



UAAlg

UNIVERSIDADE DO ALGARVE

UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS E TECNOLOGIA

MESTRADO INTEGRADO EM CIÊNCIAS FARMACÊUTICAS

Descoberta de moléculas moduladoras da interação Foxo3-DNA

Filomena Alexandra Martins do Nascimento, N^o 57594

Dissertação para obtenção do Grau de Mestre em Ciências Farmacêuticas

Faro

Outubro de 2022

UNIVERSIDADE DO ALGARVE
FACULDADE DE CIÊNCIAS E TECNOLOGIA
MESTRADO INTEGRADO EM CIÊNCIAS FARMACÊUTICAS

Descoberta de moléculas moduladoras da interação Foxo3-DNA

Filomena Alexandra Martins do Nascimento, N^o 57594

Dissertação para obtenção do Grau de Mestre em Ciências Farmacêuticas

Trabalho Realizado sob orientação:

Professor Doutor Paulo Martel

Faro

Outubro de 2022

Descoberta de moléculas moduladoras da interação Foxo3-DNA

Declaração de autoria de trabalho.

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

(Filomena Alexandra Martins do Nascimento)

(Assinatura)

Direitos de cópia ou Copyright

© Copyright: Filomena Alexandra Martins do Nascimento

Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

Agradeço à Universidade do Algarve, a todos os companheiros de curso e a todos os docentes com quem cruzei-me ao longo deste agradável percurso.

Especialmente aos que mais de perto acompanharam-me no final deste trajeto, a tutora de estágio e o meu orientador de tese, a professora Ana Serralheiro e o professor Paulo Martel. Um especial agradecimento ao meu orientador de tese que acompanhou-me durante a realização deste trabalho final, pela confiança, conhecimento, apoio e motivação transmitida, apesar de todas as dificuldades de horário entre estágio e trabalho.

A toda a minha família e amigos mais próximos, Roxane, Tatiana, Mariana, Catarina, Patrícia, Miguel, João, Davide, Xavier, Filipe e Fábio, que incentivaram-me e apoiaram-me. Um especial obrigada ao meu melhor amigo Filipe, sem ele a tese hoje não estaria terminada, pois foi o meu pilar a 100%.

Aos meus colegas do trabalho, que seguraram o trabalho na farmácia para eu concluir a minha tese de mestrado, Gonçalo, Moisés, Telma, Fernanda e Andreia.

Um grande obrigada a todos os que acompanharam-me nesta etapa da minha vida, que ouviram os meus desabafos, choros e risos. Levo a todos no meu coração, sem vocês seria muito mais difícil este percurso.

Resiliência, foco e coragem são as palavras que acompanharam-me nesta jornada fantástica.

OBRIGADA!

Resumo

As proteínas *Forkhead box* (Fox) são uma família de fatores de transcrição que desempenham papéis importantes nos mais diversos organismos. Hoje em dia sabe-se que existem no genoma humano cerca de 43 genes da família Fox. Os membros da família de proteínas Fox são importantes para um amplo espectro de processos biológicos, incluindo metabolismo, desenvolvimento, diferenciação, proliferação, apoptose, migração, invasão e longevidade. Dado que as proteínas Fox controlam esses processos essenciais de desenvolvimento e homeostáticos, não é inesperado que uma perda ou ganho da função da Fox possa alterar o destino celular e levar a problemas de saúde, tal como por exemplo a tumorigênese. Neste contexto escolheu-se estudar a proteína FoxO3a, como base para esta tese, devido à sua importância no organismo humano. O objetivo deste trabalho foi tentar encontrar moléculas moduladoras da atividade da proteína FoxO3, ou seja, moléculas que conseguem inibir ou potenciar a ligação da proteína FOXO3 ao DNA.. Para tal foram utilizadas técnicas computacionais de dinâmica molecular e de docking molecular, as quais são adequadas para este tipo de trabalho. Igualmente foram utilizadas ferramentas computacionais para a identificação de cavidades na superfície da proteína que pudessem servir de locais de ligação. A importância deste estudo reside na possibilidade de identificação de potenciais ligandos que possam vir a servir de ponto de partida para o desenvolvimento de novos fármacos, uma forma revolucionária de novas terapêuticas.

Palavras-chave: FOXO3, Desregulação, Docking, Moléculas moduladoras da atividade da FOXO3, Terapêutica alternativa, Dinâmica Molecular, Desenho computacional de fármacos, triagem virtual.

Abstract

The *Forkead box* (Fox) proteins are a family of transcription factors, which play important roles in the most diverse organisms. Today it is known that there are about 43 genes in the Fox family in the human genome. Members of this protein family are important to a broad spectrum of biological processes, including metabolism, development, differentiation, proliferation, apoptosis, migration, invasion, and longevity. Given that Fox proteins control these essential developmental and homeostatic processes, it is not unexpected that a loss or gain of function can alter cell fate and lead to health problems such as tumorigenesis. In this context, the FoxO3a protein was the subject of study for this thesis, due to its importance in the human organism. The main objective of this work was attempting to find molecules that may be able to modulate the function of FOXO3a by promoting or inhibiting its binding to DNA. In order to achieve the objective, various computational tools were used, including molecular dynamics simulations, docking, virtual screening and automatic detection of cavities (pockets) on the protein surface. The importance of this study resides in its potential to find FOXO3a ligands that could be the first step in the discovery of new drugs that could lead to successful therapies.

Keywords: FOXO3, Dysregulation, Docking that modulates FOXO3 activity, Alternative therapy, molecular dynamics, protein-ligand docking, computer aided drug design, virtual screening

Índice Geral

| | |
|--|------|
| Agradecimentos | I |
| Resumo | II |
| Abstract | III |
| Índice Geral | IV |
| Índice de Figuras | V |
| Índice de Tabelas | VIII |
| Índice de Equações | X |
| Lista de Abreviaturas | X |
| Unidades de Tempo | XI |
| Unidade de Comprimento | XII |
| Unidades de Temperatura | XII |
| Unidade de Pressão | XII |
| 1. Introdução | 1 |
| 2. Objetivo | 11 |
| 3. Enquadramento Teórico | 12 |
| 3.1. Dinâmica Molecular | 12 |
| 3.2. Descoberta de Cavidades | 21 |
| 3.3. <i>Docking</i> | 26 |
| 4. Materiais e Métodos | 32 |
| 4.1 Programas Computacionais | 32 |
| 4.2 Procedimentos Laboratoriais | 34 |
| 4.2.1 Dinâmica Molecular | 34 |
| 4.2.2 Descoberta de Cavidades | 39 |
| 4.2.3 <i>Docking</i> | 41 |
| 5. Resultados e Discussão | 42 |
| 5.1. Dinâmica molecular | 42 |
| 5.2. Descoberta de Cavidades | 60 |
| 5.3. <i>Docking</i> | 81 |
| 6. Conclusão | 97 |
| 7. Bibliografia e <i>Webgrafia</i> | 99 |
| 8. Anexos | 103 |

Índice de Figuras

| | |
|---|----|
| Figura 1.1: Estrutura tridimensional do domínio Fox com o DNA | 1 |
| Figura 1.2: Isoformas FoxO | 3 |
| Figura 1.3: Regulação das proteínas FoxO através da via fosfatidilinositol 3-quinase- proteína quinase B (PI3K-PKB) | 4 |
| Figura 1.4: Regulação das proteínas FoxO através da via fosfatidilinositol 3-quinase- proteína quinase B (PI3K-PKB) | 5 |
| Figura 1.5: Localização do gene FoxO3 no cromossoma 6 | 6 |
| Figura 1.6: Diagrama esquemático de interações entre FoxO3a-DBD e DNA | 7 |
| Figura 1.7: Estrutura tridimensional da estrutura 2K86 | 9 |
| Figura 1.8: Sobreposição da estrutura 2UZK com a estrutura 2K86 | 10 |
| Figura 3.1.9: Esquema simplificado do procedimento da dinâmica molecular | 13 |
| Figura 3.1.10: Complexo DNA-proteína numa caixa cúbica com as moléculas de água e os iões | 18 |
| Figura 3.1.11: Representação esquemática das etapas de uma simulação de dinâmica molecular | 19 |
| Figura 3.2.12: Sítio de ligação de uma proteína | 23 |
| Figura 3.3.13: Processo simplificado do <i>docking</i> | 27 |
| Figura 3.3.14: Procedimento geral do programa de <i>docking</i> | 32 |
| Figura 4.2.1.15: Esquema da realização das corridas de dinâmica molecular. | 35 |
| Figura 4.2.1.16: Passos para a realização da corrida de simulação de dinâmica molecular | 37 |
| Figura 4.2.1.17: Processos do programa FPOCKET..... | 40 |
| Figura 5.1.18: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2k86 da proteína FoxO3..... | 44 |
| Figura 5.1.19: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2UZK da proteína FoxO3 | 45 |
| Figura 5.1.20: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2UZK com o DNA (complexo) da proteína FoxO3..... | 45 |
| Figura 5.1.21: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica zero | 46 |
| Figura 5.1.22: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica um | 47 |
| Figura 5.1.23: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica dois..... | 47 |

| | |
|--|----|
| Figura 5.1.24: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica três. | 48 |
| Figura 5.1.25: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica zero | 49 |
| Figura 5.1.26: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica um | 49 |
| Figura 5.1.27: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica dois | 50 |
| Figura 5.1.28: Flutuação atômica de RMSD da hélice H1-H3 do domínio Fox DBD ligado e sem DNA da réplica três | 50 |
| Figura 5.1.29: Representação estrutura 2k86 – réplica 0 desde o tempo 0 até aos 100 nanosegundos | 52 |
| Figura 5.1.30: Representação estrutura 2k86 – réplica 0 desde o tempo 900 até aos 1000 nanosegundos | 52 |
| Figura 5.1.31: Representação estrutura 2k86 – réplica 1 desde o tempo 0 até aos 100 nanosegundos | 53 |
| Figura 5.1.32: Representação da estrutura 2k86 – réplica 1 desde o tempo 900 até aos 1000 nanosegundos | 53 |
| Figura 5.1.33: Representação estrutura 2k86 – réplica 2 desde o tempo 0 até aos 100 nanosegundos | 54 |
| Figura 5.1.34: Representação da estrutura 2k86 – réplica 2 desde o tempo 900 até aos 1000 nanosegundos | 54 |
| Figura 5.1.35: Representação estrutura 2k86 – réplica 3 desde o tempo 0 até aos 100 nanosegundos | 55 |
| Figura 5.1.36: Representação da estrutura 2k86 – réplica 3 desde o tempo 900 até aos 1000 nanosegundos | 55 |
| Figura 5.1.37: Representação estrutura 2UZK com DNA – réplica 0 desde o tempo 0 até aos 100 nanosegundos | 56 |
| Figura 5.1.38: Representação da estrutura 2UZK com DNA – réplica 0 desde o tempo 900 até aos 1000 nanosegundos. | 56 |
| Figura 5.1.39: Representação estrutura 2UZK com DNA – réplica 1 desde o tempo 0 até aos 100 nanosegundos | 57 |
| Figura 5.1.40: Representação da estrutura 2UZK com DNA – réplica 1 desde o tempo 900 até aos 1000 nanosegundos | 57 |
| Figura 5.1.41: Representação estrutura 2UZK com DNA – réplica 2 desde o tempo 0 até aos 100 nanosegundos | 58 |
| Figura 5.1.42: Representação da estrutura 2UZK com DNA – réplica 2 desde o tempo 900 até aos 1000 nanosegundos | 58 |
| Figura 5.1.43: Representação estrutura 2UZK com DNA – réplica 3 desde o tempo 0 até aos 100 nanosegundos | 59 |
| Figura 5.1.44: Representação da estrutura 2UZK com DNA – réplica 3 desde o tempo 900 até aos 1000 nanosegundos | 59 |

| | |
|--|----|
| Figura 5.2.45: Representação da proteína foxO3, na estrutura 2k86, réplica 3, conformação 10 | 75 |
| Figura 5.2.46: Representação da proteína foxO3, na estrutura 2K86, réplica 3, conformação 10 | 76 |
| Figura 5.2.47: Representação da proteína foxO3, na estrutura 2UZK, réplica 1, conformação 5..... | 77 |
| Figura 5.2.48: Representação da proteína foxO3, na estrutura 2UZK, réplica 1, conformação 5..... | 77 |
| Figura 5.2.49: Representação da proteína foxO3, na estrutura 2UZK, réplica 3, conformação 7 | 78 |
| Figura 5.2.50: Representação da proteína foxO3, na estrutura 2UZK, réplica 3, conformação 7 | 79 |
| Figura 5.2.51: Representação da proteína foxO3, na estrutura do complexo, réplica 1, conformação 5..... | 80 |
| Figura 5.2.52: Representação da proteína foxo3, na estrutura do complexo, réplica 1, conformação 1 | 80 |
| Figura 5.3.53: Histograma da distribuição das energias do docking para a cavidade 2 da conformação 10 da réplica 3 da estrutura 2k86..... | 84 |
| Figura 5.3.54: Estruturas químicas dos melhores 10 ligandos para a cavidade 2 da conformação 10 da réplica 3 da estrutura 2k86..... | 85 |
| Figura 5.3.55: Interação do melhor ligando com a proteína. Sendo o ligando a Testolactona e a proteína FoxO3, estrutura 2K86..... | 86 |
| Figura 5.3.56: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 5 da réplica 1 da estrutura 2UZK | 87 |
| Figura 5.3.57: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 5 da réplica 1 da estrutura 2UZK | 88 |
| Figura 5.3.58: Interação do melhor ligando com a proteína. Sendo o ligando a Lurasidona e a proteína FoxO3, estrutura 2UZK..... | 89 |
| Figura 5.3.59: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 7 da réplica 3 da estrutura 2UZK | 90 |
| Figura 5.3.60: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 7 da réplica 3 da estrutura 2UZK | 91 |
| Figura 5.3.61: Interação do melhor ligando com a proteína. Sendo o ligando Lumacaftor e a proteína FoxO3, estrutura 2UZK..... | 92 |
| Figura 5.3.62: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 2 da réplica 1 da estrutura do complexo | 93 |
| Figura 5.3.63: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 2 da réplica 1 da estrutura do complexo | 95 |
| Figura 5.3.64: Interação do melhor ligando com a proteína. Sendo o ligando Risdiplam e a proteína FoxO3, estruturado complexo. | 96 |

Índice de Tabelas

| | |
|--|-----|
| Tabela 3.3.1: Características físico-químicas dos diversos conceitos que podem ser utilizados para seleção de ligandos | 29 |
| Tabela 5.1.2: Estruturas e respectivas réplicas das corridas de dinâmica molecular... | 42 |
| Tabela 5.2.3: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 62 |
| Tabela 5.2.4: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 63 |
| Tabela 5.2.5: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 64 |
| Tabela 5.2.6: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 65 |
| Tabela 5.2.7: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 66 |
| Tabela 5.2.8: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 67 |
| Tabela 5.2.9: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 68 |
| Tabela 5.2.10: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 69 |
| Tabela 5.2.11: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 70 |
| Tabela 5.2.12: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 71 |
| Tabela 5.2.13: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 72 |
| Tabela 5.2.14: Parâmetros das cavidades encontradas a partir do programa FPOCKET | 73 |
| Tabela 5.3.15: Características dos melhores 10 ligandos | 84 |
| Tabela 5.3.16: Características dos melhores 10 ligandos | 87 |
| Tabela 5.3.17: Características dos melhores 10 ligandos | 90 |
| Tabela 5.3.18: Características dos melhores 10 ligandos | 94 |
| Tabela 8.1: Interações das cavidades com o DNA e a proteína no complexo réplica 0 | 103 |

| | |
|--|-----|
| Tabela 8.2: Interações das cavidades com o DNA e a proteína no complexo réplica 0 | 104 |
| Tabela 8.3: Interações das cavidades com o DNA e a proteína no complexo réplica 0 | 105 |
| Tabela 8.4: Interações das cavidades com o DNA e a proteína no complexo réplica 1 | 106 |
| Tabela 8.5: Interações das cavidades com o DNA e a proteína no complexo réplica 1 | 107 |
| Tabela 8.6: Interações das cavidades com o DNA e a proteína no complexo réplica 1 | 108 |
| Tabela 8.7: Interações das cavidades com o DNA e a proteína no complexo réplica 2 | 109 |
| Tabela 8.8: Interações das cavidades com o DNA e a proteína no complexo réplica 2 | 110 |
| Tabela 8.9: Interações das cavidades com o DNA e a proteína no complexo réplica 2 | 111 |
| Tabela 8.10: Interações das cavidades com o DNA e a proteína no complexo réplica 3 | 112 |
| Tabela 8.11: Interações das cavidades com o DNA e a proteína no complexo réplica 3 | 113 |
| Tabela 8.11: Interações das cavidades com o DNA e a proteína no complexo réplica 3 | 114 |

Índice de Equações

| | |
|---|----|
| Equação 3.1.1: Representação do campo de Forças | 14 |
| Equação 3.1.2: Potencial de Lennard-Jones | 14 |
| Equação 3.1.3: Potencial de Coulomb | 14 |
| Equação 3.1.4: Equação da aceleração da partícula | 19 |
| Equação 3.1.5: Algoritmo de Verlet | 19 |

Listas de Abreviaturas

| | |
|----------|--|
| A | Adenina |
| ADN/DNA | Ácido desoxirribonucleico |
| Arg | Arginina |
| Asn | Asparagina |
| Complexo | Cadeia A da estrutura 2UZK complexada com o DNA |
| DM | Dinâmica molecular |
| DBD | Domínio de ligação ao DNA (vem do inglês - DNA Binding Domain) |
| FOX | Família de proteínas Forkhead box (caixa cabeça de garfo) |
| FoxO | Subclasse O da família forkhead box |
| G | Guanina |
| HDACs | Histonas desacetilases |
| His | Histidina |
| Lis | Lisina |
| logP | Coeficiente de partição |
| MolWt | Massa Molecular (molecular weight) |
| MSAa | Alinhamentos de Sequência Múltiplas |
| NES | Sinal de exportação nuclear |

| | |
|----------|--|
| NLS | Sinal de localização nuclear |
| NMR | Ressonância magnética nuclear |
| NVT | Número de partículas, volume e temperatura |
| NPT | Número de partículas, pressão e temperatura |
| PB | Par de bases |
| PDB | Protein Data Bank |
| PKB | Proteína quinase B |
| PI3K-PKB | Fosfatidilinositol 3-quinase-proteína quinase B |
| PTMs | Modificações pós-tradução |
| P53 | Proteína supressora de tumores |
| RE | Recetor de estrogênio |
| RMSD | Desvio quadrático médio das posições atômicas |
| Ser | Serina |
| T | Timina |
| TAD | Domínio de transativação |
| 2UZK | Estrutura cristalina da proteína FOXO3a-DBD humana |
| 2K86 | Estrutura em solução da proteína FOXO3a-DBD humana |

Unidades de Tempo

| | |
|---------|---------------------------------------|
| Fs | 1 Femtossegundo = 10^{-15} segundos |
| Ns | 1 Nanosegundo = 10^{-9} Segundos |
| Ps | 1 Picosegundo = 10^{-12} Segundos |
| μ s | 1 Microsegundo = 10^{-6} Segundos |

Unidades de Comprimento

Å 1 Angstroms = 10^{-10} Metros

Unidades de Temperatura

°K 1 Kelvin = -272,15 Graus celsius

Unidades de Pressão

Atm 1 Atmosfera = 101325 Pascal

1. Introdução

Em 1989 foram descobertos os genes *Forkhead* na mosca *Drosophila melanogaster* ^[1] e no fígado de um rato ^[2]. A partir dessa descoberta foram encontrados mais genes *Forkhead box* (Fox) em diversas espécies, formando assim a família Fox.

As proteínas *Forkhead box* (Fox) são uma família de fatores de transcrição que desempenham papéis importantes nos mais diversos organismos. Esta superfamília é caracterizada pela sua *Forkhead box* (cabeça de garfo). A *Forkhead box* consiste em aproximadamente 100 aminoácidos que se ligam ao DNA (zona que se denomina DBD -domínio de ligação ao DNA) ^[3]. Devido à aparência de borboleta, o domínio Fox também é conhecido como domínio de 'hélice alada' (*winged helix*). A estrutura tridimensional do DBD consiste numa asa ou W1, três alfa-hélices, duas folhas beta e numa segunda asa ou W2 (figura 1.1) ^[3]. Devido à sua elevada mobilidade, a W2 geralmente não é visível nas estruturas cristalográficas das proteínas Fox.

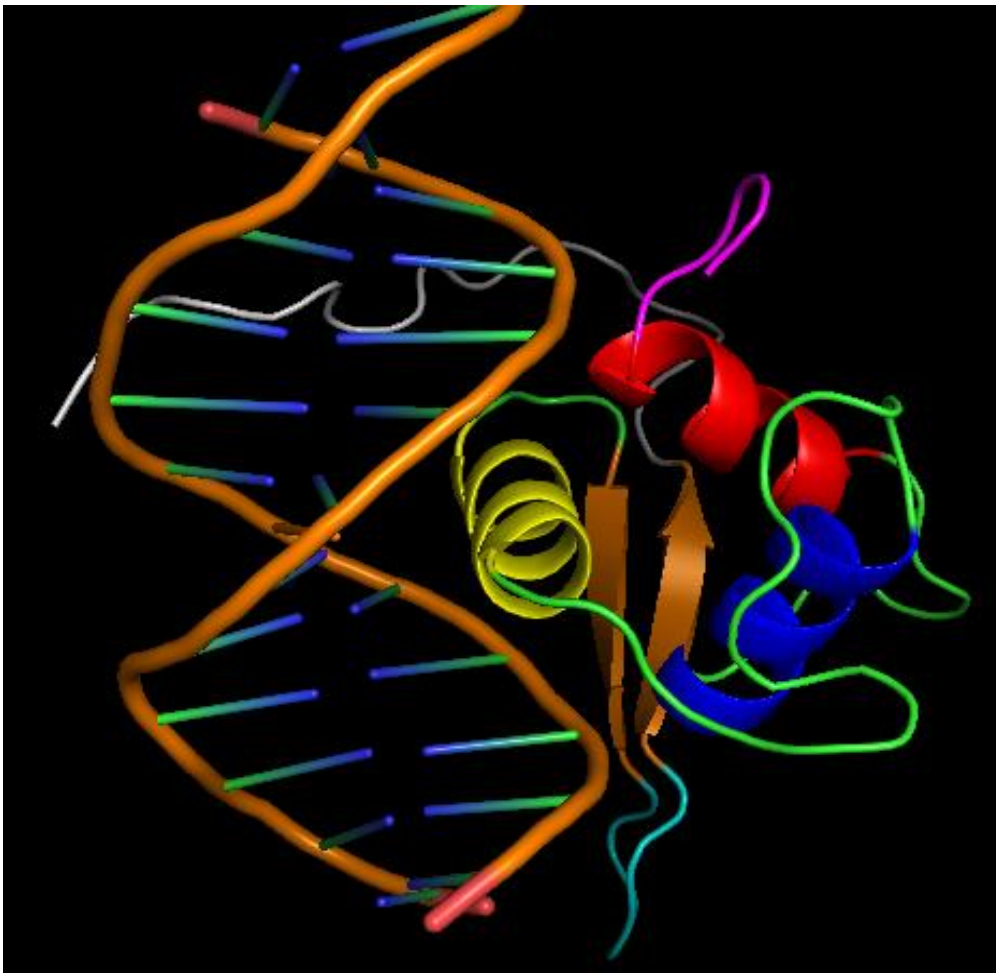


Figura 1.1: Estrutura tridimensional do domínio Fox com o DNA. A violeta encontra-se o N-terminal; vermelho a hélice 1; azul a hélice 2; amarelo a hélice 3; castanho as duas folhas beta; azul claro a asa 1; branco o c-terminal ou asa 2; verde as zonas desordenadas. Imagem original realizada no PyMOL.

Hoje em dia sabe-se que existem no genoma humano cerca de 43 genes da família Fox que dividem-se em 19 sub-famílias desde a FoxA até à FoxS [4]. A família Fox pode ser ainda dividida nas classes I e II, sendo a classe I constituída por proteínas Fox com a região c-terminal básica. Os membros desta classe são as sub-famílias A-G, I-L e a Q. Já a classe II é constituída por proteínas Fox com a região c-terminal não básica. Os membros desta classe são as restantes sub-famílias [5].

Apesar do DBD ser altamente conservado, a regulação e a função das proteínas Fox varia significativamente entre as sub-famílias, em parte devido à variação da sequência fora do DBD, permitindo a regulação diferencial e diversificação funcional. Algumas pequenas variações estruturais ocorrem também dentro do DBD, principalmente entre a segunda e a terceira hélice, em contraste com as α -hélices e folhas- β , que são altamente conservadas [6].

Os membros da família de proteínas Fox são importantes para um amplo espectro de processos biológicos, incluindo metabolismo, desenvolvimento, diferenciação, proliferação, apoptose, migração, invasão e longevidade. Dado que as proteínas Fox controlam esses processos essenciais de desenvolvimento e homeostáticos, não é inesperado que uma perda ou ganho da função da Fox possa alterar o destino celular e levar a problemas de saúde, tal como por exemplo a tumorigénese [6]. As proteínas Fox podem ativar e reprimir a expressão génica por meio do recrutamento de cofatores ou repressores, principalmente histonas desacetilases (HDACs). Além disso, as proteínas Fox interagem extensivamente com outros fatores, como a proteína supressora de tumores (p53) e recetor de estrogénio (RE). Modular a expressão génica, e a desregulação da atividade ou expressão da proteína Fox resulta em alterações nos genes alvo diretos e indiretos.

As proteínas Fox podem atuar como alvos diretos e indiretos para intervenção terapêutica. No entanto, a natureza complexa das vias de sinalização envolvendo o fator de transcrição Fox e seus papéis como supressores de tumor e oncogenes tornam estas moléculas alvos terapêuticos desafiadores. Sendo a família Fox tão importante no organismo humano é essencial o estudo dos seus membros e o avanço

do conhecimento dos mesmos, na procura de descobertas que poderão vir a suportar a prevenção ou terapêutica de certas doenças. Neste contexto escolheu-se estudar a proteína FoxO3a, como base para esta tese, devido à sua importância no organismo humano.

A FoxO3a é um fator de transcrição da sub-família O da família Fox. A sub-família O é constituída por quatro membros, a FoxO1/ FKHR, FoxO3/ FKHL1, FoxO4/ AFX e a FoxO6 [7].

Na estrutura das proteínas FoxO podem observar-se principalmente 4 domínios, o DBD do tipo *Forkhead*, um sinal de exportação nuclear (NES), um sinal de localização nuclear (NLS) e um domínio de transativação (TAD) (figura 1.2).

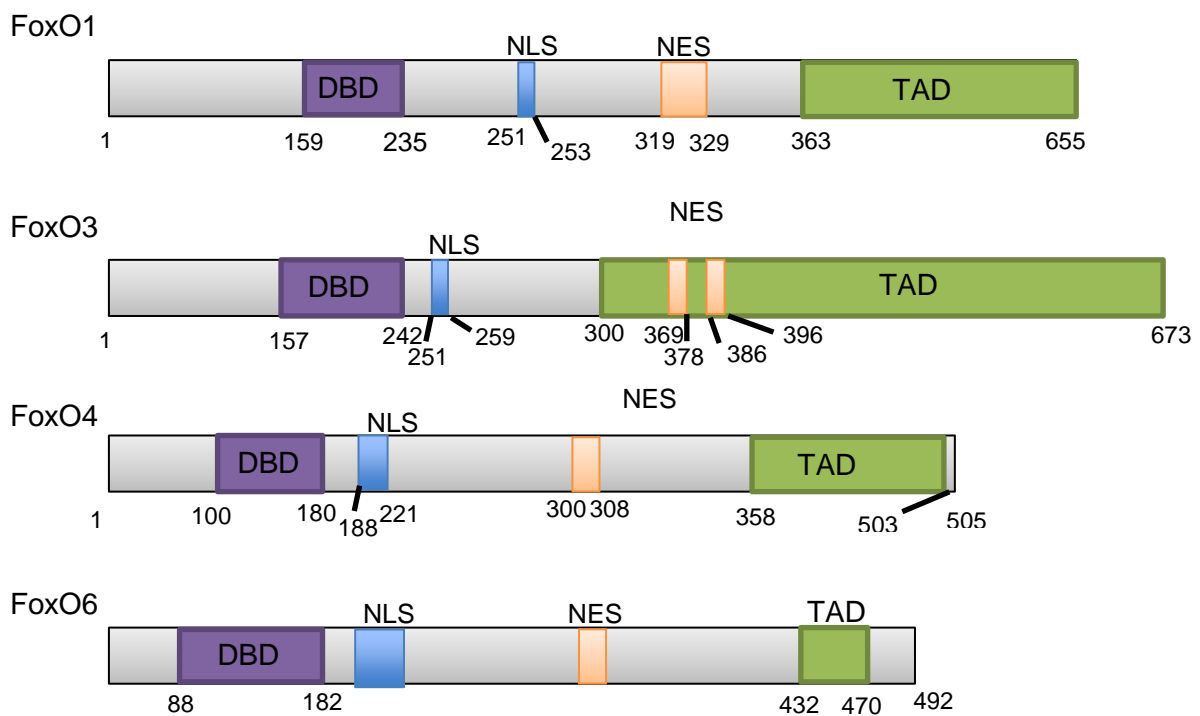


Figura 1.2: Isoformas FOXO. Imagem adaptada do artigo [8].

Para além destes 4 domínios a FoxO tem regiões desordenadas que constituem aproximadamente 75% da proteína total [8]. As proteínas FoxO regulam o metabolismo, a proliferação celular, indução da apoptose celular [9], resposta ao stress celular, sobrevivência celular, resposta aos danos do DNA [10] e longevidade. A sua atividade transcricional é regulada por diversos mecanismos pós-traducionais, incluindo fosforilação pela proteína quinase B (PKB), acetilação pela proteína de ligação ao elemento de resposta a cAMP (CBP) e ubiquitinação. Os locais de modificação pós-tradução estão localizados dentro da região básica C-terminal ou

cauda. Todas as proteínas FoxO são também reguladas pela sinalização da insulina ou do fator de crescimento através da via fosfatidilinositol 3-quinase-proteína quinase B (PI3K-PKB) (figura 1.3 e 1.4).

A fosforilação é a adição de um anião fosfato (PO_4^{3-}) a uma proteína ou outra molécula. A acetilação é a adição de um grupo acetilo (COCH_3) a uma proteína ou outra molécula. A ubiquitina é uma pequena proteína (76 aminoácidos) encontrada nas células eucariotas, a qual desempenha uma função importante na regulação do “*turnover*” proteico. Ela marca proteínas indesejadas (por exemplo proteínas mal dobradas ou “*misfolded*”) para que sejam degradadas por um complexo multiproteico denominado proteassoma.

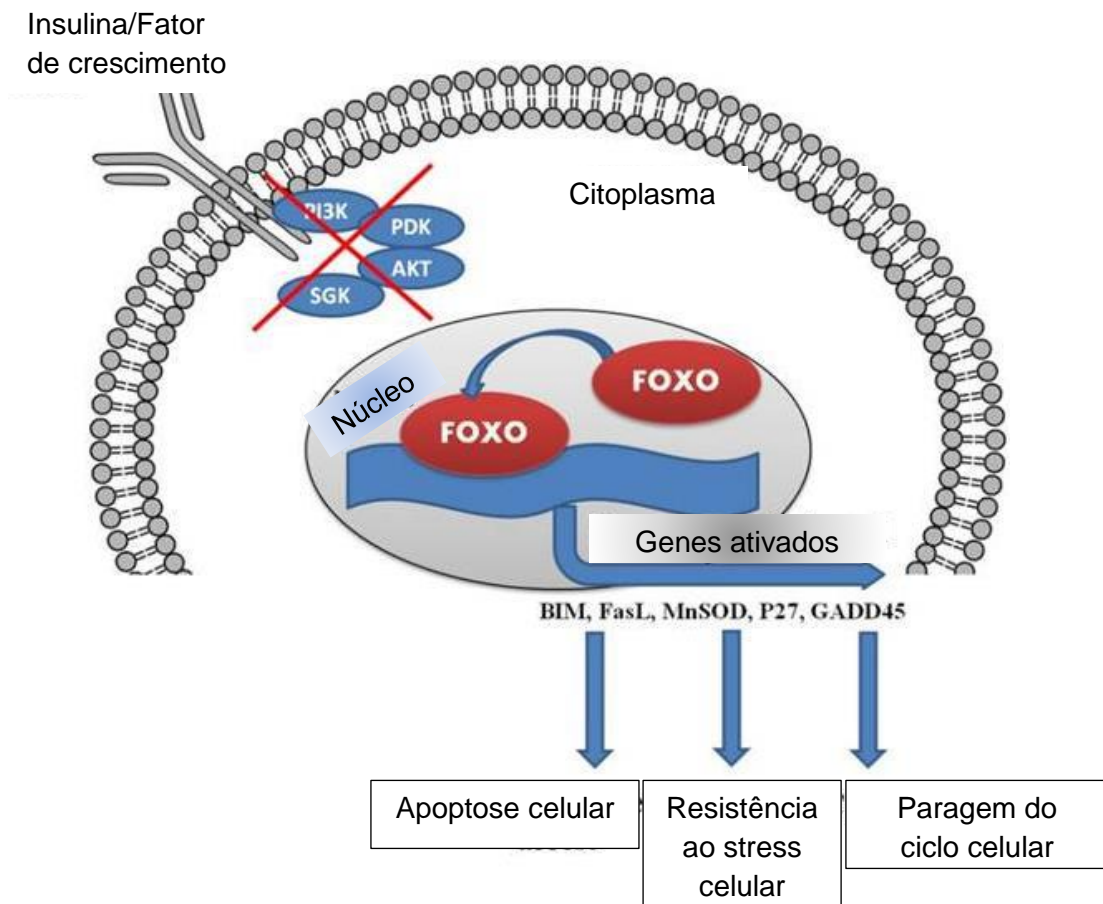


Figura 1.3: Regulação das proteínas FOXO através da via fosfatidilinositol 3-quinase-proteína quinase B (PI3K-PKB). Na ausência do sinalizador de insulina e/ou fator de crescimento, as proteínas FOXO localizam-se no núcleo e promovem a transcrição de múltiplos genes-alvo. Imagem adaptada do artigo ^[11]

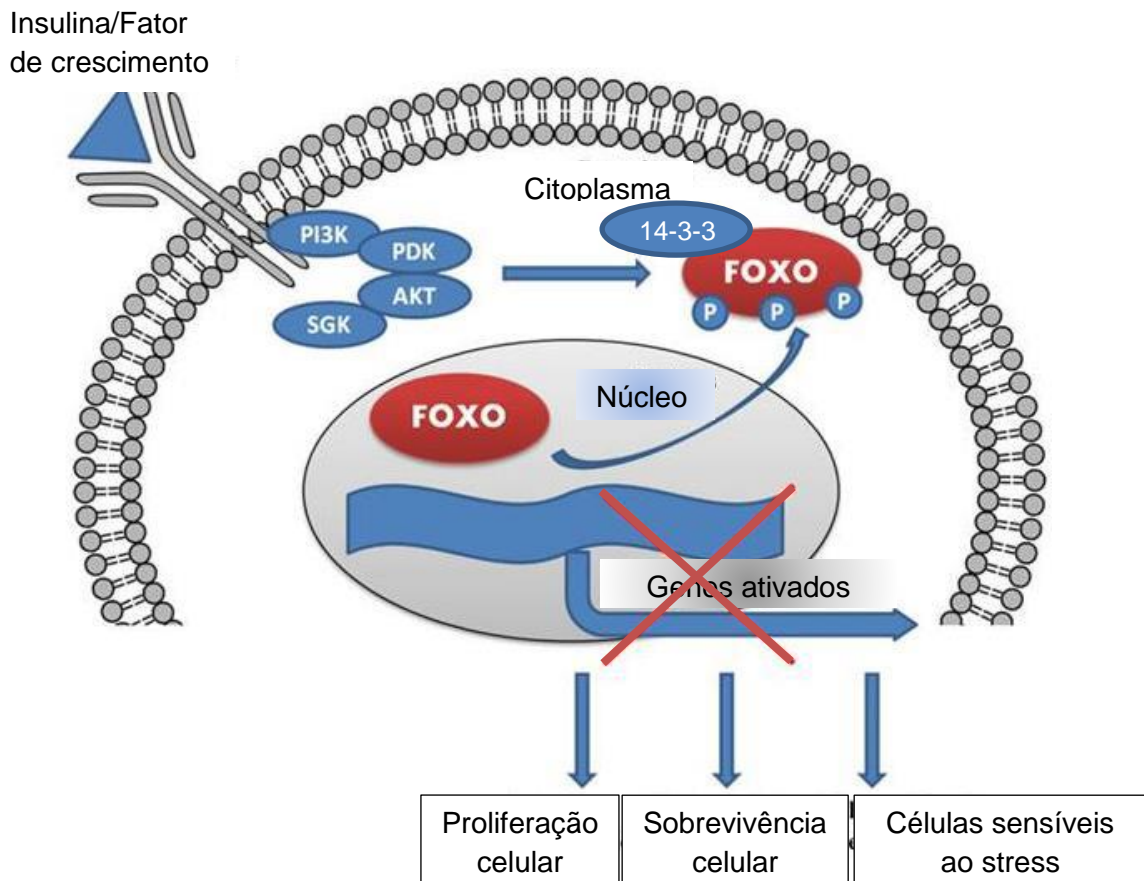


Figura 1.4: Regulação das proteínas FOXO através da via fosfatidilinositol 3-quinase-proteína quinase B (PI3K-PKB). Na presença do sinalizador de insulina e/ou fator de crescimento, as proteínas FOXO são translocadas do núcleo para o citoplasma por meio da chaperona 14-3-3, com a consequente supressão da sua atividade transcricional. Imagem adaptada do artigo ^[11].

Na figura 1.3 está ausente o sinalizador da insulina e/ou o fator do crescimento, então a proteína FoxO liga-se ao DNA no núcleo e ativa os genes de sobrevivência celular. Contrariamente na figura 1.4 encontra-se presente o sinalizador da insulina e/ou o fator do crescimento que ativa a proteína chaperona 14-3-3, que se liga a FoxO e exporta esta do núcleo para o citoplasma da célula. A proteína FoxO, quando localizada no citoplasma, não é capaz de ativar os genes de sobrevivência celular.

A fosforilação induzida pela proteína quinase B (PKB ou AKT) em três sítios regulatórios (tirosina 32, serina 253 e serina 315) gera dois sítios de ligação para a ligação das proteínas chaperonas 14-3-3, o que resulta na translocação do complexo FoxO/14-3-3 do núcleo para o citoplasma ^[12]. Além disso, a ligação das proteínas 14-3-3 à FoxO é necessária para a dissociação da proteína FoxO do DNA e evita a retranslocação da FoxO para o núcleo, mascarando a função da sequência de localização nuclear (NLS). A proteína FoxO ao sair do núcleo não ativa os genes de

proliferação celular, resposta ao stress oxidativo e sobrevivência celular. As proteínas FoxO como tantas outras proteínas encontram-se em equilíbrio no organismo humano, isto é, existe proteína FoxO livre e proteína FoxO ligada ao DNA. Uma desregulação deste equilíbrio está relacionada com vários estados patológicos associados à desregulação do ciclo celular, como o cancro, diabetes, doenças neurodegenerativas, cardiovasculares e outras [4].

A acetilação ocorre principalmente nos três resíduos de Lisina do terminal C da FoxO, Lys242, Lys245 e Lys262, os quais podem ser acetilados por CBP para diminuir a afinidade pelo DNA e aumentar a sensibilidade para a fosforilação da Ser253 por PKB.

Diversos estudos demonstram que a deleção simultânea de FoxO1, FoxO3A e FoxO4 induz uma condição progressiva de predisposição ao cancro caracterizada por linfomas tímicos e hemangiomas, enquanto a perda de genes individuais resultou em fenótipos de neoplasia menos difundidos e mais modestos, apoiando a noção de que FoxOs são supressores genuínos de tumores [13].

O gene FoxO3 ou FoxO3a encontra-se no cromossoma 6 (NC_000006.12) do genoma humano e o seu número de identificação é 2309 (figura 1.5).

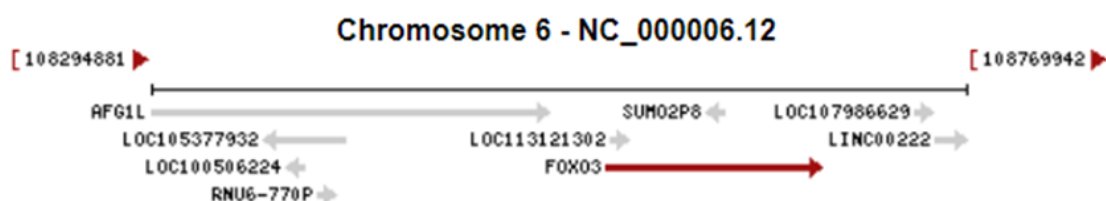


Figura 1.5: Localização do gene FoxO3 no cromossoma 6. Figura retirada da livreria nacional de medicina online [14].

A proteína FoxO3 apresenta-se no organismo humano na sua forma ligada ao DNA (estrutura com código 2UZK no PDB), e livre (estrutura com código 2K86 no PDB). A estrutura 2UZK [15] é a estrutura da forma cristalizada, obtida através do método de difracção de raios X, enquanto a estrutura 2K86 [4] é a estrutura em solução aquosa, obtida por meio da técnica de ressonância magnética nuclear (NMR). A estrutura ligada ao DNA encontra-se predominantemente no núcleo das células, enquanto a estrutura livre se encontra mais frequentemente no citoplasma, como se pode ver nas figuras 1.3 e 1.4.

A estrutura cristalina, 2UZK, com uma resolução de 2,7Å, compreende o domínio de ligação ao DNA (FoxO3a-DBD) ligado a um duplex de DNA de 13 pares de bases (pb), contendo uma sequência de ligação de consenso FoxO de 7 pb (GTAAACA). A estrutura 2UZK inicia-se no resíduo glicina 158 (GLY) e termina no resíduo serina 253 (SER). A estrutura tridimensional da 2UZK encontra-se na figura 1.1 e a hélice de reconhecimento, ou seja, a hélice que interage com as sequências de consenso do DNA é a hélice 3 (H3).

