presença da Y8 já desempenha um papel relevante. De facto, o grupo hidroxilo presente na cadeia lateral da Y demonstra acrescentar valor à estabilidade do péptido, uma vez que existe uma interação muito estável estabelecida no pep23 e no pep58. Por fim, é estabelecida uma ligação de hidrogénio entre a cadeia lateral do resíduo E897 e a cadeia principal da K3 e da F4 do pep58. Estas últimas estão de acordo com as interações descritas na literatura [17, 22, 94].

Em suma, os resultados obtidos permitem concluir que o pep23 é o péptido que estabelece um maior número de ligações de hidrogénio com o RA. As interações de hidrogénio entre a K720 e o *backbone* do aminoácido da posição 8 e a cadeia lateral do aminoácido da posição 11 são verificadas nos pep19 e no pep23. A a presença da serina na posição 11 (pep19 e pep23) demonstra ser mais favorável à interação do que a lisina nessa mesma posição (pep58). Para além disto, tanto no pep58 como no pep23 é possível verificar a interação da Y8 com a Q733, enquanto que no pep19 esta interação não ocorre pela existência de uma F nessa posição. Logo, é possível inferir que a Y na posição 8, ao invés da F, promove uma maior estabilidade e um acréscimo da afinidade do péptido para o domínio LBD do RA, uma vez que para além da Y8 estabelecer uma ligação de hidrogénio com a K720, este resíduo do péptido estabelece simultaneamente uma ligação de hidrogénio com a Q733. Esta última interação releva ser a mais estável.

4.5.5 Energia livre de ligação

O cálculo da energia livre de ligação permite determinar qual é a energia dos diferentes péptidos em interação com a proteína (i.e., RA). Através desta análise é possível avaliar qual o péptido que apresenta uma energia mais favorável a interagir e uma maior estabilidade. De um modo geral, quanto menor for a energia livre de ligação mais estável será o péptido. Os resultados obtidos da energia livre de ligação dos três sistemas estudados apresentam-se na figura 4.6. Estes resultados resultam da média das três réplicas efetuadas.

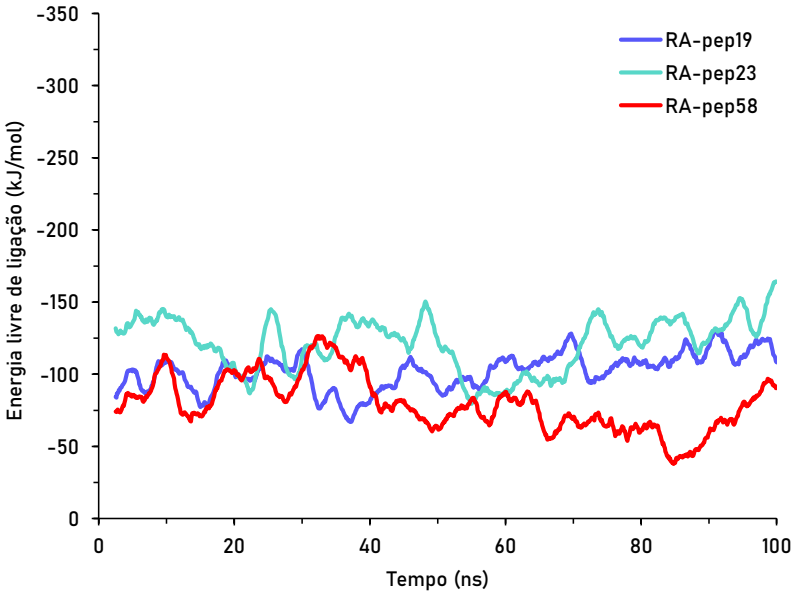


Figura 4.6: Energia livre de ligação (kJ/mol) do sistema do pep19, do pep23 e do pep58.

De acordo com a figura 4.6 verifica-se que a energia livre de ligação é mais favorável no pep23 do que no pep19 e no pep58. O valor médio da energia de ligação do sistema do pep19 é igual a -102 kJ/mol $\pm$ 21, no sistema do pep23 corresponde a -122 kJ/mol $\pm$ 26 e, por fim, no sistema do pep58 é -80 kJ/mol $\pm$ 27. De facto, no pep23 observam-se interações de hidrogénio adicionais que não ocorrem nos restantes péptidos. Em relação ao pep19, o pep23 estabelece uma ligação de hidrogénio adicional entre a Y8 e o Q33 e entre o D731 e a R5. Relativamente ao pep58, no pep23 estabelece-se uma interação de hidrogénio entre a R5 e o D731, a R2 e o E897 e a S11 e o Q733 que não são verificadas no pep58. O estabelecimento de um maior número de interações entre o pep23 e o domínio LBD do RA sugere uma maior estabilidade do péptido em interação com a proteína, o que conduz a uma menor energia livre de ligação envolvida nessa interação.

Capítulo 5

Conclusões e perspetivas futuras

O objetivo do trabalho proposto consistiu em desenvolver pequenos novos péptidos com uma maior afinidade de ligação para o RA e simultaneamente com uma melhor capacidade de internalização celular (CPPs) através da integração de diferentes ferramentas computacionais. Para atingir este objetivo, utilizou-se como ponto de partida três sequências peptídicas (Pep-A, Pep-B e Pep-C) baseadas nos motivos estruturais FxxLF, FxxLY e FxxFY, que demonstram afinidade para o RA. Partindo do Pep-A, do Pep-B e do Pep-C foram criadas as famílias de novos péptidos 1, 2 e 3, respetivamente.

As sequências dos péptidos Pep-A, Pep-B e do Pep-C foram sujeitas a três mutações, no sentido de aumentar a permeabilidade dos péptidos ao nível celular. Através do *web-server* CellPPD foram desenvolvidos novos CPPs com maior potencial ao nível celular em relação aos péptidos que lhes deram origem. Foram previstas 20 novas sequências a partir das sequências do Pep-A, do Pep-B e do Pep-C. Os resultados obtidos demonstraram que os péptidos previstos são bons CPPs, sendo que as sequências peptdídicas ricas no aminoácido arginina demonstram uma maior capacidade de penetrabilidade através da membrana celular. Das 20 sequências de cada família, foram selecionadas os 10 péptidos com uma maior eficiência de internalização celular, através do MLCPP. Após a previsão da estrutura 3D destes péptidos, estas estrutura foram submetidas a estudos de *docking* no *web-server* HADDOCK. Os três melhores péptidos advindos deste estudo foram o pep58 (RRKFRLFYKRK), o pep23 (DRRFRRLYTRS) e o pep19 (CRAFRRLFYPS) das famílias 1, 2 e 3, respetivamente. A estratégia aplicada na seleção destes três péptidos teve por base a conjugação de diferentes parâmetros: HADDOCK-*score*, RMSD, contactos polares entre o domínio LBD do RA e os péptidos, os resíduos da interface que interagem, a probabilidade associada à previsão do péptido ser um bom CPP e à sua *uptake efficiency*.

O pep19, o pep23 e o pep58 foram sujeitos a estudos de dinâmica molecular. Face aos resultados obtidos é possível concluir que as mutações efetuadas nas sequências peptídicas não alteraram as interações reportadas na literatura entre os resíduos-chave do RA (K720, Q733 e E897) e os resíduos-chave do motivo estrutural do péptido. Conclui-se ainda que existe um acréscimo de interações quando a Y está presente na posição 8 da sequência peptídica, ao invés da F, o que está concordante com a literatura [17, 22, 94]. Quer no pep23, quer no pep58 a contribuição da Y na posição 8 permite esta-

belecer uma ligação de hidrogénio adicional, em relação à F na mesma posição, bastante estável que envolve a participação da Q733 do recetor. Para além destas interações, importa mencionar a interação estabelecida pela R2 da sequência do pep19 e do pep23, e a interação da A3, da R3 e da K3 do pep19, do pep23 e do pep58, respetivamente. Estas interações apresentam uma ocupância significativa ao longo da DM e a presença dos aminoácidos das posições 2 e 3 da sequência peptdídica parece conferir potencial aos novos péptidos desenvolvidos. Verifica-se que tendencialmente a presença da serina na posição 11, ao invés da lisina, parece ser mais favorável para a interação dos péptidos com o RA. De entre os três péptidos, o pep23 é o que estabelece um maior número de interações de hidrogénio com o domínio LDB do RA. A presença das ligações de hidrogénio que envolvem a R5, a S11 e a Y8 do pep23 demonstram melhorar a capacidade de interação deste péptido com o domínio LBD do recetor. Observa-se, ainda, que este péptido é o que apresenta tendencialmente uma menor energia livre de ligação ao longo da simulação. Logo, face às análise concretizadas do RMSD, do RMSF, do SASA, das ligações de hidrogénio e da energia livre de ligação é possível inferir que o pep23 demonstra ser o péptido mais promissor.