A hélice de reconhecimento (H3) liga-se no sulco principal (*major groove*) numa orientação aproximadamente perpendicular ao eixo do DNA, fazendo inúmeras interações com a sequência de consenso FoxO. Na região entre H2 e H3, os resíduos que precedem a hélice de reconhecimento formam uma estrutura de alça (“loop”) exposta ao solvente. A asa 1 (*Wing 1*) interage com o grupo fosfato da região flanqueadora do lado 3’ da sequência de consenso de forma não específica. Na região C-terminal de FoxO3a-DBD, os resíduos após a folha beta 2 formam uma estrutura em espiral que se alinha ao sulco principal e interage com o DNA (Figura 1.6).

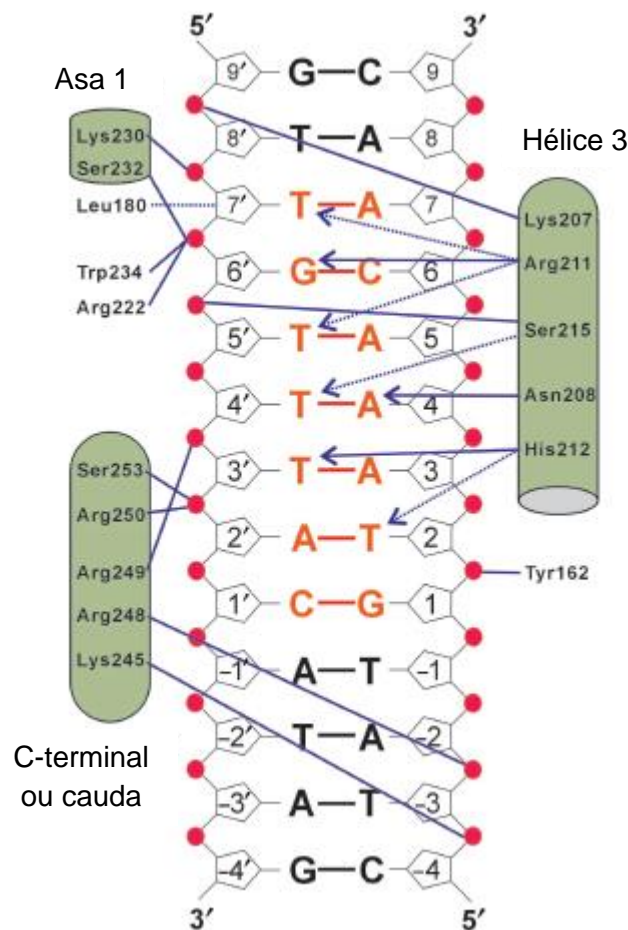


Figura 1.6: Diagrama esquemático de interações entre FOXO3a-DBD e DNA. Os grupos fosfato do DNA são representados como círculos vermelhos. Os pares de bases da sequência de ligação ao DNA do consenso FOXO, GTAAACA, estão a laranja. Os resíduos da hélice 3, da asa 1 e da bobina C-terminal que fazem contato com o DNA estão marcados em verde. As ligações de hidrogênio e os contatos de Van-der-Waals que participam do reconhecimento das bases são representados por setas com linhas a cheio e a tracejado, respectivamente. Ligações de hidrogênio e contatos de Van-der-Waals que não envolvem bases são representadas por linhas sem seta, respectivamente a cheio e tracejado. Imagem adaptada do artigo ^[15].

A hélice de reconhecimento H3 centra-se sobre a sequência consenso FoxO3, com vários dos seus resíduos formando ligações de hidrogênio, bem como contactos de Van-der-Waals, diretamente com as bases do sulco maior. Os resíduos conservados Asparagina 208 (Asn), Ser 215, Arg 211 e Histidina 212 (His) dentro de H3 são importantes para o reconhecimento de DNA. A Asn 208 liga-se à base Adenina 4 (A) através de duas ligações de hidrogênio. A Arg 211 reconhece a Timina 50 (T) e a T70 através de contatos de van-der-Waals e contribui para a especificidade do Guanina 60 (G) com uma ligação direta de hidrogênio. A cadeia lateral de His 212 projeta-se no sulco principal e reconhece as bases T2 e T30 através de um contato de van-der-Waals e uma ligação direta de hidrogênio, respetivamente. A Ser 215 reconhece a T40 através de um contato de van-der-Waals. Entre essas interações, nota-se que os grupos metil das bases de timina da sequência de ligação de consenso FoxO3 são cruciais para o reconhecimento do promotor FoxO3a-DBD. Além do reconhecimento de base, a Ser 215 também interage com o grupo fosfato de T80 para estabilizar ainda mais o complexo. A estrutura do complexo FoxO3a-DBD/DNA revela que Lys 245 interage diretamente com o grupo fosfato do DNA, mas a cadeia lateral de Lys 242 fica exposta ao solvente em vez de se ligar ao DNA. O C-terminal do DBD (W2) forma uma estrutura desordenada (“*random coil*”), a qual interage com os grupos fosfato carregados negativamente no sulco principal do DNA.

A estrutura de RMN, 2K86 inicia-se na serina 151 e termina na alanina 251. A estrutura tridimensional da 2K86 encontra-se na figura 1.7. Na figura 1.8 consta a sobreposição das duas estruturas, 2K86 e 2UZK.

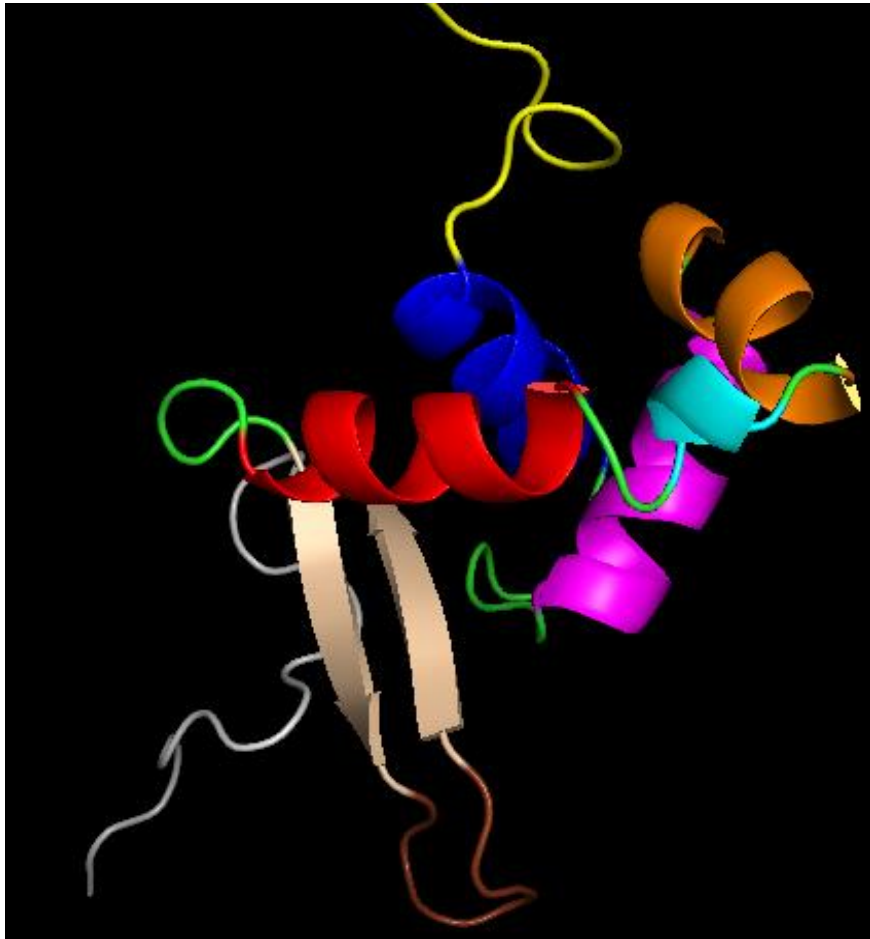


Figura 1.7: Estrutura tridimensional da estrutura 2K86. Amarelo N-terminal; azul Hélice 1; Violeta hélice 2; Laranja hélice 2; Vermelho hélice 3; Azul claro hélice 4; Castanho claro folhas beta; Castanho-escuro a assa 1; Branco o c-terminal (asa 2, W2); a verde zonas desordenadas. Imagem original realizada no PyMOL.



Figura 1.8: Sobreposição da estrutura 2UZK com a estrutura 2K86. A roxo encontra-se a estrutura 2UZK e a azul a estrutura 2K86. Presença da molécula de DNA. Imagem original realizada no PyMOL.

As principais diferenças da estrutura livre e ligada ao DNA são o c-terminal, que estão em sentidos opostos, pois a cauda da 2UZK participa na estabilização da ligação FoxO3-DNA e a presença de uma quarta hélice. Os restantes constituintes, ou seja as hélices 1,2 e 3 as folhas beta, a asa 1 e o N-terminal são conservados em ambas as estruturas e estão praticamente nos mesmos locais. A sobreposição de ambas as estruturas têm um valor de RMSD (desvio quadrático médio das posições atômicas) de 8.44 Å, ou seja ambas as estruturas são substancialmente diferentes, isto acontece devido as zonas desordenadas da proteína. Eliminando 177 átomos, que representam as zonas desordenadas, o RMSD é 2.605 Å, demonstrando que existem zonas da proteína que são conservadas e outras zonas que não se conservam, devido à diferença de ambiente em que encontram-se e da molécula de DNA estar ligada à proteína ou não. O valor do RMSD foi calculado no PyMOL, tendo em conta 575 átomos de cada estrutura.

2. Objetivo

O objetivo desta tese é encontrar moléculas moduladoras da atividade da proteína FoxO3, ou seja, moléculas que conseguem inibir ou potenciar a ligação da proteína FoxO3 ao DNA.

Uma molécula poderá inibir a ligação FoxO3-DNA por dois mecanismos distintos:

- 1) Interação com um local de ligação na FoxO3 o qual existe apenas na forma não ligada da proteína. Esta interação deslocará o equilíbrio de ligação no sentido da dissociação. A comparação das estruturas livre (2K86) e ligada (2UZK) da FoxO3 é um primeiro passo nesta abordagem.
- 2) Interação com locais de ligação na interface FoxO3-DNA, interrompendo ou bloqueando interações cruciais para a formação do complexo. Os segmentos essenciais de ligação são a Hélice de reconhecimento (H3), o C-terminal ou cauda e a asa 1, como referenciados na figura 1.7.

Do mesmo modo, uma molécula potenciadora da ligação FoxO3-DNA poderá ligar-se a um local exclusivo da estrutura ligada, ou atuar diretamente na interface FoxO3-DNA, estabilizando-a.

Considerando a importância do papel que a FoxO3 aparenta desempenhar em múltiplos mecanismos regulatórios e patológicos, a modelação da sua estrutura, função e ligação ao DNA reveste-se de grande importância. A possibilidade de modular a interação do domínio de ligação ao DNA (DBD) da FoxO3 por ação de ligandos, em particular, apresenta-se como uma via promissora no desenvolvimento de futuras terapêuticas. Para tal, torna-se necessário estudar a estrutura e dinâmica desta proteína por meio de simulações computacionais, tentar encontrar potenciais cavidades que possam permitir a ligação seletiva de ligandos à forma livre ou ligada da FoxO3 e identificação desses ligandos, através de *docking* molecular.

Os estudos de dinâmica molecular e de *docking* são adequados para este tipo de trabalho, pois permitem a observação e análise das diferenças estruturais entre as formas da proteína ligada e não-ligada ao DNA, a identificação de eventuais cavidades

ou bolsas de interação de ligandos presentes na superfície de uma ou outra das estruturas e a identificação de potenciais ligandos que possam servir de ponto de partida para desenvolvimento de novos fármacos ou avaliação de fármacos já existentes para outras patologias que possam ser reusados neste contexto.

Embora tenham sido já desenvolvidos estudos de índole semelhante ^{[15][16][17]} para outras proteínas da família Fox, as características e importância da FoxO3 numa variedade de processos celulares justificam um estudo aprofundado de pesquisa de potenciais moduladores da sua ação. É expectável o aparecimento de trabalhos semelhantes focando outros membros da família, se considerarmos a potencial importância de múltiplos membros da família FoxO enquanto mediadores de variadas patologias.

3. Enquadramento teórico

3.1 Dinâmica Molecular

As simulações de dinâmica molecular (DM) tiveram início nos anos 1950, sendo que a DM é uma técnica computacional que simula a evolução temporal de sistemas de átomos e moléculas sujeitos à ação de forças resultantes de um campo de potencial que modela as interações inter-atômicas do sistema em estudo. As simulações de DM têm sido usadas extensivamente para auxiliar e complementar o planeamento racional de novos fármacos através do computador, por meio do estudo da estrutura, função, complexação e interações das biomoléculas no meio. O estudo da dinâmica molecular é essencial para uma posterior pesquisa de conformações e cavidades moleculares, complementado com estudos de docking computacional para um melhor planeamento racional de novos fármacos ^[18].

A metodologia da DM é fundamentada nos princípios da Mecânica Clássica ^[19], tendo como bases as leis da mecânica Newtoniana e o cálculo das forças a partir de um potencial interatômico de pares (campo de forças). A DM é uma extensão da Mecânica Molecular, onde o comportamento dinâmico de um sistema molecular é simulado através da integração numérica das equações de movimento.

Segundo a mecânica molecular:

- O sistema molecular é descrito por uma função designada campo de forças que representa a energia do sistema como função da posição dos centros atômicos (aproximação de Born-Oppenheimer).
- Os átomos são descritos como pontos com uma determinada carga (cargas pontuais) e massa.
- As ligações covalentes são tratadas como “molas” (aproximação harmónica), com uma constante de força que depende dos átomos ligados e da ordem da ligação.
- As ligações não-covalente podem ser eletrostáticas (interações entre cargas pontuais), ou forças de Van-der-Waals (termos de Lennard-Jones)
- São usados termos adicionais para fixar os valores dos ângulos de ligação, torções (ângulos dihedro) e outros termos de energia.

Existem diversos programas computacionais que podem ser usados para os cálculos de dinâmica molecular, mas todos se baseiam no mesmo método. De modo simplificado encontra-se na figura 3.1.9 um esquema do procedimento da dinâmica molecular.

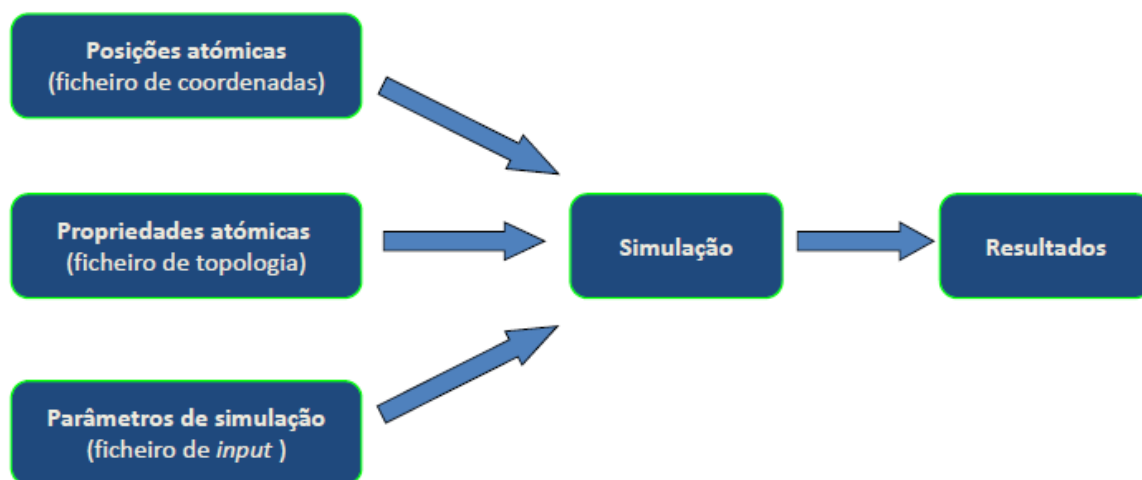


Figura 3.1.9: Esquema simplificado do procedimento da dinâmica molecular.

Primeiramente retira-se o ficheiro em formato *PDB* do site *Protein Data Bank* e antes da simulação de dinâmica molecular tem que programar-se certos parâmetros, por exemplo, o tipo de campo de forças, a forma da caixa, o tamanho da caixa, o tipo de solvente em solução, os iões, o *timestep* e o *timestep compressed*. Após retirar o ficheiro *pdb* e escolha do campo de forças os ficheiros de coordenadas de topologia

e de *input* são gerados automaticamente. Estes ficheiros são essenciais para a corrida de DM, pois contêm as informações sobre os aminoácidos que constituem a proteína e as suas coordenadas espaciais.

Ainda antes da simulação a molécula dentro da caixa tem que passar por um processo de minimização de energia e duas fases de equilibração. Após todo este processo dá-se a simulação de dinâmica molecular e obtêm-se os seus resultados.

O campo de força empírico é conhecido como uma função de energia potencial onde permite que a energia potencial total do sistema, $V(r)$, seja calculada a partir da estrutura tridimensional (3D) do sistema molecular. O $V(r)$ é descrito como a soma de vários termos de energia, incluindo os termos para átomos ligados (comprimentos e ângulos de ligação, ângulos diedros) e os termos para átomos não ligados (interações de Van-der-Waals e de Coulomb). Um típico campo de força é representado pela Equação 3.1.1:

$$V(r) = \sum V_l + \sum V_\theta + \sum V_{vdW} + \sum V_{elet}$$

Equação 3.1.1

Onde, V_l é a energia de estiramento da ligação em relação a seu valor de equilíbrio (ou ideal), o V_θ é a energia de deformação do ângulo de ligação em relação a seu valor de equilíbrio, o V_ϕ é a energia devido à torção em torno de uma ligação, o V_{vdW} representa a energia das interações de Van-der-Waals e o V_{elet} representa as energias de atração ou repulsão eletrostática entre duas cargas.

A energia das interações de Van-der-Waals e as energias de atração ou repulsão eletrostáticas entre duas cargas são representados respectivamente pelos potenciais de Lennard-Jones (Equação 3.1.2) e de Coulomb (Equação 3.1.3).

$$V_{vdW} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

Equação 3.1.2

$$V_{elet} = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_{rr_{ij}}}$$

Equação 3.1.3

Na Equação 3.1.2, ϵ_{ij} é a profundidade do potencial entre a barreira atrativa e a repulsiva, e σ_{ij} é a distância finita na qual o potencial interpartícula é zero. Ambos são parâmetros ajustados experimentalmente ou por cálculos teóricos. No caso das interações eletrostáticas (Equação 3.1.3), q_i e q_j correspondem à magnitude das cargas pontuais de cada átomo, r_{ij} a distância entre as cargas, ϵ_0 a permissividade do espaço livre e ϵ_r a constante dielétrica relativa do meio. As cargas pontuais são geralmente obtidas por cálculos quânticos *ab initio* e ajustes empíricos.

Vários campos de forças foram desenvolvidos pela comunidade científica de forma independente e com conjuntos de parâmetros específicos e geralmente não transportáveis entre campos de forças. Alguns incluem outros termos para descrever especificamente as ligações de hidrogénio ou para acoplar oscilações entre ângulos e comprimentos de ligação, com o objetivo de se obter uma melhor concordância com espectros vibracionais e outras propriedades experimentais de conjuntos de moléculas de teste. A fiabilidade dos resultados obtidos em simulações com um determinado campo de forças está totalmente dependente da parametrização rigorosa desse mesmo campo de forças. Além disso, o conjunto de moléculas de teste usado na parametrização condiciona fortemente a aplicabilidade do mesmo. Assim, a escolha do campo de forças depende, em grande parte, do sistema a ser estudado e das propriedades que serão investigadas. No caso de sistemas biomoleculares, os campos de força mais utilizados são *CHARMM*, *GROMOS*, *AMBER*, *OPLS*, *CVFF*, entre outros ^[19].

Numa simulação de dinâmica molecular são definidos, para além do campo de forças a utilizar, uma série de outros parâmetros que descrevem a geometria e propriedades do sistema, bem como outras propriedades do próprio algoritmo de simulação. Dois deles são a forma e o tamanho da caixa, sendo que este último deverá ser suficientemente grande para evitar artefactos na fronteira do sistema. As condições de fronteira são frequentemente estabelecidas pela escolha de valores fixos para as extremidades da caixa (o que pode causar artefactos) ou pela utilização de condições de fronteira periódicas nas quais um lado da caixa de simulação interage com o outro lado da mesma, artificialmente criando um sistema infinito. Para uma simulação em solução aquosa, o número de moléculas de água aumenta com o tamanho da caixa de simulação, e também o peso computacional, tamanho dos

ficheiros produzidos e tempo de cálculo. Assim, simulações podem ser feitas em caixas com diferentes geometrias, tais como cúbicas, octaédricas truncadas, dodecaédricas ou de prisma hexagonal. A geometria da caixa é normalmente escolhida de acordo com a geometria do sistema macromolecular a simular, por forma a minimizar o número de moléculas de solvente necessárias e reduzir o peso computacional da simulação.

Por outro lado, para simular moléculas em solvente é necessária a escolha entre solventes explícitos, os quais são modelos determinados a partir de mecânica molecular, mecânica quântica e resultados experimentais e são definidos tal como as restantes partículas na simulação, ou solventes implícitos que representam o solvente como um meio contínuo, geralmente mais aplicável a líquidos. Partículas explícitas de solvente (como os modelos da água *TIP3P*, *SPC/E* e *SPC-f*) têm de ser calculados exhaustivamente por via do campo de forças, enquanto os solventes implícitos utilizam uma aproximação de campo médio para o cálculo das forças. A utilização de um solvente explícito é computacionalmente exigente e requer a inclusão de cerca de dez vezes mais partículas na simulação. Contudo, a granularidade e viscosidade de um solvente explícito é essencial para reproduzir certas propriedades das moléculas de soluto, sendo particularmente importante para reproduzir a cinética dos processos em estudo.

De forma a garantir a eletroneutralidade do sistema molecular, adicionam-se iões positivos ou negativos ao sistema em número suficiente para neutralizar a carga da macromolécula em estudo. Esta neutralização é obrigatória quando se usa o método PME (*particle-mesh Ewald*)^[20] para o cálculo das interações electrostáticas de longa distância.

A minimização de energia é uma técnica computacional que visa encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema de interesse. A minimização prévia do sistema a simular garante a geometria adequada e uma menor força sobre os átomos.

Os últimos parâmetros a serem adicionados são o *timestep* e o *timestep compressed*. O *timestep* é o intervalo de tempo entre as avaliações do potencial resultante da discretização do tempo empregue nos algoritmos de integração das

equações do movimento das partículas. Este intervalo de tempo deve ser pequeno, o suficiente para evitar erros de discretização (ou seja, menor do que o período correspondente à frequência vibracional mais alta no sistema). Um valor típico do *timesteps* para simulações de DM é na ordem de 1-2 femtosegundo (10^{-15} s), mas depende do campo de forças escolhido e também das velocidades e frequências de vibração dos átomos do sistema. Quanto mais altas forem as velocidades das moléculas, menor terá de ser o valor de *timestep*, para evitar variações significativas da conformação das moléculas entre *timesteps* consecutivos. A descrição do tempo numa simulação computacional é necessariamente descontínua, sendo necessário o uso de intervalos de tempo muito pequenos para aproximar a natureza contínua das trajetórias e não perder conformações intermédias. Por outro lado, o *timestep compressed* é um parâmetro que indica o intervalo de tempo (em número de *timesteps*) entre escritas sucessivas das coordenadas e outras propriedades do sistema, ao longo da dinâmica molecular (o termo “*compressed*” resultado do facto de que os dados são escritos num formato comprimido). Os documentos escritos poderão depois ser analisados e reunidos de modo a criar uma representação contínua da simulação.

Na fase 1 da equilibração a quantidade de substância (N), o volume (V) e a temperatura (T) são mantidos constantes, ou seja a simulação é realizada no ensemble canónico (NVT). Por vezes as simulações neste ensemble também se designam por dinâmica molecular a temperatura constante (CTMD). No *ensemble* NVT a temperatura é mantida constante por via de trocas de energia com um termóstato (reservatório de calor, ou banho térmico, virtual). Existem vários algoritmos capazes de definir o banho térmico virtual, por forma a assegurar a manutenção de uma temperatura aproximadamente constante. Técnicas populares de controlo da temperatura incluem o reescalamento das velocidades, o termóstato de Nosé-Hoover^{[21][22]}, correntes de Nosé-Hoover, o termóstato de Berendsen, o termóstato de Andersen e a dinâmica de Langevin. É de salientar que o termóstato de Berendsen pode introduzir o efeito de “cubo de gelo voador” (*flying ice cube*), em que toda a energia do sistema é escoada para os graus de liberdade rotacionais e translacionais do sistema, com o conseqüente “congelamento” da estrutura. Este problema é particularmente comum se as simulações forem realizadas em vácuo.

A fase 2 da equilibração é realizada num ensemble isotérmico-isobárico (NPT), em que quantidade de substância (N), pressão (P) e temperatura (T) são mantidos constantes. Em adição ao termóstato, é necessário um baróstato para manter constante a pressão sobre o sistema. Estas condições são as que melhor correspondem às condições laboratoriais de um recipiente aberto à temperatura e pressão ambiente.

Na figura 3.1.10 é possível observar a proteína pronta para a simulação.

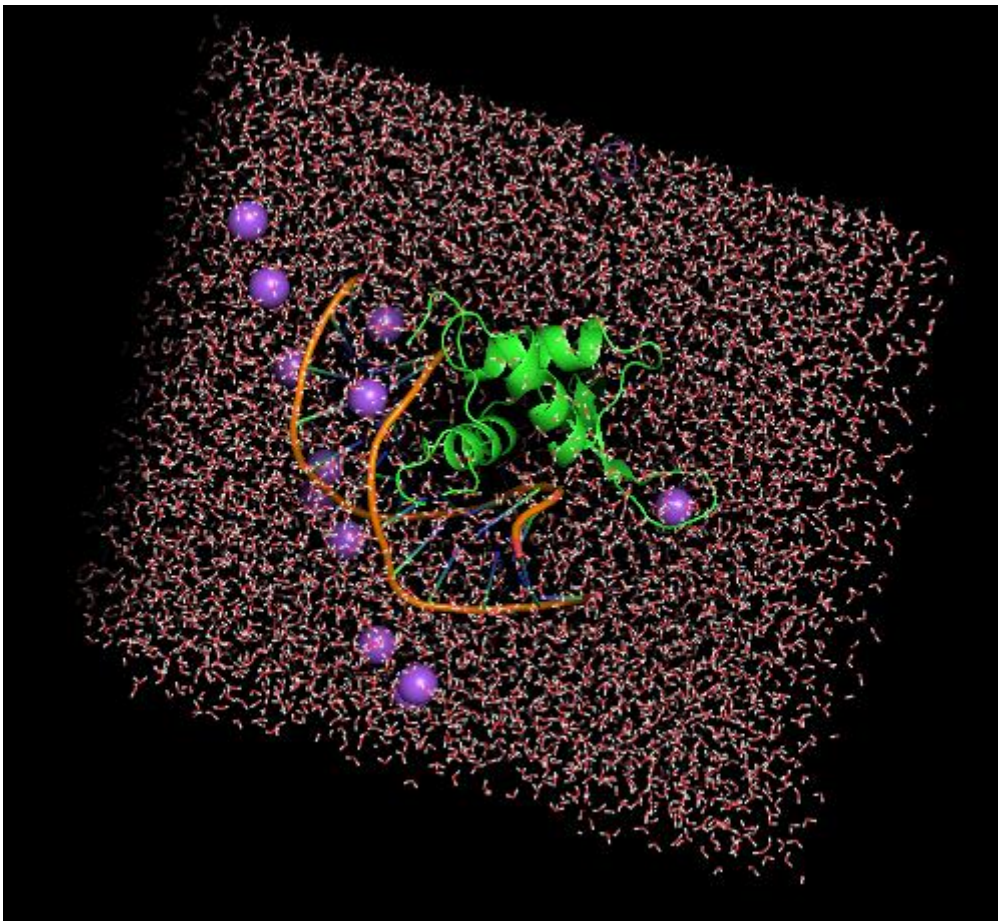


Figura 3.1.10: Complexo DNA-proteína numa caixa cúbica com as moléculas de água e os iões. Os iões estão representados a roxo; a proteína a verde; a vermelho e branco encontram-se as moléculas de água; a laranja a cadeia do DNA. Imagem original feita no PyMOL

Após todos estes processos e uma vez definido o campo de forças pode-se passar à simulação. Durante a simulação calculam-se as forças que atuam sobre cada átomo, derivando a energia potencial (obtida do campo de força escolhido) em relação às posições desses mesmos átomos. A partir da equação da aceleração da partícula (equação 3.1.4) e integrando as equações de movimento podem obter-se as velocidades, cuja integral, por sua vez, fornece a mudança de posição de cada átomo.

Com as novas posições e velocidades de cada partícula, obtêm-se as energias potenciais e cinética do sistema. Aplicando-se sucessivamente esse procedimento, obtêm-se o que se denomina de “trajetória”, que nada mais é do que o conjunto de posições e velocidades de cada partícula ao longo do tempo. Um dos métodos mais utilizados em dinâmica molecular para integrar as equações de movimento é o algoritmo de Verlet, que utiliza as posições e acelerações dos átomos no tempo t e as posições do passo anterior, $r(t-\Delta t)$, para determinar as novas posições no tempo $t+\Delta t$, de acordo com a equação 3.1.5 [15]. Na figura 3.1.11 encontra-se o esquema da etapa de simulação da DM, descrita anteriormente.

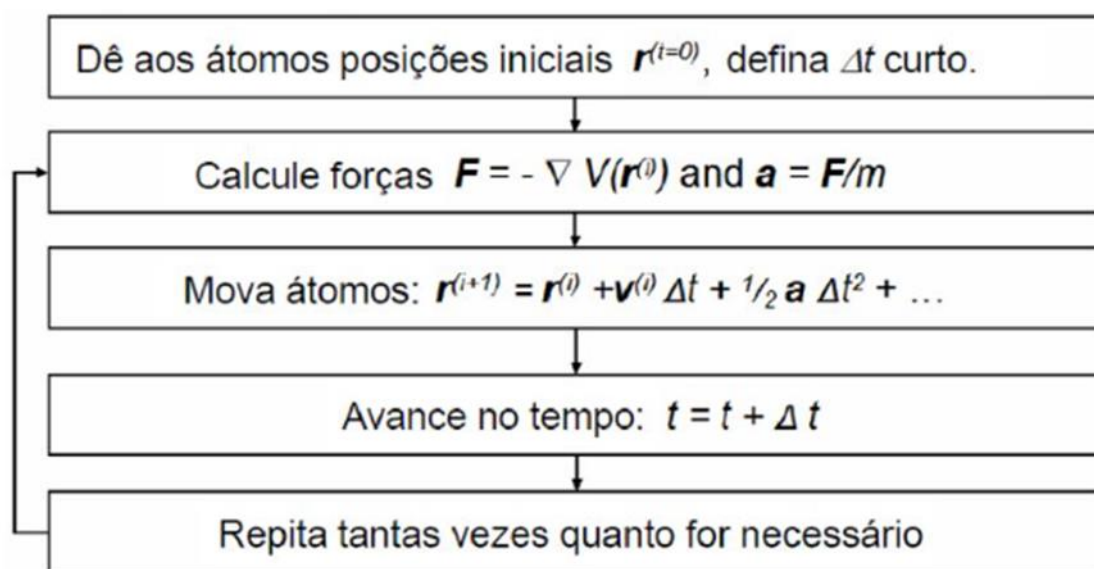


Figura 3.1.11: Representação esquemática das etapas de uma simulação de dinâmica molecular.

$$F_i(t) = -\frac{\partial V(r_i)}{\partial r_i} \quad \text{Equação 3.1.4}$$

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \quad \text{Equação 3.1.5}$$

A descrição da realidade física é forçosamente aproximada nos métodos de simulação, conduzindo a limitações mais ou menos importantes na descrição das propriedades do sistema. No caso presente, a descrição dos vários potenciais interatômicos é simplificada de forma mais ou menos grosseira para tornar viável a realização de simulações em sistemas com elevado número de átomos (como as macromoléculas biológicas). As ligações de hidrogénio intramoleculares, as interações de Van-der-waals e as forças eletrostáticas de longo alcance são exemplos

de potenciais com descrição aproximada. As ligações de hidrogénio intramoleculares não são explicitamente incluídas na maioria dos campos de forças atuais, mas são descritos como interações de Coulomb de cargas atômicas pontuais. Esta é uma aproximação algo grosseira pois as ligações de hidrogénio têm em parte um carácter quântico de ligação covalente. Além disso, as interações eletrostáticas são geralmente calculadas utilizando a constante dielétrica do vácuo, ainda que o meio aquoso envolvente tenha uma constante dielétrica muito superior (embora a utilização da constante dielétrica macroscópica para pequenas distâncias interatómicas seja questionável). As cargas do sistema são geralmente representadas como pontos, ignorando efeitos de polarização e interações multipolo de ordem superior. Por fim, as interações de Van-der-Waals em DM são geralmente descritas pelos potenciais de Lennard-Jones (teoria de Fritz London), embora em rigor estes se apliquem apenas no vácuo. Porém, todos os tipos de forças de Van-der-Waals são de origem eletrostática e portanto dependem das propriedades dielétricas do ambiente. A medição direta das forças atrativas entre diferentes materiais (na forma da constante de Hamaker) demonstra que "A interação entre hidrocarbonetos em meio aquoso é cerca de 10% da sua atracção no vácuo". Normalmente negligencia-se a dependência das forças de Van-der-Waals em relação ao ambiente nas simulações, mas esta pode ser incluída se na simulação se utilizarem campos de força com cargas polarizáveis. Outro problema que estas simulações apresentam é o cálculo das interações eletrostáticas de longo alcance, as quais são fortemente condicionadas pelo meio ambiente, tornando necessário usar métodos como o *Particle Mesh Ewald* (PME) [20]. Esta abordagem é necessária, porque o simples cálculo das interações entre cargas da proteína usando como meio ambiente o vácuo não leva em consideração o efeito da polarização das moléculas de água e da sua re-orientação, a qual num sistema físico real ocorre num volume muito maior do que aquele que é definido pela caixa de simulação. Ou seja, o contributo das moléculas de água que estão para lá do volume da caixa não pode ser desprezado, e tem um efeito significativo sobre as partículas carregadas dentro da caixa. O método de PME é um processo matemático de aproximar o contributo desse volume praticamente infinito de água que circunda a molécula de proteína. Este cálculo torna as simulações consideravelmente mais pesadas, mas é geralmente aceite que o sistema se comporta de uma forma fisicamente mais realista. Para simulações usando PME define-se um parâmetro de cut-off, o qual é uma distância (em nanómetros) abaixo da qual as interações entre

cargas são calculadas pela aplicação direta da equação de Coulomb a cada par de cargas. A contribuição energética de cargas a uma distância superior ao cut-off é calculada pelo método PME.

O resultado obtido da simulação de dinâmica molecular é uma trajetória do sistema de partículas, ou seja, é um conjunto de configurações que estas mesmas partículas (átomos e moléculas) vão adotando ao longo do tempo.

Os tempos de simulação de sistemas biológicos, tais como proteínas ou fragmentos de DNA em solução aquosa, são geralmente da escala do nano ao microssegundo. Apesar de parecer um intervalo de tempo muito curto, os resultados dessas simulações podem fornecer dados fiáveis já que os eventos de interesse ocorrem na mesma escala temporal. Ainda que o intervalo de tempo do sistema físico a simular esteja na escala dos nanossegundos, para sistemas com elevado número de átomos (como é o caso das proteínas e outras macromoléculas biológicas) o tempo computacional “real” destas simulações é geralmente de horas a semanas, resultando ainda na geração de ficheiros de dados com gigabytes de tamanho. Desta forma, torna-se necessário o acesso a sistemas computacionais de capacidade e performance elevada para levar a cabo este tipo de estudos.

A técnica da dinâmica molecular é uma metodologia adequada para este tipo de trabalho, pois permite a identificação e análise das diferenças estruturais entre a proteína ligada e não-ligada ao DNA para um conjunto alargado de conformações que permitem uma visão dinâmica do problema.

3.2 Descoberta de cavidades ou sítio de ligação

As interações proteína-ligando são cruciais para todos os organismos vivos, e uma única alteração na sequência de aminoácidos de um local de interação pode levar a doenças devastadoras. Assim sendo, o estudo de sítios de ligação é relevante em múltiplas áreas da investigação biomédica, incluindo mecanismos de doenças, planeamento de fármacos e regulação fisiológica. Para possibilitar este estudo, foram desenvolvidos diversos algoritmos, ferramentas e *softwares* fiáveis e eficientes capazes de detetar cavidades em recetores proteicos.

Após descoberta dos sítios de ligação é possível fazerem-se estudos e design de um inibidor para silenciar a função de uma proteína ou de uma molécula potenciadora da função da proteína, sendo estas duas técnicas muito comuns na terapia farmacêutica.

Em bioquímica e biologia molecular, um sítio de ligação/cavidade/bolso é uma região na superfície de uma macromolécula, com potencial capacidade de ligar uma molécula mais pequena, à qual se dá normalmente o nome de “ligando”.

Já o sítio ativo é normalmente parte de um sítio de ligação e é uma combinação de resíduos de aminoácidos (e por vezes cofatores) que estão diretamente envolvidos no mecanismo de catálise enzimática. Assim, a ligação de um ligando ao centro ativo desencadeia uma reação química na qual se forma um produto, enquanto no sítio de ligação se pode ligar um ligando, mas não ocorre necessariamente nenhuma reação química. Os sítios de ligação são bastante importantes, pois podem servir de locais de inibição ou potenciadores da função da macromolécula. Estes sítios possuem frequentemente complementaridade de forma ou outras propriedades com os seus ligandos, de modo a garantir a especificidade e afinidade. A ligação do ligando à macromolécula é frequentemente, mas nem sempre, acompanhada por uma mudança conformacional que pode afetar a função da proteína (ajuste induzido) [23].

Os ligandos podem ser outras proteínas, ácidos nucleicos, substratos enzimáticos, segundos mensageiros, hormonas, moduladores alostéricos ou fármacos e podem ligar-se de forma reversível ou irreversível à macromolécula. Se os ligandos formarem ligações não covalentes, como por exemplo pontes de hidrogénio, interações eletrostáticas e ligações de Van-der-Waals, a sua dissociação da macromolécula é muito mais fácil e estamos na presença de uma ligação reversível. As ligações irreversíveis são mais raras, ocorrendo geralmente através de ligações covalentes entre a macromolécula e o ligando. Os ligandos podem ser inibidores ou potenciadores da função da macromolécula, ou seja, se o ligando se ligar a um determinado sítio de ligação pode inibir a função da proteína e se outro ligando se ligar a um outro sítio de ligação já poderá ser um potenciador da função da macromolécula. Este aspeto justifica a importância do estudo dos sítios de ligação das macromoléculas e os seus ligandos, pois esse estudo pode conduzir à descoberta de

ligandos capazes de interferir numa doença ou numa terapêutica, como por exemplo o caso dos beta-bloqueadores.

Nas doenças cardiovasculares, utilizam-se os medicamentos da classe dos betabloqueadores para tratar pacientes com hipertensão. Os beta-bloqueadores (β -bloqueadores) são agentes anti-hipertensivos que bloqueiam a ligação das hormonas adrenalina e noradrenalina aos recetores β_1 e β_2 no coração e nos vasos sanguíneos, ou seja descobriu-se um conjunto de moléculas que competiam pelo sítio de ligação das hormonas que causam a hipertensão arterial. Essas moléculas foram levadas a ensaios clínicos e hoje em dia são considerados medicamentos para tratar a hipertensão arterial. Todo este processo leva muitos anos a ser concluído, iniciando-se com uma etapa de identificação da doença, do mecanismo desta e seleção de uma molécula alvo. Após a identificação do alvo, fazem-se então estudos dos sítios de ligação e posterior *docking* desses sítios de ligação com uma variedade de ligandos por forma a descobrir moléculas capazes de ligar-se ao sítio de ligação e potencialmente mitigar ou inibir totalmente o curso da doença em estudo.

Na figura 3.2.12 é possível observar um sítio de ligação.

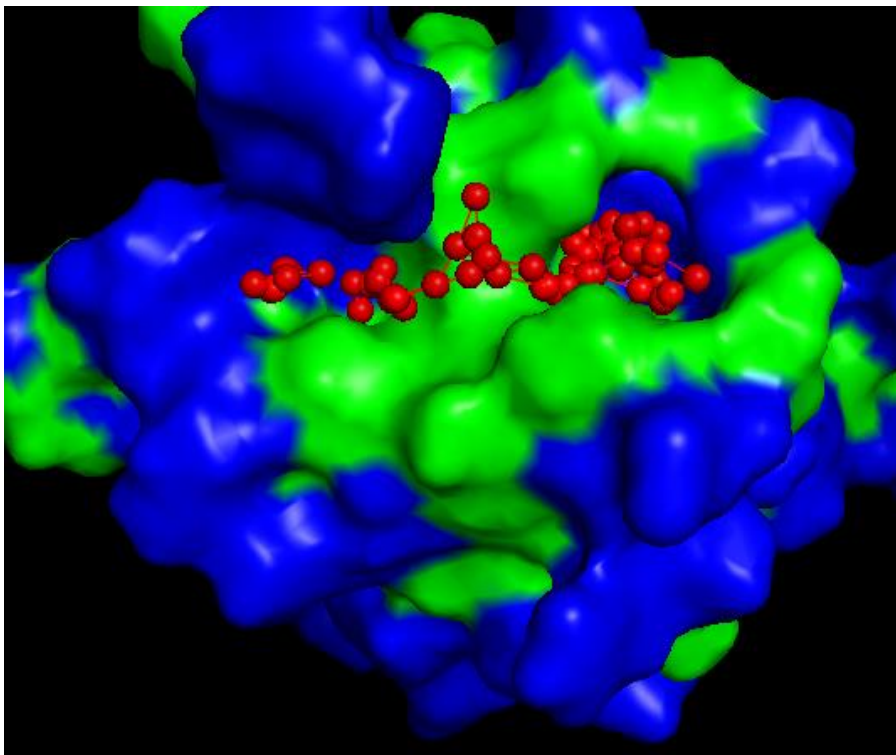


Figura 3.2.12: Sítio de ligação da FOXO3. Representação da proteína em superfície. A verde encontra-se a zona hidrofóbica da proteína e a azul o resto da proteína. As esferas a vermelho denotam o volume de uma potencial cavidade/sítio de ligação prevista pelo *software* FPocket. Imagem original, obtida através do programa FPocket e visualizada no Pymol.

Alta afinidade e especificidade são duas características fulcrais e imprescindíveis dos ligandos. Uma afinidade elevada é essencial para que o ligando se ligue ao recetor no local em estudo e produza o efeito desejado, minimizando a ligação a outros locais do alvo ou outras macromoléculas, o que causaria possivelmente efeitos colaterais indesejados e/ou inatividade da ação esperada. A especificidade e afinidade são duas características que determinam a força da ligação química e a natureza do ligando que irá ligar-se. A especificidade desempenha um papel muito importante no reconhecimento e ligação de um ligando com uma molécula alvo, ou seja, se ocorrer uma pequena mudança estrutural ou mutação no sítio de ligação de uma proteína, isso pode resultar numa perda de ligação porque o ligando deixa de ser capaz de reconhecer a forma 3D ou o arranjo espacial da proteína.

Em um sistema real, a interação de um ligando com o seu sítio de ligação não depende apenas da proteína alvo, mas também de outros fatores ambientais, como temperatura, natureza do solvente, pH e presença de iões, cofatores ou outras moléculas. O tamanho, profundidade e hidrofobicidade da cavidade/sítio de ligação também são importantes para a ligação com o ligando.

Os *softwares* para deteção de cavidades são desenvolvidos para diferentes propósitos, utilizando diferentes conjuntos de dados para treino e teste, e diferentes abordagens metodológicas. Sendo assim, deve ser feita uma escolha criteriosa do *software* a utilizar, tendo em conta o sistema em estudo. Os métodos de deteção de cavidade podem ser classificados nas seguintes categorias:

- Métodos baseados em sequência ou previsões evolutivas
- Métodos baseados em similaridade estrutural
- Método baseado em geometria ou previsões baseada em estrutura
- Métodos baseados em energia

O primeiro método a ser descrito foi o método baseado em sequência ou evolutivos preveem o sítio de ligação usando apenas sequências em formação. Exploram a informação evolutiva contida em Alinhamentos de Sequências Múltiplas (MSAs) de sequências homólogas e extraem um subconjunto de resíduos que mostram um alto grau de conservação. A suposição por trás dessa ideia é que a pressão evolutiva agindo sobre resíduos funcionalmente importantes reduzirá a sua variabilidade numa

família de proteínas. Apesar da sua utilidade para inferir resíduos funcionalmente importantes, todos os métodos baseados em sequência sofrem da limitação fundamental de não serem capazes de discriminar entre resíduos que são conservados como parte de um sítio de ligação de resíduos que são cruciais para a estabilidade, regulação ou *folding*. Por outras palavras, enquanto os resíduos de ligação são geralmente conservados numa família de proteínas, a conservação por si só nem sempre é um critério suficientemente específico para identificar um sítio de ligação, uma vez que os resíduos podem ser conservados por outras razões que não a ligação. Os métodos baseados em sequência também não fornecem informações geométricas e físico-químicas sobre o sítio de ligação, como área, volume, forma e propriedades de interação molecular. Para superar essas limitações, outras abordagens foram concebidas que levam explicitamente em consideração as informações estruturais [24].

Os métodos baseados em similaridade estrutural pressupõem que a similaridade de sequência leva à similaridade estrutural. Ou seja, o sítio de ligação em uma proteína pode ser previsto pela comparação de proteínas semelhantes com sítios de ligação conhecidos [19]. Uma limitação deste método é o estar dependente de haver estrutura de proteínas semelhante à proteína em estudo com ligandos conhecidos.

Nos métodos baseados em geometria as cavidades são detetadas na geometria do recetor, ou seja, a maior cavidade do recetor é a cavidade a qual se liga o ligando. *Laskowski et al*, verificaram que 83% das vezes a cavidade a qual se liga o ligando é a maior do recetor. Ou seja este método baseia-se no tamanho, área ou volume da cavidade [19]. Apesar da sua utilidade para a identificação de sítios de ligação, as abordagens geométricas têm dificuldade em identificar sítios de ligação que não sejam bolsas profundas. Além disso, as abordagens geométricas são incapazes de distinguir diferentes tipos de sítios, como hidrofóbicos versus polares, o que pode fornecer informações adicionais sobre a possível função de uma proteína [25].

Os métodos de identificação de cavidades de ligação baseados em energia resolvem este problema usando sondas mutáveis para detetar as diferentes bolsas de ligação. Ao contrário dos métodos anteriormente descritos, os métodos baseados em energia não detetam apenas a maior cavidade do recetor, sendo por isso mais

amplamente utilizados e fiáveis. Estes métodos consideram o potencial de interação entre proteína e ligando. Locais energeticamente favoráveis indicam a localização da proteína recetora onde um ligando pode interagir e ligar-se. São geralmente usados três tipos de sondas diferentes, para localizar o sítio hidrofóbico, os locais aceitador de ligações de hidrogénio e locais dadores de ligações de hidrogénio [25].

Após deteção dos sítios de ligação, é importante caracterizar esse sítio, determinando por exemplo o pKa, flexibilidade, tamanho, forma, hidrofobicidade, hidrofiliidade e drogabilidade da bolsa e dos grupos químicos que a formam. Estes fatores são importantes para posterior seleção e testagem de ligandos ao sítio ativo (exemplo: não se irá testar um ligando de grandes dimensões numa cavidade de pequenas dimensões).

A drogabilidade de uma pequena molécula foi definida primeiramente por Hopkins e Groom como a sua capacidade de modular uma proteína relacionada com uma doença específica [26].

No curso do estudo de um recetor, após identificar as cavidades é importante rever num banco de dados se essas cavidades estão presentes/são semelhantes as cavidades encontradas noutras proteínas, de modo a validar a cavidade (especificidade). Uma boa base de dados para essa comparação é o *sc-PDB*, por exemplo, mas existem mais bases de dados com este propósito.

3.3 Docking

A Docagem Molecular, também conhecida como Acoplamento molecular, Ancoragem molecular ou *Docking* é um procedimento computacional que tenta prever o modo de ligação não-covalente de macromoléculas ou, mais frequentemente, de uma macromolécula (recetor) e uma pequena molécula (ligando) de forma eficiente. O objetivo do *docking* é alcançar uma conformação otimizada para o sistema proteína-ligando e sua orientação relativa, de modo a que a energia livre do sistema seja mínima [27].

O *docking* é atualmente bastante utilizado, para as mais diversas finalidades, por exemplo:

- Realizar triagem (*screening*) virtual de grandes bibliotecas de compostos, classificar resultados e propor hipóteses estruturais de como os ligandos se ligam aos alvos, o que é de extrema importância para a otimização de leads.
- Aumenta a probabilidade de encontrar novos inibidores de proteínas-chave de patógenos, ao prever quão bem um composto pode ligar-se ao alvo, onde ele pode ligar-se, e que tipos de interações poderia formar no local de ligação.
- Identificar candidatos promissores para testes de laboratório, o que poupa substancialmente o tempo e custo do trabalho laboratorial.
- Desenho de fármacos com base na estrutura, devido à capacidade de prever a conformação de ligação de ligandos de moléculas pequenas ao local de ligação do alvo desejado.
- Caracterização do comportamento da ligação recetor-ligado, a qual desempenha um papel importante no *design* racional de fármacos, bem como no elucidar de processos bioquímicos fundamentais.

Existem duas abordagens que são particularmente populares na comunidade de *docking* molecular. Uma abordagem usa uma técnica de correspondência que descreve a proteína e o ligando como superfícies complementares (Figura 3.3.13). A segunda abordagem simula o processo de ancoragem real no qual as energias de interação entre ligando-proteína são calculadas. Ambas as abordagens têm vantagens significativas, bem como algumas limitações.

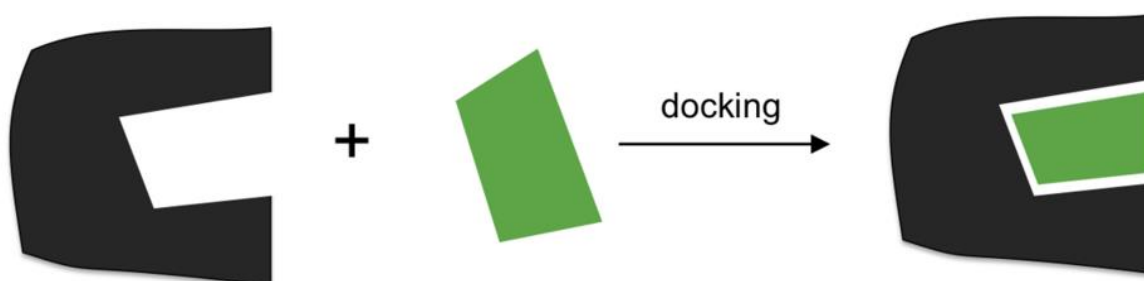


Figura 3.3.13: Processo simplificado do *docking*. Molécula alvo a preto, ligando a verde e o produto final após *docking* (Molécula alvo-Ligando). Imagem consultada no dia 20/8/22 as 20h, retirada da wikipedia commons, de acordo com a licença Creative Commons Attribution-Share Alike 4.0 International [28].

Os métodos de correspondência geométrica/complementaridade de forma descrevem a proteína e o ligando como um conjunto de características que os tornam acopláveis. Esses recursos podem incluir descritores de superfície molecular/superfície complementar. Neste caso, a superfície molecular do recetor é descrita em termos de sua área de superfície acessível ao solvente e a superfície molecular do ligando é descrita em termos de sua descrição de superfície correspondente. A complementaridade entre as duas superfícies equivale à descrição de correspondência de forma que pode ajudar a encontrar a pose complementar de encaixar o alvo e as moléculas do ligando. Outra abordagem é descrever as características hidrofóbicas da proteína usando voltas nos átomos da cadeia principal. Ainda outra abordagem é usar uma técnica de descritor de forma de Fourier. As abordagens baseadas em complementaridade de forma são tipicamente rápidas e robustas, mas geralmente não conseguem modelar movimentos ou a dinâmica conformacional do complexo ligando/proteína com precisão, embora desenvolvimentos recentes permitam que esses métodos investiguem a flexibilidade do ligando. Os métodos de complementaridade de forma podem varrer rapidamente vários milhares de ligandos em questão de segundos e realmente descobrir se eles podem se ligar ao sítio ativo da proteína e geralmente são escaláveis até mesmo para interações proteína-proteína. Eles também são muito mais acessíveis a abordagens baseadas em farmacóforos, uma vez que usam descrições geométricas dos ligandos para encontrar a ligação ideal [25].

Como ponto de partida do *docking*, utilizam-se estruturas não ligadas do recetor e possíveis ligandos. Essas estruturas que podem ser obtidas por uma variedade de processos: via experimental (cristalografia de raios X, NMR), simulações de dinâmica molecular, de modelação por homologia, entre outros), que se encontram em diversas bases de dados como por exemplo o *Pubchem*, *Protein Data Bank*, *DrugBank*, entre outros.

Antes de se poder realizar o *docking* tem que se escolher, conhecer e estudar a proteína recetora, pois se o *docking* for realizado após uma avaliação de dinâmica molecular e das potenciais cavidades da proteína recetora, este terá em conta a mobilidade da proteína e irá haver uma seleção prévia das partes importantes da

estrutura para a corrida de *docking*, o que diminuirá o tempo da corrida, pois não se realiza o *docking* em toda a proteína.

Outro ponto importante antes do *docking* é a seleção dos ligandos, pois existem vários bilhões de ligandos e nem todos podem ser testados em tempo significativo, por este motivo é que é tão importante haver uma seleção prévia dos ligandos a serem testados. De modo a selecionar os melhores ligandos podemos usar alguns conceitos, como por exemplo, a regras dos cinco, líder-similar, fragmento-similar, fármaco-similar. A regra dos 5 (Lipinski) ^[29], fragmento-similar, líder-similar e fármaco-similar, que tem em conta o peso molecular, número de aceitadores ou doadores de ligações de hidrogénios e o coeficiente de partição calculado para o sistema 1-octanol/água e são amplamente utilizados como filtro molecular para projetos de planeamento e *design* de fármacos devido a sua simplicidade conceitual e facilidade de aplicação. De acordo com a regra dos cinco, a ocorrência de baixa absorção após administração por via oral, ou de baixa permeabilidade, é mais provável para compostos que apresentam uma combinação de duas ou mais propriedades da tabela 3.3.1. Os ligandos para serem bons fragmentos/líderes/fármacos similares tem que respeitar os valores presentes na tabela 3.3.1 ^[30].

Tabela 3.3.1: Características físico-químicas dos diversos conceitos que podem ser utilizados para seleção de ligandos.

| Propriedade | Regra dos 5 | Fragmento-similar | Líder-similar | Fármaco-similar |
|--|-------------|-------------------|---------------|-----------------|
| Peso Molecular | > 500 | < 250 | < 350 | < 500 |
| Números de doadores de ligações de hidrogénio (soma de grupos OH e NH) | > 5 | 2.5 | 3.5 | < 5 |
| Número de aceitadores de ligações de hidrogénio (soma de átomos O e N) | >1 0 | < 3 | < 4 | < 5 |
| Coeficiente de partição calculado para o sistema 1-octanol/água (clog P) | >5 | <6 | < 8 | < 10 |

Para além das propriedades físico-químicas também se pode olhar a reatividade química e a diversidade estrutural dos ligandos. Tendo em conta estas questões aquando da escolha dos ligandos, aumentam-se as hipóteses de sucesso na identificação de novos compostos bioativos.

Existem diversos programas computacionais para realização de *docking*, desde o desenvolvimento do primeiro algoritmo em 1980, mas independentemente do programa utilizado, o processo de *docking* envolve dois passos: (1) predição da conformação, posição e orientação do ligando dentro dos sítios de ligação e (2) avaliação da afinidade de ligação. Estes dois passos têm que ter em conta a flexibilidade do recetor, o que é um aspeto essencial, pois pode ocorrer uma mudança conformacional significativa após a ligação do ligando ao recetor [28].

1) Os algoritmos disponíveis para a pesquisa conformacional, têm em consideração os parâmetros estruturais dos ligandos, os graus de liberdade de torção (diédrico), translacional e rotacional, que são modificados progressivamente. Existem métodos de busca sistemáticos e estocásticos, cada algoritmo de *docking* tem um desses métodos no seu algoritmo. Métodos de busca sistemáticos promovem pequenas variações nos parâmetros estruturais, alterando gradativamente a conformação dos ligandos. O algoritmo sonda a paisagem de energia do espaço conformacional e, após vários ciclos de busca e avaliação, converge para a solução de energia mínima, correspondendo ao modo de ligação mais provável. Embora o método seja eficaz em explorar o espaço conformacional, ele pode convergir para um mínimo local ao invés do mínimo global.

Os métodos estatísticos realizam a busca conformacional modificando aleatoriamente os parâmetros estruturais dos ligandos. Para isso, o algoritmo gera conjuntos de conformações moleculares e preenche uma ampla faixa do cenário energético. Essa estratégia evita aprisionar a solução final em um mínimo de energia local e aumenta a probabilidade de encontrar um mínimo global. Como o algoritmo promove uma ampla cobertura do cenário energético, o custo computacional associado a este procedimento é uma limitação importante.

2) De modo a avaliar a afinidade de ligação, utilizam-se funções de pontuação para estimar a energia de ligação dos complexos ligando-recetor. Essa energia é dada

pela constante de ligação (K_d) e pela energia livre de gibbs. As funções de energia são categorizadas em 3 grupos: funções baseadas em campos de forças, empíricas e baseadas em conhecimento. A previsão de energia de ligação é realizada avaliando os fenómenos físico-químicos mais importantes envolvidos na ligação ligando-recetor, ou seja, interações intermoleculares, dessolvatação e efeitos entrópicos. Quanto maior for o número de parâmetros físico-químicos avaliados, maior a precisão da função de pontuação e maior é o custo computacional envolvidos.

As funções de pontuação baseadas no campo de forças estimam a energia de ligação somando as contribuições de termos ligados (alongamento de ligação, flexão de ângulo e variação diedro) e não ligados (interações eletrostáticas e Van-der-Waals) em uma função mestre geral, usando a mecânica clássica.

Nas funções de pontuação empírica cada termo da função descreve um tipo de evento físico envolvido na ligação ligando-recetor (ligações de hidrogénio, interações iónicas e apolares, efeitos de dessolvatação e entrópicos). As funções empíricas são mais rápidas que as funções baseadas no campo de forças.

Funções de pontuação “*knowledge-based*” usam potenciais de energia estimados a partir da frequência com que dois átomos diferentes são encontrados dentro de uma determinada distância no conjunto de dados. Esta informação é usada para construir uma função empírica de energia.

Cada função de pontuação tem as suas virtudes e limitações, portanto o uso simultâneo de diferentes metodologias de pontuação tem sido cada vez mais usado como forma de obter uma pontuação de consenso.

Na figura 3.3.14 encontra-se o procedimento geral de como funcionam os programas de *docking*. Primeiro seleciona-se a proteína recetora e os ligandos e colocam-se ambos no programa. O programa vai encaixando os ligandos um por um na cavidade da molécula e calculando as suas energias mínimas globais.

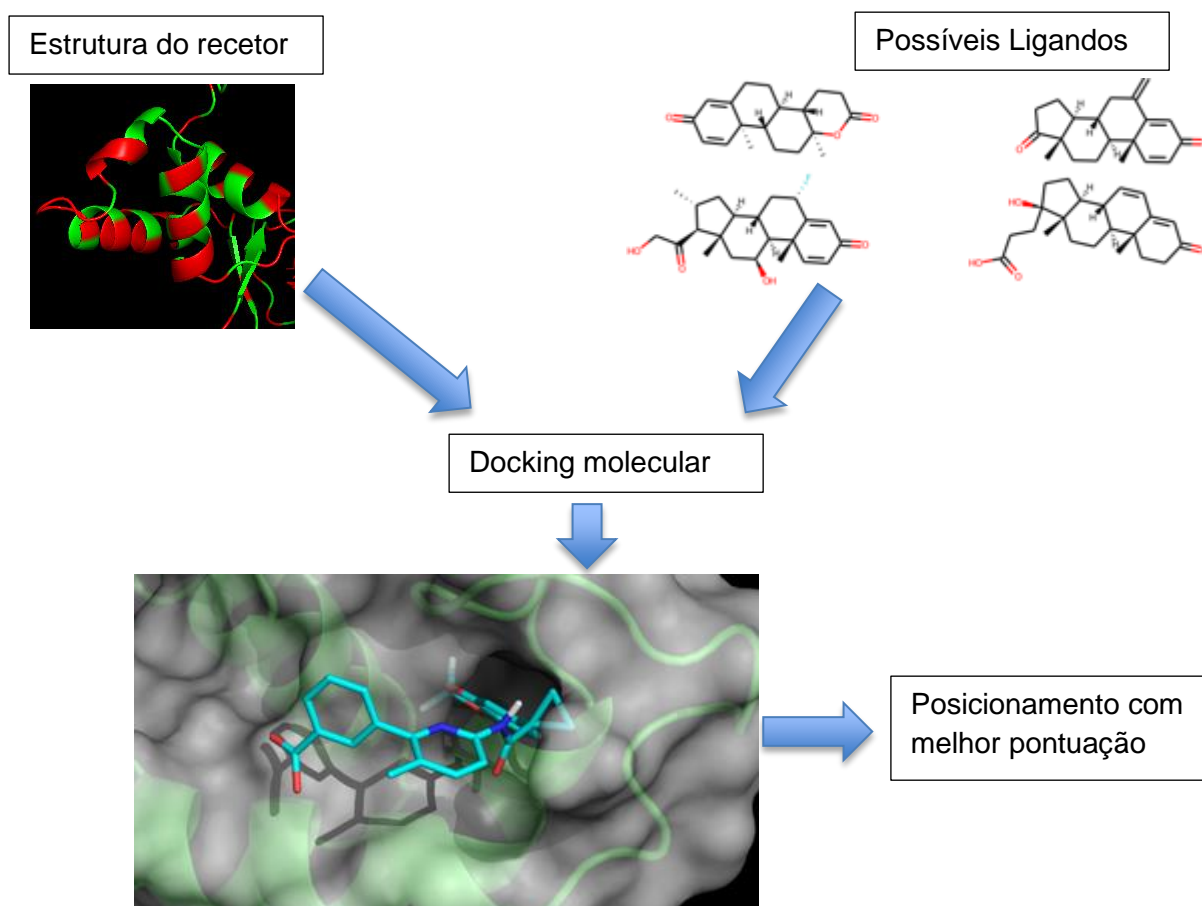


Figura 3.3.14: Procedimento geral do programa de *docking*.

Sendo o *docking* um método *in silico*, os resultados finais do *docking* têm que ser analisados *in vitro*, de modo a confirmar se os ligandos são verídicos ou se são falsos-positivos. Somente após os estudos *in vitro*, passa-se aos estudos *in vivo*, de modo a assegurar eficácia e segurança do composto. Por último, se tudo estiver de acordo com as normas o composto passa para o mercado para ser comercializado. Todo este processo é dispendioso a nível temporal e monetário. Dependendo da complexidade do problema estes estudos *in silico* podem levar meses e os *in vitro* e *in vivo* até a comercialização, demoram anos (entre 3 e 15 Anos).

4. Materiais e métodos

4.1 Programas computacionais

Foram utilizados 4 programas computacionais, o visualizador molecular PyMOL (versão 2.4.0) ^[31], a suite de modelação e dinâmica molecular, o GROMACS (versão 2018.1) ^[32], um *software* de deteção de cavidades de proteínas, o FPOCKET ^[33] e o *software* de *docking* molecular, o Autodock Vina ^{[34][35]}.

O PyMOL é um *software* utilizado para visualização de estruturas moleculares e que permite fazer análises e cálculos de propriedades moleculares e comparação de diferentes estruturas. O PyMOL foi criado por Warren Lyford e lançado no ano 2000.

O GROMACS é um *software* usado para o cálculo de trajetórias de dinâmica molecular, minimização de energia, pesquisa e análise conformacional. O ano do lançamento do Gromacs foi em 1991.

O Fpocket é um algoritmo de código livre e bastante rápido de detecção de cavidades de proteínas, baseado no algoritmo de Tesselação de Voronoi. Este algoritmo pertence à classe dos métodos geométricos de detecção de cavidades/sítios de ligação.

O AutoDock Vina é um *software* de *docking* computacional ligando-proteína, que realiza *docking* rígido ou flexível do ligando e recetor.

Também foi utilizado a base de dados de estruturas *Protein Data Bank* [36], para pesquisa e guardar as estruturas da proteína FOXOa3 e a base de dados *DrugBank* [37], de modo a selecionar os ligando para as corridas de *docking*.

O *Protein Data Bank* é um banco de dados de livre acesso contendo a quase totalidade das estruturas 3D de proteínas e ácidos nucleicos. Esses dados experimentais são geralmente obtidos através da difração de raios X, ressonância magnética nuclear e crio-microscopia eletrônica por diversos físicos, biólogos e bioquímicos de todo o mundo. O *Protein Data Bank* tem neste momento 196.779 [22 de outubro de 2022] estruturas no formato PDB. Tal como o *Protein Data Bank*, o *DrugBank* é um banco de dados de livre acesso, mas focado nos dados químicos e farmacológicos dos medicamentos. O *DrugBank* regista informações sobre a substância ativa, estrutura química e via metabólica dos mais diversos medicamentos aprovados, em estudo e mesmo dos retirados do mercado. Neste momento o *DrugBank* tem um total 14752 moléculas pequenas que incluem vários tipos de fármacos e drogas, desde os aprovados até aos retirados do mercado [25 de outubro de 2022]. Possui também informação sobre os alvos (quando conhecidos) de muitas dessas moléculas.

4.2 Procedimentos Laboratoriais

4.2.1 Dinâmica Molecular

Para a realização da dinâmica molecular seguiu-se o esquema representado na figura 4.2.1.15.

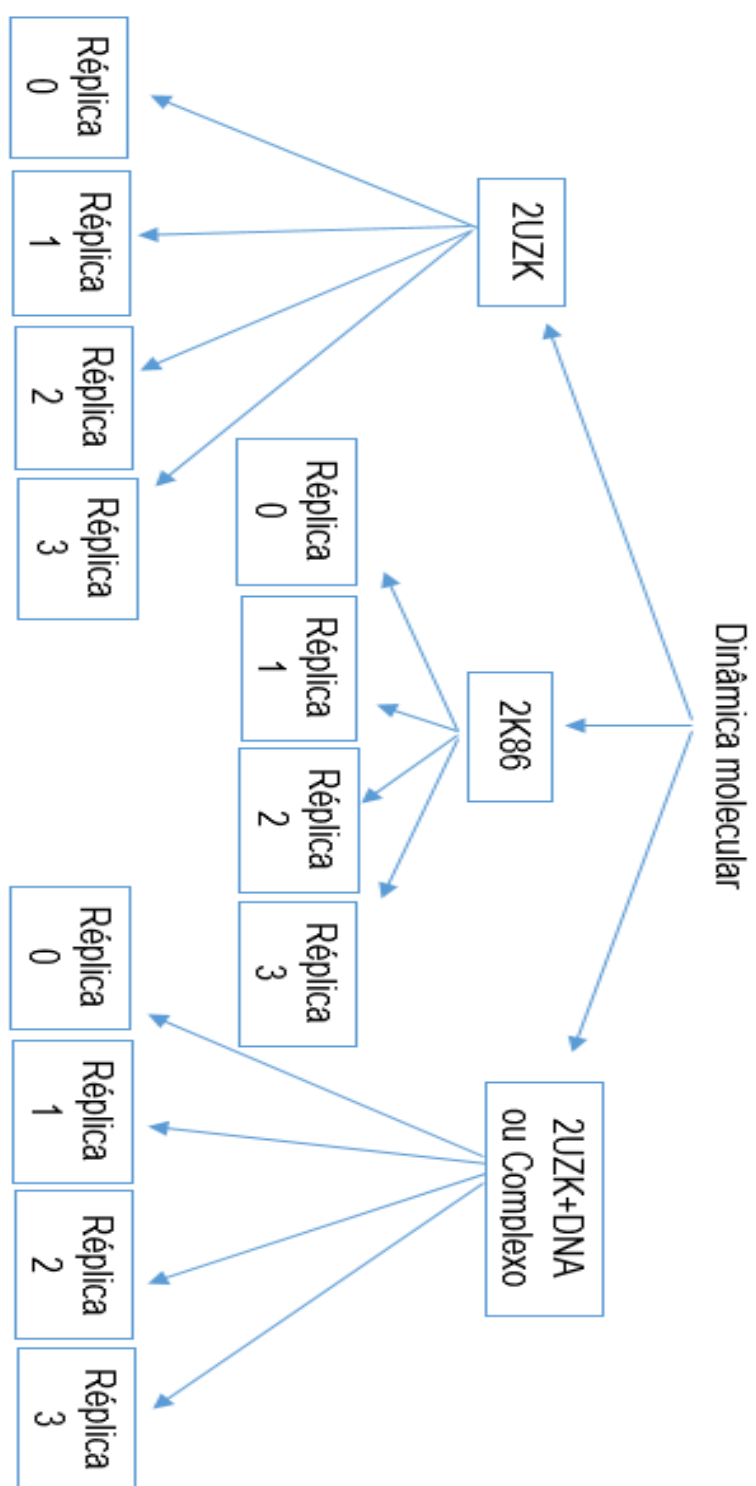


Figura 4.2.1.15: Esquema da realização das corridas de dinâmica molecular.

As simulações de dinâmica molecular foram realizadas partir de 3 estruturas da FoxO3: estrutura cristalográfica da proteína ligada a um oliómero de DNA duplex (entrada 2UZK do *PDB*), a mesma estrutura cristalográfica com a molécula de DNA removida (entrada 2UZK no *PDB*) e uma estrutura da proteína livre em solução aquosa obtida por ressonância magnética nuclear (entrada 2K86 do *PDB*). Todas as estruturas foram retiradas do site do *Protein Data Bank*. Para cada uma das 3 estruturas foram realizadas 4 réplicas de simulação, de modo a aumentar a amostragem e verificar a consistência dos resultados obtidos.

Na primeira das duas estruturas obtidas da entrada 2UZK do *Protein Data Bank*, foram eliminadas as moléculas de água e uma das duas cópias do dímero DNA-proteína presentes no ficheiro *PDB*. Para a segunda estrutura obtida da 2UZK, foi eliminada também a cadeia de DNA, de modo a obter um sistema com a proteína livre.

No caso da estrutura 2K86 teve de se escolher uma das 20 conformações presentes no ficheiro *PDB* (as estruturas obtidas por RMN normalmente contém múltiplas conformações). Optou-se, neste caso, por escolher a primeira configuração presente no ficheiro.

Após estes ficheiros estarem processados, seguiram-se os seguintes passos (Figura 4.1.2.16).

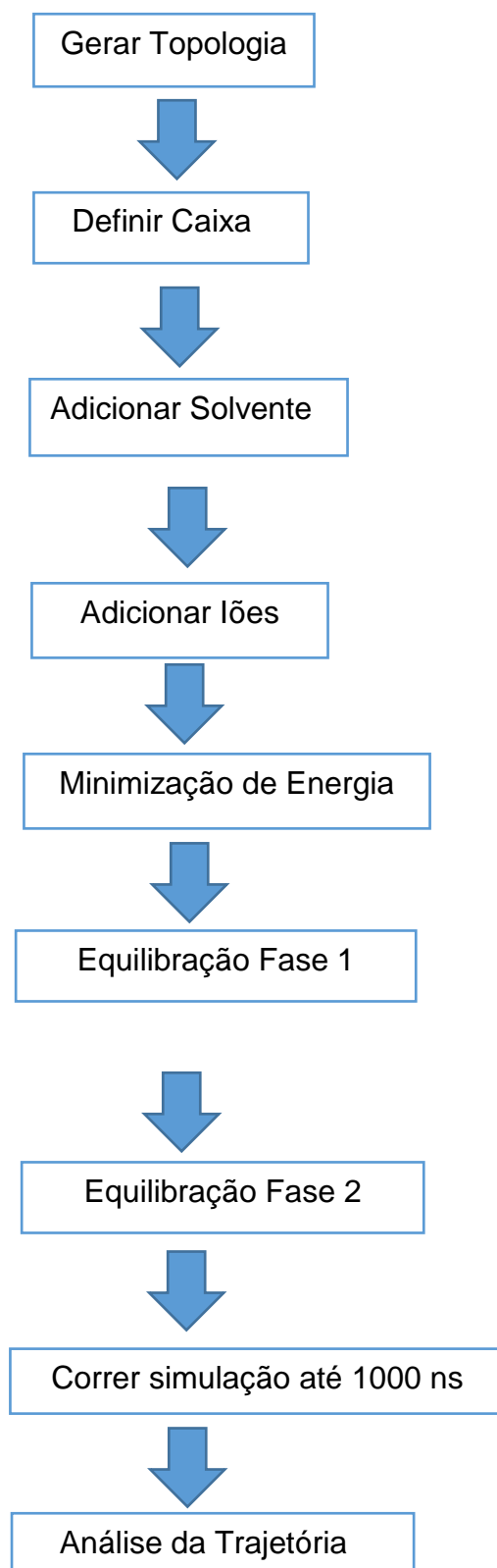


Figura 4.1.2.16: Passos para a realização da corrida de simulação de dinâmica molecular.

Para todas as simulações de dinâmica molecular das diferentes estruturas a ferramenta *pdb2gmx* do programa GROMACS foi usada para produzir topologias e ficheiros de coordenadas para as 3 estruturas anteriormente processadas. O campo

de forças usado foi o AMBER14sb, que é um campo de forças “*all-atom*”, significando que contém parâmetros explícitos para todos os átomos dos sistemas em estudos. As moléculas foram em seguida colocadas numa caixa, que foi preenchida com moléculas de água do tipo TIP3P [38], para solvatar o sistema. De modo a garantir a neutralidade das caixas de simulação, foram adicionados iões Cl⁻ e Na⁺ em quantidade suficiente para neutralizar a carga da proteína. O sistema foi de seguida minimizado. Após a minimização ocorre a fase de equilibração, que consiste em duas simulações consecutivas, uma no *ensemble* de número de moles, volume e temperatura constantes (NVT) e outra no *ensemble* de número de moles, pressão e temperatura constantes (NPT), usando um termóstato de Anderson e um baróstato de Parrinello-Raman para fixar a pressão e temperatura a 300 kelvin e 1 atmosfera. Nas simulações de equilibração os átomos pesados do sistema foram fixados através de uma força de “*restraining*”, e os comprimentos das ligações químicas envolvendo átomos de hidrogénio foram constrangidos com o algoritmo *LINCS*. A análise da pressão e temperatura ao longo da simulação permite verificar a qualidade da equilibração. Após a equilibração, iniciou-se a corrida de produção, removendo a força de “*restraining*” de forma a deixar livres todos os átomos da proteína, mas para o caso do complexo DNA-Proteína manteve-se a força de “*restraining*” nos átomos de fósforo em ambas as cadeias de DNA (cadeia E e B), mantendo as condições do ensemble NPT. As simulações de produção foram corridas em parcelas de 10ns, com escrita das coordenadas do sistema a cada 10 picosegundos (ps). Após término da corrida é necessário concatenar os vários segmentos numa trajetória contínua, a qual é guardada num ficheiro em formato *PDB*, que pode ser visualizado no programa PyMOL. Para efetuar esta concatenação utilizaram-se as ferramentas *trajcat*, *trajconv* e *fitprogressive* do GROMACS. As conformações foram gravadas de 10 em 10 ps, ou seja num total de 1000 nanosegundos (o tempo das simulações) correspondem a 100000 configurações da proteína. Assim sendo, o ficheiro final contém 100000 conformações da proteína sem as moléculas de água.

Todo o processo acima foi igual para todas as estruturas, exceto:

- Geometria das caixas de simulação: para o complexo DNA-proteína (2UZK) utilizou-se uma caixa cúbica com espaço de 1 Angstroms (Å) entre a molécula e a parede da caixa, para a estrutura isolada (2UZK)

foi usada uma caixa octaédrica com espaço de 1,5 Angstroms (Å) entre a molécula e a parede da caixa e por último a estrutura 2K86 foi colocada numa caixa octaédrica com espaço de 1 Angstroms (Å) entre a molécula e a parede da caixa.

- Para as réplicas 1, 2 e 3, variaram-se as condições iniciais da simulação através de rotações das coordenadas do sistema no *software* PyMOL e uso de diferentes valores para o parâmetro “*gen_seed*”, o qual controla a geração aleatória de velocidades iniciais do *ensemble* NVT, para as várias réplicas.

Todas as réplicas 1 têm uma rotação de 45° em todas as coordenadas (x,y,z), a réplicas 2 têm uma rotação de 90° e as réplicas 3 têm uma rotação de 135°, relativamente as coordenadas do ficheiro original retirado do *Protein Data Bank*. Quanto ao *gen_seed* temos em todas as réplicas 0 o valor -1, nas réplicas 1 o valor de 654321, nas réplicas 2 o valor 123456 e por fim nas réplicas 3 o valor 43265. Desta forma garante-se simultaneamente variabilidade e reprodutibilidade das condições das simulações de dinâmica molecular.

4.2.2 Descoberta de Cavidades ou Sítio de Ligação

As simulações de dinâmica molecular anteriormente descritas produziram 4 réplicas de 1000 nanosegundos da proteína FoxO3 ligada ao DNA (complexo) e da proteína FoxO3 livre, a partir da forma cristalizada (2uzk) e da forma em solução (RMN) (2k86). Dado que as conformações da proteína são escritas a cada 10 ps, 1000 nanosegundos correspondem a 100000 configurações da proteína, o que é um número excessivamente elevado para ser alvo de uma análise detalhada. De modo a analisar-se um número significativo de configurações para estudar o comportamento da proteína utilizou-se a ferramenta *trjconv* para reduzir o número de conformações para 21 conformações, correspondendo a uma amostragem de 50 em 50 ns. O intervalo de 50 ns produz conformações que, sendo relativamente próximas, cobrem o espaço conformacional obtido durante a simulação e permitem uma análise viável).

Essas 21 conformações de cada réplica foram submetidas ao programa computacional *FPOCKET*, por onde passaram pelo seguinte processo ^[17] (Figura 4.2.2.17).

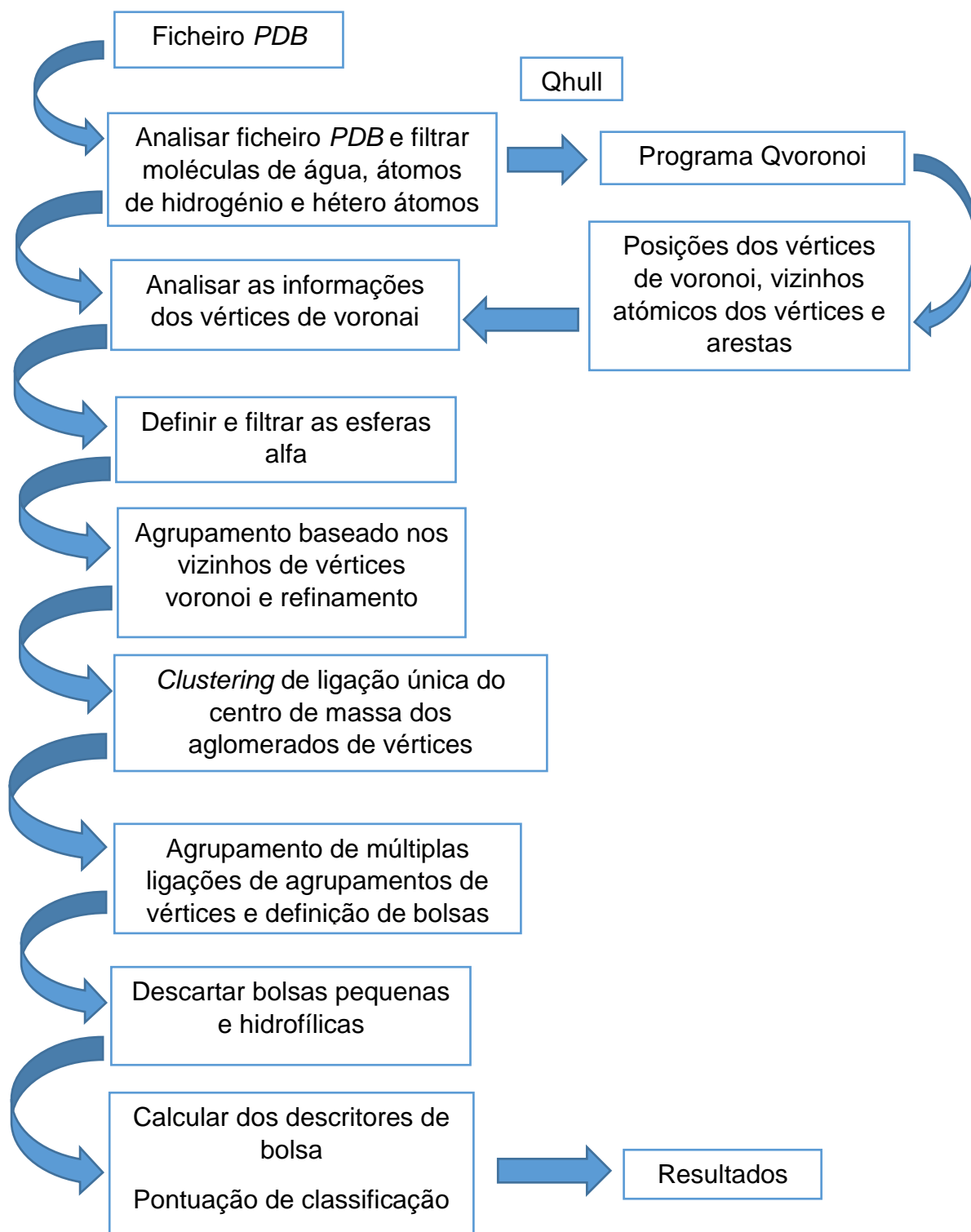


Figura 4.2.2.17: Processos do programa FPOCKET.

4.2.3 Docking Molecular

Após a seleção visual das melhores cavidades através do PyMOL, estas foram usadas numa análise de triagem virtual (VS, *virtual screening*), usando o programa computacional *Autodock Vina* e vários conjuntos de estruturas de moléculas pequenas. Depois da identificação das melhores cavidades na molécula do recetor é necessário selecionar o conjunto de ligandos a testar no *docking* molecular. A coleção de ligandos do banco de dados de fármacos *DrugBank* foi usada para o *screening* virtual, incluindo 11951 pequenas moléculas. Dessas 11951 moléculas foram retiradas as moléculas com mais de 20 torções, por serem ligandos de grandes dimensões, como menos potencial farmacológico e para os quais os cálculos de *docking* molecular são substancialmente mais demorados. Assim sendo das 11951 moléculas foram utilizadas 10550 moléculas para realização do *docking*.

As estruturas foram obtidas a partir do *DrugBank* em formato sdf 2D, pelo que foi necessário produzir a sua conversão em estrutura 3D. Este processo foi levado a cabo com o *software* Gypsum-DL [39][40]. Em seguida foi necessário proceder à conversão do formato sdf 3D no formato “pdbqt” usado como input do *software* Autodock Vina, sendo para tal usado o *software* OpenBabel [41].

Após a preparação dos ligantes, utilizou-se o *software* PyMOL com o *plugin* Autodock-Vina para definir e centrar a grelha de *docking* no local de cada cavidade em estudo. Neste processo a estrutura do recetor (FoxO3a) é também convertida em formato pdbqt para usar como ficheiro inicial para o Autodock Vina. O *plugin* grava as estruturas do recetor e ligando em formato pdbqt, assim como o ficheiro de configuração contendo as definições necessárias para correr o Autodock Vina. A figura 3.3.14 referenciada acima representa o procedimento geral do programa de *docking*, depois do procedimento referido acima.

Após a análise visual de todas as cavidades através do PyMOL, selecionaram-se as melhores cavidades em cada réplica, pois haviam algumas cavidades que não eram importantes para o estudo de *docking* e outras eram repetidas ao longo da réplica e/ou estrutura. Deste modo foram selecionadas 4 cavidades interessantes para o estudo de *docking* rígido. Sendo as seguintes cavidades:

- Cavidade 2 na conformação 10 da réplica 3 da estrutura 2K86

- Cavidade 1 na conformação 5 da réplica 1 da estrutura 2UZK
- Cavidade 1 na conformação 7 da réplica 3 da estrutura 2UZK
- Cavidade 2 na conformação 2 da réplica 1 do complexo (2UZK+DNA)

5. Resultados e Discussão

5.1 Dinâmica Molecular

A primeira fase do trabalho realizado na tese consistiu na realização de simulações de dinâmica molecular (DM), nas formas livre e em complexo com o DNA da proteína FoxO3a. As simulações de DM permitem mostrar conformações das várias estruturas que poderão ser consideravelmente diferentes das estruturas experimentais que servem de ponto de partida. Nesta análise procura-se detetar diferenças de estrutura e dinâmica entre as formas livre e ligada da FoxO3 e também mostrar conformações que permitam encontrar potenciais cavidades/sítio de ligação.