Futuramente, como um método de validação dos resultados *in silico*, os três péptidos mais promissores resultantes dos estudos computacionais (pep19, pep23 e o pep58) serão sintetizados através do método de síntese de péptidos em fase sólida (SPPS). Posteriormente, será avaliada a afinidade de ligação desses péptido para o RA com recurso à técnica denominada por ressonância plasmónica de superfície (SPR). Assim, é possível estabelecer uma correlação dos resultados computacionais obtidos com os estudos *in vitro*. Na figura 5.1 encontra-se um esquema representativo das etapas concretizadas no presente trabalho e das perspetivas futuras do trabalho a desenvolver.

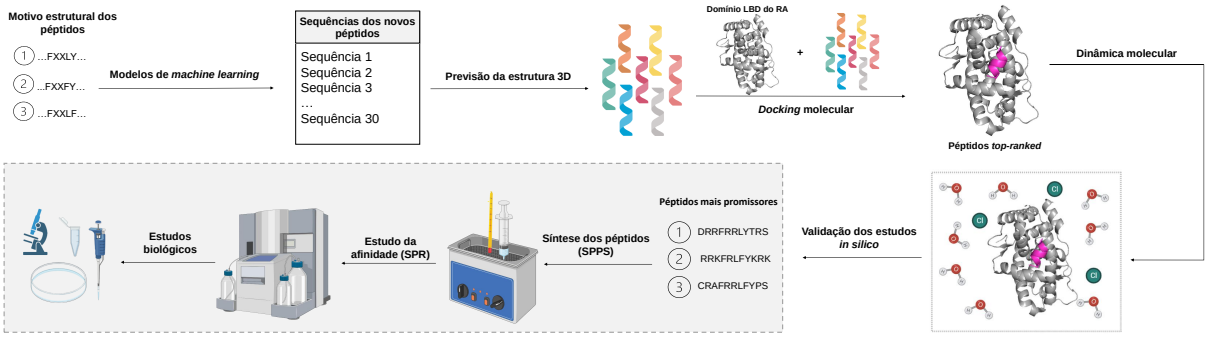


Figura 5.1: Conjunto de etapas desenvolvidas no presente trabalho e perspetivas futuras.

Bibliografia

[1] M. Brumec, M. Sobočan, I. Takač, and D. Arko. Clinical implications of androgen-positive triple-negative breast cancer. *Cancers*, 13(7), 2021.

[2] L. Yin, J.-J. Duan, X.-W. Bian, and S.-c. Yu. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research*, 22(1):1–13, 2020.

[3] M. Ensenyat-Mendez, P. Llinàs-Arias, J. I. Orozco, S. Íñiguez-Muñoz, M. P. Salomon, B. Sesé, M. L. DiNome, and D. M. Marzese. Current triple-negative breast cancer subtypes: Dissecting the most aggressive form of breast cancer. *Frontiers in Oncology*, 11, 2021.

[4] F. Cardoso. *100 perguntas Chave no Cancro da Mama*. 2nd edition. Permanyer Portugal, 2017.

[5] A. Mina, R. Yoder, and P. Sharma. Targeting the androgen receptor in triple-negative breast cancer: Current perspectives. *OncoTargets and Therapy*, 10:4675–4685, 2017.

[6] Y. M. Lee, M. H. Oh, J. H. Go, K. Han, and S. Y. Choi. Molecular subtypes of triple-negative breast cancer: understanding of subtype categories and clinical implication. *Genes and Genomics*, 42 (12):1381–1387, 2020.

[7] M. Núñez Abad, S. Calabuig-Fariñas, M. Lobo de Mena, M. José Godes Sanz de Bremond, C. García González, S. Torres Martínez, J. Á. García-García, V. Iranzo González-Cruz, and C. Camps Herrero. Update on systemic treatment in early triple negative breast cancer. *Therapeutic Advances in Medical Oncology*, 13:1–18, 2021.

[8] L. M. Sutton, D. Cao, V. Sarode, K. H. Molberg, K. Torgbe, B. Haley, and Y. Peng. Decreased androgen receptor expression is associated with distant metastases in patients with androgen receptor-expressing triple-negative breast carcinoma. *American Journal of Clinical Pathology*, 138 (4):511–516, 2012.

[9] N. A. Bagegni, Y. Tao, and F. O. Ademuyiwa. Clinical outcomes with neoadjuvant versus adjuvant chemotherapy for triple negative breast cancer: A report from the national cancer database. *PLoS One*, 14(9):e0222358, 2019.

[10] H. S. Walter and S. I. Ahmed. Targeted therapies in cancer. *Surgery (Oxford)*, 39(4):202–207, 2021.

[11] Y. T. Lee, Y. J. Tan, and C. E. Oon. Molecular targeted therapy: Treating cancer with specificity. *European Journal of Pharmacology*, 834:188–196, 2018.

[12] S. P. Leelananda and S. Lindert. Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry*, 12:2694–2718, 2016.

[13] S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi. Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research*, 38(9):1686–1701, 2015.

[14] X. Lin, X. Li, and X. Lin. A review on applications of computational methods in drug screening and design. *Molecules*, 25(6):1375, 2020.

[15] L. Gerratana, D. Basile, G. Buono, S. De Placido, M. Giuliano, S. Minichillo, A. Coinu, F. Martorana, I. De Santo, L. Del Mastro, M. De Laurentiis, F. Puglisi, and G. Arpino. Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treatment Reviews*, 68(April):102–110, 2018.

[16] T. Proverbs-Singh, J. L. Feldman, M. J. Morris, K. A. Autio, and T. A. Traina. Targeting the androgen receptor in prostate and breast cancer: several new agents in development. *Endocrine-related cancer*, 22(3):R87–R106, 2015.

[17] Androgen receptor: Structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, 36(1):3–23, 2015.

[18] H. Derakhshankhah and S. Jafari. Cell penetrating peptides: A concise review with emphasis on biomedical applications. *Biomedicine & Pharmacotherapy*, 108:1090–1096, 2018.

[19] C.-L. Hsu, Y.-L. Chen, S. Yeh, H.-J. Ting, Y.-C. Hu, H. Lin, X. Wang, and C. Chang. The use of phage display technique for the isolation of androgen receptor interacting peptides with *(F/W)XXL(F/W)* and *FXXLY* new signature motifs. *Journal of Biological Chemistry*, 278(26):23691–23698, 2003.

[20] E. Hur, S. J. Pfaff, E. S. Payne, H. Grøn, B. M. Buehrer, R. J. Fletterick, and U. Schibler. Recognition and accommodation at the androgen receptor coactivator binding interface. *PLoS biology*, 2(9): e274, 2004.

[21] B. He, J. A. Kemppainen, and E. M. Wilson. F XX LF and W XX LF Sequences Mediate the NH 2 -terminal Interaction with the Ligand Binding Domain of the Androgen Receptor *. *Journal of Biological Chemistry*, 275(30):22986–22994, 2000.