Na tabela 5.1.2 estão representadas as estruturas e as réplicas realizadas nas corridas de dinâmica molecular.

Tabela 5.1.2: Estruturas e respetivas réplicas das corridas de dinâmica molecular.

| Estrutura | Caixa | Réplica | Tempo de simulação (ns) |
|------------------------|---|---------|-------------------------|
| 2UZK-DNA (complexo) | Cúbica com 1 Å entre a molécula e a parede da caixa | 0 | 1000 |
| | | 1 | |
| | | 2 | |
| | | 3 | |
| 2UZK | Octaédrica com 1,5 Å entre a molécula e a parede da caixa | 0 | |
| | | 1 | |
| | | 2 | |
| | | 3 | |
| 2K86 | Octaédrica com 1 Å entre a molécula e a parede da caixa | 0 | |
| | | 1 | |
| | | 2 | |
| | | 3 | |

As diferentes estruturas iniciais da proteína foram colocadas em caixas com dimensões e geometrias diferentes, de forma a otimizar o número de moléculas de solvente na caixa (o tempo de simulação é crucialmente afetado pelo número de moléculas de solvente presentes na caixa de simulação). Para uma mesma dimensão, a geometria de octaedro truncado corresponde a um número significativamente menor de moléculas de solvente no sistema. A estrutura do complexo apresenta maiores dimensões devido à presença do oligómero de DNA, pelo que é vantajoso reduzir o número de moléculas de solvente, usando uma geometria octaédrica para a caixa de simulação. Esta geometria em nada interfere com a qualidade das simulações, apenas influencia a sua duração.

A distância entre cada molécula de proteína e a parede da caixa correspondente varia no decorrer da simulação, devido às variações de posição e conformação da molécula. Este valor de distância entre a molécula e a parede da caixa deve ser controlado e mantido acima de um valor mínimo, para que a molécula não interaja com as suas imagens nas cópias vizinhas da caixa (numa simulação com condições fronteira periódicas, existem infinitas réplicas virtuais da caixa de simulação, ver introdução teórica), o que iria introduzir artefactos e destruir o realismo físico da simulação. A ferramenta *mindist* do GROMACS foi usada para monitorizar a distância mínima entre a proteína e as suas imagens e garantir que a mesma não descia abaixo do *cut-off* definido para as interações não covalentes. Para garantir este resultado, as dimensões das caixas foram escolhidas de modo a que para estruturas do complexo FoxO3a/DNA e estrutura em solução 2k86 fossem de 1 Å e de 1,5 Å para a estrutura 2UZK. No final dos 1000 nanosegundos, e dado que se escolheu amostrar conformações do sistema a cada 10ps, obtêm-se um ficheiro com um total de 100000 conformações da proteína. Para tornar o número de conformações mais maneável em termos de análise, foi gerado uma versão sub-amostrada da trajetória, com a ferramenta *trajconv* e a opção "--skip 50", a qual grava uma conformação de 50 em 50 conformações da trajetória original. Assim o ficheiro final, utilizado para posteriores análises têm configurações de 500 em 500 picosegundos, totalizando no final 20001 conformações para uma trajetória de 1 μ s = 1000 nanosegundos.

De forma a avaliar a evolução e conservação da estrutura do sistema ao longo de cada simulação de DM, é usual calcular o desvio quadrático médio (RMSD) relativo

a uma conformação de referência, tipicamente a conformação experimental. O RMSD (desvio quadrático médio – *Root Mean Square Deviation*) é uma medida do grau de semelhança estrutural de duas configurações moleculares, consistindo no desvio médio (quadrático) entre as posições dos átomos correspondentes de duas moléculas sobrepostas. Foi calculado o RMSD de todas as réplicas da mesma estrutura, de modo a verificar se a corrida de dinâmica molecular se tinha mantido ou se teriam ocorrido diferenças consideráveis. As imagens abaixo (5.1.18-5.1.20) representam as 4 réplicas das três estruturas utilizadas no trabalho.

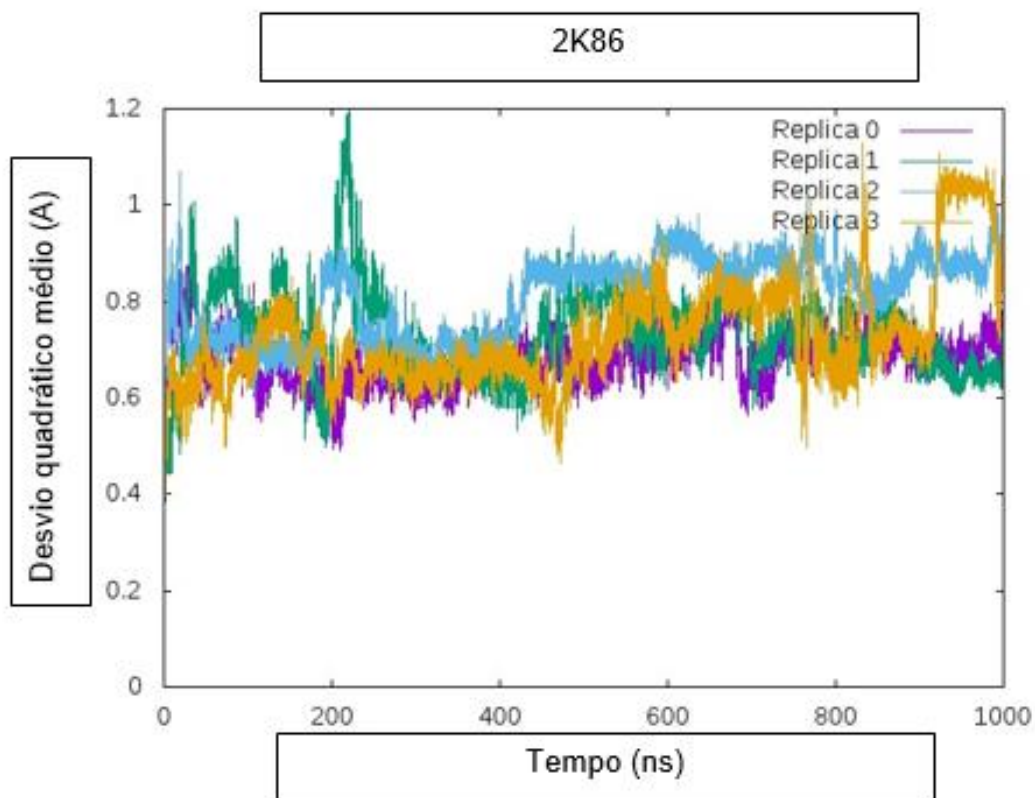


Figura 5.1.18: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2k86 da proteína FoxO3.

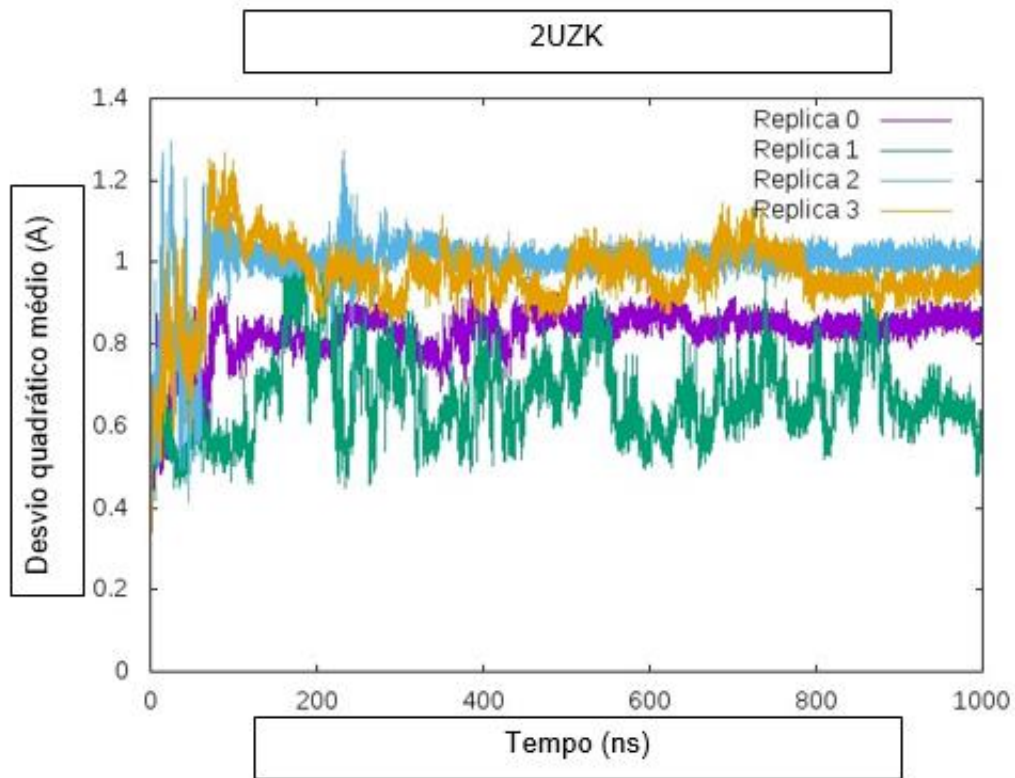


Figura 5.1.19: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2UZK da proteína FoxO3.

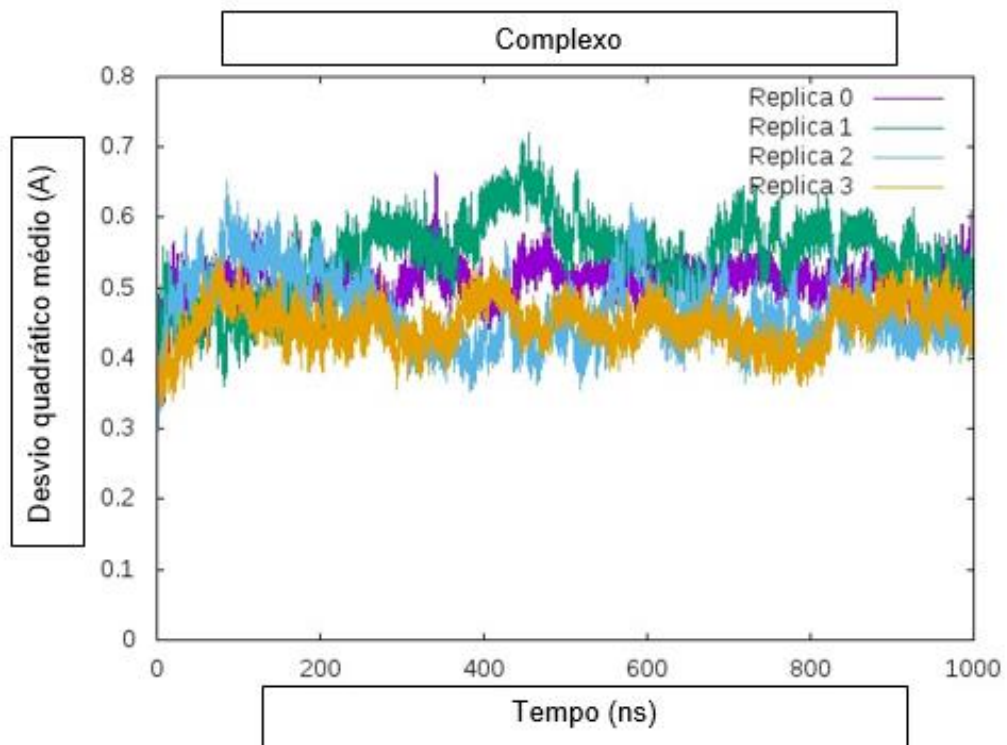


Figura 5.1.20: Desvio quadrático médio ao longo do tempo para as diversas réplicas da estrutura 2UZK com o DNA (complexo) da proteína FoxO3.

O RMSD é uma medida de distância, a qual permite avaliar os desvios em relação a uma conformação de referência (neste caso as estruturas cristalográficas que serviram de ponto de partida para as simulações de DM) Valores de RMSD mais baixos implicam menores desvios das conformações da simulação relativamente à estrutura de referência e são uma indicação que a simulação é capaz de preservar as características estruturais da molécula em estudo. Observando as figuras 5.1.18, 5.1.19 e 5.1.20, verificamos que as trajetórias da Figura 5.1.20 têm valores de RMSD mais baixos e que as réplicas mantêm-se mais estáveis ao longo do tempo, o que poderá dever-se à proteína estar ligada à molécula de DNA, conferindo uma maior rigidez e restrição dos movimentos.

Foi também calculado o RMSD da configuração 2K86 comparativamente ao complexo, utilizando os seguintes conjuntos de átomos: proteína total e hélices H1 e H3. A escolha das hélices H1 e H3 foi motivada pela maior instabilidade observada para a hélice H2, a qual invalida esta última como medida da conformação global de cada proteína. As figuras a baixo demonstram o RMSD da proteína ligada ao DNA (complexo) e da proteína livre de DNA (2k86) para todas as réplicas (figura 5.1.21 até 5.1.24).

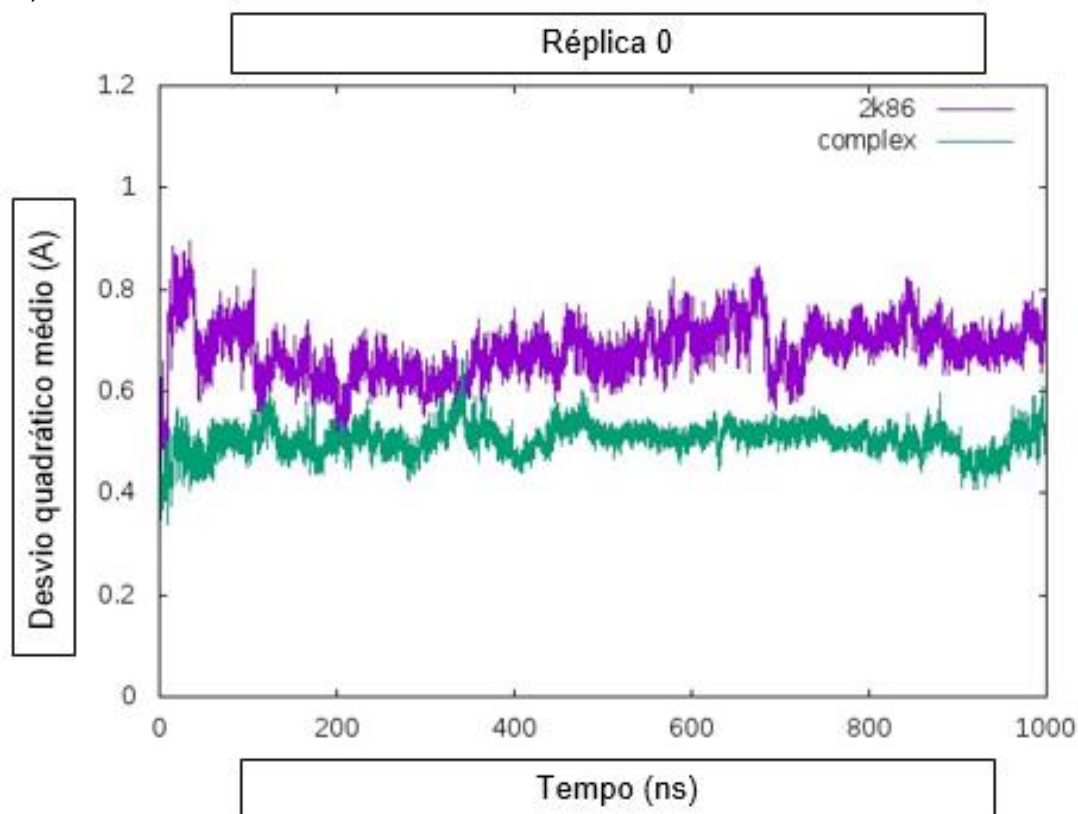


Figura 5.1.21: Flutuação atômica de RMSD do domínio FoxO3 DBD ligada e sem DNA da réplica zero.

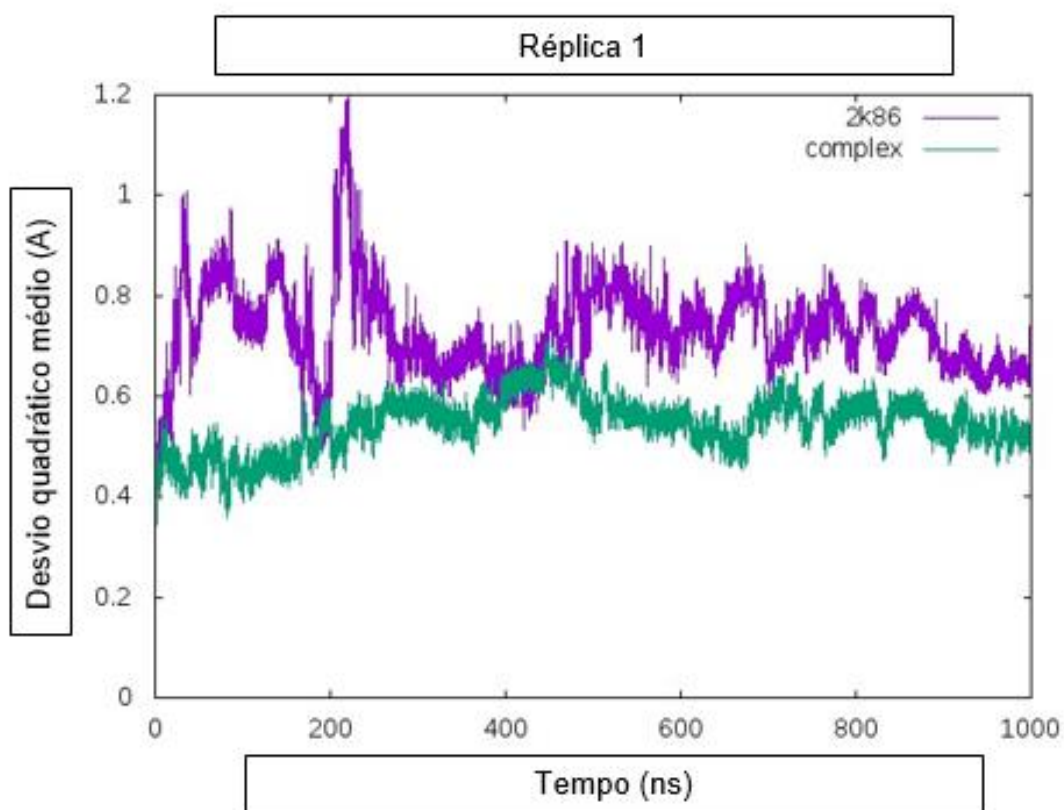


Figura 5.1.22: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica um.

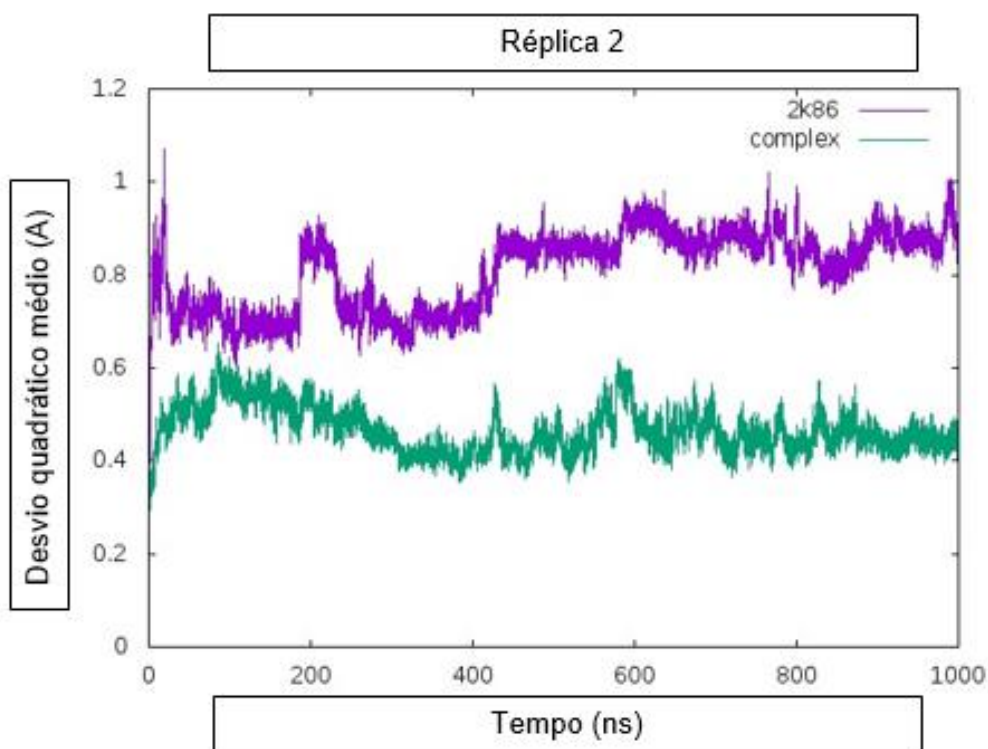


Figura 5.1.23: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica dois.

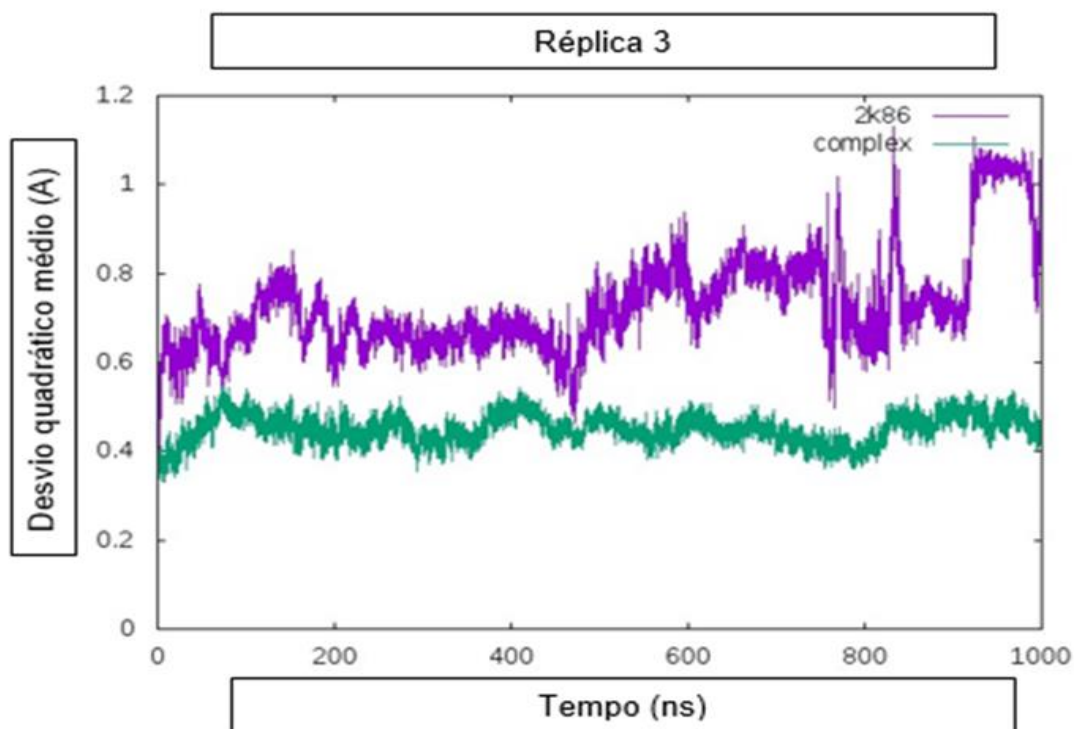


Figura 5.1.24: Flutuação atômica de RMSD do domínio FoxO3 DBD ligado e sem DNA da réplica três.

Observando as figuras de 5.1.21 a 5.1.24 verificamos que a estrutura livre exibe valores de RMSD mais elevados que a estrutura em complexo. Esta observação indica que a estrutura do DBD em complexo apresenta maior estabilidade e rigidez, previsível dado o efeito estabilizador da interação com o DNA, restringindo a movimentação da proteína, tal como na Figura 5.1.20. As largas flutuações do RMSD total observadas em algumas das réplicas parecem poder ser atribuídas às zonas mais flexíveis da estrutura, tais como a cauda e o loop, e também a variações da orientação relativa das hélices. Para testar esta hipótese, foi calculado o RMSD somente das hélices H1 e H3 para todas as réplicas (figuras 5.1.25 até 5.1.28) da estrutura 2K86 e 2UZK.

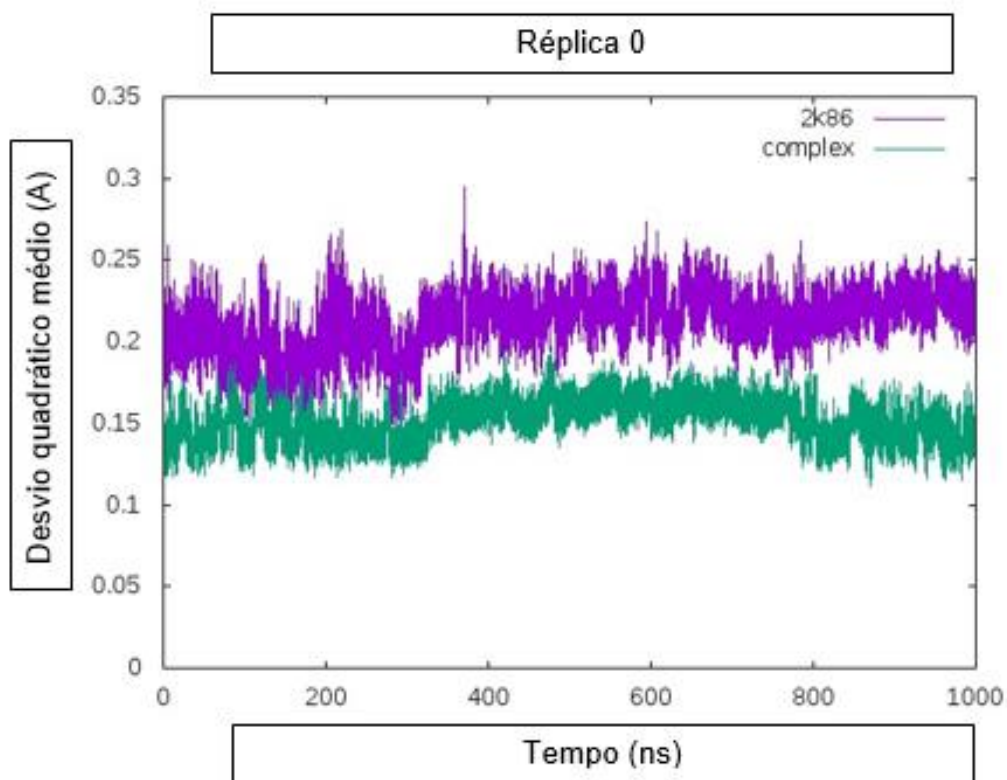


Figura 5.1.25: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica zero.

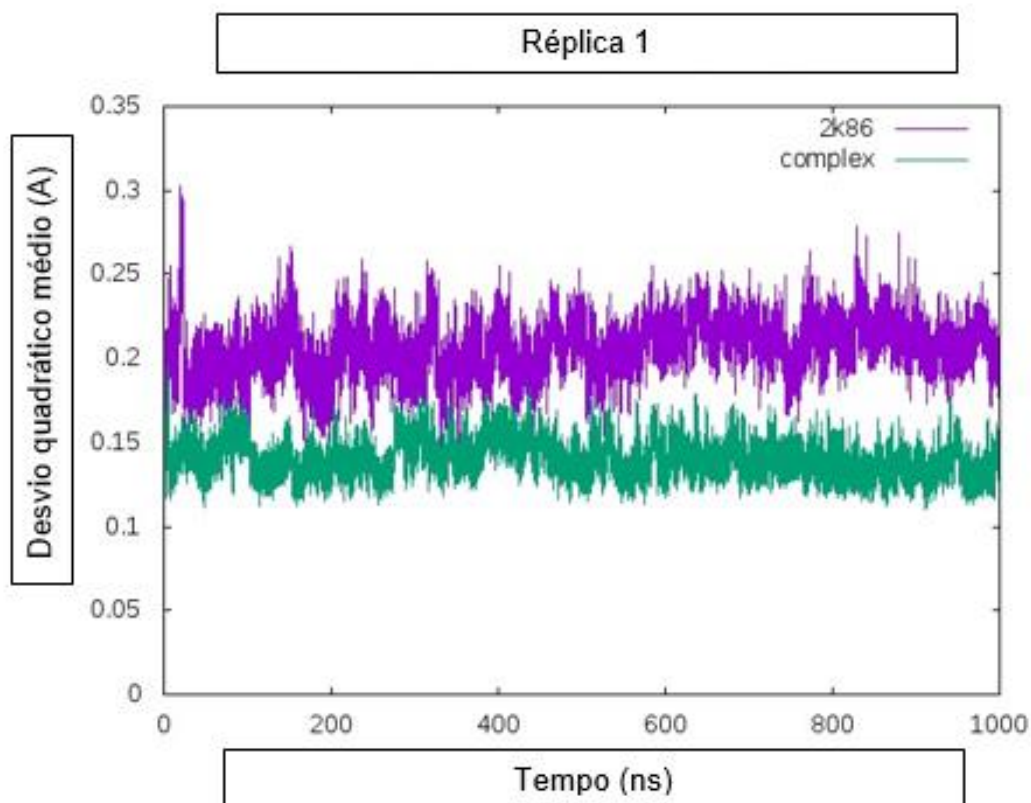


Figura 5.1.26: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica um.

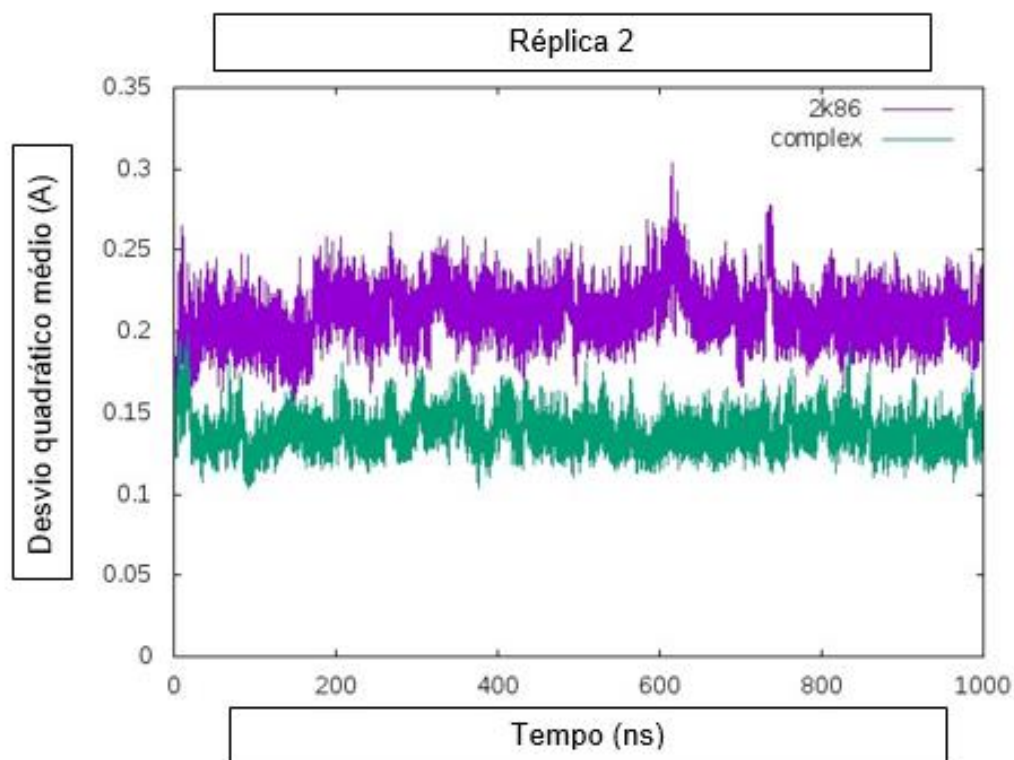


Figura 5.1.27: Flutuação atômica de RMSD da hélice H1-H3 do domínio FoxO3 DBD ligado e sem DNA da réplica dois.

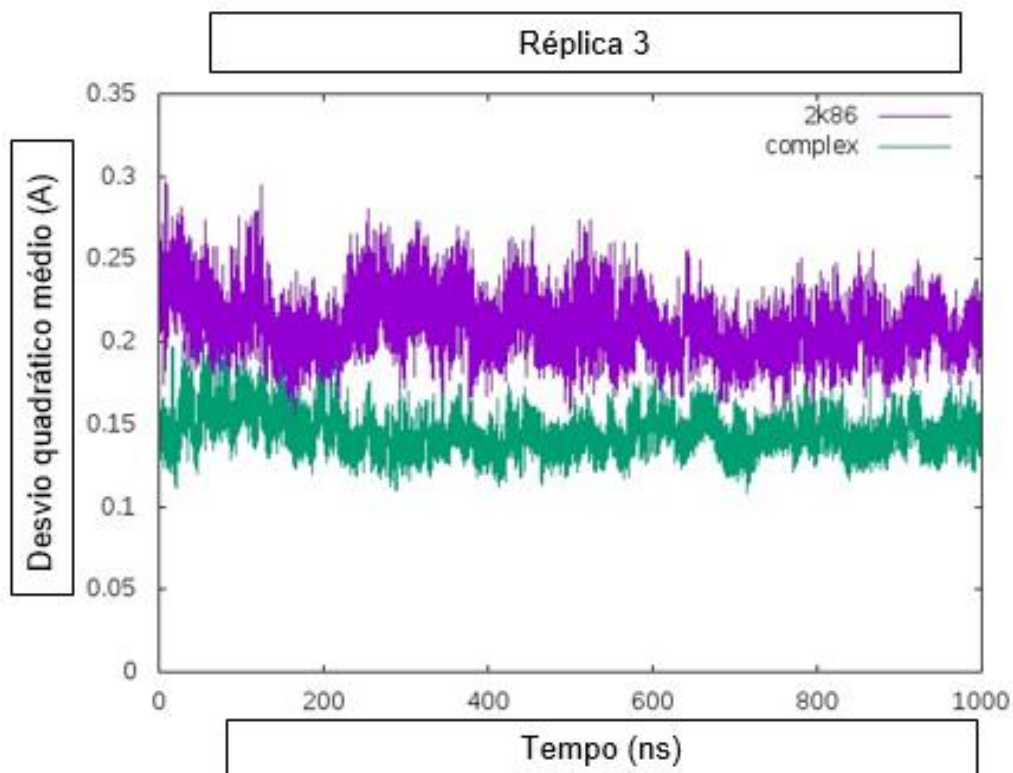


Figura 5.1.28: Flutuação atômica de RMSD da hélice H1-H3 do domínio Fox DBD ligado e sem DNA da réplica três.

Embora se mantenha uma diferença clara entre os valores de RMSD da estrutura livre e complexo quando o RMSD é calculado apenas para as hélices H1 e H3, esta diferença é muito mais constante e não se observam praticamente flutuações significativas do RMSD da estrutura livre. Estes resultados apontam para uma conservação tanto da conformação das hélices H1 e H3, como da sua orientação relativa. A conservação da hélice H3 é um ponto particularmente importante, pois é esta a hélice envolvida na interação com o “*major groove*” das sequência de DNA reconhecida pela FOXO3.

O RMSD foi calculado tendo em conta a estrutura 2k86, pois é a estrutura livre em solução da FoxO3, obtida através do método de RMN, estando provavelmente mais próxima da estrutura real da proteína em solução aquosa. A estrutura 2UZK sem DNA foi estudada para se comparar com a estrutura 2k86 e avaliar se estas convergiram para uma estrutura semelhante. Os resultados quanto a este ponto não foram conclusivos, sendo provavelmente necessário realizar simulações mais longas para o efeito.

Como os ficheiros finais das simulações de dinâmica molecular estão em formato PDB e contêm 2001 conformações, a sua análise seria difícil no PyMOL, deste modo foi preciso utilizar as ferramentas do gromacs, *do_dssp* e *xpm2ps*, que demonstram graficamente as estruturas que a proteína adota ao longo da simulação. As figuras abaixo (5.1.29 – 5.1.44) representam as estruturas das diversas conformações químicas da proteína e das diversas réplicas do tempo zero aos cem nanosegundos e do tempo novecentos a mil nanosegundos. Deste modo verificou-se a evolução inicial e final da simulação, conseguindo comparar as mudanças que ocorreram durante a simulação de dinâmica molecular, se essas diferenças são significativas ou não e de que modo influenciam a molécula.

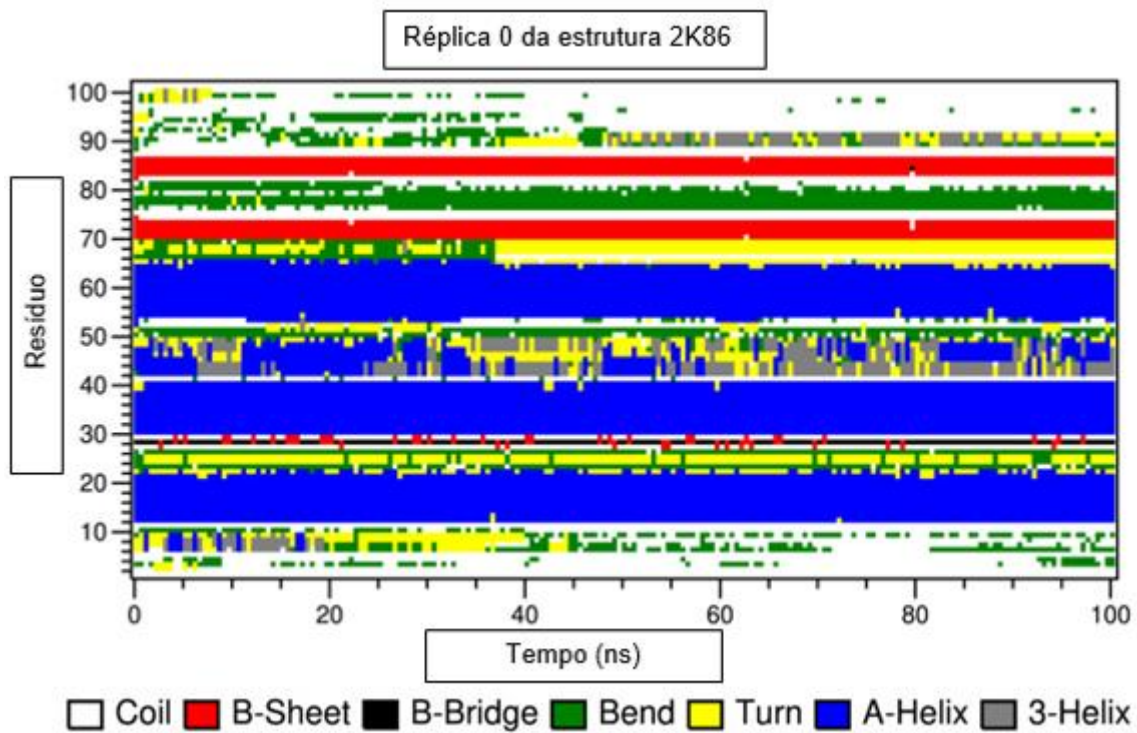


Figura 5.1.29: Representação estrutura 2k86 – réplica 0 desde o tempo 0 até ao s 100 nanosegundos.

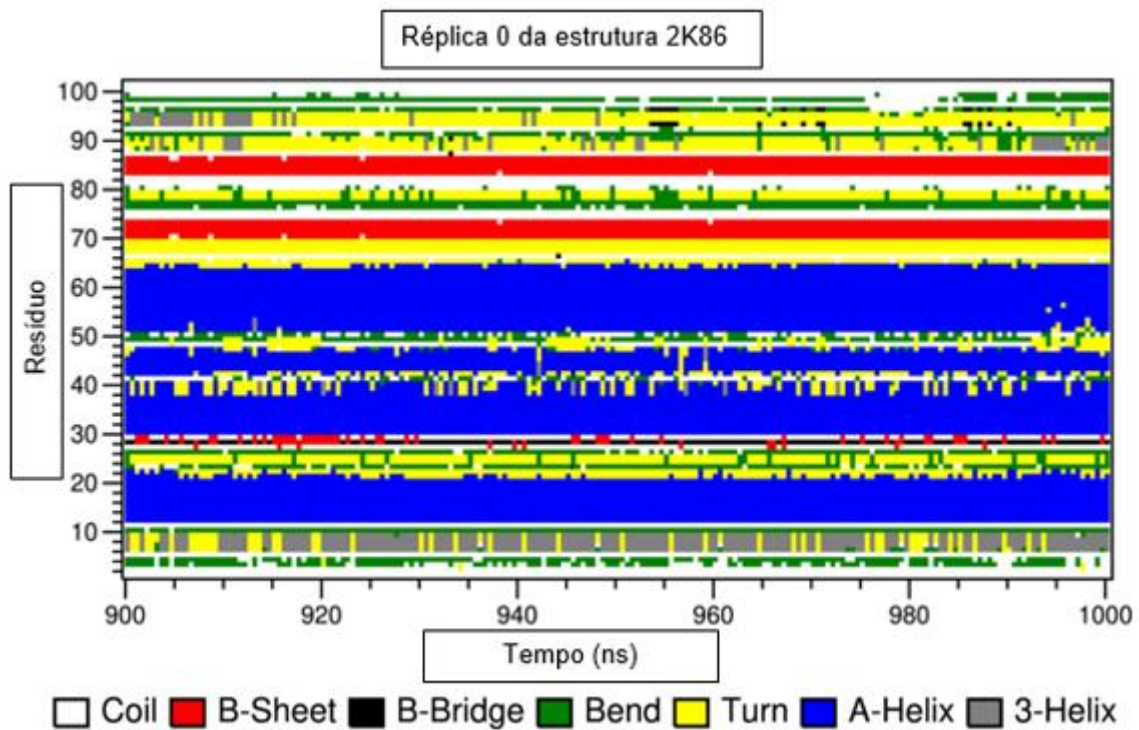


Figura 5.1.30: Representação estrutura 2k86 – réplica 0 desde o tempo 900 até aos 1000 nanosegundos.

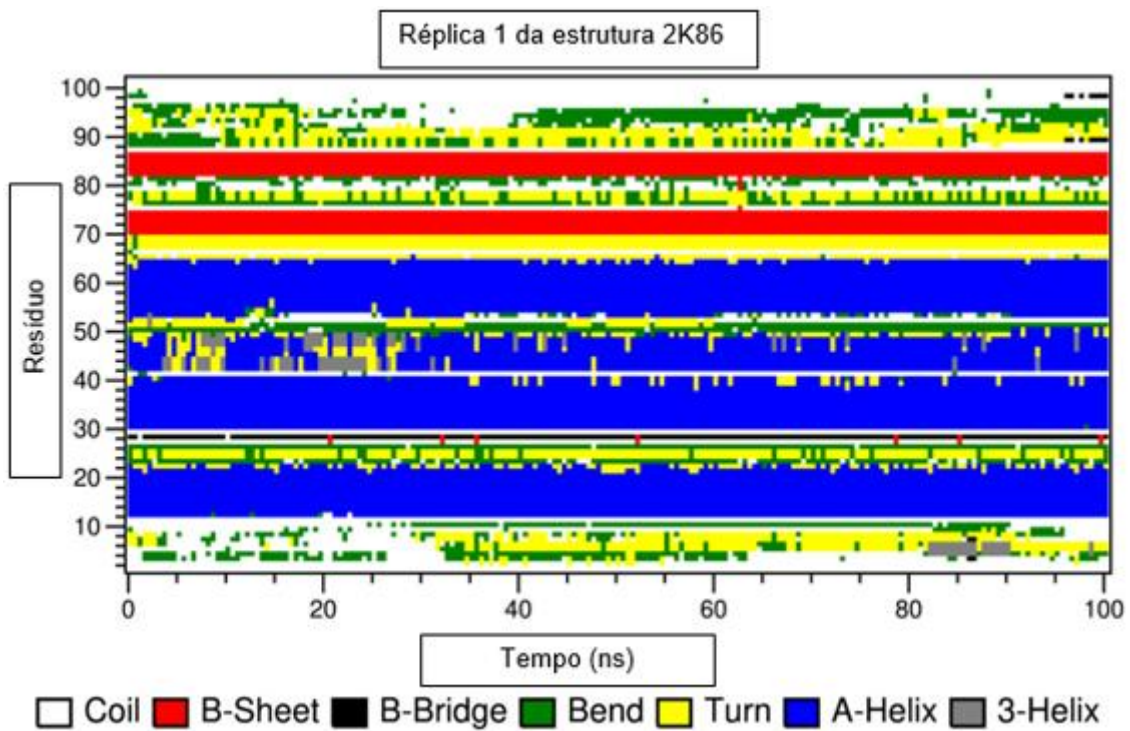


Figura 5.1.31: Representação estrutura 2k86 – réplica 1 desde o tempo 0 até aos 100 nanosegundos.

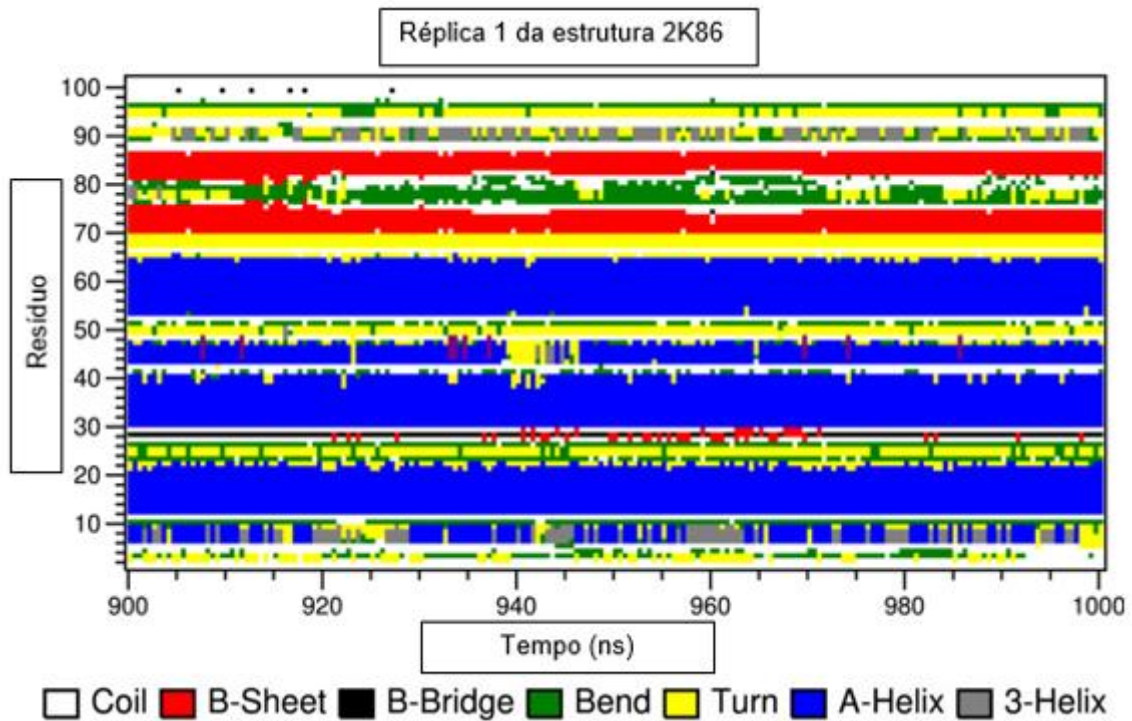


Figura 5.1.32: Representação da estrutura 2k86 – réplica 1 desde o tempo 900 até aos 1000 nanosegundos.

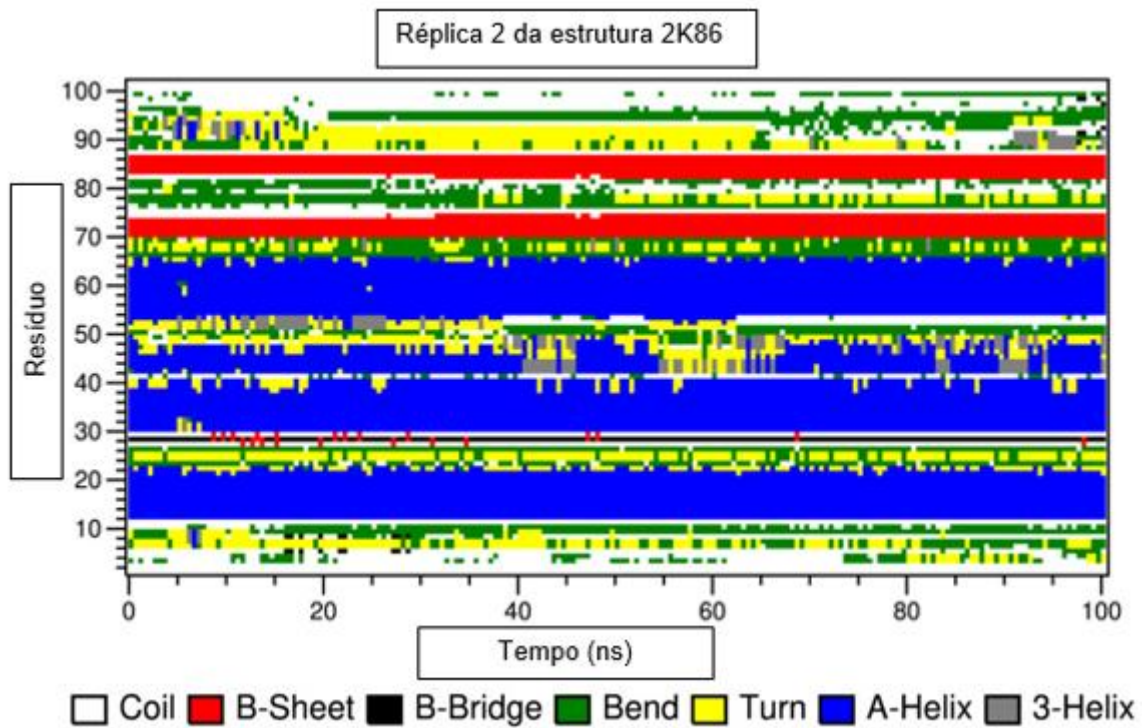


Figura 5.1.33: Representação estrutura 2k86 – réplica 2 desde o tempo 0 até aos 100 nanosegundos.

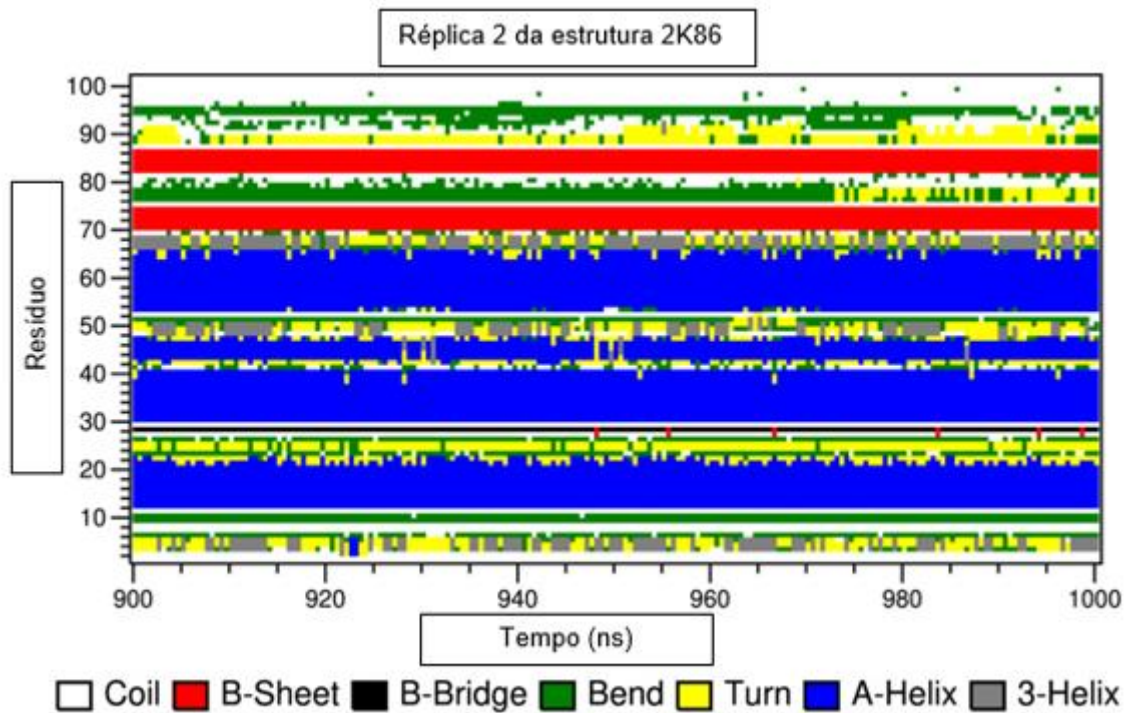


Figura 5.1.34: Representação da estrutura 2k86 – réplica 2 desde o tempo 900 até aos 1000 nanosegundos.

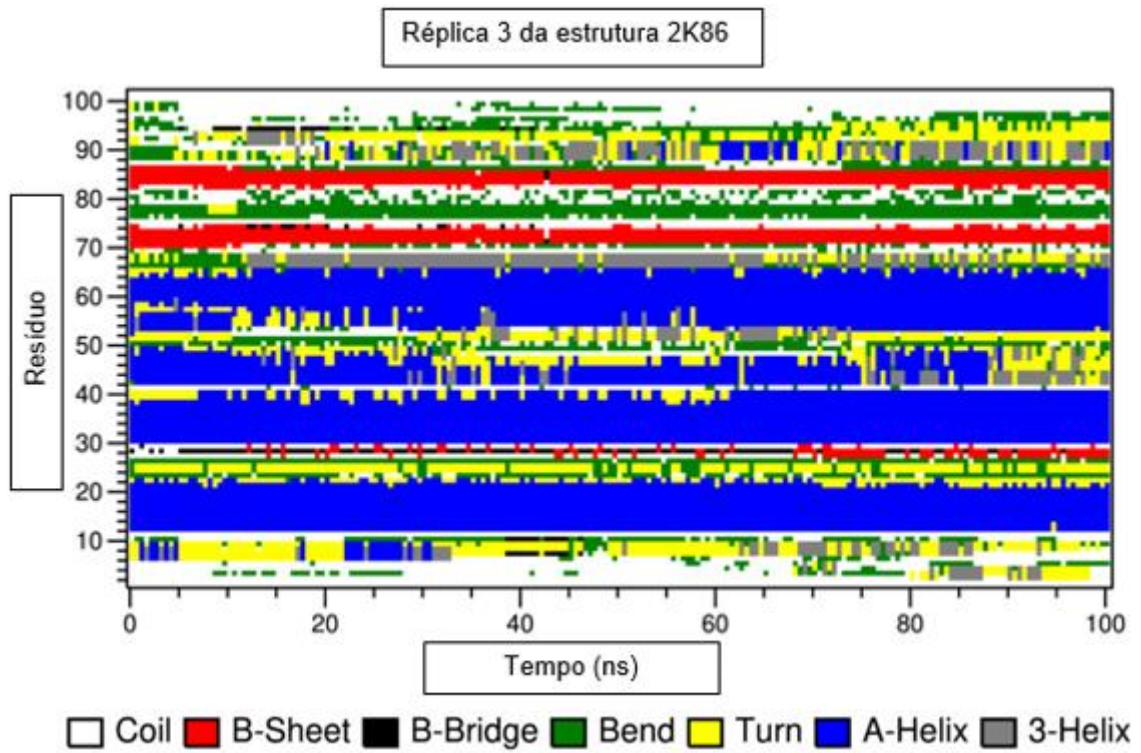


Figura 5.1.35: Representação estrutura 2k86 – réplica 3 desde o tempo 0 até aos 100 nanosegundos.

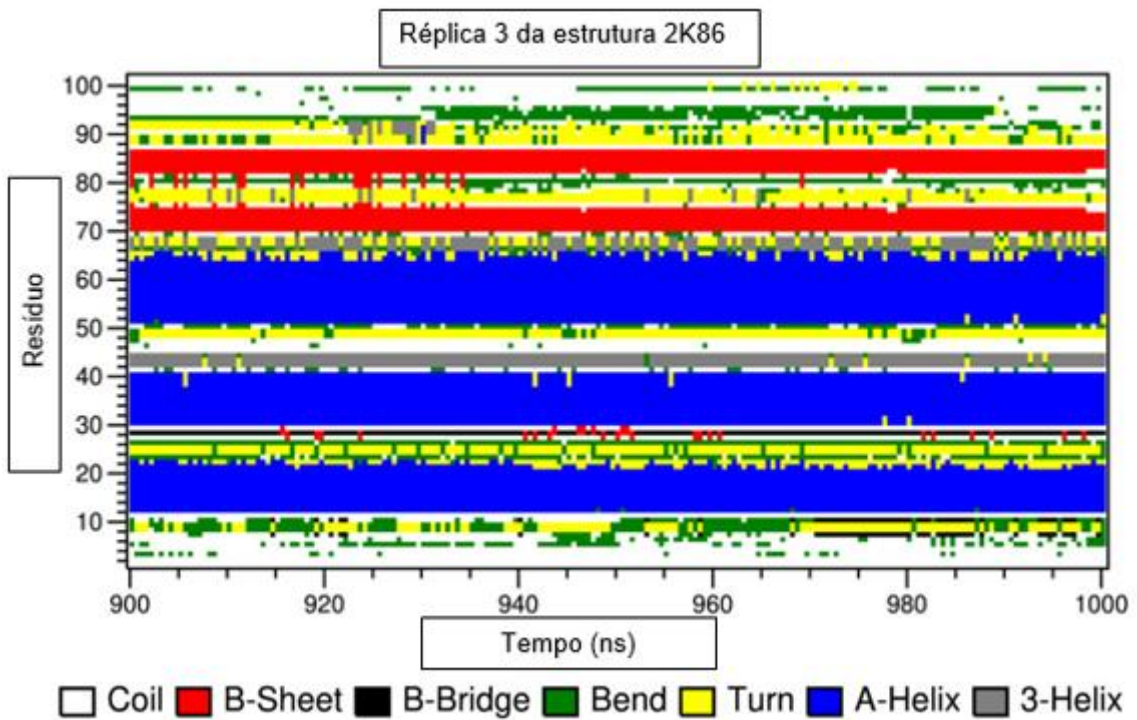


Figura 5.1.36: Representação da estrutura 2k86 – réplica 3 desde o tempo 900 até aos 1000 nanosegundos.

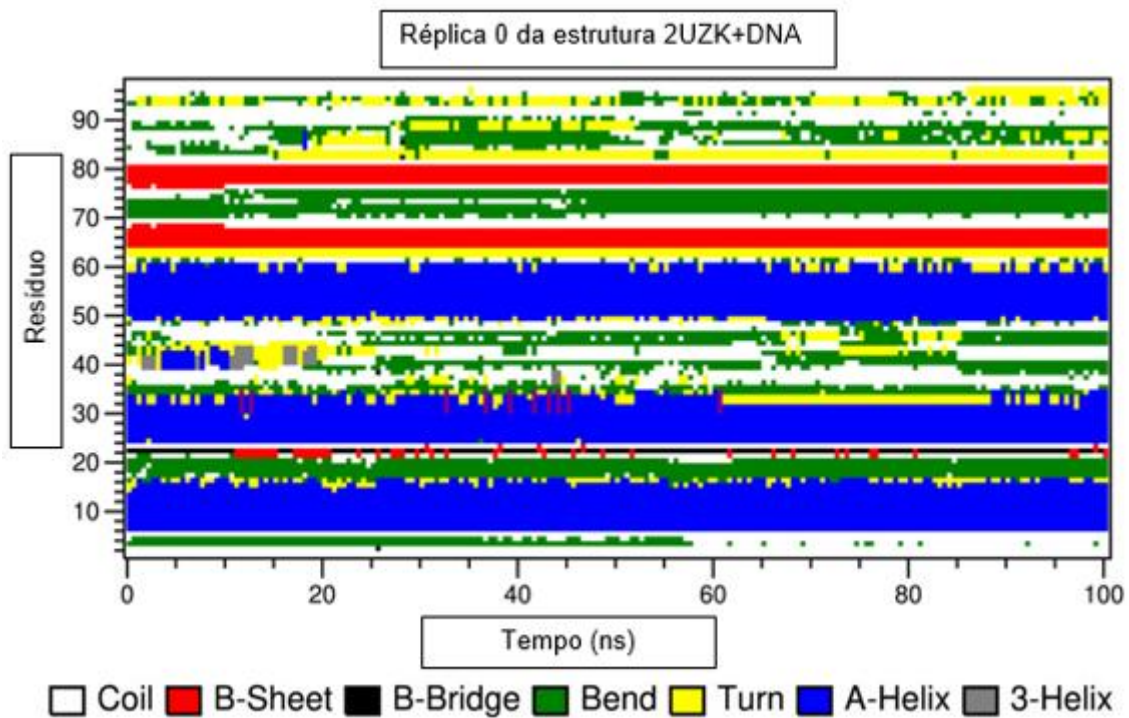


Figura 5.1.37: Representação estrutura 2UZK com DNA – réplica 0 desde o tempo 0 até aos 100 nanosegundos.

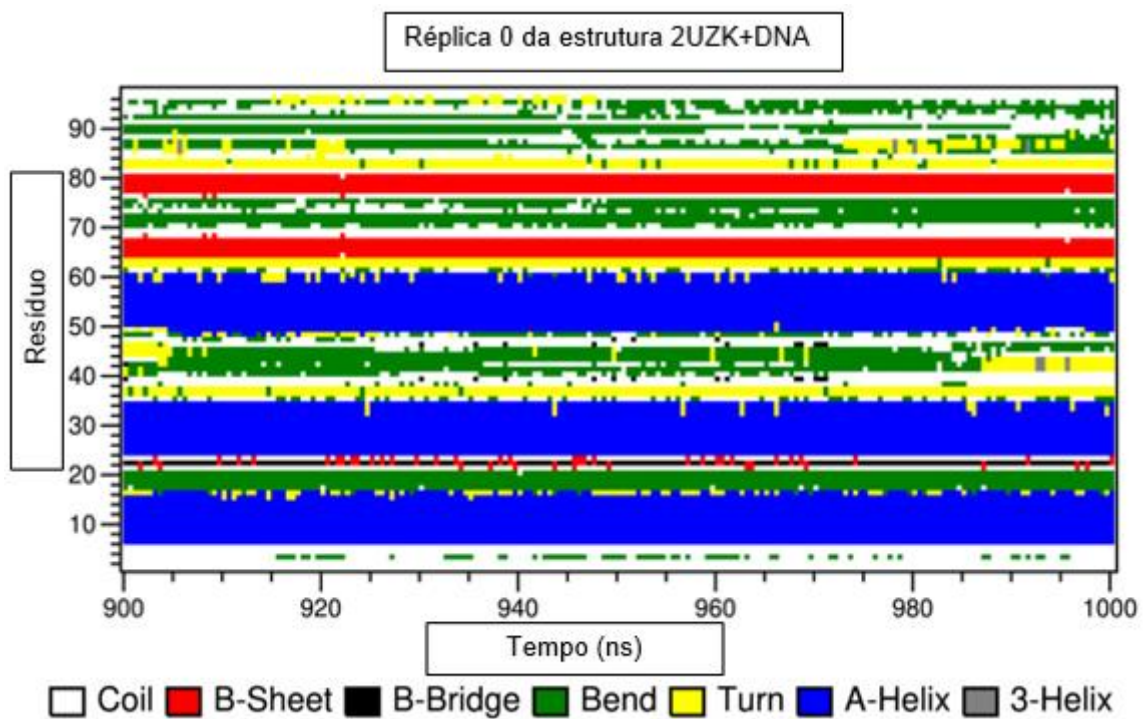


Figura 5.1.38: Representação da estrutura 2UZK com DNA – réplica 0 desde o tempo 900 até aos 1000 nanosegundos.

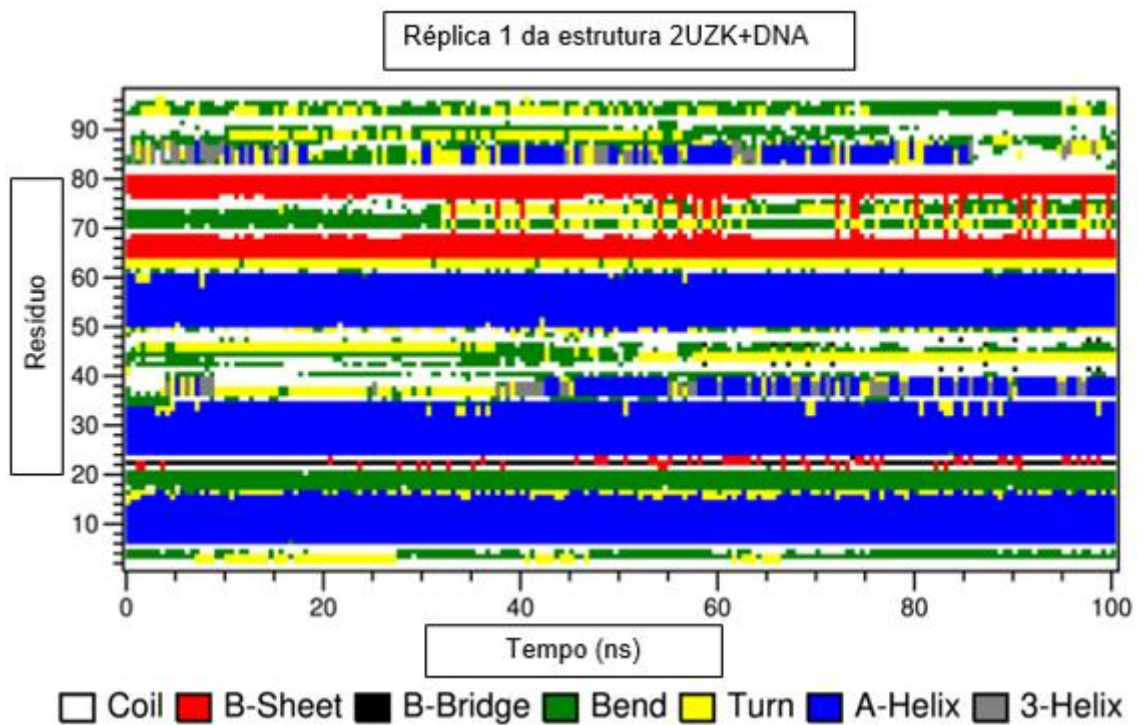


Figura 5.1.39: Representação estrutura 2UZK com DNA – réplica 1 desde o tempo 0 até aos 100 nanosegundos.

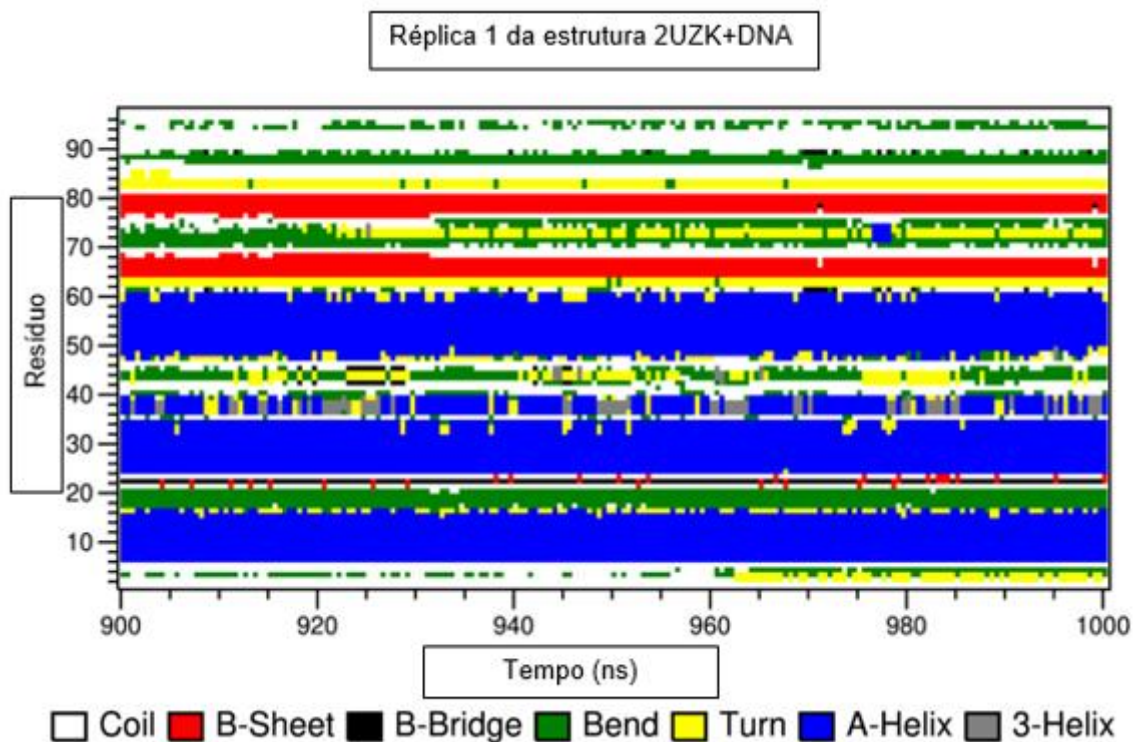


Figura 5.1.40: Representação da estrutura 2UZK com DNA – réplica 1 desde o tempo 900 até aos 1000 nanosegundos.

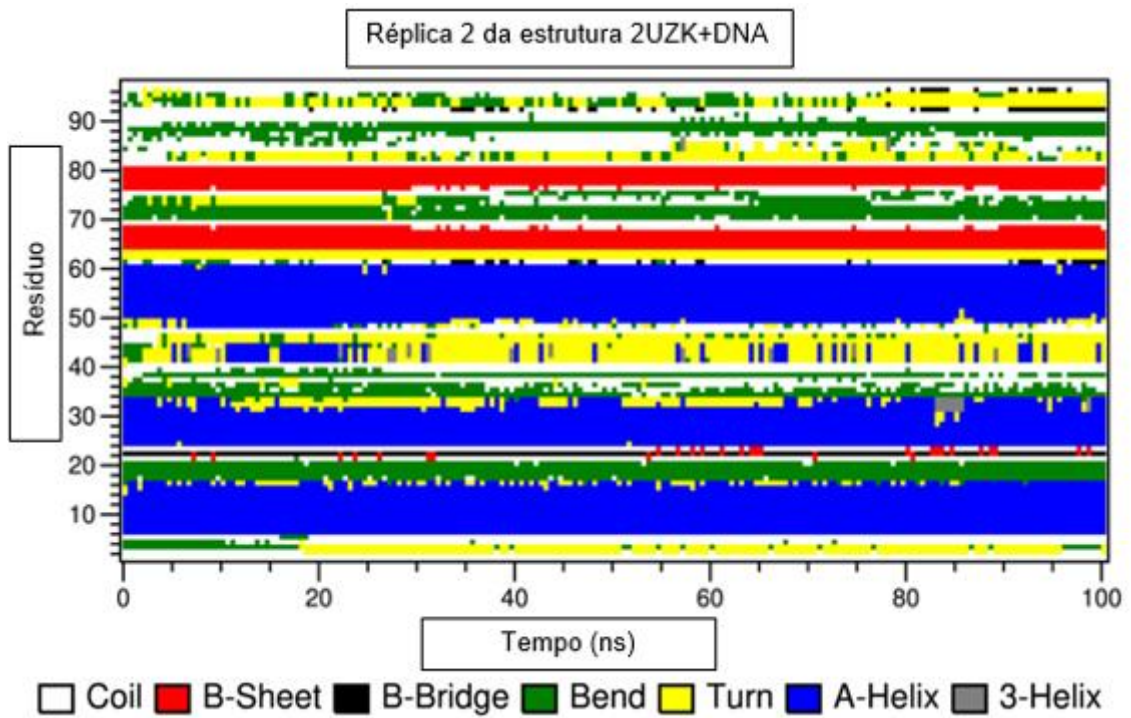


Figura 5.1.41: Representação estrutura 2UZK com DNA – réplica 2 desde o tempo 0 até aos 100 nanosegundos.

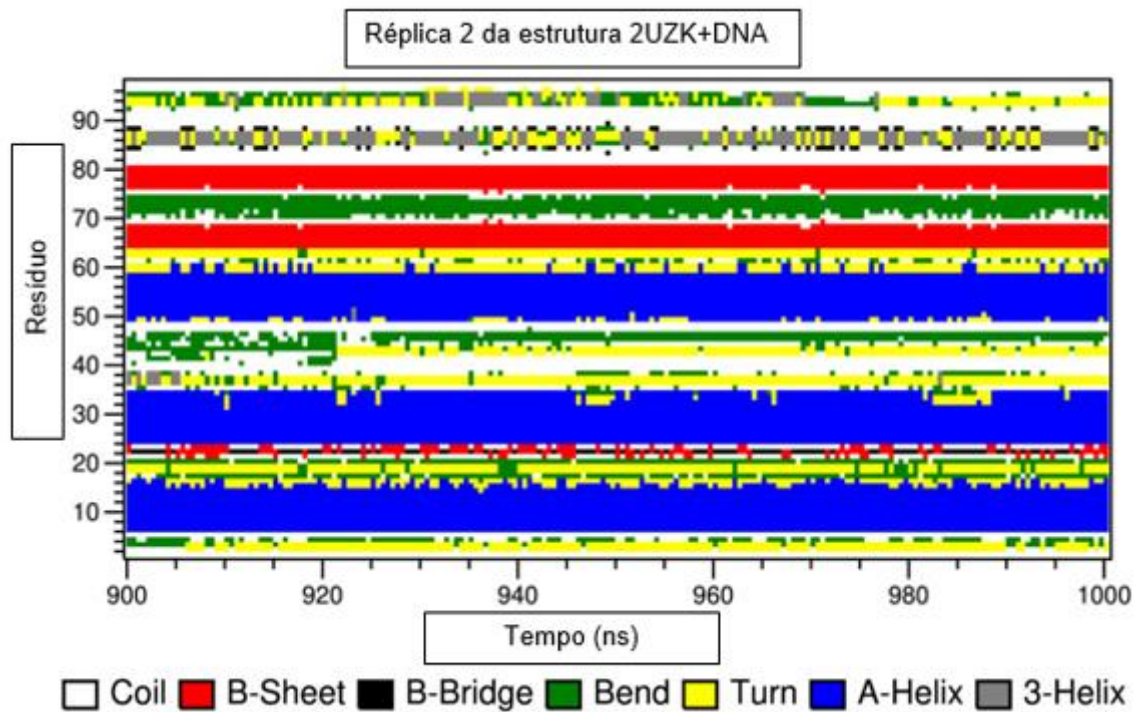


Figura 5.1.42: Representação da estrutura 2UZK com DNA – réplica 2 desde o tempo 900 até aos 1000 nanosegundos.

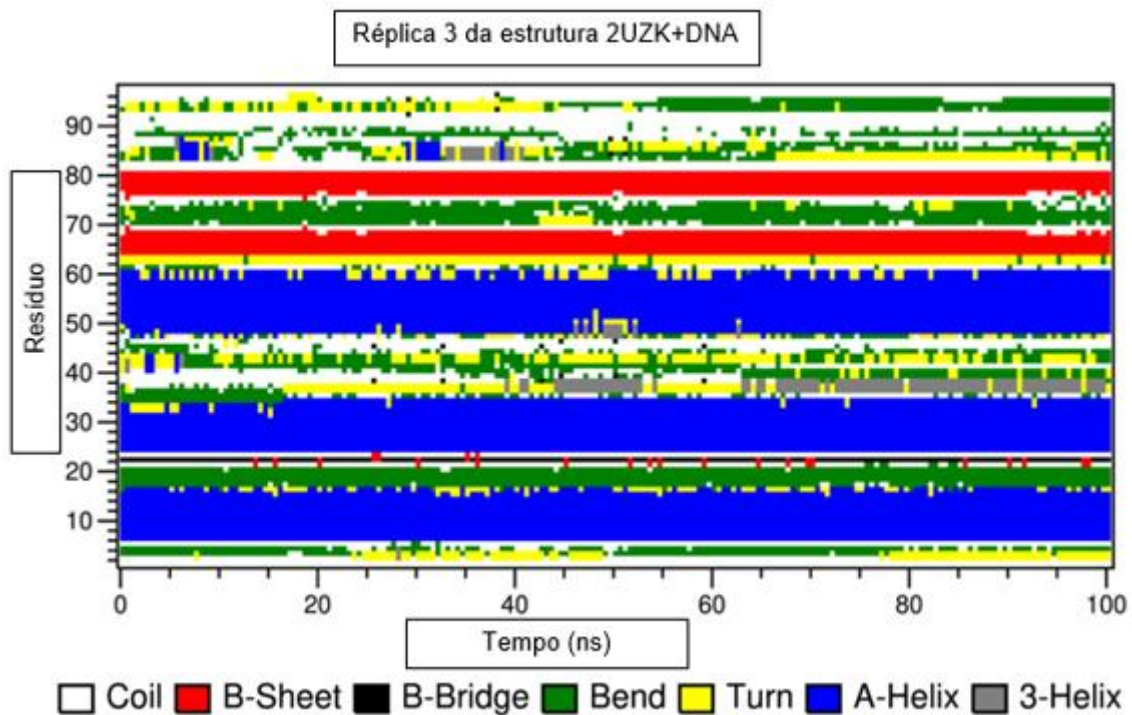


Figura 5.1.43: Representação estrutura 2UZK com DNA – réplica 3 desde o tempo 0 até aos 100 nanosegundos.

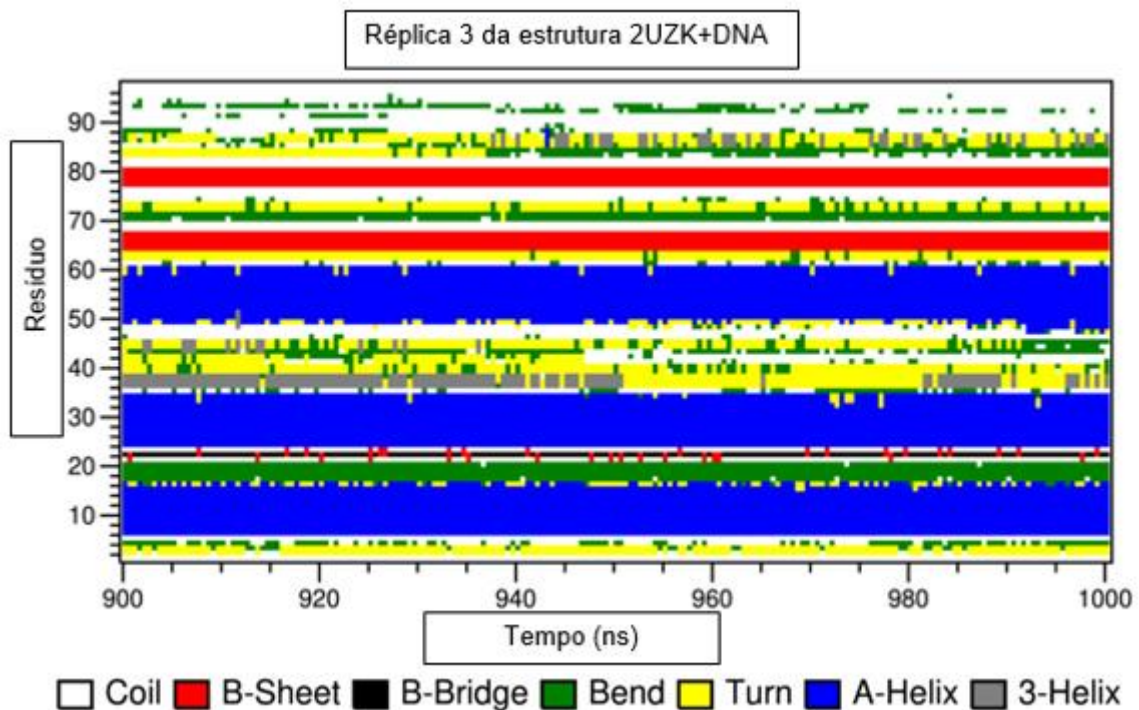


Figura 5.1.44: Representação da estrutura 2UZK com DNA – réplica 3 desde o tempo 900 até aos 1000 nanosegundos.

A ferramenta *do_dssp* lê o arquivo da trajetória e calcula a estrutura secundária para cada conformação da trajetória. O *xpm2ps* converte o documento xv feito pela ferramenta *do_dssp* em *postscript*.

Em todas as figuras verifica-se que a estrutura básica mantém-se ao longo de toda a simulação, ou seja, as hélices alfa, as folhas beta e a asa 2. Vendo-se assim que a estrutura tem zonas rígidas que mantém-se e vê-se que não houve destruição da proteína ao longo dos 1000 ns da simulação, o que permite verificar que a simulação foi bem realizada e que se pode continuar com o estudo. (1000 ns já representam uma escala temporal considerável para a evolução da conformação de uma proteína em condições fisiológicas).

As únicas variações significativas ao longo das simulações são as zonas desordenadas da proteína FoxO3, o *loop* e a cauda. Estas variações são uma consequência natural da flexibilidade intrínseca da FoxO3 e não representam qualquer problema em termos da qualidade simulação, pelo contrário contribuem para a sua validação.

Na estrutura 2K86 observa-se claramente a organização da hélice 4, a qual não está presente nas simulações tendo como ponto de partida a estrutura 2uzk. A hélice 4 está localizada entre os resíduos 42 e 50.

5.2 Descoberta de cavidades

Para a descoberta de cavidades foram utilizadas 21 conformações, amostradas de cada trajetória por meio da ferramenta *trajconv* com as opções “*-sep -dt 50000*”, guardando assim uma conformação a cada 50 nanosegundos do documento inicial com 100000 conformações, obtendo-se no final 21 conformações em ficheiros separados.

Essas 21 conformações foram extraídas das corridas de dinâmica molecular para todas as estruturas e correspondentes réplicas.

De seguida, estas conformações foram submetidas ao programa FPOCKET para deteção de potenciais cavidades. Após a corrida do programa, este gera uma

pasta com diversos ficheiros contendo informação sobre as diversas cavidades encontradas.

A análise dos resultados relativos à deteção de cavidades teve de ser feita sem perder de vista o objetivo desta tese: encontrar moléculas moduladoras da atividade da proteína FoxO3, ou seja, moléculas que conseguem inibir ou potenciar a ligação da proteína FoxO3 ao DNA. O primeiro passo na descoberta dessas moléculas moduladoras terá necessariamente que ser a identificação das potenciais cavidades onde estas se possam ligar.

Uma cavidade para ser considerada potencialmente inibidora da ligação tem que existir exclusivamente na forma livre da proteína ou existir numa zona crucial de ligação ao DNA, impedindo assim, ou dificultando, a ligação da proteína ao DNA.

Já uma cavidade potenciadora da ligação FoxO3-DNA deveria existir apenas na estrutura ligada, ou atuar diretamente na interface FoxO3-DNA, estabilizando-a.

Deste modo, organizou-se toda a informação relevante nas seguintes tabelas. Nestas temos representados diversos parâmetros relativos a cada cavidade encontrada; conformação onde foi encontrada, pontuação da cavidade, pontuação de drogabilidade, volume e centro de massa da cavidade.

As cavidades são definidas pelos conjuntos de esferas alfa. As cavidades são distinguidas pelo número de resíduos (esferas alfa), a melhor cavidade é sempre a primeira e a que tem maior número de resíduos ^[42].

A pontuação da cavidade é a probabilidade de esta ser um sítio de ligação de uma molécula pequena. A pontuação de drogabilidade representa o quão “drogável” é o local de ligação, ou seja, estima a sua capacidade de ligar-se a pequenas moléculas semelhantes a fármacos. Em termos simples, uma pontuação de drogabilidade 0 ou próxima de 0, indica que a cavidade tem pouco potencial para a ligação de moléculas com as características químicas típicas de um fármaco. Se a pontuação for superior a 0.5, considera-se existir uma possibilidade significativa de ligação de moléculas semelhantes a fármacos na cavidade (note-se que estes são valores indicativos apenas) ^[43]. As previsões de drogabilidade são importantes para evitar alvos intratáveis e para concentrar os esforços de descoberta de medicamentos

em locais que oferecem melhores perspectivas, o que faz consequentemente que seja necessário menos tempo de estudo.