[22] C.-l. Hsu, J.-s. Liu, P.-l. Wu, H.-h. Guan, S.-d. Yeh, W.-l. Ma, and C.-j. Chen. Identification of a new androgen receptor (AR) co-regulator BUD31 and related peptides to suppress wild-type and mutated AR-mediated prostate cancer growth via peptide screening and *X-ray* structure analysis. 8, 2014.

[23] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021.

[24] M. Abotaleb, P. Kubatka, M. Caprnda, E. Varghese, B. Zolakova, P. Zubor, R. Opatrilova, P. Kruzliak, P. Stefanicka, and D. Büsselberg. Chemotherapeutic agents for the treatment of metastatic breast cancer: An update. *Biomedicine & Pharmacotherapy*, 101:458–477, 2018.

[25] M. Kalia. Biomarkers for personalized oncology: Recent advances and future challenges. *Metabolism*, 64(3):S16–S21, 2015.

[26] *American Cancer Society*. Breast cancer facts & figures 2019–2020. *Am Cancer Soc*, pages 1–44, 2019.

[27] L. Wang, S. Zhang, and X. Wang. The metabolic mechanisms of breast cancer metastasis. *Frontiers in Oncology*, 10:2942, 2021.

[28] G. Turashvili and E. Brogi. Tumor heterogeneity in breast cancer. *Frontiers in medicine*, 4:227, 2017.

[29] M. Tao, T. Song, W. Du, S. Han, C. Zuo, Y. Li, Y. Wang, and Z. Yang. Classifying breast cancer subtypes using multiple kernel learning based on omics data. *Genes*, 10(3):200, 2019.

[30] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso. Breast cancer. *Nature reviews Disease primers*, 5(1):1–31, 2019.

[31] P. H. Tan, I. Ellis, K. Allison, E. Brogi, S. B. Fox, S. Lakhani, A. J. Lazar, E. A. Morris, A. Sahin, R. Salgado, et al. The 2019 world health organization classification of tumours of the breast. 2020.

[32] P. Tang, M. M. Bui, and Y. Peng. Immunohistochemistry in breast cancer. In *Practical Breast Pathology*, pages 173–192. Springer, 2019.

[33] L. J. Grimm and M. A. Mazurowski. Breast cancer radiogenomics: Current status and future directions. *Academic Radiology*, 27(1):39–46, 2020.

[34] L. J. McGhan, A. E. McCullough, C. A. Protheroe, A. C. Dueck, J. J. Lee, R. Nunez-Nateras, E. P. Castle, R. J. Gray, N. Wasif, M. P. Goetz, J. R. Hawse, T. J. Henry, M. T. Barrett, H. E. Cunliffe, and B. A. Pockaj. Androgen receptor-positive triple negative breast cancer: A unique breast cancer subtype. *Annals of Surgical Oncology*, 21(2):361–367, 2014.

[35] J. Mehanna, F. G. Haddad, R. Eid, M. Lambertini, and H. R. Kourie. Triple-negative breast cancer: current perspective on the evolving therapeutic landscape. *International journal of women's health*, 11:431, 2019.

[36] D. D. Singh and D. K. Yadav. Tnbc: potential targeting of multiple receptors for a therapeutic breakthrough, nanomedicine, and immunotherapy. *Biomedicines*, 9(8):876, 2021.

[37] B. D. Lehmann, J. A. Bauer, X. Chen, M. E. Sanders, A. B. Chakravarthy, Y. Shyr, and J. A. Pietenpol. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation*, 121(7):2750–2767, 7 2011.

[38] B. D. Lehmann, B. Jovanović, X. Chen, M. V. Estrada, K. N. Johnson, Y. Shyr, H. L. Moses, M. E. Sanders, and J. A. Pietenpol. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PloS one*, 11(6):e0157368, 2016.

[39] M. Hubalek, T. Czech, and H. Müller. Biological subtypes of triple-negative breast cancer. *Breast Care*, 12(1):8–14, 2017.

[40] N. I. of Health et al. National cancer institute (NCI)-Dictionary of cancer terms: OS. (Consultado em: 05/11/2021), .

[41] N. I. of Health et al. National cancer institute (NCI)-Dictionary of cancer terms: RFS. (Consultado em: 05/11/2021), .

[42] A. Anestis, I. Zoi, A. G. Papavassiliou, and M. V. Karamouzis. Androgen receptor in breast cancer—clinical and preclinical research insights. *Molecules*, 25(2):358, 2020.

[43] A. Losurdo, R. De Sanctis, B. Fernandes, R. Torrisi, G. Masci, E. Agostinetto, W. Gatzemeier, V. Errico, A. Testori, C. Tinterri, et al. Insights for the application of tils and ar in the treatment of tnbc in routine clinical practice. *Scientific Reports*, 10(1):1–9, 2020.

[44] R. A. Davey and M. Grossmann. Androgen receptor structure, function and biology: from bench to bedside. *The Clinical Biochemist Reviews*, 37(1):3, 2016.

[45] C. M. Venema, R. D. Bense, T. G. Steenbruggen, H. H. Nienhuis, S.-Q. Qiu, M. van Kruchten, M. Brown, R. M. Tamimi, G. A. Hospers, C. P. Schröder, et al. Consideration of breast cancer subtype in targeting the androgen receptor. *Pharmacology & therapeutics*, 200:135–147, 2019.

[46] M. Chen, Y. Yang, K. Xu, L. Li, J. Huang, and F. Qiu. Androgen receptor in breast cancer: From bench to bedside. *Frontiers in Endocrinology*, page 573, 2020.

[47] E. R. Azhagiya Singam, P. Tachachartvanich, M. A. La Merrill, M. T. Smith, and K. A. Durkin. Structural dynamics of agonist and antagonist binding to the androgen receptor. *The Journal of Physical Chemistry B*, 123(36):7657–7666, 2019.

[48] P. M. Levine, M. J. Garabedian, and K. Kirshenbaum. Targeting the androgen receptor with steroid conjugates: miniperspective. *Journal of medicinal chemistry*, 57(20):8224–8237, 2014.

[49] N. Al-Othman, M. Ahram, and M. Alqaraleh. Role of androgen and microrna in triple-negative breast cancer. *Breast disease*, 39(1):15–27, 2020.

[50] F. Claessens, S. Denayer, N. Van Tilborgh, S. Kerkhofs, C. Helsen, and A. Haelens. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuclear receptor signaling*, 6, 2008.

[51] X. Yu, P. Yi, R. A. Hamilton, H. Shen, M. Chen, C. E. Foulds, M. A. Mancini, S. J. Ludtke, Z. Wang, and B. W. O'Malley. Structural insights of transcriptionally active, full-length androgen receptor coactivator complexes. *Molecular cell*, 79(5):812–823, 2020.

[52] D. Li, W. Zhou, J. Pang, Q. Tang, B. Zhong, C. Shen, L. Xiao, and T. Hou. A magic drug target: androgen receptor. *Medicinal Research Reviews*, 39(5):1485–1514, 2019.

[53] M. Nadal, S. Prekovic, N. Gallastegui, C. Helsen, M. Abella, K. Zielinska, M. Gay, M. Vilaseca, M. Taulès, A. B. Houtsmuller, et al. Structure of the homodimeric androgen receptor ligand-binding domain. *Nature communications*, 8(1):1–14, 2017.

[54] S. Sakkiah, R. Kusko, B. Pan, W. Guo, W. Ge, W. Tong, and H. Hong. Structural changes due to antagonist binding in ligand binding pocket of androgen receptor elucidated through molecular dynamics simulations. *Frontiers in pharmacology*, 9:492, 2018.

[55] X. Kong, E. Xing, T. Zhuang, P.-K. Li, and X. Cheng. Mechanistic insights into the allosteric inhibition of androgen receptors by binding function 3 antagonists from an integrated molecular modeling study. *Journal of Chemical Information and Modeling*, 61(7):3477–3494, 2021.

[56] B. He, J. T. Minges, L. W. Lee, and E. M. Wilson. The fxxlf motif mediates androgen receptor-specific interactions with coregulators. *Journal of Biological Chemistry*, 277(12):10226–10235, 2002.