O volume da cavidade é calculado por meio de um algoritmo de Monte-Carlo que calcula o volume total ocupado por todas as esferas alfa na cavidade [33].

O centro de massa representa o centro da cavidade, este parâmetro é necessário para a corrida de *docking*.

As seguintes tabelas (5.2.3 – 5.2.14) apenas demonstram os melhores valores, ou seja, as pontuações de drogabilidade superiores a 0.5, sendo que se optou por não considerar drogabilidades abaixo deste valor.

- 2k86 réplica 0

Tabela 5.2.3: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 2 | 0.299 | 0.769 | 558.116 | 31.49 | 41.65 | 26.34 |
| | 6 | 0.247 | 0.942 | 892.801 | 32.43 | 42.66 | 27.65 |
| | 7 | 0.380 | 0.586 | 326.915 | 29.39 | 42.72 | 28.32 |
| | 8 | 0.331 | 0.829 | 1413.671 | 29.52 | 43.26 | 30.11 |
| | 9 | 0.437 | 0.795 | 378.778 | 30.13 | 40.21 | 26.32 |
| | 12 | 0.415 | 0.751 | 242.559 | 29.58 | 40.50 | 30.34 |
| | 16 | 0.336 | 0.738 | 393.093 | 29.96 | 41 | 27.93 |
| 2 | 4 | 0.325 | 0.782 | 592.607 | 36.22 | 40.96 | 26.44 |

A cavidade 1 não é muito interessante porque em todas as conformações estão localizadas no segmento flexível da asa 2 e é provável que não esteja presente em uma estrutura em solução.

A cavidade 2 também está localizada na região da asa flexível 2.

Apesar de a pontuação da cavidade e a pontuação de drogabilidade nas cavidades serem aceitáveis, a localização das mesmas não são interessantes para este estudo, pois são zonas flexíveis e que possivelmente não se mantêm ao longo

do tempo. Neste tipo de estudo procuram-se cavidades que sejam suficientemente estáveis e conservadas ao longo das variações conformacionais da proteína de forma a permitir a associação estável de ligandos que possam modular a sua atividade. Assim sendo, nesta réplica não se detetou nenhuma cavidade interessante para a análise de *docking*.

- 2k86 réplica 1

Tabela 5.2.4: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidades | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|-----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 1 | 0.106 | 0.631 | 455.608 | 30.62 | 68.32 | 28.41 |
| | 8 | 0.267 | 0.531 | 422.855 | 29.94 | 47.27 | 27.23 |
| | 9 | 0.463 | 0.938 | 549.121 | 43.51 | 61.00 | 38.23 |
| | 12 | 0.234 | 0.515 | 573.907 | 32.17 | 42.06 | 25.26 |
| | 13 | 0.299 | 0.613 | 507.386 | 32.46 | 41.91 | 28.94 |
| 2 | 2 | 0.141 | 0.576 | 353.687 | 38.18 | 35.92 | 33.72 |

A cavidade 1 na conformação 9 tem a maior pontuação de drogabilidade e a oportunidade mais interessante para um ligando modular a interação DNA-proteína, uma vez que fica próximo hélice 3, com potencial para interagir com DNA e proteína, talvez interrompendo o reconhecimento do DNA pela FoxO3a. Uma outra possibilidade é que o sítio de ligação seja mascarado pelo DNA, tornando a afinidade do ligando muito maior para a proteína não ligada.

As outras cavidades ficam novamente na cauda altamente flexível da asa 2 e, portanto, não apresentam uma grande promessa para encontrar ligandos altamente específicos.

A cavidade 1 na conformação 9 revela-se assim como a mais promissora para o estudo de *docking*.

- 2k86 réplica 2

Tabela 5.2.5: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Conformação | Pontuação | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-------------|-----------|----------------------------|--------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 0 | 0.466 | 0.515 | 374.266 | 30.54 | 48.66 | 35.81 |
| | 3 | 0.319 | 0.557 | 475.340 | 29.28 | 46.02 | 35.16 |
| | 5 | 0.239 | 0.968 | 871.082 | 37.77 | 42.38 | 39.24 |
| | 8 | 0.161 | 0.600 | 435.936 | 44.39 | 63.64 | 37.49 |
| | 11 | 0.536 | 0.863 | 327.053 | 32.23 | 48.14 | 37.84 |
| | 15 | 0.327 | 0.535 | 270.098 | 37.93 | 61.51 | 39.39 |
| | 16 | 0.266 | 0.584 | 296.675 | 36.32 | 62.50 | 38.73 |
| | 18 | 0.257 | 0.756 | 344.835 | 35.84 | 64.61 | 38.07 |
| | 19 | 0.298 | 0.685 | 636.087 | 42.37 | 44.94 | 40.22 |
| 2 | 17 | 0.181 | 0.700 | 314.051 | 37.04 | 62.97 | 38.28 |
| 3 | 1 | 0.348 | 0.682 | 374.483 | 45.69 | 61.74 | 34.36 |

A cavidade 1 encontra-se, nas conformações 0,3,5,8,11,19, entre a cauda C-terminal flexível e a hélice H1. Como tal, o seu interesse é discutível uma vez que a cauda é uma zona desordenada e varia muito a sua orientação espacial. Apesar da conformação 5 ter maior pontuação de drogabilidade encontra-se numa zona flexível, o que não se adequa ao estudo.

A cavidade 1 nas conformações 15,16 e 18 encontra-se na zona flexível N-terminal antes da hélice H1. A cavidade 2 na conformação 17 encontra-se nesta mesma zona.

Já a cavidade 3 na conformação 1 é muito mais interessante, pois está entre o N-terminal e a hélice H3 (a hélice de reconhecimento). Os ligandos alojados nesta cavidade poderão eventualmente interromper o reconhecimento do DNA pela FoxO3a ou seja, poderão funcionar como inibidores da ligação da molécula ao DNA.

- 2k86 réplica 3

Tabela 5.2.6: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|--------|-------|
| | | | | | | | |
| 1 | 9 | 0.315 | 0.771 | 702.242 | 4.17 | -27.31 | 37.54 |
| | 14 | 0.317 | 0.571 | 569.578 | 4.29 | -18.27 | 36.61 |
| 2 | 10 | 0.351 | 0.887 | 553.896 | 13.5 4 | -4.48 | 27.16 |

A cavidade 1 nas conformações 9 e 14 novamente não é tão interessante porque é formada principalmente pela cauda da asa 2. No entanto, a cavidade 2 na conformação 10 parece mais promissora, pois interage com a hélice 3 e a região N-terminal/início da hélice 1 e tem a maior pontuação de drogabilidade. Logo esta cavidade revela-se promissora para o estudo de *docking* molecular, pois poderá ser um local de ligação para um potencial da ligação FoxO3-DNA.

Verifica-se que ao longo das simulações das diferentes réplicas, algumas cavidades aparecem no mesmo local, por exemplo réplica 1 - conformação 9, réplica 2 - conformação 1 e réplica 3 conformação 10. Esta constância na observação de uma cavidade através das diferentes réplicas torna-a automaticamente mais interessante como potencial local de ligação estável.

Desta forma a cavidade 2, na conformação 10 foi selecionada para o estudo de *docking*.

A estrutura cristalográfica 2k86 (retirada diretamente do *site protein data bank*, pré-simulação), não revelou a presença de nenhuma cavidade com pontuação de drogabilidade superior a 0.5.

- 2UZK réplica 0

Tabela 5.2.7: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 1 | 0.334 | 0.554 | 359.217 | 37.73 | 48.07 | 20.13 |
| | 6 | 0.438 | 0.897 | 516.615 | 33.12 | 49.24 | 21.68 |
| | 7 | 0.209 | 0.726 | 473.535 | 40.29 | 59.54 | 45.15 |
| | 16 | 0.222 | 0.507 | 721.990 | 38.93 | 52.39 | 19.42 |
| 2 | 8 | 0.231 | 0.530 | 318.035 | 38.02 | 46.62 | 20.87 |
| 6 | 10 | 0.136 | 0.624 | 411.409 | 28.95 | 48.30 | 25.25 |

A cavidade 1 na conformação 1 interage com o *loop* entre a hélice 2 e a hélice 3. Esta região do *loop* corresponde à forma desorganizada de hélice 4 prevalente nas simulações da estrutura 2K86. Não está claro se deve ser favorecido nas simulações ligadas ao DNA.

A cavidade 1 na conformação 6 está inserida entre as hélices 1 e 2 e parcialmente a cauda N-terminal. Tem algum potencial, pois não é muito dependente das regiões mais flexíveis.

A cavidade 1 na conformação 7 é criada principalmente pela cauda da asa 2 e como tal não é muito interessante.

A cavidade 1 na conformação 16 é semelhante a cavidade 1 na conformação 6, na medida em que fica inserida entre as hélices 1 e 2, e parcialmente a hélice 3. Pode ser potencialmente interessante.

A cavidade 2 na conformação 8 fica no mesmo local que a cavidade 1 na conformação 1, no *loop* desordenado entre as hélices 2 e 3. Pode ser interessante apenas como recurso de ligação ao DNA.

A cavidade 6 na conformação 10 é criada pela cauda flexível do terminal N, logo não é interessante para o estudo.

Assim sendo, a cavidade 1 na conformação 16 e 6 é uma potencial cavidade para o estudo do docking.

- 2UZK réplica 1

Tabela 5.2.8: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 0 | 0.387 | 0.677 | 293.390 | 34.47 | 44.46 | 27.08 |
| | 4 | 0.412 | 0.652 | 287.593 | 28.24 | 43.90 | 19.16 |
| | 5 | 0.486 | 0.822 | 370.768 | 38.68 | 50.08 | 21.08 |
| | 11 | 0.291 | 0.707 | 240.798 | 32.00 | 56.16 | 20.62 |
| | 15 | 0.299 | 0.739 | 735.287 | 29.30 | 58.36 | 24.17 |
| | 17 | 0.277 | 0.572 | 624.105 | 29.52 | 59.37 | 20.36 |
| 2 | 2 | 0.279 | 0.731 | 504.717 | 40.52 | 45.32 | 22.89 |
| | 12 | 0.179 | 0.612 | 467.746 | 31.48 | 58.08 | 24.45 |
| | 14 | 0.291 | 0.515 | 410.712 | 30.57 | 56.23 | 23.29 |

A cavidade 1 na conformação 0 é parcialmente formada pelo laço flexível N-terminal, como tal, não é potencialmente interessante para o estudo.

A cavidade 1 na conformação 4 está na região do laço flexível entre as hélices 2 e 3, o que pode ser interessante.

A cavidade 1 na conformação 5 está no lado exposto da hélice 2, parecendo ser um estável candidato a um bolso de licitação.

A cavidade 1 na conformação 11 e 17 fica no *loop* desordenado entre as hélices 2 e 3, tal como a cavidade 2 na conformação 12 e 14. Esta cavidade é um possível candidato ao *docking*, pois mantém-se estável ao longo da simulação e apresenta valores altos de pontuação de drogabilidade. Esta cavidade pode ser um inibidor da ligação da proteína FoxO3 ao DNA.

A cavidade 1 na conformação 15 está localizada no grupo do *loop* desordenado entre hélices 2 e 3, e próximo ao terminal N flexível.

A cavidade 2 na conformação 2 está entre as hélices 1 e 2 e parcialmente revestida pelo segmento N-terminal flexível.

Assim sendo a cavidade 1, na conformação 5 foi selecionada para o estudo de *docking*.

- 2UZK réplica 2

Tabela 5.2.9: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 0 | 0.382 | 0.720 | 389.003 | 32.09 | 48.62 | 38.90 |
| | 6 | 0.253 | 0.502 | 877.603 | 30.36 | 51.31 | 21.44 |
| | 7 | 0.381 | 0.720 | 406.350 | 51.28 | 60.60 | 33.04 |
| | 8 | 0.277 | 0.588 | 692.456 | 35.14 | 42.18 | 31.20 |
| | 20 | 0.301 | 0.684 | 645.534 | 50.48 | 61.28 | 32.95 |
| 2 | 19 | 0.184 | 0.699 | 471.994 | 32.59 | 41.10 | 32.71 |

A cavidade 1 na conformação 0 localiza-se na fenda entre as hélices 1 e 2, o que aparenta ser um bom local de ligação.

A cavidade 1 na conformação 6 situa-se entre as hélices 2 e 3, ladeado pelo *loop* desordenado, o que pode ser potencialmente interessante.

A cavidade 1 na conformação 7 é formada principalmente pelo *loop* desordenado e pela asa 2, pelo que não é tão interessante.

A cavidade 1 na conformação 8 está na hélice 1 e cauda N-terminal. Esta cavidade é potencialmente interessante, embora parte do revestimento possa ser instável.

A conformação 19 e 20 não são interessantes para o estudo devido, pois encontram-se em zonas flexíveis da proteína, nomeadamente cauda do N terminal e a asa 2, respetivamente.

- 2UZK réplica 3

Tabela 5.2.10: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 7 | 0.372 | 0.801 | 405.605 | 29.12 | 44.95 | 35.09 |
| | 8 | 0.311 | 0.514 | 564.207 | 29.44 | 46.42 | 36.26 |
| | 10 | 0.210 | 0.748 | 970.221 | 31.05 | 45.52 | 36.18 |
| | 11 | 0.148 | 0.634 | 664.192 | 29.91 | 44.38 | 35.41 |
| | 14 | 0.247 | 0.813 | 871.179 | 30.38 | 46.54 | 37.40 |
| | 15 | 0.282 | 0.574 | 592.699 | 27.00 | 47.67 | 35.16 |
| 2 | 18 | 0.346 | 0.733 | 600.392 | 29.85 | 44.46 | 37.38 |
| | 0 | 0.179 | 0.567 | 182.171 | 28.27 | 51.66 | 35.92 |
| | 19 | 0.277 | 0.502 | 418.497 | 44.54 | 47.98 | 38.95 |

A cavidade 1 na conformação 7, 8, 11, 14, 15, 18 aparenta ser interessante pois está localizada perto da hélice 3 e o laço desordenado conectando para a hélice 2, o que pode interferir na ligação ao DNA. Esta cavidade é um bom candidato ao estudo pelo *docking*, pois mantém-se ao longo da trajetória, tem uma boa pontuação de drogabilidade e vai de encontro com o objetivo do estudo.

A cavidade 1 na conformação 10 encontra-se na mesma localização das conformações acima, mas esta cavidade é consideravelmente maior que as anteriores, volume 970,221. A cavidade encontra-se numa localização interessante, mas devido ao seu tamanho pode ser instável e alocar ligandos pouco seletivos e específicos. A cavidade 2 na conformação 0 inserida entre a hélice 1 e o *loop* desordenado entre hélice 2 e 3. A maior parte do revestimento do bolso vem do terminal N cauda flexível, então não muito interessante.

A cavidade 2 na conformação 19 está na face oposta à hélice 3, inserida entre as hélices 1 e 2, o que a torna uma cavidade potencialmente inibidora da ligação ao DNA.

A estrutura cristalográfica 2UZK (pré-dinâmica molecular) também apresenta uma cavidade com pontuação de drogabilidade superior a 0,5. A cavidade 1 encontra-se entre o terminal N e o *loop*.

Portanto a cavidade 1, na conformação 7 foi selecionada para o estudo de *docking*.

No caso da estrutura do complexo (2uzk+DNA) antes de analisarmos a pontuação de drogabilidade da cavidade, foi necessário verificar se a cavidade apenas interagiu com o DNA, com a proteína ou com ambos. As cavidades mais importantes neste contexto são aquelas que interagem exclusivamente com a proteína, ou em simultâneo com a proteína e DNA. As cavidades que interagem apenas com o DNA não são importantes neste estudo. Através de um *script* foi possível ver se a cavidade interagiu com os aminoácidos da proteína e as bases dos ácidos nucleicos do DNA.

Neste estudo foram ignorados todos as cavidades localizados exclusivamente no DNA. São importantes as cavidades que interagem com o DNA e a proteína e aqueles que surgem apenas na proteína ligada ao DNA, não aparecendo na sua forma livre.

- Complexo réplica 0

Tabela 5.2.11: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 11 | 0.568 | 0.970 | 1593.920 | 44.13 | 45.33 | 32.26 |
| | 14 | 0.323 | 0.684 | 1699.408 | 45.86 | 45.35 | 32.35 |
| | 18 | 0.147 | 0.607 | 419.910 | 29.69 | 40.03 | 34.97 |
| | 20 | 0.559 | 0.661 | 1837.306 | 33.74 | 44.87 | 26.73 |
| 2 | 2 | 0.247 | 0.722 | 747.790 | 30.12 | 41.74 | 36.55 |
| | 7 | 0.116 | 0.681 | 524.102 | 31.25 | 41.84 | 34.68 |

A cavidade 1 e 2 nas mais diversas conformações alojam ligandos que ligam-se simultaneamente à proteína foxo3 e ao DNA como se pode verificar nas tabelas 8.1 a 8.3 do anexo I.

A cavidade 1 na conformação 18 e a cavidade 2 na conformação 2 e 7 estão entre a H3 e o DNA, o que é bastante interessante para potenciar a ligação da proteína-DNA.

Já a cavidade 1 na conformação 11,14 e 20 são pouco interessantes, uma vez que se ligam maioritariamente ao DNA e a parte da proteína a que se ligam é uma zona flexível. A cavidade 1 nas conformações 11 e 14 encontra-se no entre o DNA e a asa 2, enquanto que a cavidade 1 conformação 20 está entre o DNA e o terminal N.

- Complexo réplica 1

Tabela 5.2.12: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 2 | 0.484 | 0.757 | 1214.327 | 35.63 | 47.42 | 36.51 |
| | 3 | 0.169 | 0.813 | 871.370 | 33.09 | 35.42 | 43.83 |
| | 5 | 0.240 | 0.721 | 595.605 | 31.29 | 32.65 | 44.03 |
| | 9 | 0.269 | 0.777 | 599.405 | 31.93 | 30.33 | 43.43 |
| | 11 | 0.243 | 0.773 | 755.555 | 32.67 | 29.47 | 44.78 |
| | 12 | 0.235 | 0.547 | 648.813 | 33.01 | 29.26 | 43.57 |
| | 13 | 0.280 | 0.790 | 726.932 | 33.20 | 29.07 | 44.34 |
| | 14 | 0.370 | 0.803 | 444.228 | 30.40 | 30.70 | 44.59 |
| | 15 | 0.168 | 0.556 | 262.825 | 31.68 | 29.86 | 42.10 |
| | 18 | 0.251 | 0.838 | 897.106 | 31.88 | 28.77 | 44.71 |
| 2 | 19 | 0.368 | 0.687 | 691.068 | 30.36 | 30.31 | 45.00 |
| | 7 | 0.366 | 0.550 | 698.860 | 33.33 | 30.65 | 44.48 |
| | 8 | 0.385 | 0.739 | 425.564 | 30.18 | 32.75 | 44.15 |
| | 10 | 0.254 | 0.876 | 393.540 | 31.51 | 32.14 | 42.30 |
| | 17 | 0.139 | 0.695 | 292.896 | 31.94 | 30.30 | 42.38 |

A cavidade 1 e 2 nas mais diversas conformações alojam ligandos que se ligam simultaneamente à proteína FOXO3 e ao DNA como se pode verificar nas tabelas 8.4 a 8.6 do anexo II.

A cavidade 1 na conformação 2 encontra-se maioritariamente no DNA e uma pequena parte na asa 2, o que torna esta cavidade pouco interessante.

Em todas as outras conformações, as cavidades encontram-se entre a hélice H3, a cauda e o DNA, isto faz com que estas cavidades 1 e 2 sejam interessantes para a ligação de potenciadores da interação da proteína-DNA.

- Complexo réplica 2

Tabela 5.2.13: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 6 | 0.322 | 0.960 | 716.147 | 40.71 | 37.07 | 43.96 |
| | 11 | 0.309 | 0.596 | 1262.318 | 33.18 | 47.83 | 28.10 |
| 2 | 5 | 0.342 | 0.858 | 797.071 | 34.94 | 45.41 | 41.36 |
| | 7 | 0.392 | 0.849 | 497.452 | 36.57 | 38.68 | 43.82 |
| 3 | 0 | 0.111 | 0.716 | 337.694 | 30.42 | 24.28 | 42.89 |

As cavidades 1 e 2 nas conformações 5, 6, 7 e 11 poderão alojar ligandos que se ligam simultaneamente ao DNA e a proteína FoxO3. A conformação 0, por outro lado, liga-se apenas à proteína FoxO3. Estes dados podem ser consultados nas tabelas 8.7 a 8.9 do anexo III.

As cavidades 1 e 2 nas conformações 6, 5 e 7 encontram-se entre a hélice H3, a cauda e o DNA.

A cavidade 1 na conformação 11 está maioritariamente no DNA e um pouco na cauda da proteína a qual, sendo uma região bastante flexível, torna esta cavidade pouco relevante.

A cavidade 3 na conformação 0 encontra-se entre as cadeias beta, a hélice H1 e o *loop*. Esta cavidade é interessante pois se existir na conformação ligada da proteína e não na sua conformação livre, poderá ser uma cavidade potenciadora da interação proteína-DNA.

A cavidade na conformação 5 é maior e está mais deslocada para a cauda da proteína e para o DNA do que nas conformações 6 e 7, as quais são menores e estão mais deslocadas para a hélice H3, sendo assim mais interessantes. A cavidade na conformação pode ser uma cavidade potenciadora da ligação DNA- proteína.

- Complexo réplica 3

Tabela 5.2.14: Parâmetros das cavidades encontradas a partir do programa *FPOCKET*.

| Cavidade | Confor mação | Pontua ção | Pontuação de drogabilidade | Volume (Å ³) | Centro de massa | | |
|----------|-----------------|---------------|-------------------------------|-----------------------------|-----------------|-------|-------|
| | | | | | | | |
| 1 | 0 | 0.115 | 0.745 | 368.515 | 24.39 | 28.43 | 44.17 |
| | 1 | 0.269 | 0.609 | 258.643 | 34.38 | 42.89 | 39.25 |
| | 10 | 0.350 | 0.859 | 1099.399 | 44.81 | 36.76 | 25.63 |
| 2 | 9 | 0.123 | 0.520 | 259.056 | 33.88 | 41.56 | 40.36 |
| | 11 | 0.199 | 0.891 | 498.011 | 37.25 | 49.35 | 34.37 |
| | 17 | 0.241 | 0.889 | 436.972 | 27.28 | 29.18 | 43.45 |
| | 20 | 0.386 | 0.622 | 259.835 | 32.29 | 47.58 | 34.27 |

As cavidades 1 e 2 nas conformações 1, 9, 10, 11, 17 e 20 poderão alojar ligandos que se liguem simultaneamente ao DNA e a proteína FoxO3. A conformação 0, por outro lado, interage apenas com a proteína. Estes dados podem ser consultados nas tabelas 8.10 a 8.12 do anexo IV.

A cavidade 1 na conformação 10 encontra-se maioritariamente no DNA e um pouco na cauda da proteína FoxO3, o que não é importante. Esta cavidade é observada nas várias réplicas de dinâmica molecular, o que demonstra a reprodutibilidade dos resultados, e dá mais peso às conclusões retiradas do estudo (as réplicas foram iniciadas a partir de estados distintos e mesmo assim parecem convergir para a exibição de propriedades estruturais comuns). Esta cavidade

observa-se sempre na conformação 1, 11 e 14 da réplica 0, na conformação 2 da réplica 1, na conformação 11 da réplica 2 e por último na conformação 10 na réplica 3.

A cavidade 1 na conformação 1 e 9 encontra-se entre a hélice H3, o DNA e a cauda.

A cavidade 2 na conformação 11 e 20 encontra-se entre o DNA e a cauda, o que pode ser interessante para potenciar a ligação do DNA à proteína, pois um potencial ligando poderá assim interagir com ambas as moléculas, fortalecendo a sua ligação. Sabe-se que a cauda é uma zona flexível da proteína FoxO3, mas esta mantém-se estável ao longo da simulação de dinâmica molecular, pois tem um papel importante na estabilização e na ligação da proteína FoxO3 ao DNA.

A cavidade 1 e 2 na conformação 0 e 17, respetivamente, encontra-se entre o terminal N e o *loop*. Esta cavidade podia ser interessante se apenas fosse observada na estrutura ligada ao DNA. Esta cavidade foi observada aparece na estrutura 2UZK sem DNA, pré-simulação de MD.

A cavidade que se encontra entre a hélice H3, o DNA e a cauda mantém-se em todas as réplicas e em várias conformações da proteína durante a simulação de dinâmica molecular, o que torna ainda mais evidente a importância desta cavidade para a descoberta de ligando potenciadores da ligação proteína FoxO3 ao DNA e a própria fiabilidade das corridas de dinâmica molecular.

A estrutura do complexo pré-dinâmica molecular também apresentou uma cavidade com mais de 0,5 de drogabilidade. Essa cavidade é a mesma que está presente na conformação 0 e 17 da réplica 3 do complexo.

Encontraram-se cerca de 20 cavidades interessantes para este estudo, mas dadas as limitações de tempo o estudo foi focado em apenas 4 cavidades, a saber:

- Cavidade 2 na conformação 10 da réplica 3 da estrutura 2K86
- Cavidade 1 na conformação 5 da réplica 1 da estrutura 2UZK
- Cavidade 1 na conformação 7 da réplica 3 da estrutura 2UZK
- Cavidade 2 na conformação 2 da réplica 1 do complexo (2UZK+DNA)

A cavidade 2 na conformação 10 da réplica 3 da estrutura 2K86 interage com a hélice 3 e com a região N-terminal (início da hélice 1). Esta cavidade tem uma boa pontuação de drogabilidade, uma boa pontuação, ou seja profundidade da cavidade e um bom volume, ligandos de maiores dimensões tendem a ser menos seletivos e específicos para o local de ligação de interesse, pelo que as cavidades de dimensões extremas não deverão ser usadas neste estudo. Esta cavidade tem 0,351 de pontuação, 0,887 de pontuação de drogabilidade e 553,89 Å³ de volume. A cavidade é ilustrada nas Figuras 5.2.45 e 5.2.46, estando a proteína representada respetivamente como superfície e “cartoon”.

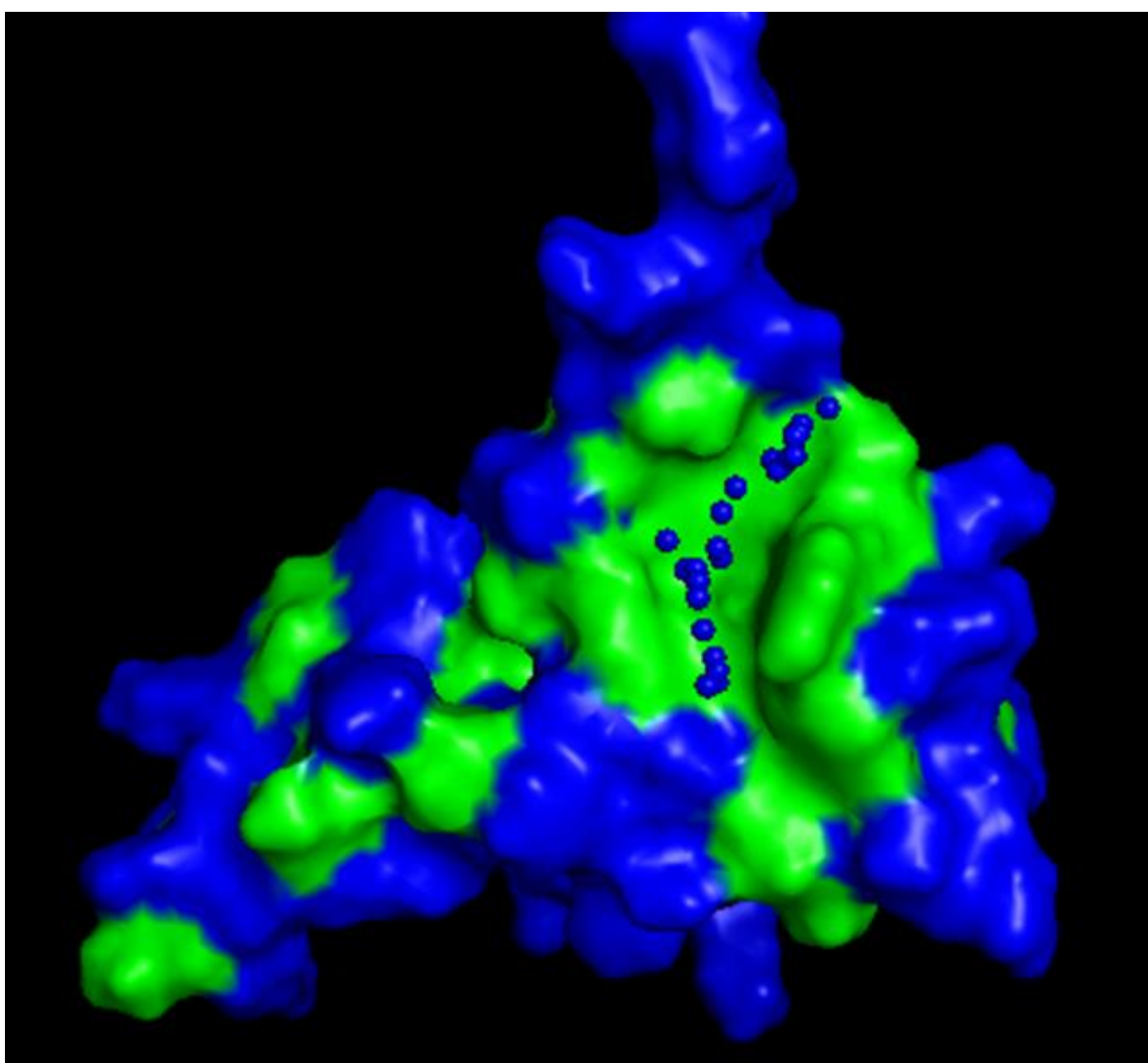


Figura 5.2.45: Representação da proteína foxO3, na estrutura 2k86, réplica 3, conformação 10. Representação da cavidade 2, na forma da proteína em superfície. A azul encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a azul representam o local da cavidade.

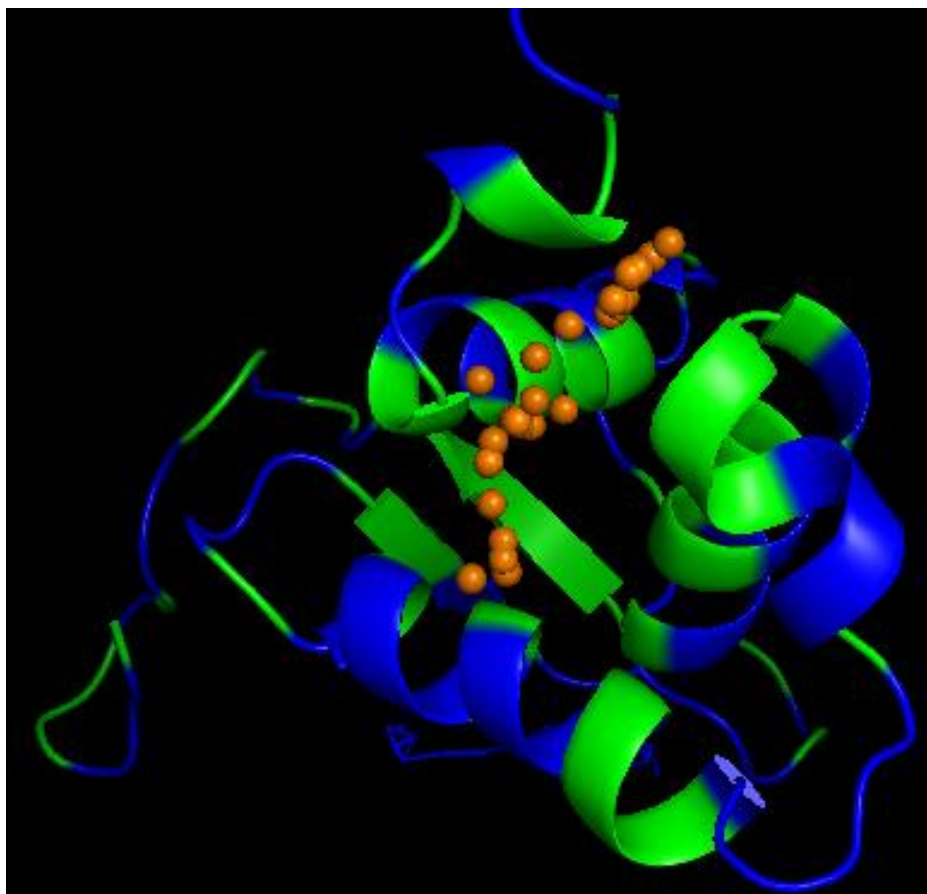


Figura 5.2.46: Representação da proteína foxO3, na estrutura 2K86, réplica 3, conformação 10. Representação da cavidade 2, na forma da proteína em superfície. A azul encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a laranja representam o local da cavidade.

Esta cavidade é uma boa candidata a alojar ligandos inibidores da ligação da proteína ao DNA, pois se o ligando for suficientemente grande vai impedir a ligação da hélice de reconhecimento ao DNA e assim deslocamos o equilíbrio da proteína ligada para a proteína desligada. Esta cavidade também está presente em outras réplicas, demonstrando reprodutibilidade do estudo. E é observada também em réplicas das trajetórias de 2k86 e 2UZK. A cavidade foi selecionada para *docking* de entre as várias observadas neste local com base nas suas propriedades comparativamente às outras conformações.

A cavidade 1 na conformação 5, da estrutura 2UZK- réplica 1, está no lado exposto da hélice 2, parecendo ser um estável candidato a uma bolsa de ligação. Esta cavidade tem uma boa pontuação (0,486), uma boa pontuação de drogabilidade (0,822) e um bom volume (370,768 Å³).

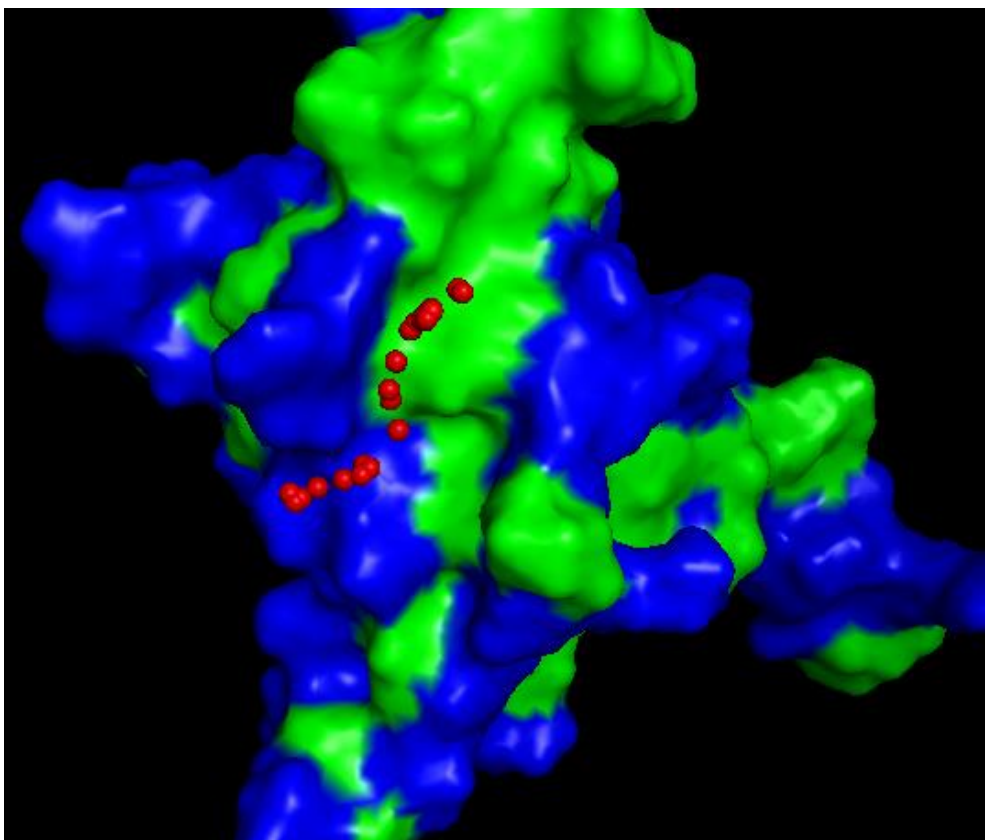


Figura 5.2.47: Representação da proteína foxO3, na estrutura 2UZK, réplica 1, conformação 5. Representação da cavidade 1. A azul encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a vermelho representam o local da cavidade.

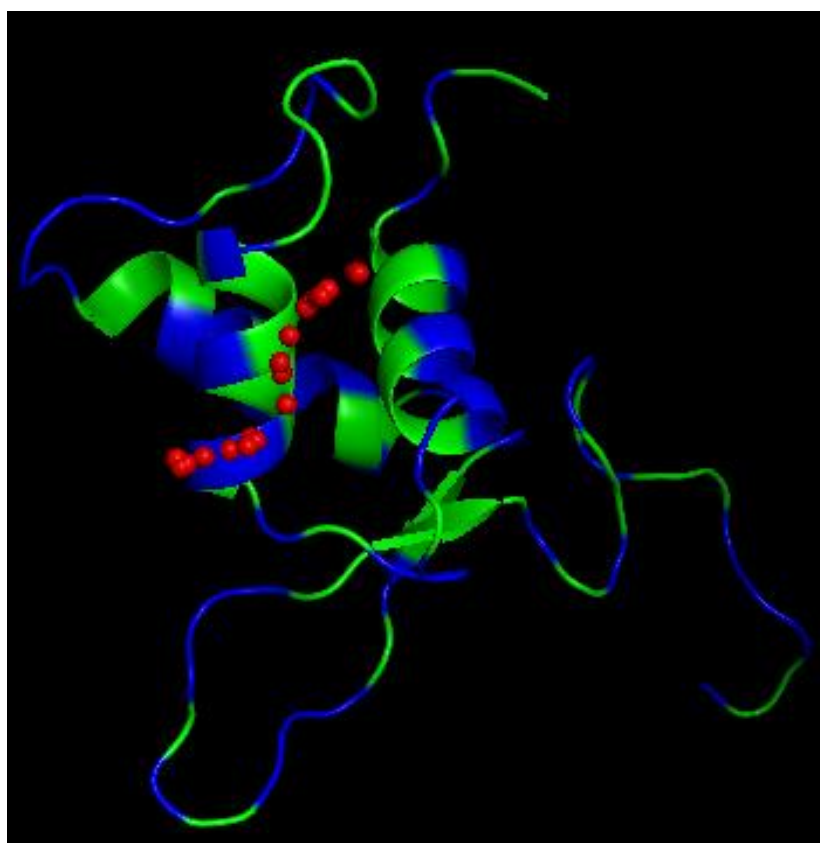


Figura 5.2.48: Representação da proteína foxO3, na estrutura 2UZK, réplica 1, conformação 5. Representação da cavidade 1. A azul encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a vermelho representam o local da cavidade.

Esta cavidade aparentemente não é constante nas outras conformações e réplicas, mas é uma cavidade interessante para o estudo.

A cavidade 1 na conformação 7 da réplica 3 da estrutura 2UZK está localizada perto da hélice 3 e o laço desordenado conectando para a hélice 2, o que pode interferir na ligação ao DNA. Esta cavidade é um bom candidato ao estudo pelo docking, pois mantém-se ao longo da trajetória como se vê nas conformações 7, 8, 11, 14, 15, 18 desta réplica. Esta cavidade tem pontuação (0,372), pontuação de drogabilidade (0,514) e volume (405,605 Å³), o que vai de encontro com o objetivo do estudo.

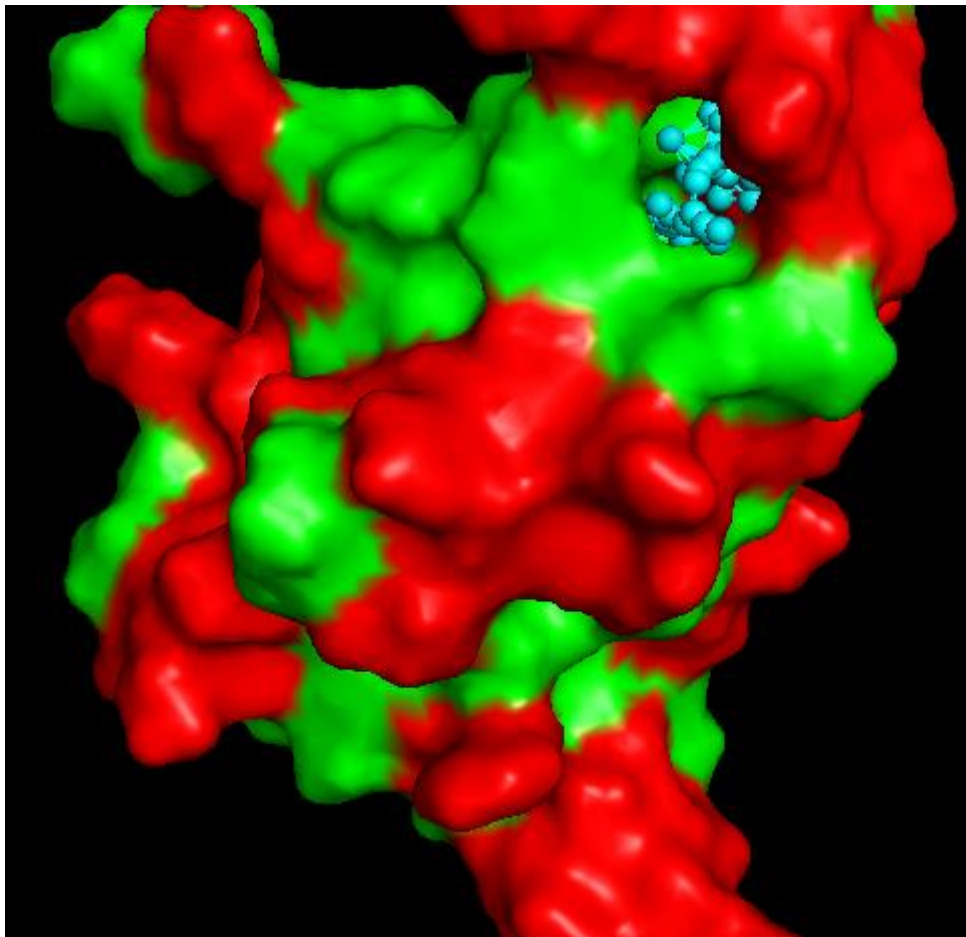


Figura 5.2.49: Representação da proteína foxO3, na estrutura 2UZK, réplica 3, conformação 7. Representação da cavidade 1. A vermelho encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a azul representam o local da cavidade.

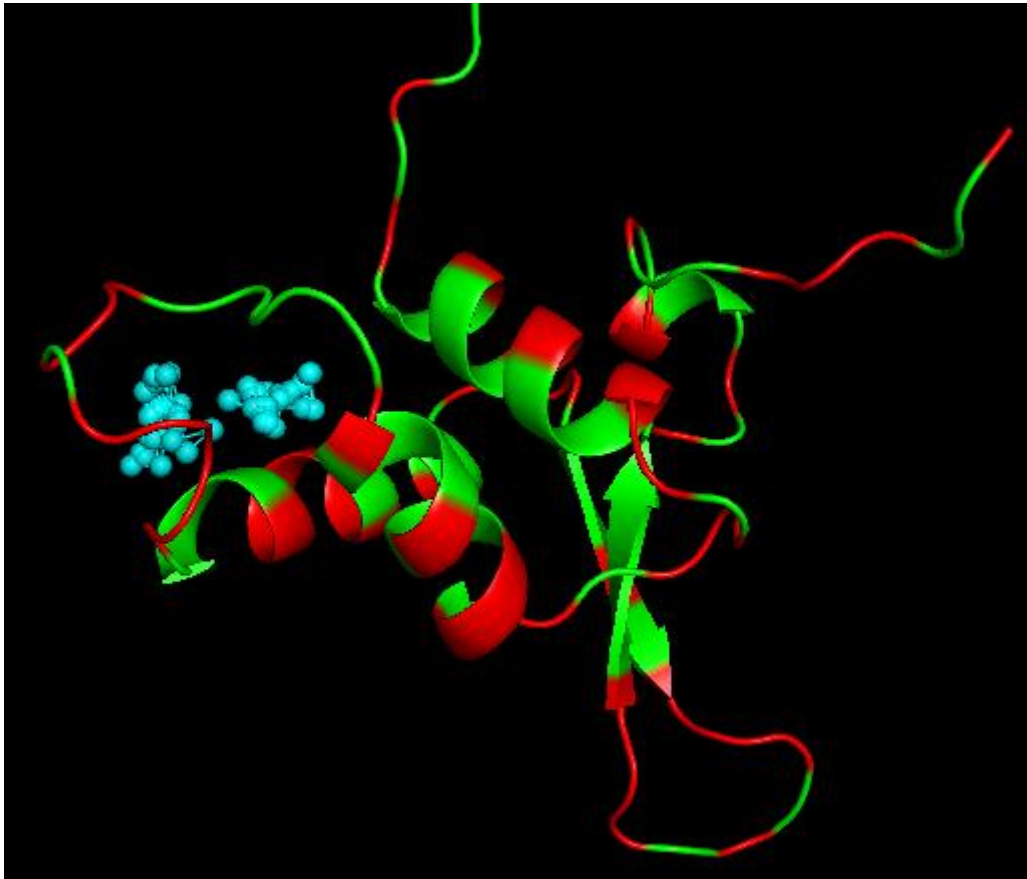


Figura 5.2.50: Representação da proteína foxO3, na estrutura 2UZK, réplica 3, conformação 7. Representação da cavidade 1. A vermelho encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a azul representam o local da cavidade.

Esta cavidade aparecia noutras réplicas e conformações, mas escolheu-se a réplica 3, conformação 7 devido às suas características.

A cavidade 2 na conformação 2 da réplica 1 do complexo (2UZK+DNA) poderá alojar ligandos que se ligam simultaneamente à proteína FoxO3 e ao DNA, localizando-se maioritariamente entre a hélice H3, o DNA e um segmento da cauda. Esta cavidade pode ser potenciadora da ligação da proteína ao DNA. Esta cavidade tem uma pontuação de 0,157, uma pontuação de drogabilidade 0,750 e um volume de 339,68 Å³.

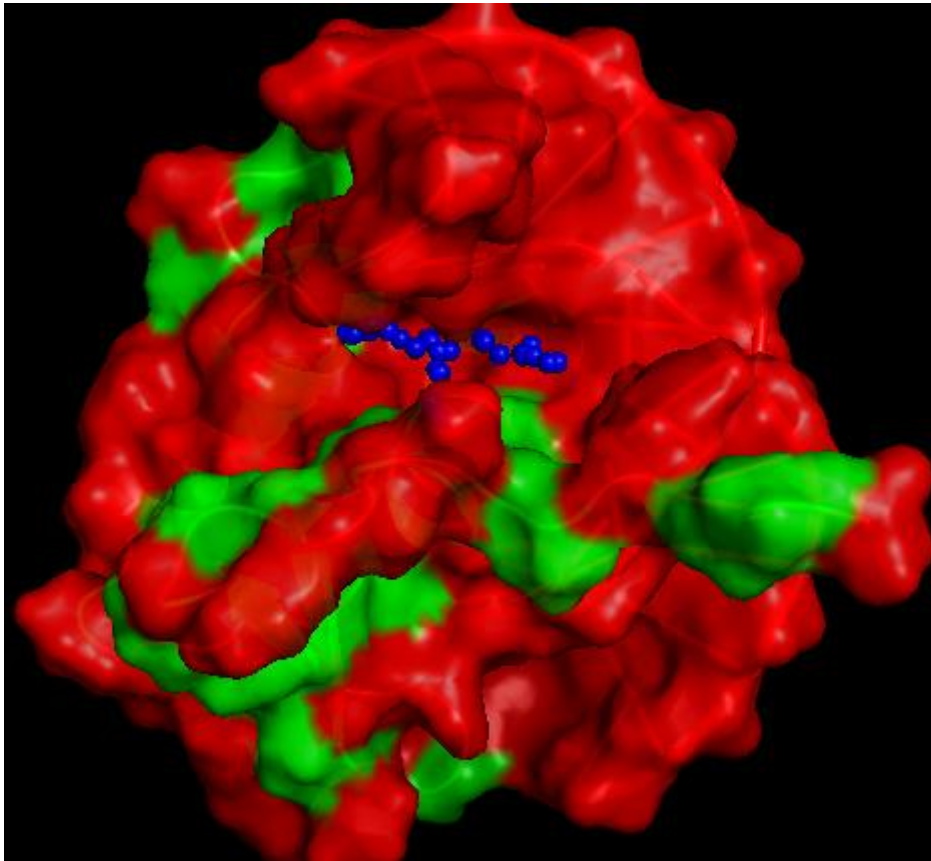


Figura 5.2.51: Representação da proteína foxO3, na estrutura do complexo, réplica 1, conformação 5. Representação da cavidade 1. A vermelho encontram-se as zonas hidrofílicas e a verde as zonas hidrofóbicas. As esferas a azul representam o local da cavidade.

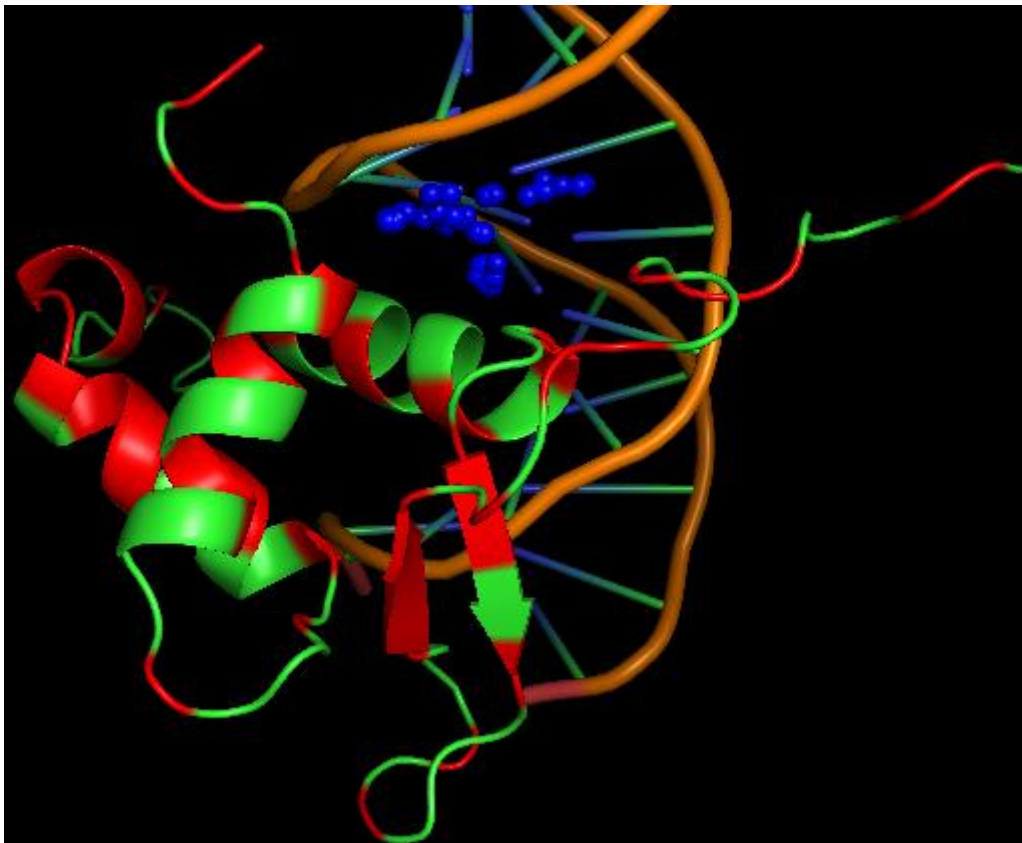


Figura 5.2.52: Representação da proteína foxo3, na estrutura do complexo, réplica 1, conformação 1. Representação da cavidade 2. A verde encontram-se as zonas hidrofílicas e a vermelho as zonas hidrofóbicas. As esferas a azul representam o local da cavidade.

As duas primeiras cavidades são mais a superfície da proteína e as duas últimas são mais internas, tanto que se vê nas imagens.

5.3 Docking Molecular

Uma coleção de 10550 moléculas pequenas foi retirada da base de dados *DRUGBANK*. Nas moléculas que constam desta base de dados inclui-se uma larga fração de compostos aprovados para uso como fármacos ou que se encontram em fase de investigação e testes. Assim, potenciais “hits” desta coleção de moléculas poderão ter uma via mais curta até à sua utilização como inibidores da FoxO3, visto já terem sido aprovados para outras indicações (estratégia conhecida como “*drug repurposing*”). Esta coleção de compostos foi utilizada num estudo de “*screening*” virtual, no qual é realizado o *docking* computacional de cada um dos ligandos às cavidades dos alvos que se pretendem estudar. Para este estudo foi usado o *software* de “*docking*” Autodock Vina. Como resultado do protocolo de triagem de alto rendimento (“*virtual high-throughput screening*”), obtém-se uma lista de compostos ordenados pela energia de interação com a cavidade, calculada pela função de energia (campo de forças) do Autodock Vina. Estes resultados foram organizados numa tabela da qual constam outras propriedades de cada composto, incluindo o código de acesso do *DRUGBANK*, a posição, a energia, a massa molecular, a carga, coeficiente de partição (logP), número de torções e número de átomos pesados.

A carga é referente ao estado de ionização predominante a pH neutro e pode ser positiva, negativa ou neutra.

A energia listada é a energia de *docking* (*binding*) calculada através da “*scoring function*” do Autodock Viva - quanto mais negativa, maior será a estabilização da molécula na cavidade. Os compostos estão ordenados na tabela por energia crescente, começando com os compostos mais estáveis.

O coeficiente de partição (logP) é a medida diferencial da solubilidade dos compostos químicos em dois solventes imiscíveis, geralmente uma fase orgânica

apolar e uma fase aquosa. Este coeficiente é calculado como o logaritmo da razão das concentrações do composto nas duas fases, geralmente octanol (solvente orgânico) e água. O $\log P$ dá-nos uma medida do carácter hidrofílico ou hidrofóbico da molécula. Um valor de $\log P$ positivo e elevado denota uma solubilidade mais alta em octanol do que em água, o que reflete o carácter lipofílico/hidrofóbico do composto, enquanto um valor negativo implica a hidrofiliicidade da molécula. Um valor igual a zero significa que a molécula é igualmente solúvel nas duas fases. Alguns estudos indicam que a absorção ótima de um medicamento ocorre para valores de $\log P$ na vizinhança de 1.8.

Da lista ordenada de “hits” (potenciais ligandos) produzida pelo *software* de *docking*, foram estudadas as 100 moléculas com energias de *docking* mais negativas (interação mais estável com a cavidade), e simultaneamente com estado de “aprovadas” (pela FDA). As moléculas aprovadas pela FDA são importantes pois estas moléculas já são fármacos aprovados e comercializados, ou seja as propriedades físico-químicas, farmacocinética (ADMET) e de segurança, já foram testadas e aprovadas e pode-se usar esses ligandos para outros objetivos terapêuticos, com mais segurança e menos investigação e investimentos iniciais. Os ligandos muito grandes são pouco específicos e seletivos para a proteína e têm piores propriedades farmacocinéticas, razão pela qual foram excluídas dos resultados moléculas com mais de 40 átomos de carbono. Também se teve em conta as moléculas que não respeitam a regra dos 5 de Lipinski. A regra de Lipinski dá uma indicação de boa biodisponibilidade oral, baseando-se em duas das propriedades mostradas na tabela 3.3.1 ($\log P$ e peso molecular), e também no número de ligações de hidrogénio que o composto pode formar como dador ou aceitador.

De acordo com a regra dos cinco, é esperada uma melhor absorção por via oral para compostos que possuam uma massa molecular inferior a 500, um valor de $\log P$ inferior a 5, sejam dadores de um máximo de 5 ligações de hidrogénio e aceptadores de um máximo 19 ligações de hidrogénio. Estas propriedades foram calculadas para os compostos estudados a partir da sua informação estrutural, usando a livreria de *Python* para *Cheminformatica RDkit*^[43]. A coleção de moléculas do *DRUGBANK* inclui moléculas “aprovadas”, “em investigação” e “experimentais”. As moléculas experimentais encontram-se em estudo e já demonstraram ser promissoras, mas

ainda não entraram na fase de ensaios clínicos ^[37]. As moléculas em investigação são moléculas promissoras e que já se encontram na fase de ensaios clínicos, e por último as moléculas aprovadas são aquelas que, após passarem pela fase de estudos experimentais e ensaios clínicos e tendo sido comprovado o seu efeito terapêutico, receberam a aprovação pela FDA (ou outra entidade reguladora) para a comercialização e já se encontram no mercado ^{[37][37]}.

Assim sendo os critérios de exclusão foram:

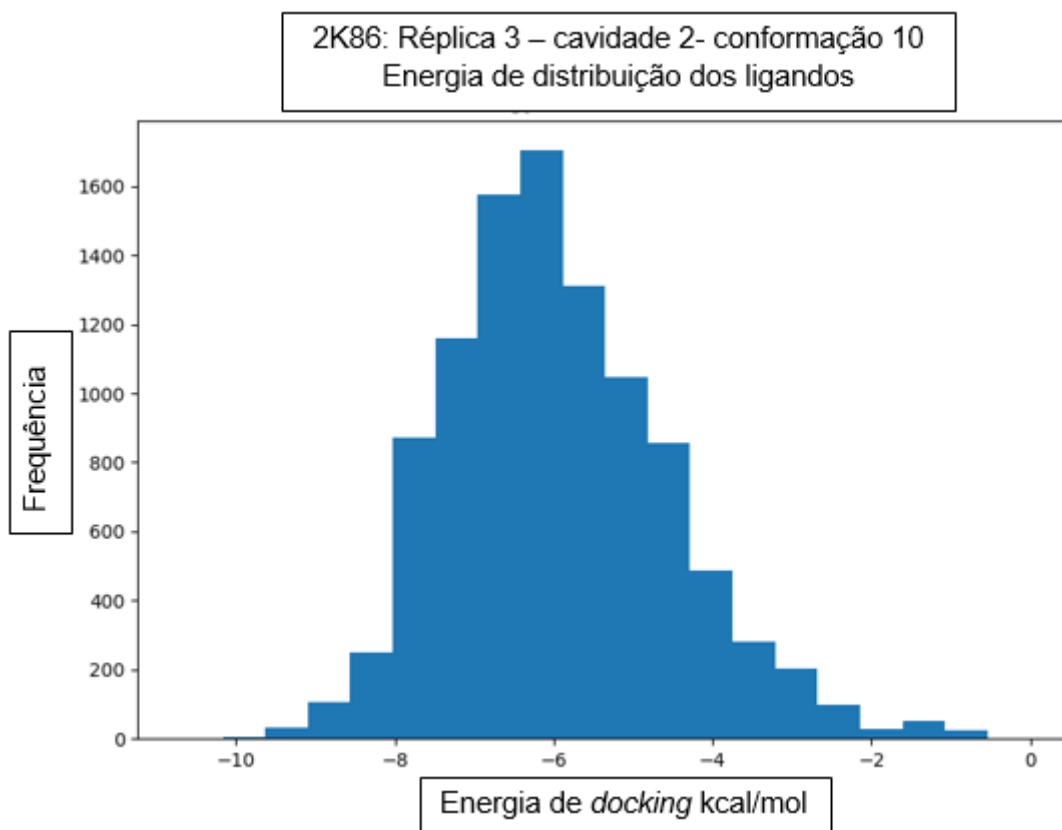
- Moléculas com mais de 40 átomos de carbono
- Moléculas que ainda não foram aprovadas

Em todas as réplicas e conformações selecionadas olhou-se para os 100 primeiros ligandos aprovados e após esta seleção descartaram-se os ligandos que não seguiam a regra dos 5 de Lipinski.

1. Estrutura 2K86 - Réplica 3 - Cavity 2 - Conformação 10

Das 10550 moléculas usadas para o *docking*, apenas 10097 moléculas ligaram-se à proteína FoxO3. Dessas 10097 moléculas verifica-se uma variação da energia de ligação desde o valor zero até ao valor -10,7 kcal/mol. Na Figura 5.3.53 pode ver-se um histograma da distribuição de energia dos 10097 “hits”, estando o máximo da curva em cerca de -6 kcal/mol.

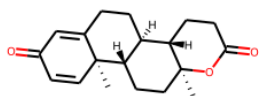
As primeiras 100 moléculas da lista de resultados foram analisadas manualmente, e selecionaram-se as 40 moléculas que respeitam a regra dos 5. Na tabela 5.3.15 encontram-se os melhores 10 ligandos e as suas características físicas químicas, estando representadas na figura 5.3.54 as estruturas químicas das mesmas. Na figura 5.3.55. representa-se a estrutura do complexo FoxO3-DB00894, o ligando com o melhor score (mais negativo) da lista.



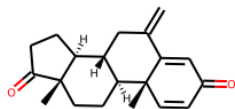
5.3.53: Histograma da distribuição das energias do docking para a cavidade 2 da conformação 10 da réplica 3 da estrutura 2k86.

Tabela 5.3.15: Características dos melhores 10 ligandos.

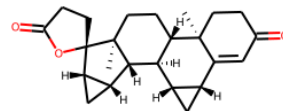
| Ligando | ID | Energia (kcal/mol) | Massa molecular | Carga | logP | Torção | Átomos pesados |
|---------------|---------|--------------------|-----------------|-------------------------|------|--------|----------------|
| Testolactona | DB00894 | -9.2 | 300.397 | -2.08×10^{-16} | 3.58 | 0 | 22 |
| Exemestano | DB00990 | -9.0 | 296.409 | -9.71×10^{-17} | 4.02 | 0 | 22 |
| Lurasidone | DB08815 | -8.9 | 493.696 | 0.99 | 3.67 | 4 | 35 |
| Eltrombopag | DB06210 | -8.9 | 441.467 | -0.99 | 1.38 | 5 | 33 |
| Drospirenona | DB01395 | -8.9 | 366.500 | -2.01×10^{-16} | 4.30 | 0 | 27 |
| Fluocortolone | DB08971 | -8.9 | 376.467 | -2.77×10^{-16} | 2.63 | 4 | 27 |
| Boldenone | DB01541 | -8.8 | 286.414 | 4.37×10^{-16} | 3.65 | 1 | 21 |
| Netarsudil | DB13931 | -8.7 | 453.541 | 9.43×10^{-16} | 4.90 | 10 | 34 |
| Vibegron | DB14895 | -8.7 | 444.534 | 5.55×10^{-17} | 2.77 | 8 | 33 |
| Canrenoic | DB09015 | -8.7 | 357.469 | -1 | 2.55 | 4 | 26 |



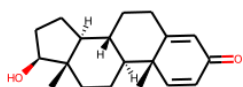
Testolactone (DB00894)



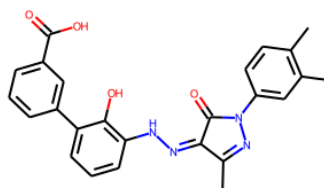
Exemestane (DB00990)



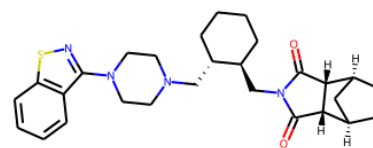
Drospirenone (DB01395)



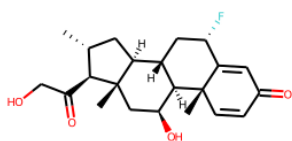
Boldenone (DB01541)



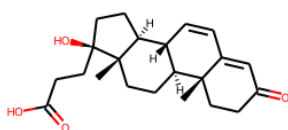
Eltrombopag (DB06210)



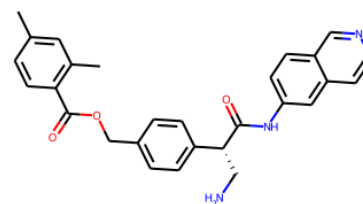
Lurasidone (DB08815)



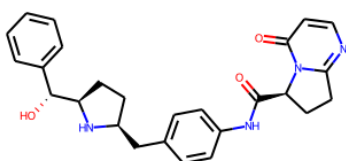
Fluocortolone (DB08971)



Canrenoic acid (DB09015)

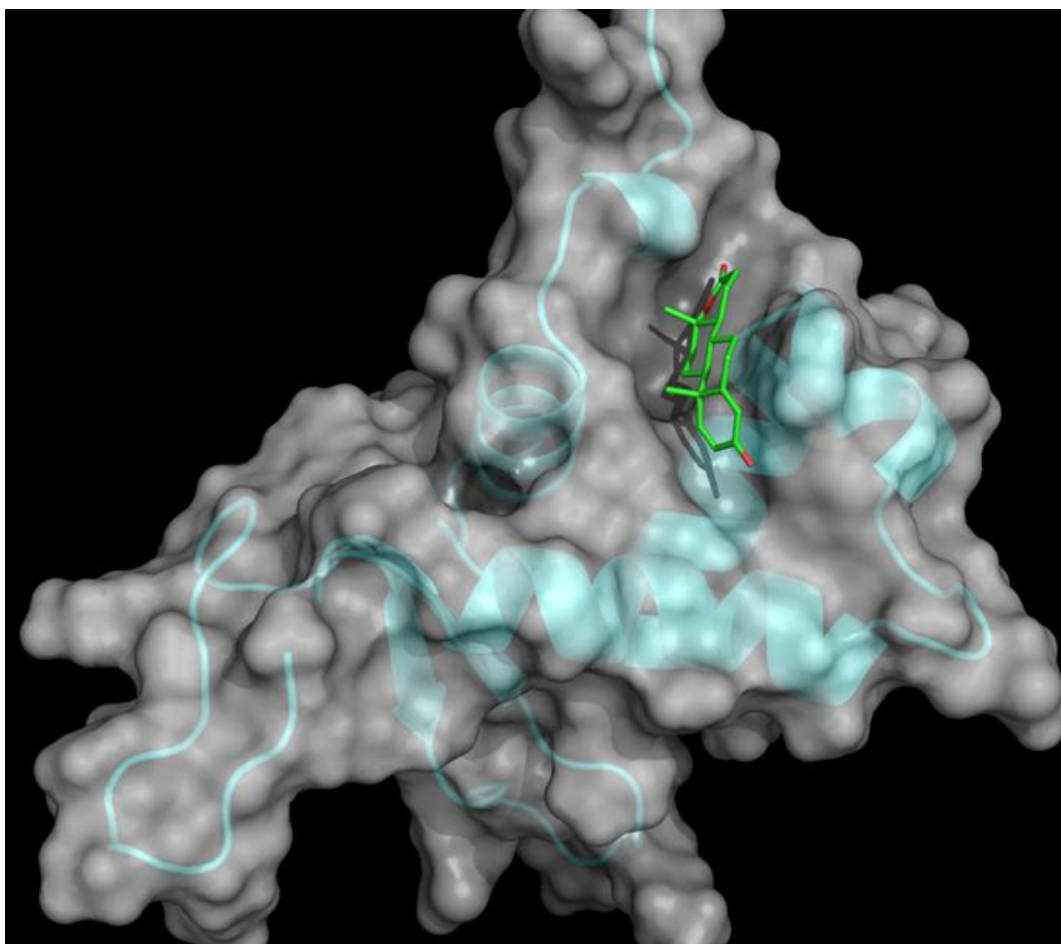


Netarsudil (DB13931)



Vibegron (DB14895)

5.3.54: Estruturas químicas dos melhores 10 ligandos para a cavidade 2 da conformação 10 da réplica 3 da estrutura 2k86.

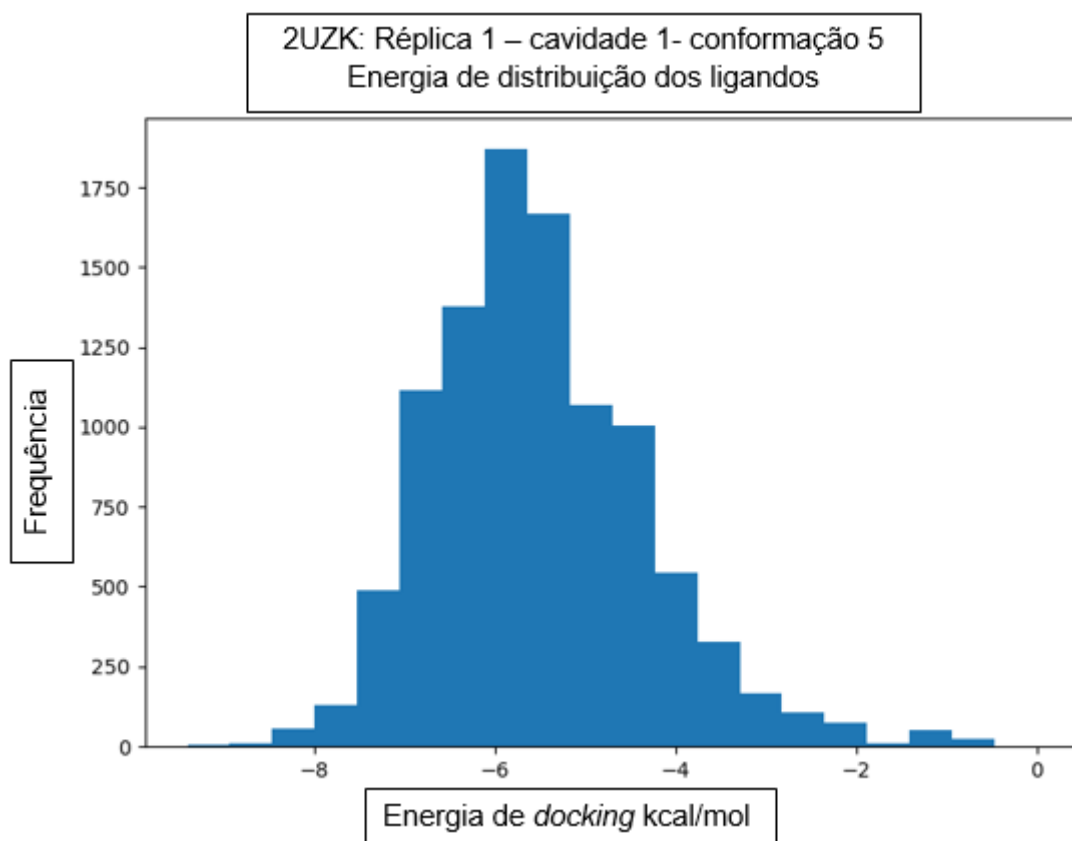


5.3.55: Interação do melhor ligando com a proteína. Sendo o ligando a Testolactona e a proteína FoxO3, estrutura 2K86.

2. Estrutura 2UZK - Réplica 1 - Caverna 1 - Conformação 5

Das 10097 moléculas que se ligaram à proteína, verifica-se uma variação da energia de ligação desde o valor zero até ao valor -9,4 kcal/mol.

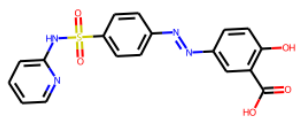
Após a análise manual das primeiras 100 moléculas dos resultados e exclusão das que não respeitam a regra dos 5, ficam apenas 32 moléculas. Na tabela 5.3.16 encontram-se os melhores 10 ligandos e as suas características físicas químicas, estando representadas na figura 5.3.56 o histograma da distribuição das energias de *docking*, na figura 5.3.57 as estruturas físicas e químicas das mesmas. Na figura 5.3.58 representa-se a estrutura do complexo FoxO3- DB08815, o ligando com o melhor score (mais negativo) da lista.



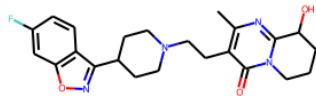
5.3.56: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 5 da réplica 1 da estrutura 2UZK.

Tabela 5.3.16: Características dos melhores 10 ligandos

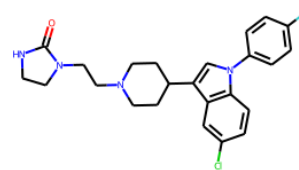
| Ligando | ID | Energia (kcal/mol) | Massa molecular | Carga | logP | Torções | Átomos pesados |
|---------------|---------|--------------------|-----------------|------------------------|------|---------|----------------|
| Lurasidona | DB08815 | -8 | 493.696 | 0.99 | 3.67 | 4 | 35 |
| Grapiprant | DB12836 | -7.7 | 491.616 | 4.16×10^{-17} | 4.34 | 8 | 35 |
| Lumacaftor | DB09280 | -7.7 | 451.404 | 1 | 3.41 | 5 | 33 |
| Tepotinib | DB15133 | -7.7 | 492.582 | 5.82×10^{-16} | 3.41 | 3 | 33 |
| Risdiplam | DB15305 | -7.6 | 402.481 | 1 | 1.38 | 2 | 30 |
| Eltrombopag | DB06210 | -7.6 | 441.467 | 0.99 | 1.38 | 5 | 33 |
| Paliperidona | DB01267 | -7.5 | 426.491 | 8.32×10^{-17} | 3.08 | 5 | 31 |
| Sulfasalazina | DB00795 | -7.5 | 396.384 | -2 | 1.73 | 6 | 28 |
| Larotrectinib | DB14723 | -7.5 | 428.442 | 0 | 2.94 | 3 | 31 |
| Sertindole | DB06144 | -7.5 | 440.949 | 1.38×10^{-16} | 4.62 | 5 | 31 |



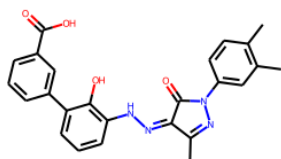
Sulfasalazine (DB00795)



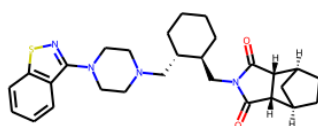
Paliperidone (DB01267)



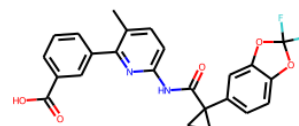
Sertindole (DB06144)



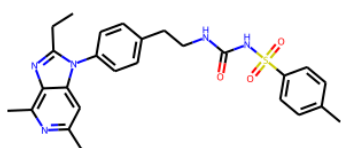
Etrambopag (DB06210)



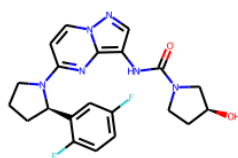
Lurasidone (DB08815)



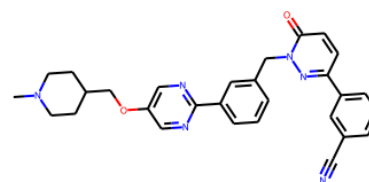
Lumacaftor (DB09280)



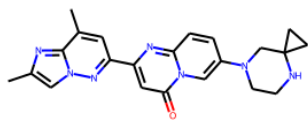
Grapiprant (DB12836)



Larotrectinib (DB14723)

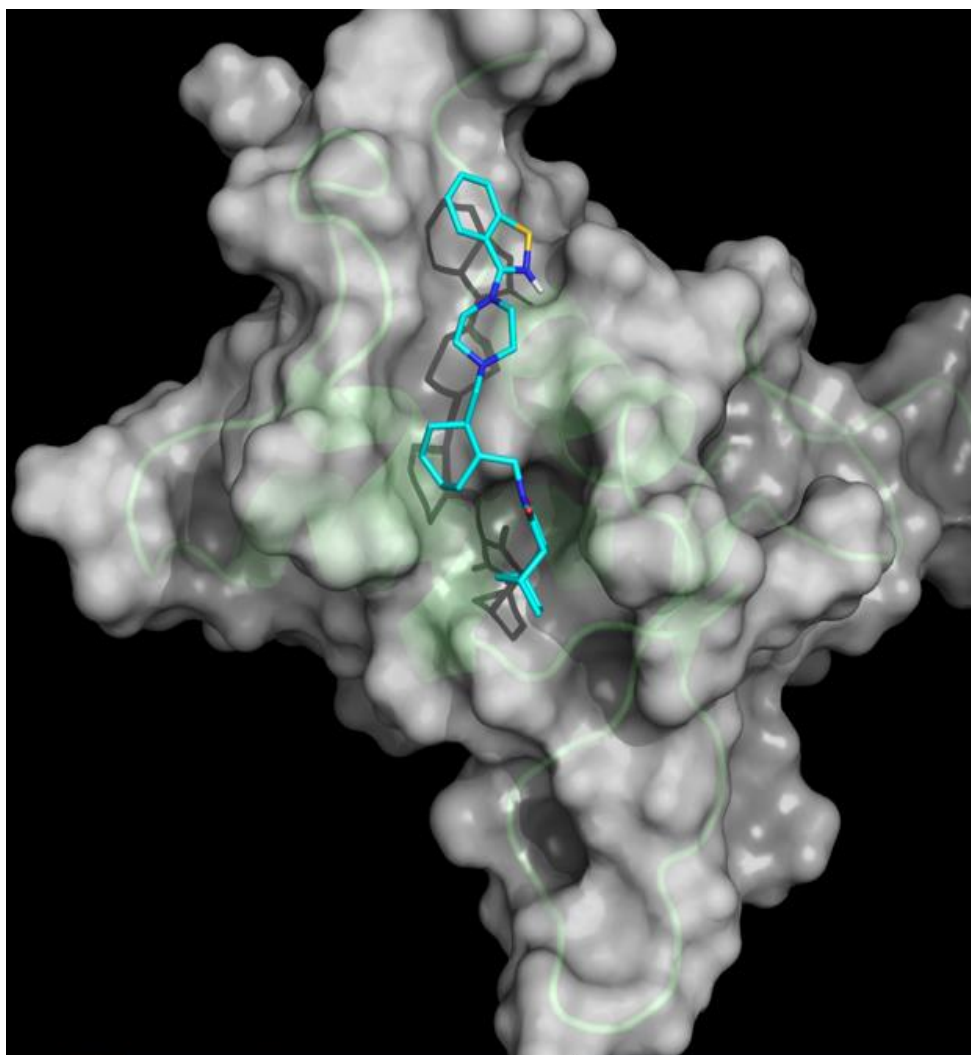


Tepotinib (DB15133)



Risdiplam (DB15305)

5.3.57: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 5 da réplica 1 da estrutura 2UZK.

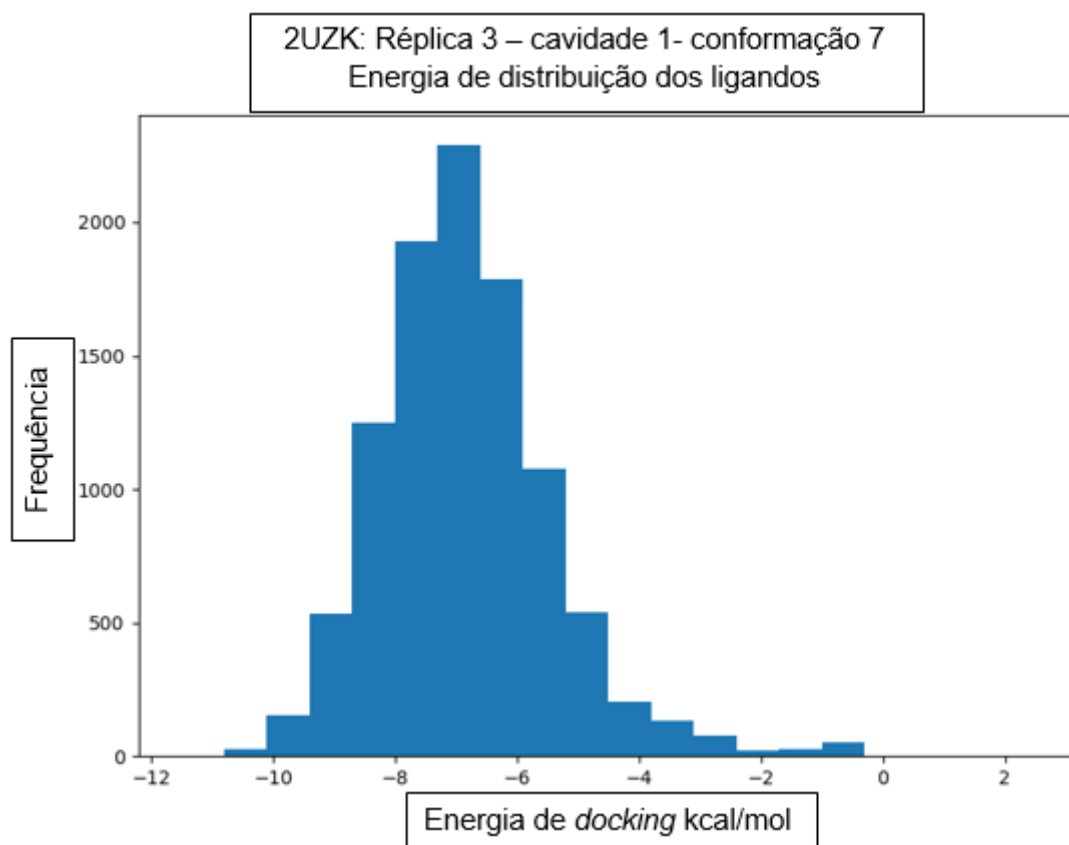


5.3.58: Interação do melhor ligando com a proteína. Sendo o ligando a Lurasidona e a proteína FoxO3, estrutura 2UZK.

3. 2UZK - réplica 3 - Cavidade 1 - Conformação 7

Nesta estrutura obteve-se um total de 10105 moléculas ligadas à proteína FoxO3, com uma variação da energia de ligação desde o valor 2,5 até ao valor -11,5 kcal/mol. Obteve-se um total de 56 moléculas das 100 estudadas.

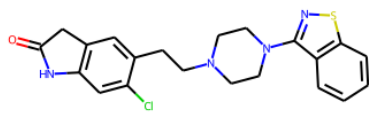
Na Figura 5.3.59 pode ver-se um histograma da distribuição de energia dos “hits”, estando o máximo da curva em cerca de -8 kcal/mol. Na tabela 5.3.17 encontram-se os melhores 10 ligandos e as suas características físicas químicas, estando representadas na figura 5.3.60 as estruturas físicas e químicas das mesmas. Na figura 5.3.61 representa-se a estrutura do complexo FoxO3-DB09280, o ligando com o melhor score (mais negativo) da lista.



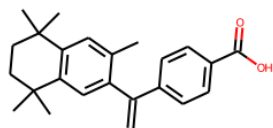
5.3.59: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 7 da réplica 3 da estrutura 2UZK.