[57] H. J. Dubbink, R. Hersmus, C. S. Verma, H. A. Van Der Korput, C. A. Berrevoets, J. Van Tol, A. C. Ziel-Van Der Made, A. O. Brinkmann, A. C. Pike, and J. Trapman. Distinct recognition modes of FXXLF and LXXLL motifs by the androgen receptor. *Molecular Endocrinology*, 18(9):2132–2150, 2004.

[58] D. Tesauro, A. Accardo, C. Diaferia, V. Milano, J. Guillon, L. Ronga, and F. Rossi. Peptide-based drug-delivery systems in biotechnological applications: recent advances and perspectives. *Molecules*, 24(2):351, 2019.

[59] F. Araste, K. Abnous, M. Hashemi, S. M. Taghdisi, M. Ramezani, and M. Alibolandi. Peptide-based targeted therapeutics: Focus on cancer treatment. *Journal of controlled release*, 292:141–162, 2018.

[60] I. W. Hamley. Small bioactive peptides for biomaterials design and therapeutics. *Chemical reviews*, 117(24):14015–14041, 2017.

[61] M. C. Oliveira and J. D. Correia. Biomedical applications of radioiodinated peptides. *European Journal of Medicinal Chemistry*, 179:56–77, 2019.

[62] L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, G. Shao, X. Wang, R. Wang, and C. Fu. Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1):1–27, 2022.

[63] B. G. de la Torre and F. Albericio. Peptide therapeutics 2.0, 2020.

[64] D. M. Copolovici, K. Langel, E. Eriste, and U. Langel. Cell-penetrating peptides: design, synthesis, and applications. *ACS nano*, 8(3):1972–1994, 2014.

[65] A. Gautam, K. Chaudhary, R. Kumar, A. Sharma, P. Kapoor, A. Tyagi, and G. P. Raghava. In silico approaches for designing highly effective cell penetrating peptides. *Journal of translational medicine*, 11(1):1–12, 2013.

[66] A. Gautam, K. Chaudhary, R. Kumar, and G. P. S. Raghava. Computer-aided virtual screening and designing of cell-penetrating peptides. In *Cell-penetrating peptides*, pages 59–69. Springer, 2015.

[67] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe. Computational methods in drug discovery. *Pharmacological reviews*, 66(1):334–395, 2014.

[68] M. Kontoyianni. Docking and virtual screening in drug discovery. In *Proteomics for drug discovery*, pages 255–266. Springer, 2017.

[69] W. Yu and A. D. MacKerell. Computer-aided drug design methods. In *Antibiotics*, pages 85–106. Springer, 2017.

[70] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, et al. The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, 58(6):899–907, 2002.

[71] S. A. Hollingsworth and R. O. Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

[72] N. S. Pagadala, K. Syed, and J. Tuszynski. Software for molecular docking: a review. *Biophysical reviews*, 9(2):91–102, 2017.

[73] S. J. De Vries, M. Van Dijk, and A. M. Bonvin. The HADDOCK web server for data-driven biomolecular docking. *Nature protocols*, 5(5):883–897, 2010.

[74] A. Saponaro, V. Maione, A. M. Bonvin, and F. Cantini. Understanding docking complexes of macromolecules using haddock: the synergy between experimental data and computations. *Bio-protocol*, 10(20), 2020.

[75] C. Geng, S. Narasimhan, J. P. Rodrigues, and A. M. Bonvin. Information-driven, ensemble flexible peptide docking using haddock. In *Modeling Peptide-Protein Interactions*, pages 109–138. Springer, 2017.

[76] J. Beaufays, L. Lins, A. Thomas, and R. Brasseur. In silico predictions of 3d structures of linear and cyclic peptides with natural and non-proteinogenic residues. *Journal of Peptide Science*, 18 (1):17–24, 2012.

[77] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. The I-TASSER suite: protein structure and function prediction. *Nature methods*, 12(1):7–8, 2015.

[78] Y. Zhang. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9(1):1–8, 2008.

[79] M. M. Gromiha, R. Nagarajan, and S. Selvaraj. Protein structural bioinformatics: an overview. 2019.

[80] M. Dorn, M. B. e Silva, L. S. Buriol, and L. C. Lamb. Three-dimensional protein structure prediction: Methods and computational strategies. *Computational biology and chemistry*, 53:251–276, 2014.

[81] J. Lee, P. L. Freddolino, and Y. Zhang. Ab initio protein structure prediction. In *From protein structure to function with bioinformatics*, pages 3–35. Springer, 2017.

[82] J. Maupetit, P. Derreumaux, and P. Tuffery. PEP-FOLD: an online resource for *de novo* peptide structure prediction. *Nucleic acids research*, 37(suppl_2):W498–W503, 2009.

[83] A. Roy, A. Kucukural, and Y. Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, 5(4):725–738, 2010.

[84] V. D. Mouchlis, A. Afantitis, A. Serra, M. Fratello, A. G. Papadiamantis, V. Aidinis, I. Lynch, D. Greco, and G. Melagraki. Advances in de novo drug design: from conventional to machine learning methods. *International journal of molecular sciences*, 22(4):1676, 2021.

[85] B. Manavalan, S. Subramaniyam, T. H. Shin, M. O. Kim, and G. Lee. Machine-learning-based prediction of cell-penetrating peptides and their uptake efficiency with improved accuracy. *Journal of proteome research*, 17(8):2715–2726, 2018.

[86] M. Wiederstein and M. J. Sippl. Prosa-Web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 35(suppl_2):W407–W410, 2007.

[87] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[88] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. PDB2PQR: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research*, 32(suppl_2): W665–W667, 2004.

[89] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[90] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[91] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno. gmx_mmpbsa: A new tool to perform end-state free energy calculations with gromacs. *Journal of Chemical Theory and Computation*, 17(10):6281–6291, 2021. PMID: 34586825.

[92] W. Humphrey, A. Dalke, and K. Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[93] G. Guidotti, L. Brambilla, and D. Rossi. Cell-penetrating peptides: from basic research to clinics. *Trends in pharmacological sciences*, 38(4):406–424, 2017.

[94] B. He, R. T. Gampe, A. J. Kole, A. T. Hnat, T. B. Stanley, G. An, E. L. Stewart, R. I. Kalman, J. T. Minges, and E. M. Wilson. Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Molecular Cell*, 16(3):425–438, 2004. ISSN 10972765.

Apêndice A

Estrutura 3D dos péptidos

A.1 Validação da estrutura dos péptidos (ProSa-*web*)

Figura A.1: *Z-score* da estrutura 3D e visualização molecular da energia dos resíduos dos péptidos iniciais (Pep-A, Pep-B e Pep-C).

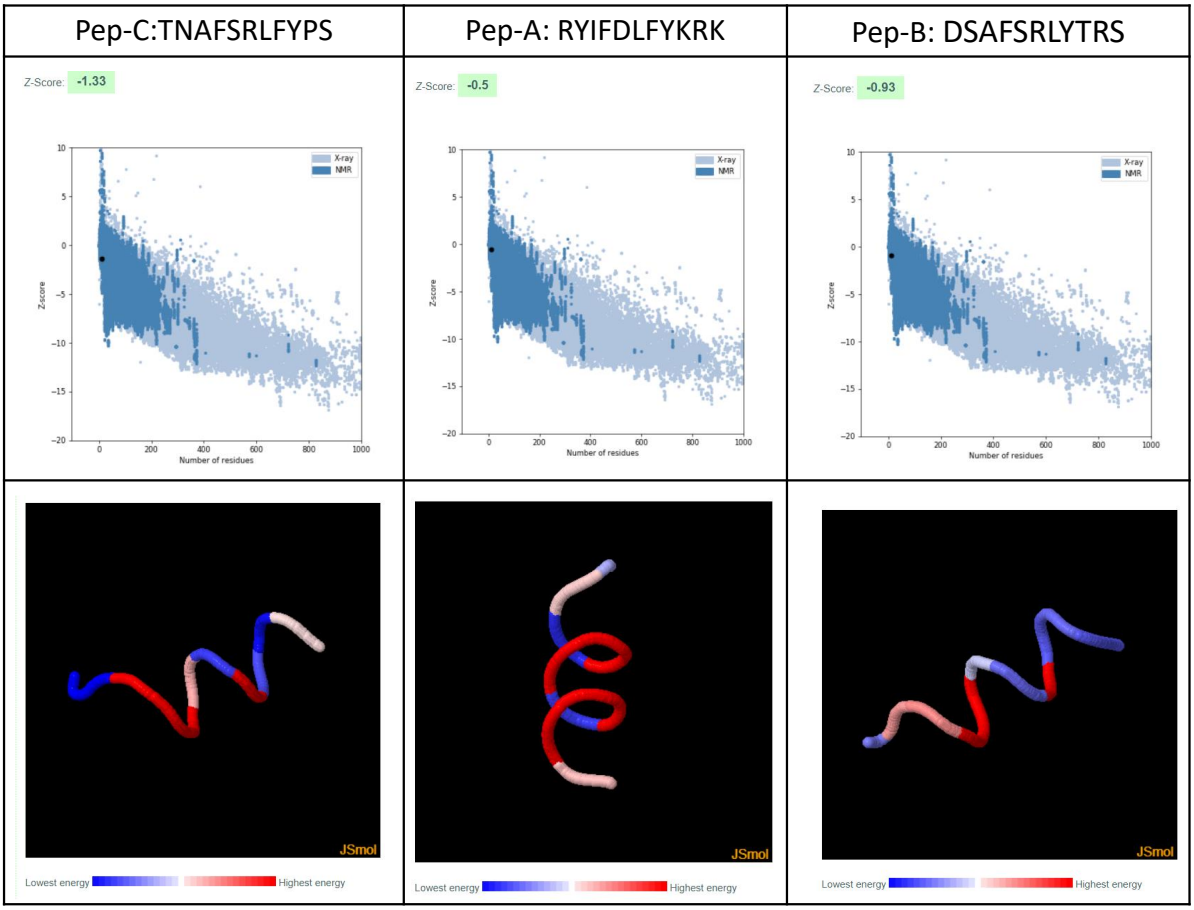


Figura A.2: Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 1.

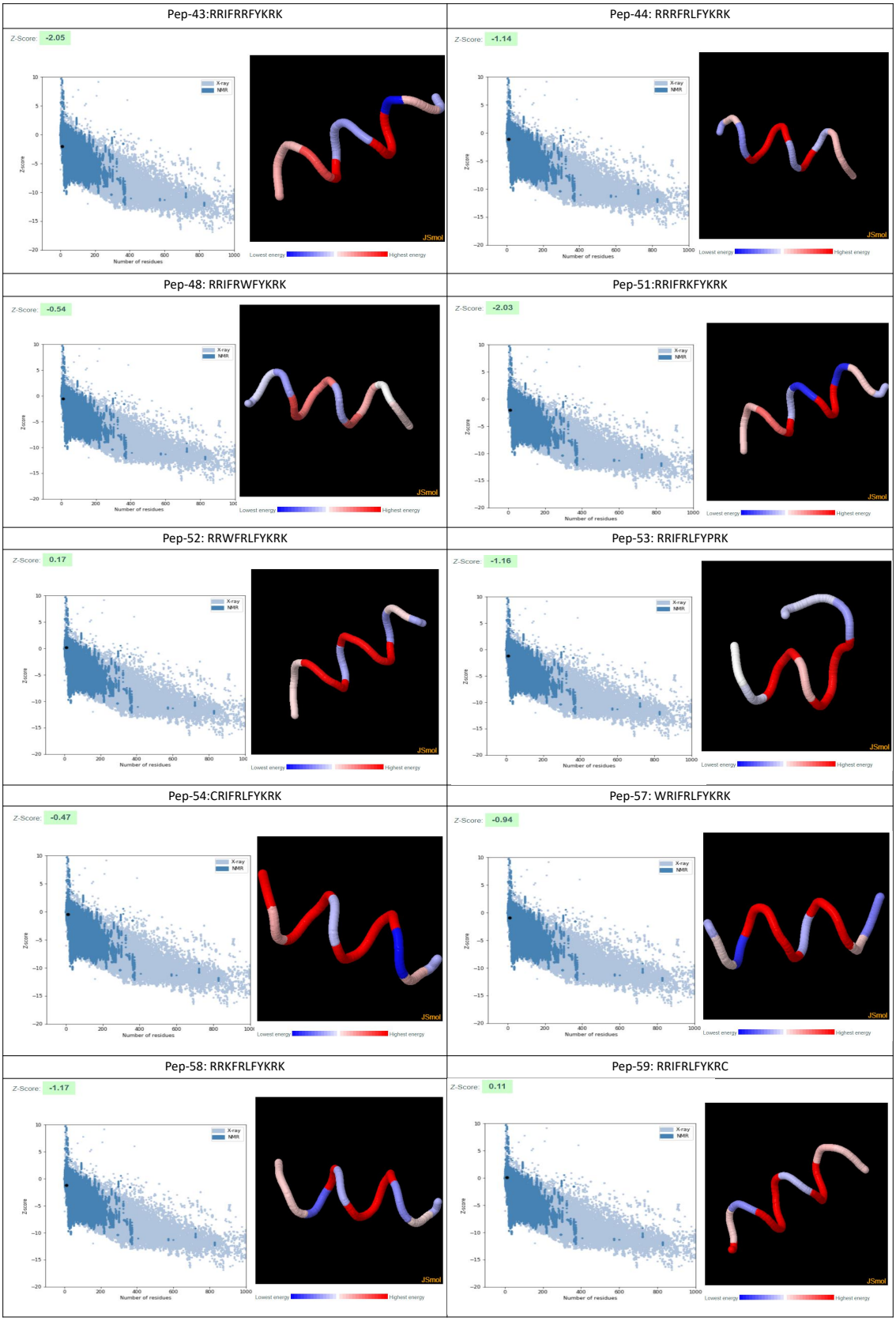


Figura A.3: Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 2.

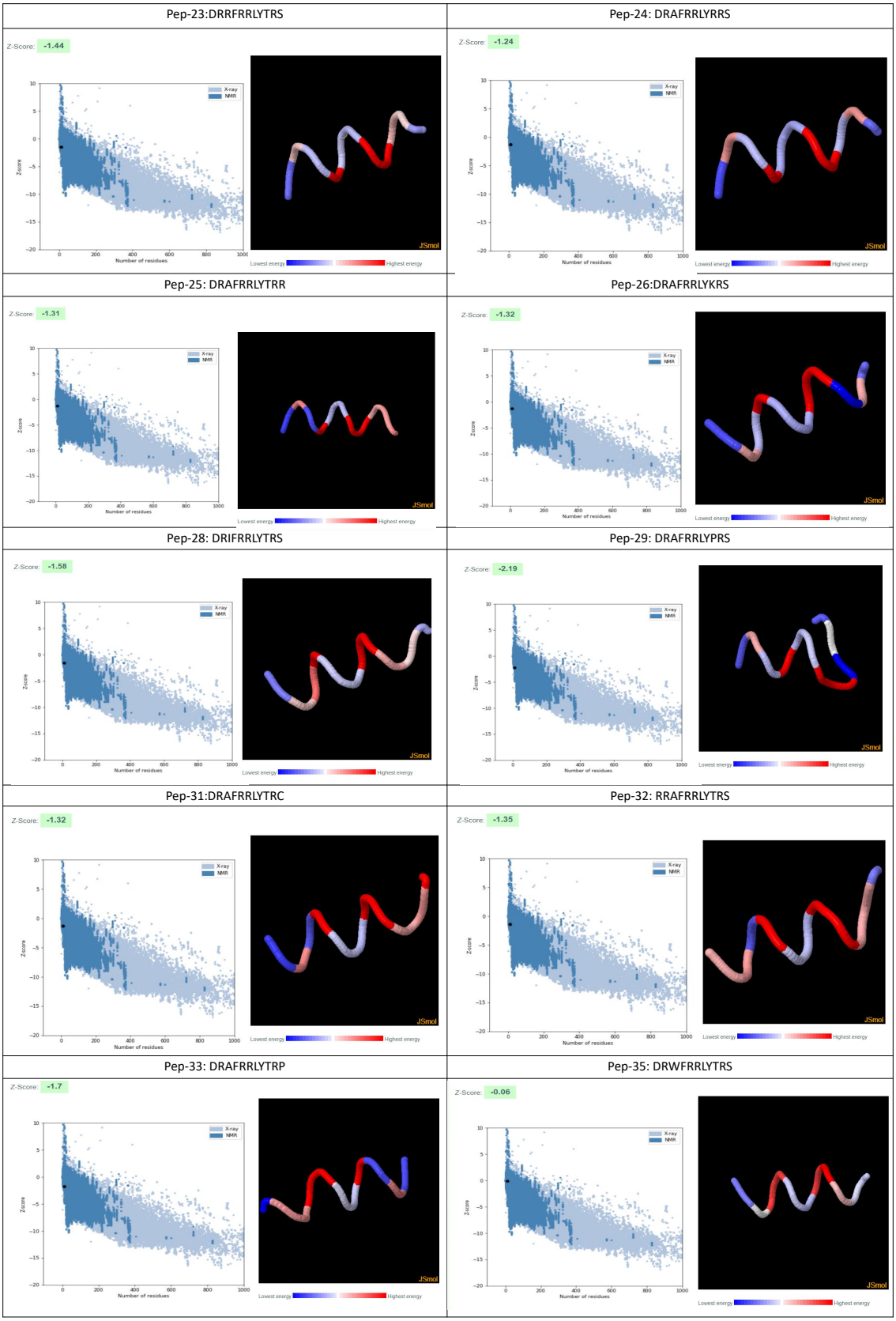
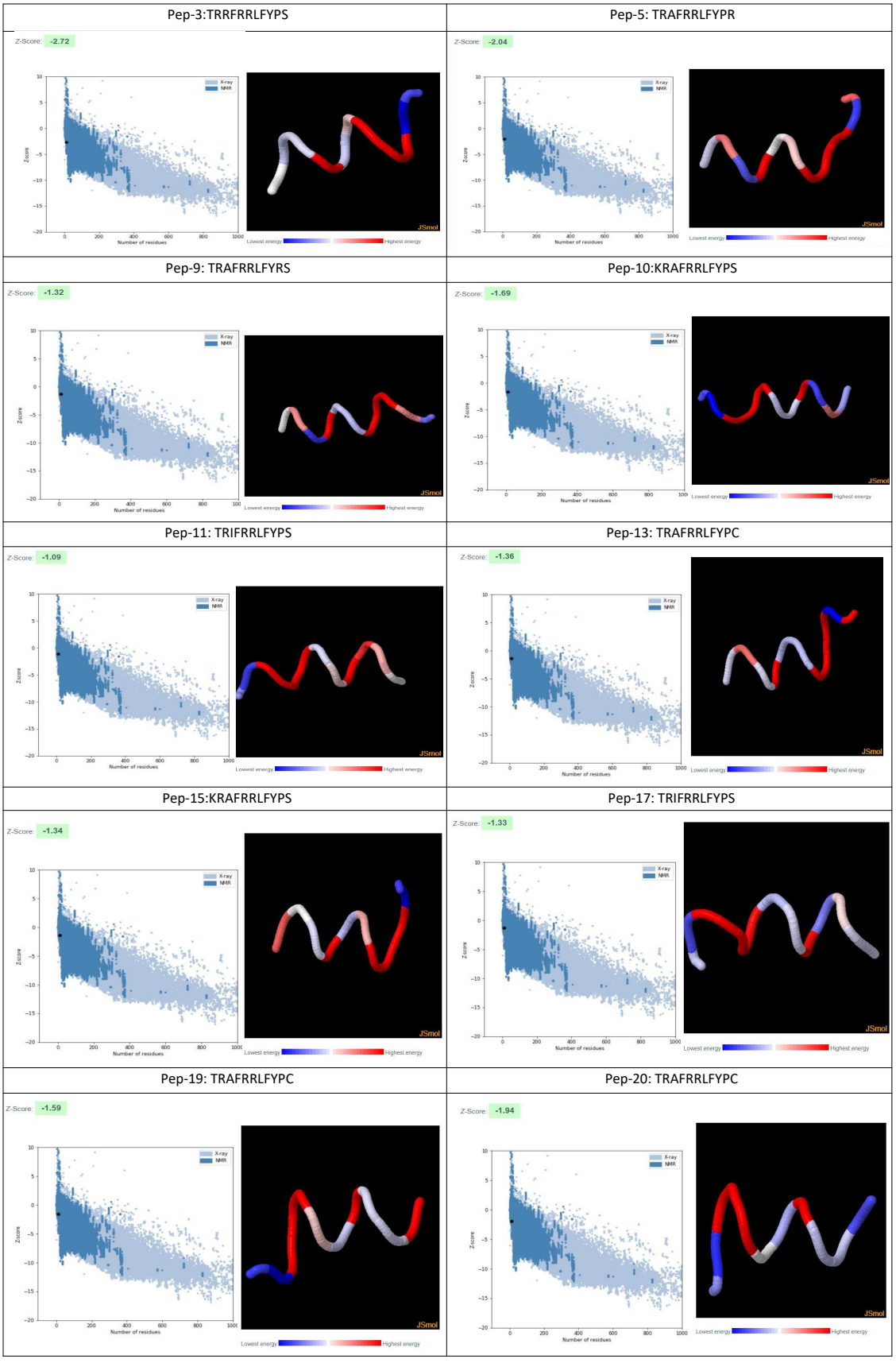


Figura A.4: Z-*score* e visualização molecular da energia dos resíduos dos novos péptidos da família 3.



A.2 Estruturas 3D finais dos péptidos (I-TASSER e PEP-FOLD)

Figura A.5: Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após a previsão do estado de protonação pelo PDB2PQR dos péptidos da família 1.

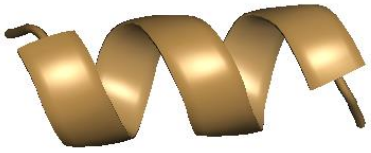
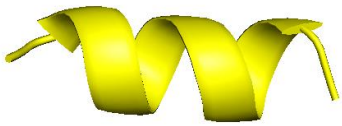
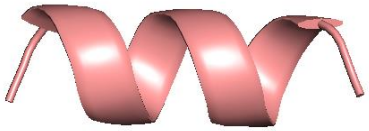
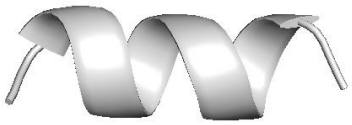
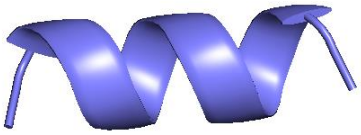
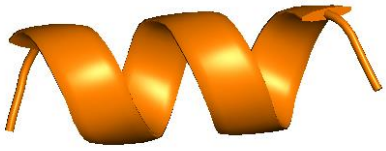
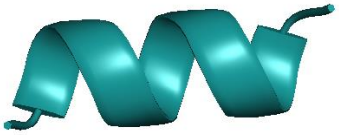
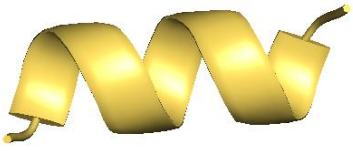
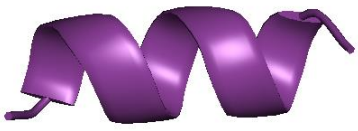
<table>
<tr><td colspan="2">Pep-A-RYIFDLFYKRK</td></tr>
<tr><td colspan="2">  </td></tr>
<tr><td>Pep43-RRIFRRFYKRK</td><td>Pep44-RRRFRLFYKRK</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>Pep48-RRIFRWFYKRK</td><td>Pep51-RRIFRKFYKRK</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>Pep52-RRWFRLFYKRK</td><td>Pep53-RRIFRLFYPRK</td></tr>
<tr><td>  </td><td> </td></tr>
<tr><td>Pep54-CRIFRLFYKRK</td><td>Pep57-WRIFRLFYKRK</td></tr>
<tr><td>  </td><td> </td></tr>
<tr><td>Pep58-RRKFRLFYKRK</td><td>Pep59-RRIFRLFYKRC</td></tr>
<tr><td>  </td><td>  </td></tr>
</table>

Figura A.6: Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após a previsão do estado de protonação pelo PDB2PQR dos péptidos da família 2.

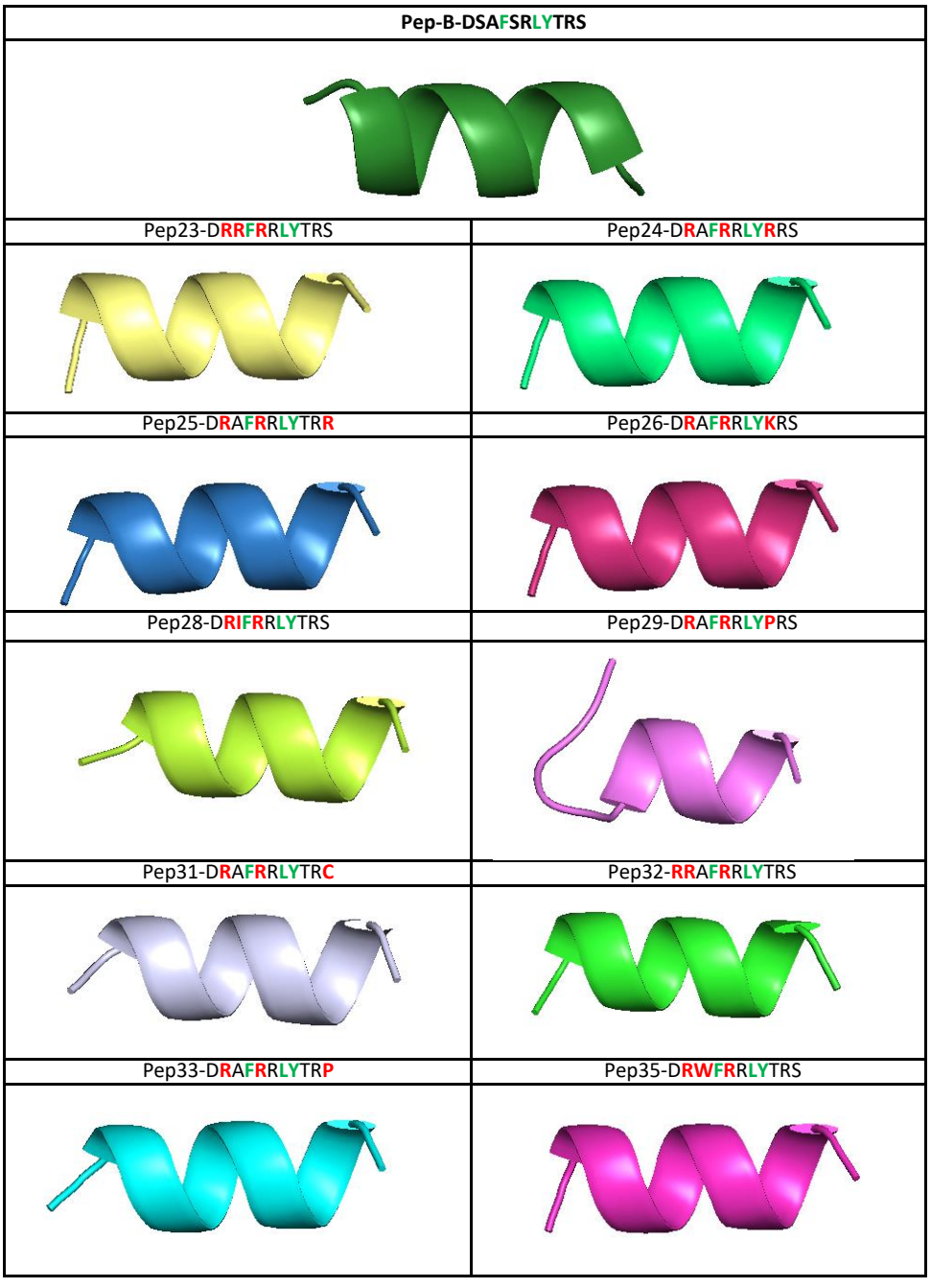
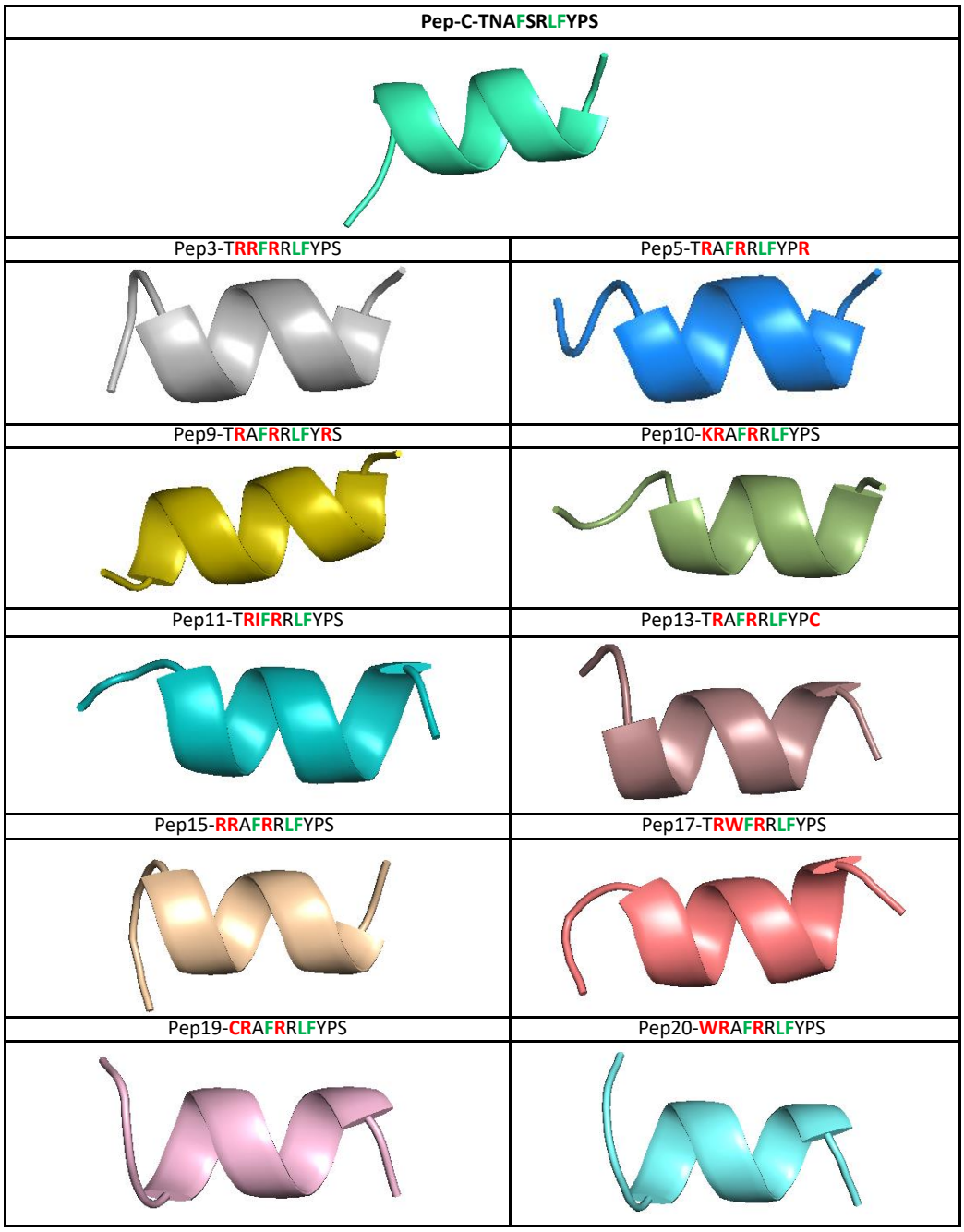


Figura A.7: Estrutura 3D resultante da previsão do I-TASSER/PEP-FOLD após previsão do estado de protonação pelo PDB2PQR dos péptidos da família 3.



Apêndice B

Dinâmica Molecular

B.1 RMSD dos complexos e dos péptidos

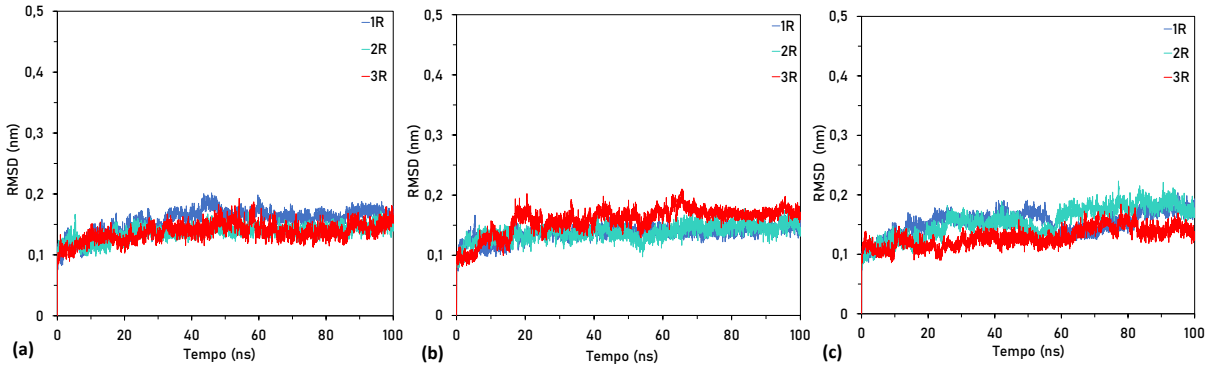


Figura B.1: Representação do RMSD dos complexos, calculado em relação aos átomos Cα. (a) RMSD do complexo recetor-pep19; (b) RMSD do complexo recetor-pep23; (c) RMSD do complexo recetor-pep58.

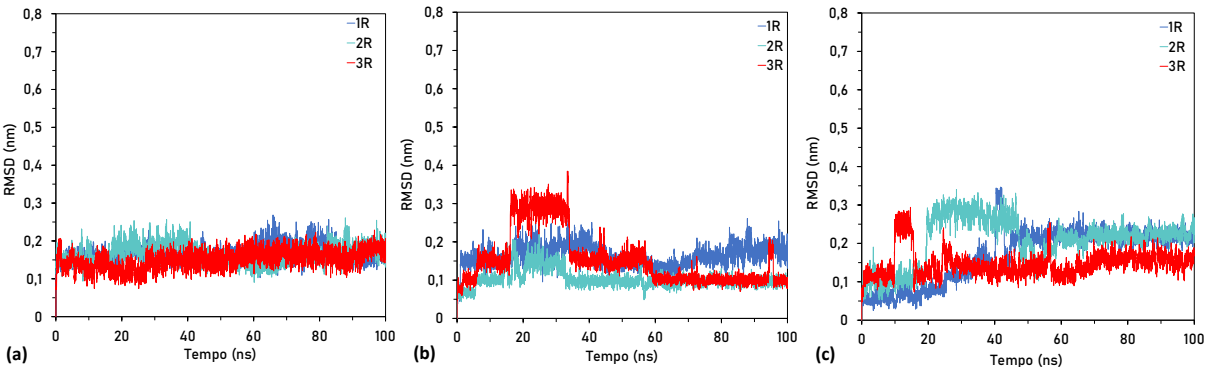


Figura B.2: Representação do RMSD dos péptidos, calculado em relação aos átomos Cα.(a) RMSD do pep19; (b) RMSD do pep23; (c) RMSD do pep58.

B.2 SASA

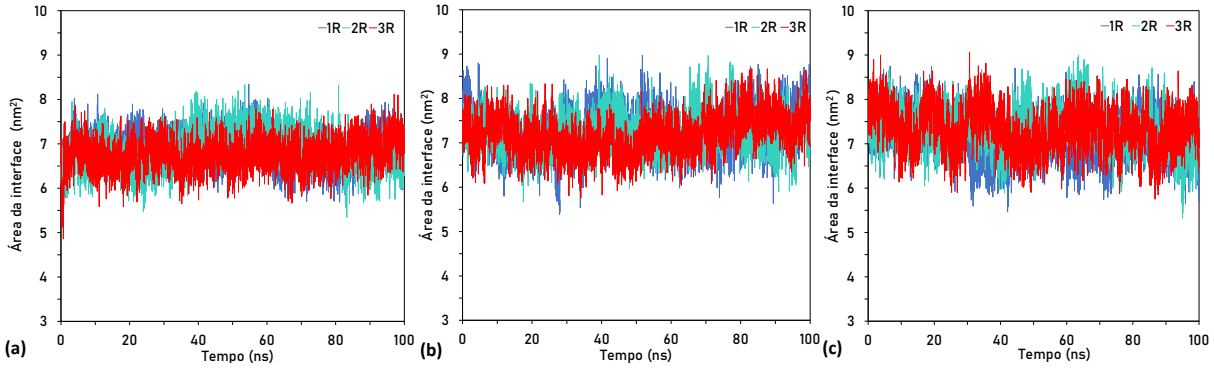


Figura B.3: Área da interface dos péptidos. (a) Área da interface do pep19; (b) Área da interface do pep23; (c) Área da interface do pep58.

B.3 Energia livre de ligação

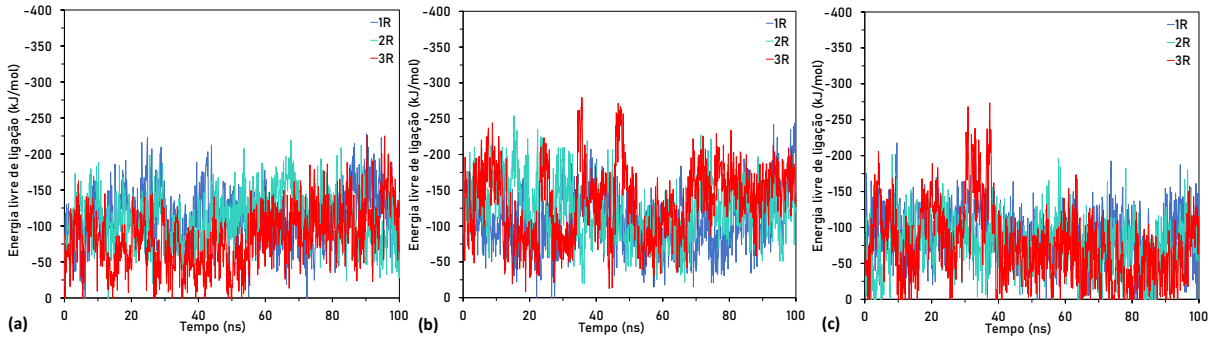


Figura B.4: Energia livre de ligação (kJ/mol) do sistemas. (a) Energia livre de ligação do sistema do pep19; (b) Energia livre de ligação do sistema do pep23; (b) Energia livre de ligação do sistema do pep58.