Tabela 5.3.17: Características dos melhores 10 ligandos

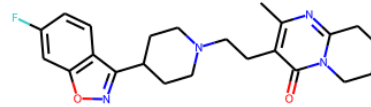
| Ligando | ID | Energia (kcal/mol) | Massa Molecular | Carga | logP | Torções | Átomos pesados |
|--------------|----------|--------------------|-----------------|-------------------------|------|---------|----------------|
| Lumacaftor | DB09 280 | -10.8 | 451.404 | -1 | 3.41 | 5 | 33 |
| Oxazepam | DB00 842 | -10.3 | 286.717 | 4.99×10^{-16} | 2.44 | 2 | 20 |
| Ziprasidone | DB00 246 | -10.2 | 412.945 | -7.21×10^{-16} | 3.80 | 3 | 28 |
| Bexaroteno | DB00 307 | -10.1 | 347.477 | -0.99 | 4.76 | 3 | 26 |
| Flunarizine | DB04 841 | -9.9 | 405.511 | 1 | 3.96 | 6 | 30 |
| Paliperidona | DB01 267 | -9.9 | 426.491 | 8.31×10^{-17} | 3.08 | 5 | 31 |
| Risperidona | DB00 734 | -9.9 | 411.500 | 0.99 | 3.00 | 4 | 30 |
| Bictegravir | DB11 799 | -9.9 | 447.368 | -1.99 | 1.78 | 4 | 32 |
| Olaparib | DB09 074 | -9.7 | 434.470 | -1.66×10^{-16} | 0.55 | 5 | 32 |
| Tecovirimat | DB12 020 | -9.7 | 375.325 | -1 | 3.18 | 4 | 27 |



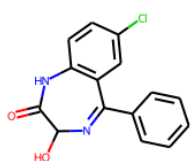
Ziprasidone (DB00246)



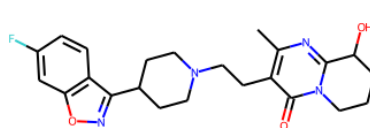
Bexarotene (DB00307)



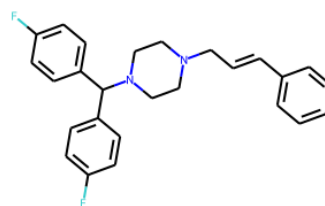
Risperidone (DB00734)



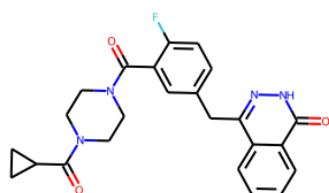
Oxazepam (DB00842)



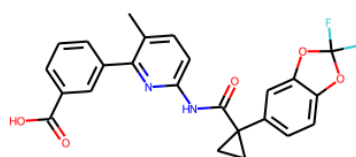
Paliperidone (DB01267)



Flunarizine (DB04841)



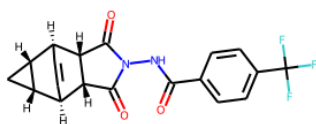
Olaparib (DB09074)



Lumacaftor (DB09280)

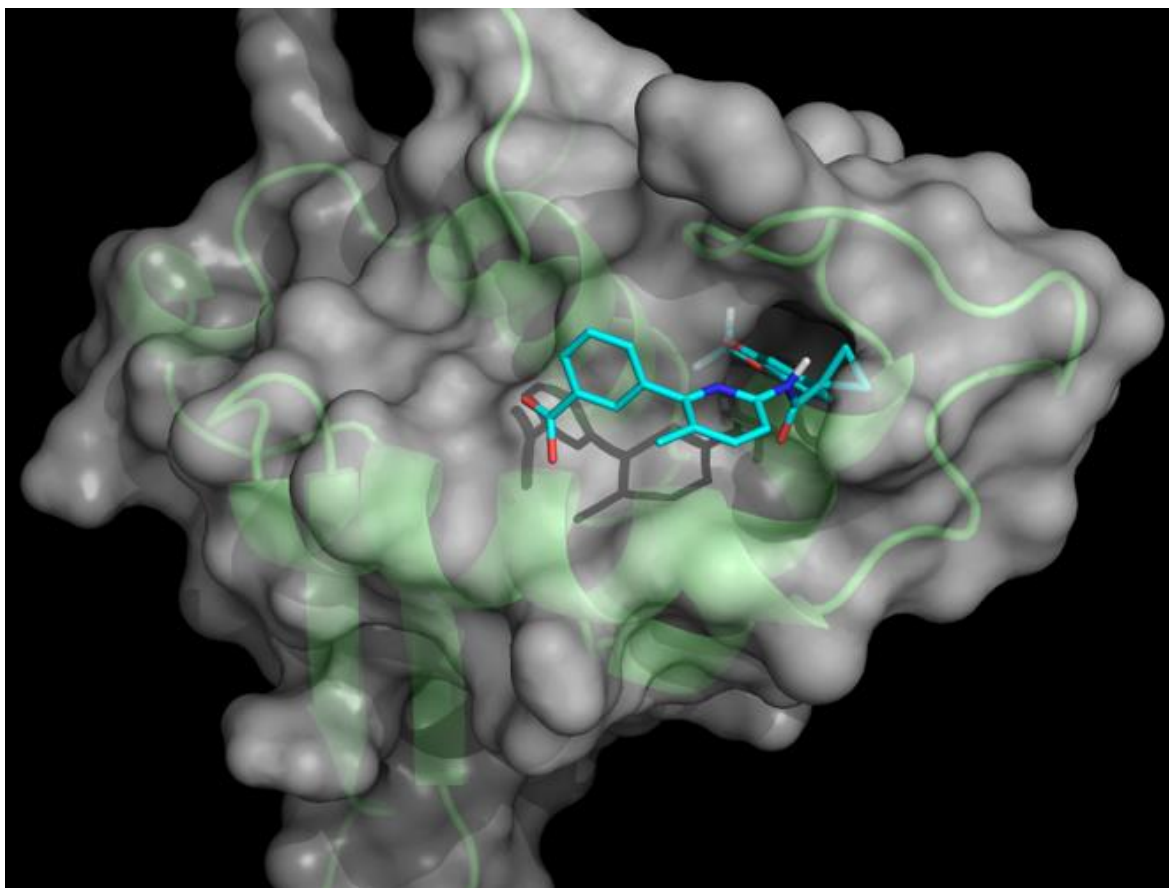


Bictegavir (DB11799)



Tecovirimat (DB12020)

5.3.60: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 7 da réplica 3 da estrutura 2UZK.

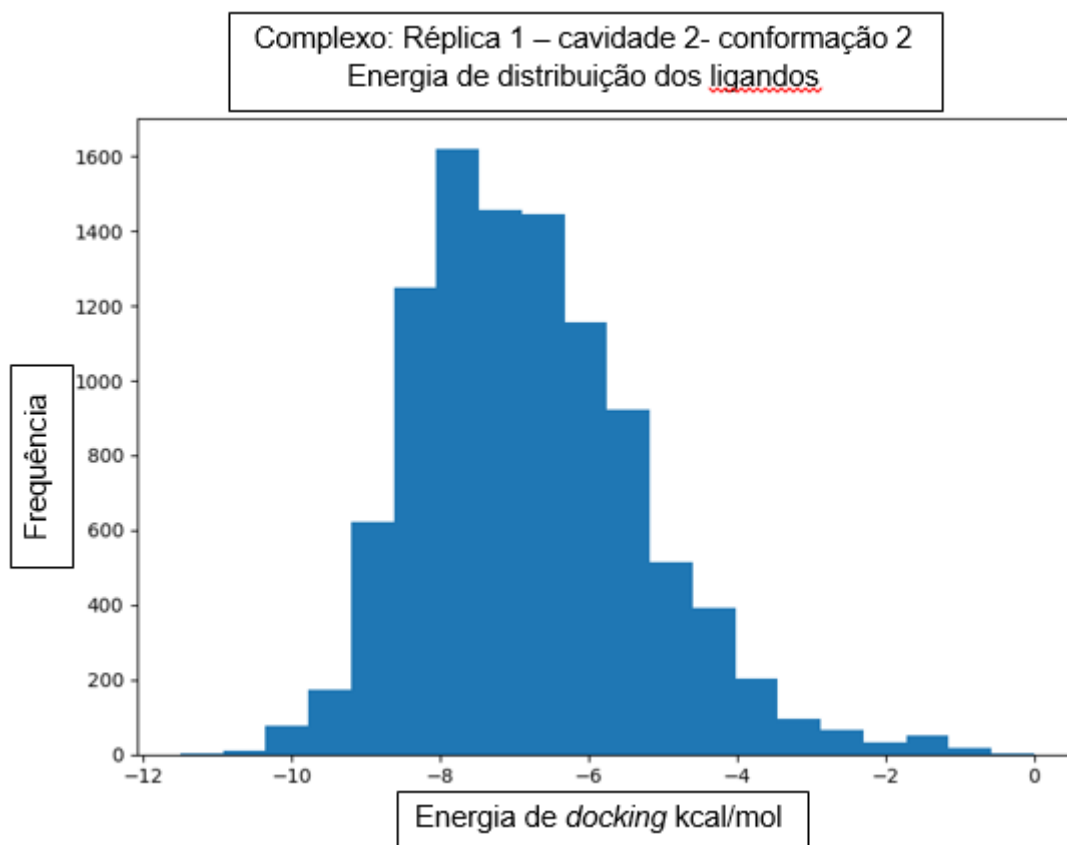


5.3.61: Interação do melhor ligando com a proteína. Sendo o ligando Lumacaftor e a proteína FoxO3, estrutura 2UZK.

4. Complexo réplica 1 cavidade 2 conformação 2

Das 10097 moléculas que se ligaram à proteína, verifica-se uma variação da energia de ligação desde o valor zero até ao valor -11,5 kcal/mol. Neste caso o histograma da distribuição de scores tem o seu máximo um pouco mais abaixo, num score de aproximadamente -8.0 kcal/mol, diminuindo a distância aos hits de menor energia.

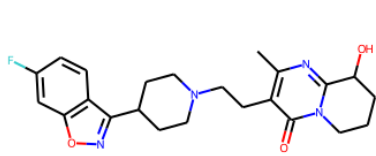
Totalizando 31 moléculas finais seleccionadas, tendo em conta os parâmetros acima descritos nas outras cavidades. Na figura 5.3.62 esta representado o histograma da distribuição das energias do *docking* molecular. Na tabela 5.3.18 encontram-se os melhores 10 ligandos e as suas características físicas químicas, estando representadas na figura 5.3.63 as estruturas físicas e químicas das mesmas, na figura 5.3.64 a estrutura do complexo FoxO3- DB15305, o ligando com o melhor score (mais negativo) da lista.



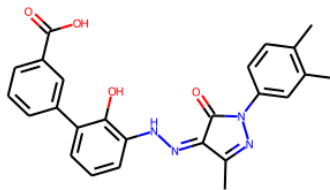
5.3.62: Histograma da distribuição das energias do docking para a cavidade 1 da conformação 2 da réplica 1 da estrutura do complexo.

Tabela 5.3.18: Características dos melhores 10 ligandos

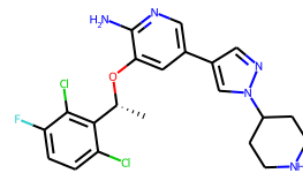
| Ligando | ID | Energia (kcal/mol) | Massa molecular | Carga | logP | Torção | Átomos pesados |
|--------------|----------|--------------------|-----------------|-------------------------|------|--------|----------------|
| Risdiplam | DB15305 | -9.9 | 402.281 | 1 | 1.38 | 2 | 30 |
| Avapritinib | DB15233 | -9.9 | 498.569 | 6.25×10^{-16} | 2.60 | 4 | 37 |
| Lumacaftor | DB09280 | -9.6 | 451.404 | -1 | 3.41 | 5 | 33 |
| Eltrombopag | DB06210 | -9.4 | 441.467 | -0.99 | 1.38 | 5 | 33 |
| Capmatinib | DB1760 | -9.3 | 412.427 | 4.30×10^{-16} | 3.63 | 5 | 31 |
| Talazoparib | DB11760 | -9.3 | 380.357 | -7.63×10^{-16} | 1.67 | 2 | 28 |
| Dasabuvir | DB09183 | -9.3 | 491.568 | 2 | 4.24 | 6 | 35 |
| Paliperidona | DB01267 | -9.3 | 426.491 | 8.23×10^{-17} | 3.08 | 5 | 31 |
| Crizotinib | DB008865 | -9.2 | 451.352 | 1 | 4.45 | 5 | 30 |
| Tecovirimat | DB12020 | -9.2 | 375.325 | -1 | 3.18 | 4 | 27 |



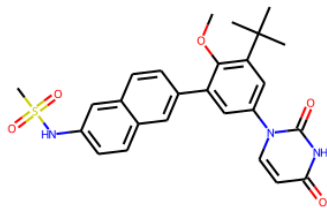
Paliperidone (DB01267)



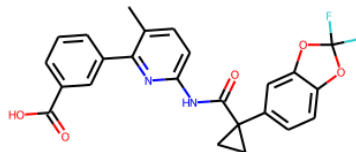
Eltrombopag (DB06210)



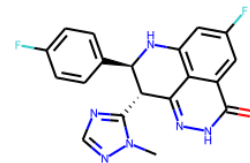
Crizotinib (DB08865)



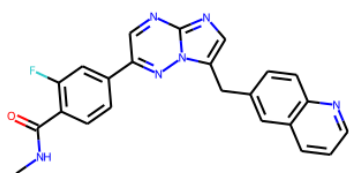
Dasabuvir (DB09183)



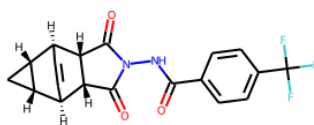
Lumacaftor (DB09280)



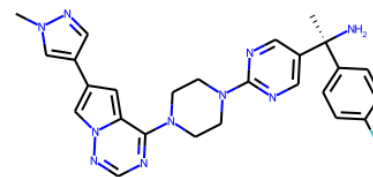
Talazoparib (DB11760)



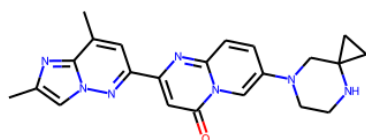
Capmatinib (DB11791)



Tecovirimat (DB12020)

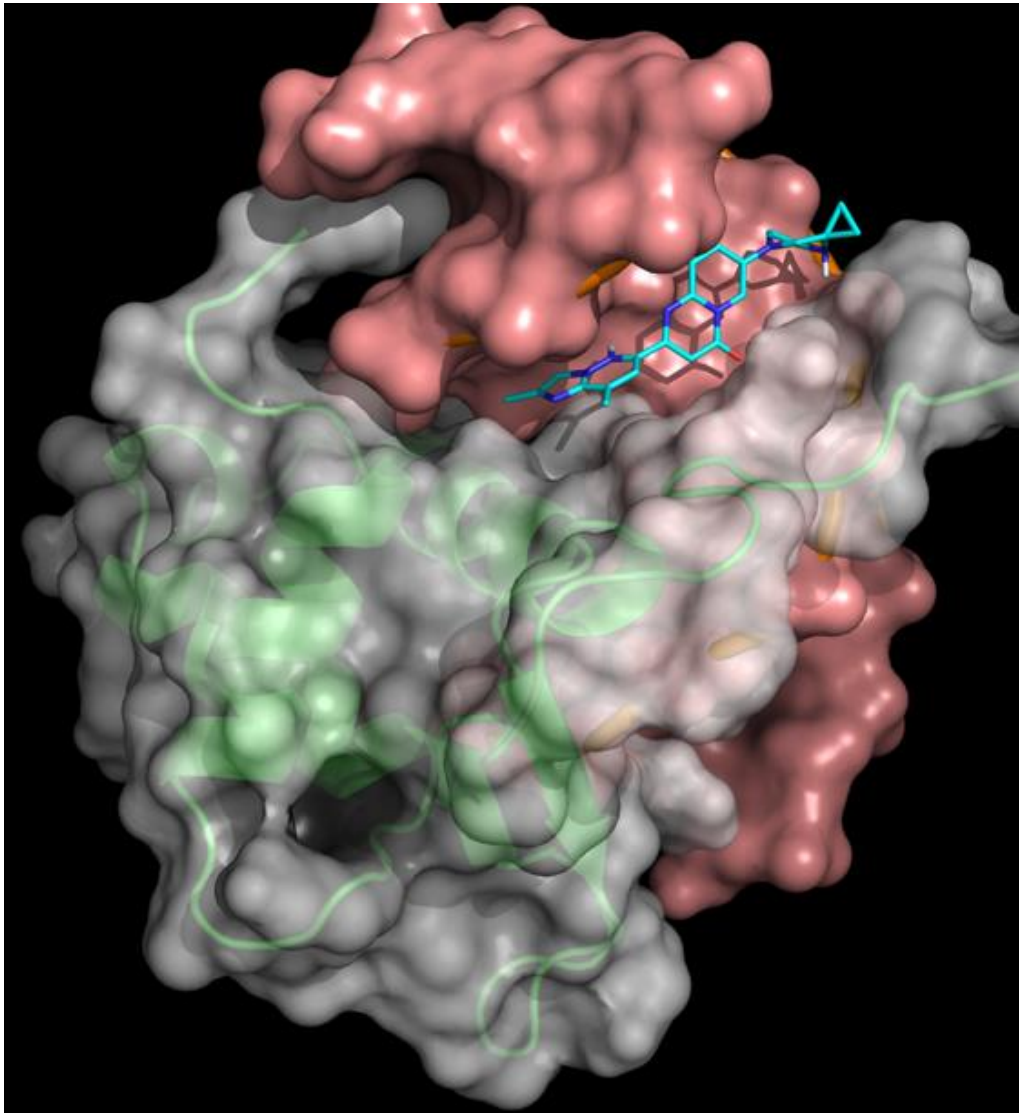


Avapritinib (DB15233)



Risdiplam (DB15305)

5.3.63: Estruturas químicas dos melhores 10 ligandos para a cavidade 1 da conformação 2 da réplica 1 da estrutura do complexo.



5.3.64: Interação do melhor ligando com a proteína. Sendo o ligando Risperidone e a proteína FoxO3, estruturado complexo.

Os histogramas nas figuras 5.3.53/56/59/62 representam a distribuição das energias para os dockings das diversas estruturas e cavidades. Os valores próximos do mínimo destacam-se da média das energias, tendo deste modo o maior potencial para serem significativos. Por este, motivo olhou-se para os valores extremos ou seja, os mais negativos.

Como se pode ver nas tabelas 5.3.15 ao 5.3.18 alguns ligando repetem-se nas diferentes cavidades, como por exemplo a paliperidona, o eltrombopag e Lurasidona. Naturalmente, estes ligandos são previsivelmente menos específicos, pelo que seria desejável focar-nos em moléculas que são únicas em cada lista. Note-se, no entanto,

que algumas das moléculas repetidas apresentam variações de energia de ligação muito grandes de cavidade para cavidade.

Selecionado um pequeno conjunto de compostos para cada cavidade, o próximo passo poderia ser testar experimentalmente a afinidade de ligação destes para com o domínio FoxO3. Uma das técnicas que poderia ser utilizadas para este propósito seria um ensaio baseado em fluorescência na presença e na ausência do DNA. Só assim se têm a percepção de que como se ligaria o ligando (potenciador ou inibidor da ligação DNA-FoxO3 e se teria uma boa especificidade para a proteína).

6. Conclusão

O presente estudo permitiu fazer uma análise comparativa das estruturas 2k86 e 2uzk da proteína FoxO3a. Foram comparadas as propriedades dinâmicas e conformacionais das proteínas livres e ligadas ao DNA por meio de simulações de dinâmica molecular, e as conformações extraídas da DM foram usadas para identificar potenciais cavidades capazes de ligar moléculas pequenas. Finalmente as cavidades assim identificadas foram usadas num estudo de *screening* virtual com vista à identificação de potenciais moléculas pequenas capazes de interagir com a FoxO3a e modular a sua capacidade de ligação ao DNA. A comparação da DM evidenciou a maior estabilidade da proteína em complexo com o DNA, o que aponta para a possibilidade de existirem diferenças conformacionais capazes de serem exploradas no sentido de encontrar compostos capazes de se ligar a uma ou outra forma.

Neste trabalho foram produzidas 4 réplicas de cada simulação de dinâmica molecular, o que permitiu avaliar a reprodutibilidade e a constância das características estruturais e dinâmicas dos vários modelos no decurso das simulações. Os resultados obtidos oferecem suporte ao realismo físico e fiabilidade das estruturas simuladas.

Também foram encontradas cerca de 20 cavidades com potencial interessante para modular a atividade da proteína FoxO3, embora o estudo tenha sido focado em apenas 4 cavidades. A análise de outras cavidades poderia fornecer informações adicionais importantes para a modelação da proteína.

Nas corridas de *docking* molecular (*virtual screenig*) utilizou-se uma coleção alargada de compostos extraídos da base de dados *Drugbank*, com vista à

identificação de potenciais ligandos, nas quais detetou-se ligandos exclusivos para o complexo e para a proteína livre. Como por exemplo no caso do complexo: avaprinib, capmatinib e talazoparib. Descobriram-se também ligandos iguais para diferentes conformações da proteína, o que levaria primeiramente a estudar os ligandos exclusivos para a conformação livre e ligada ao DNA, pois os ligandos que se ligam a múltiplas cavidades são previsivelmente menos específicos na sua interação com a proteína.

Embora este estudo tenha contribuído para um melhor conhecimento da estrutura e interação da FoxO3a com potenciais ligandos, existem vários aspetos importantes que seria necessário abordar para se conseguir um trabalho completo:

Comparação dos resultados obtidos com os de outras ferramentas de docking, de forma a obter resultados de consenso que tornariam o estudo mais fiável

Uso de outras bases de dados com distintas e mais alargadas coleções de moléculas ligandos. (mais diversidade no estudo)

Uso e comparação dos resultados de diferentes ferramentas de deteção de cavidades de forma a ter indicações mais fiáveis sobre as cavidades realmente significativas e que mereceriam ser exploradas.

Determinação experimental da afinidade de ligação ao domínio FoxO3 de pelo menos alguns dos compostos identificados no *screening virtual*. Uma das técnicas que poderia ser utilizadas para este efeito seria um ensaio baseado em fluorescência na presença e na ausência do DNA. A validação experimental é essencial para qualquer estudo de docking computacional.

Realização de simulações de dinâmica molecular com a FoxO3 em complexo com os melhores ligandos com intuito de avaliar a estabilidade das conformações previstas pelo docking.

Cálculo da energia livre de *binding* dos melhores complexos pelo método MM-PBSA (dinâmica molecular combinada com cálculos de electrostática contínua)

Cálculo e re-ordenação (*re-scoreing*) das energias de ligação dos melhores ligandos com funções de energia diferentes da função usada pelo *software* Autodock Vina.

Apesar das muitas possibilidades de estudos adicionais com a FoxO3, o presente trabalho já contribuiu para o avanço do conhecimento sobre o comportamento da proteína FoxO3 no organismo humano e obter algumas pistas para a modulação da sua atividade.

Assim sendo, pode concluir-se que se cumpriu parcialmente o objetivo do trabalho, pois foram encontrados vários ligandos para a possível modulação da proteína FoxO3. Faltou o teste experimental da capacidade de modulação dos melhores compostos como forma de validação final do trabalho.

7. Bibliografia e Webgrafia

- [1] Weigel D, Jurgens G, Kuttner F, Seifert E and Jackie H: The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell*, 1989, 57(4), 645-658.
- [2] Lai E, Prezioso VR, Smith E, Litvin O, Costa RH, Darnell JE. HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev*. 1990 Aug;4(8):1427-36.
- [3] Lehmann OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS. Fox's in development and disease. *Trends Genet* 2003 Jun;19(6):339-44.
- [4] Wang F, Marshall CB, Yamamoto K, Li GY, Plevin MJ, You H, et al. Biochemical and Structural Characterization of an Intramolecular Interaction in FOXO3a and Its Binding with p53. *J Mol Biol*. 2008 Dec 19;384(3):590-603.
- [5] Katoh M, Katoh M. Human Fox Gene Family (Review). *Int J Oncol*. 2004 Nov;25(5):1495-500.
- [6] Myatt SS, Lam EWF. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer*. 2007 Nov 7(11):847-859.
- [7] Hornsveld M, Smits LMM, Meerlo M, Van Amersfoort M, Koerkamp MJAG, Van Leenen D, et al. Tumor Biology and Immunology FOXO Transcription Factors Both Suppress and Support Breast Cancer Progression. *Cancer Res*. 2018 May 1;78(9):2356-2369.
- [8] Calissi G, Lam EW, Link W. Therapeutic strategies targeting FOXO transcription factors. *Nat Rev Drug Discov*. 2021 Jan;20(1):21-38.

- [9] Chen YF, Pandey S, Day CH, Chen YF, Jiang AZ, Ho TJ, Chen RJ, Padma VV, Kuo WW, Huang CY. Synergistic effect of HIF-1 α and FoxO3a trigger cardiomyocyte apoptosis under hyperglycemic ischemia condition. *J Cell Physiol.* 2018 Apr;233(4):3660-3671.
- [10] Fluteau A, Ince PG, Minett T, Matthews FE, Brayne C, Garwood CJ, Ratcliffe LE, Morgan S, Heath PR, Shaw PJ, Wharton SB, Simpson JE; MRC Cognitive Function Ageing Neuropathology Study Group. The nuclear retention of transcription factor FOXO3a correlates with a DNA damage response and increased glutamine synthetase expression by astrocytes suggesting a neuroprotective role in the ageing brain. *Neurosci Lett.* 2015 Nov 16;609:11-7.
- [11] Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO Signaling Pathways as Therapeutic Targets in Cancer. *Int J Biol Sci.* 2017 Jul 6;13(7):815-827.
- [12] Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, et al. Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor. *Cell.* 1999 Mar 19;96(6):857-68.
- [13] Wang, F., Marshall, C. B. & Ikura, M. Forkhead followed by disordered tail: the intrinsically disordered regions of FOXO3a. *Intrinsically Disord Proteins.* 2015 3;3(1) e1056906.
- [14] Tsai KL, Sun YJ, Huang CY, Yang JY, Hung MC, Hsiao CD. Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.* 2007;35(20):6984-94.
- [15] Hou L, Zhu D, Liang Y, Tian X, Li L, Wang P, Zhu L, Weng X, Wang Y, Li Y, Wu T, Wang J, Meng X. Identification of a specific peptide binding to colon cancer cells from a phage-displayed peptide library. *Br J Cancer.* 2018 Jan;118(1):79-87.
- [16] Hagenbuchner J, Obsilova V, Kaserer T, Kaiser N, Rass B, Psenakova K, Docekal V, Alblova M, Kohoutova K, Schuster D, Aneichyk T, Vesely J, Obexer P, Obsil T, Ausserlechner MJ. Modulating FOXO3 transcriptional activity by small, DBD-binding molecules. *Elife.* 2019 Dec 4;8:e48876.
- [17] Sun Y, Ai X, Hou J, Ye X, Liu R, Shen S, et al. Integrated discovery of FOXO1-DNA stabilizers from marine natural products to restore chemosensitivity to anti-EGFR-based therapy for metastatic lung cancer. *Mol Biosyst.* 2017;13(2):330–7.
- [18] Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem Rev.* Maio de 2006;1589-1615
- [19] Namba AM, Da Silva VB, Da Silva CHTP. Dinâmica molecular: Teoria e aplicações em planejamento de fármacos. *Ecl. Quím., São Paulo,* 33(4): 13-24, 2008.
- [20] T. Darden, D. York, L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *Journal of Chemical Physics,* 98(12), 1993, 10089-10092.

- [21] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble, *Molecular Physics*, 52:2, 1984, 255-268.
- [22] Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys*. 1985 Mar;31(3):1695-1697.
- [23] Avashthi H, Srivastava A, Singh DB. Cavity/Binding Site Prediction Approaches and Their Applications. *Comput Drug Des*. 2020;
- [24] Ghersi D, Sanchez R. Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *J Struct Funct Genomics*. 2011 Jul;12(2):109-17.
- [25] Zheng X, Gan L, Wang E, Wang J. Pocket-based drug design: Exploring pocket space. *AAPS J*. 2013 Jan;15(1):228-41.
- [26] Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002 Sep;1(9):727-30.
- [27] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and
- [28] [https://en.wikipedia.org/wiki/Docking_\(molecular\)#/media/File:Docking_representation_2.png](https://en.wikipedia.org/wiki/Docking_(molecular)#/media/File:Docking_representation_2.png). Acedido a 15/5/22
- [29] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (March 2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". *Advanced Drug Delivery Reviews*. 46 (1–3): 3–26
- [30] Ferreira RS, Glaucius O, Andricopulo AD. Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos. *Quim. Nova*, Vol. 34, No. 10, 2011, 1770-1778.
- [31] The PyMOL Molecular Graphics System, Version 2.4.0 Schrödinger, LLC. DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*, 40(1), 82-92.
- [32] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 2015, 19-25.
- [33] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009 Jun 2;10:168.
- [34] Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*.
- [35] Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*.

- [36] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
- [37] Wishart, David S., et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." Nucleic acids research 46.D1 (2018): D1074-D1082.
- [38] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics, 79(2), 926-935.
- [39] Ropp, Patrick J., Jacob O. Spiegel, Jennifer L. Walker, Harrison Green, Guillermo A. Morales, Katherine A. Milliken, John J. Ringe, and Jacob D. Durrant. (2019) "Gypsum-DL: An Open-source Program for Preparing Small-molecule Libraries for Structure-based Virtual Screening." Journal of Cheminformatics 11:1.
- [40] Ropp PJ, Kaminsky JC, Yablonski S, Durrant JD (2019) Dimorphite-DL: An open-source program for enumerating the ionization states of drug-like small molecules. J Cheminform 11:14.
- [41] O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. Journal of cheminformatics, 3(1), 1-14.
- [42] Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. J Med Chem. 2010 Aug 12;53(15):5858-67.
- [43] Landrum G, Tosco P, Kelley B, Ric, sriniker, gedeck, et al. rdkit/rdkit: 2022_03_5 (Q1 2022) Release. 4 de Agosto de 2022.

8. Anexos

Anexo I – COMPLEXO RÉPLICA 0

Tabela 8.1: Interações das cavidades com o DNA e a proteína no complexo réplica 0

| Cavidade | Conformação 0 | Conformação 1 | Conformação 2 | Conformação 3 | Conformação 4 | Conformação 5 | Conformação 6 |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | DNA | DNA | DNA+F OXO3 | FOXO | DNA+F OXO3 | DNA+F OXO3 | DNA+FOX O3 |
| 2 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA | DNA+FOX O3 |
| 3 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA | DNA | DNA+F OXO3 | FOXO3 |
| 4 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | FOXO3 |
| 5 | FOXO3 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | DNA+F OXO3 | DNA |
| 6 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 |
| 7 | DNA+F OXO3 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | FOXO3 |
| 8 | DNA+F OXO3 | DNA | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | DNA |
| 9 | DNA+F OXO3 | FOXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | DNA+FOX O3 |
| 10 | FOXO3 | DNA+F OXO3 | FOXO3 | DNA+F OXO3 | | DNA+F OXO3 | DNA+FOX O3 |
| 11 | DNA | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | | DNA | DNA+FOX O3 |
| 12 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | | DNA+F OXO3 | DNA+FOX O3 |
| 13 | FOXO3 | DNA+F OXO3 | DNA | DNA | | | FOXO3 |
| 14 | DNA+F OXO3 | DNA+F OXO3 | DNA+F OXO3 | FOXO3 | | | DNA+FOX O3 |
| 15 | DNA+F OXO3 | FOXO3 | FOXO3 | DNA+F OXO3 | | | DNA+FOX O3 |
| 16 | DNA+F OXO3 | | | DNA+F OXO3 | | | DNA+FOX O3 |
| 17 | FOXO3 | | | | | | |
| 18 | DNA+F OXO3 | | | | | | |
| 19 | FOXO3 | | | | | | |

Tabela 8.2: Interações das cavidades com o DNA e a proteína no complexo réplica 0

| Cavidade | Conformação 7 | Conformação 8 | Conformação 9 | Conformação 10 | Conformação 11 | Conformação 12 | Conformação 13 |
|-----------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | DNA+FOXO3 | FOXO3 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA |
| 2 | DNA+FOXO3 | DNA | FOXO3+DNA | FOXO3+DNA | FOXO3 | FOXO3+DNA | FOXO3+DNA |
| 3 | DNA+FOXO3 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | DNA | FOXO3+DNA | FOXO3 |
| 4 | DNA | FOXO3+DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3+DNA | FOXO3 |
| 5 | FOXO3 | FOXO3+DNA | DNA | DNA | FOXO3 | FOXO3 | FOXO3+DNA |
| 6 | FOXO3 | FOXO3+DNA | FOXO3 | FOXO3 | FOXO3+DNA | FOXO3 | FOXO3+DNA |
| 7 | FOXO3 | FOXO3 | FOXO3+DNA | FOXO3+DNA | Foxo3 | FOXO3 | FOXO3 |
| 8 | DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3 |
| 9 | DNA+FOXO3 | FOXO3+DNA | Foxo3 | FOXO3+DNA | Foxo3 | FOXO3+DNA | FOXO3 |
| 10 | FOXO3 | FOXO3 | FOXO3+DNA | DNA | FOXO3+DNA | | FOXO3 |
| 11 | FOXO3 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | | |
| 12 | DNA+FOXO3 | FOXO3 | | Foxo3 | FOXO3+DNA | | |
| 13 | DNA+FOXO3 | FOXO3+DNA | | | FOXO3 | | |
| 14 | DNA+FOXO3 | FOXO3+DNA | | | | | |
| 15 | DNA+FOXO3 | FOXO3+DNA | | | | | |

Tabela 8.3: Interações das cavidades com o DNA e a proteína no complexo réplica 0

| Cavidade | Conformação 14 | Conformação 15 | Conformação 16 | Conformação 17 | Conformação 18 | Conformação 19 | Conformação 20 |
|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 4 | DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 5 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 6 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 |
| 7 | DNA | Foxo3 | FOXO3 +DNA | DNA | FOXO3 | FOXO3 | FOXO3+DNA |
| 8 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 |
| 9 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 |
| 10 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3+DNA |
| 11 | FOXO3 | FOXO3 | | | FOXO3 | DNA | FOXO3+DNA |
| 12 | FOXO3 +DNA | | | | FOXO3 +DNA | FOXO3 +DNA | |
| 13 | FOXO3 +DNA | | | | FOXO3 +DNA | FOXO3 +DNA | |
| 14 | | | | | FOXO3 +DNA | FOXO3 +DNA | |
| 15 | | | | | Foxo3 | DNA | |
| 16 | | | | | DNA | FOXO3 +DNA | |
| 17 | | | | | | FOXO3 +DNA | |

ANEXO II – COMPLEXO RÉPLICA 1

Tabela 8.4: Interações das cavidades com o DNA e a proteína no complexo réplica 1

| Cavidade | Conformação 0 | Conformação 1 | Conformação 2 | Conformação 3 | Conformação 4 | Conformação 5 | Conformação 6 |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3+DNA |
| 3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 4 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 5 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 |
| 6 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3+DNA |
| 7 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3+DNA |
| 8 | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 9 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | DNA | FOXO3 |
| 10 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 | FOXO3+DNA |
| 11 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | DNA | FOXO3 |
| 12 | FOXO3 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 13 | FOXO3 | FOXO3 | FOXO3 +DNA | | FOXO3 | FOXO3 | |
| 14 | FOXO3 | FOXO3 | FOXO3 +DNA | | FOXO3 3 | FOXO3 +DNA | |
| 15 | FOXO3 | FOXO3 | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | |
| 16 | FOXO3 +DNA | FOXO3 +DNA | | | FOXO3 +DNA | | |
| 17 | FOXO3 +DNA | FOXO3 +DNA | | | | | |

Tabela 8.5: Interações das cavidades com o DNA e a proteína no complexo réplica 1

| Cavidade | Conformação 7 | Conformação 8 | Conformação 9 | Conformação 10 | Conformação 11 | Conformação 12 | Conformação 13 |
|-----------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 |
| 3 | FOXO3 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 4 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | DNA | DNA | FOXO3 | FOXO3 |
| 5 | FOXO3 +DNA | DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 6 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 |
| 7 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 8 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | DNA | FOXO3 +DNA | FOXO3+DNA |
| 9 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3+DNA |
| 10 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 11 | FOXO3 +DNA | FOXO3 | Foxo3 | FOXO3 +DNA | FOXO3 +DNA | | FOXO3 |
| 12 | | FOXO3 +DNA | FOXO3 +DNA | | | | FOXO3+DNA |
| 13 | | FOXO3 +DNA | FOXO3 | | | | |
| 14 | | | FOXO3 +DNA | | | | |
| 15 | | | FOXO3 | | | | |

Tabela 8.6: Interações das cavidades com o DNA e a proteína no complexo réplica 1

| Cavidade | Conformação 14 | Conformação 15 | Conformação 16 | Conformação 17 | Conformação 18 | Conformação 19 | Conformação 20 |
|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 3 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 4 | DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 5 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3+DNA |
| 6 | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 7 | FOXO3 | DNA | DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3 |
| 8 | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 |
| 9 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 10 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 11 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 |
| 12 | FOXO3 +DNA | | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 |
| 13 | FOXO3 +DNA | | FOXO3 +DNA | | FOXO3 | FOXO3 +DNA | FOXO3 |
| 14 | FOXO3 +DNA | | FOXO3 | | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 15 | FOXO3 | | | | | | FOXO3 |
| 16 | FOXO3 +DNA | | | | | | FOXO3+DNA |
| 17 | FOXO3 | | | | | | FOXO3 |

ANEXO III – COMPLEXO RÉPLICA 2

Tabela 8.7: Interações das cavidades com o DNA e a proteína no complexo réplica 2

| Cavidade | Conformação 0 | Conformação 1 | Conformação 2 | Conformação 3 | Conformação 4 | Conformação 5 | Conformação 6 |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3+DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | DNA | DNA | FOXO3 | FOXO3 |
| 4 | FOXO3 | DNA | DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3+DNA |
| 5 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | DNA |
| 6 | FOXO3 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 7 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 | DNA | FOXO3 +DNA | FOXO3 |
| 8 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | DNA | FOXO3+DNA |
| 9 | FOXO3 +DNA | FOXO3 | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3+DNA |
| 10 | FOXO3 +DNA | FOXO3 | | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 |
| 11 | FOXO3 | DNA | | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 12 | FOXO3 +DNA | FOXO3 +DNA | | DNA | | | FOXO3+DNA |
| 13 | FOXO3 | FOXO3 | | FOXO3 +DNA | | | FOXO3 |
| 14 | FOXO3 +DNA | FOXO3 | | | | | FOXO3+DNA |
| 15 | FOXO3 +DNA | FOXO3 | | | | | |
| 16 | | FOXO3 | | | | | |

Tabela 8.8: Interações das cavidades com o DNA e a proteína no complexo réplica 2

| Cavida de | Confor mação 7 | Confor mação 8 | Confor mação 9 | Confor mação 10 | Confor mação 11 | Confor mação 12 | Confor mação 13 |
|----------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | DNA | DNA | FOXO3 +DNA | DNA | FOXO3 +DNA |
| 2 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 3 | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | DNA | FOXO3 +DNA |
| 4 | FOXO3 | DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | Foxo3 |
| 5 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 6 | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 7 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA |
| 8 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA |
| 9 | FOXO3 +DNA | DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA |
| 10 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | Foxo3 |
| 11 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 12 | DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | | DNA |
| 13 | FOXO3 +DNA | | FOXO3 | FOXO3 +DNA | | | FOXO3 |
| 14 | DNA | | FOXO3 | FOXO3 | | | FOXO3 +DNA |
| 15 | FOXO3 | | FOXO3 +DNA | FOXO3 | | | |
| 16 | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 | | | |
| 17 | | | | FOXO3 +DNA | | | |
| 18 | | | | FOXO3 | | | |

Tabela 8.9: Interações das cavidades com o DNA e a proteína no complexo réplica 2

| Cavidade | Conformação 14 | Conformação 15 | Conformação 16 | Conformação 17 | Conformação 18 | Conformação 19 | Conformação 20 |
|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3+DNA |
| 2 | FOXO3 +DNA | DNA | DNA | DNA | FOXO3 | FOXO3 +DNA | FOXO3+DNA |
| 3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 4 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 5 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 |
| 6 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 7 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3 | FOXO3+DNA |
| 8 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 9 | FOXO3 +DNA | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | DNA |
| 10 | FOXO3 +DNA | FOXO3 | | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 11 | | FOXO3 | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 12 | | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 |
| 13 | | FOXO3 +DNA | | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3+DNA |
| 14 | | | | | FOXO3 +DNA | FOXO3 +DNA | FOXO3+DNA |
| 15 | | | | | FOXO3 +DNA | | FOXO3+DNA |
| 16 | | | | | FOXO3 +DNA | | FOXO3+DNA |
| 17 | | | | | FOXO3 | | FOXO3+DNA |
| 18 | | | | | FOXO3 +DNA | | |

Tabela 8.10: Interações das cavidades com o DNA e a proteína no complexo réplica 3

| Cavidade | Conformação 0 | Conformação 1 | Conformação 2 | Conformação 3 | Conformação 4 | Conformação 5 | Conformação 6 |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | FOXO3 | FOXO3+DNA | DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA |
| 2 | FOXO3+DNA | FOXO3+DNA | DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | DNA |
| 3 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA |
| 4 | FOXO3+DNA | DNA | DNA | FOXO3 | DNA | FOXO3 | FOXO3+DNA |
| 5 | FOXO3+DNA | DNA | FOXO3+DNA | DNA | FOXO3 | DNA | FOXO3 |
| 6 | FOXO3 | FOXO3+DNA | FOXO3 | FOXO3 | DNA | FOXO3 | FOXO3+DNA |
| 7 | FOXO3+DNA | FOXO3 | FOXO3 | DNA | FOXO3 | FOXO3+DNA | FOXO3+DNA |
| 8 | FOXO3+DNA | FOXO3+DNA | FOXO3 | FOXO3 | FOXO3+DNA | FOXO3+DNA | DNA |
| 9 | DNA | DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3 | FOXO3+DNA |
| 10 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3 | FOXO3 | FOXO3+DNA |
| 11 | FOXO3 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3 |
| 12 | FOXO3 | DNA | FOXO3 | FOXO3 | FOXO3+DNA | FOXO3 | DNA |
| 13 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | | FOXO3 | FOXO3 |
| 14 | DNA | FOXO3+DNA | FOXO3+DNA | | | FOXO3+DNA | FOXO3 |
| 15 | FOXO3 | FOXO3+DNA | FOXO3+DNA | | | FOXO3+DNA | FOXO3+DNA |
| 16 | FOXO3+DNA | FOXO3+DNA | FOXO3+DNA | | | | DNA |
| 17 | FOXO3 | DNA | | | | | |

Tabela 8.11: Interações das cavidades com o DNA e a proteína no complexo réplica 3

| Cavidade | Conformação 7 | Conformação 8 | Conformação 9 | Conformação 10 | Conformação 11 | Conformação 12 | Conformação 13 |
|-----------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | FOXO3 +DNA | DNA | DNA | FOXO3 +DNA | DNA | DNA | DNA |
| 2 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | DNA | FOXO3 +DNA | DNA | FOXO3 |
| 4 | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 5 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA |
| 6 | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 | DNA | DNA | FOXO3 +DNA |
| 7 | DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA |
| 8 | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 9 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 10 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 11 | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA |
| 12 | FOXO3 +DNA | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 |
| 13 | FOXO3 | | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 | FOXO3 +DNA |
| 14 | FOXO3 | | FOXO3 | FOXO3 +DNA | FOXO3 | | FOXO3 +DNA |
| 15 | FOXO3 +DNA | | FOXO3 +DNA | | FOXO3 +DNA | | |
| 16 | FOXO3 +DNA | | | | FOXO3 +DNA | | |
| 17 | FOXO3 +DNA | | | | FOXO3 +DNA | | |

Tabela 8.12: Interações das cavidades com o DNA e a proteína no complexo réplica 3

| Cavi dade | Confor mação 14 | Confor mação 15 | Confor mação 16 | Confor mação 17 | Confor mação 18 | Confor mação 19 | Confor mação 20 |
|----------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 1 | FOXO3 +DNA | DNA | FOXO3 +DNA | DNA | DNA | DNA | DNA |
| 2 | DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA |
| 3 | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA |
| 4 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 +DNA |
| 5 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | DNA | DNA | FOXO3 +DNA | FOXO3 +DNA |
| 6 | FOXO3 | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 7 | FOXO3 +DNA | FOXO3 +DNA | DNA | FOXO3 | FOXO3 | FOXO3 +DNA | FOXO3 |
| 8 | FOXO3 | FOXO3 +DNA | FOXO3 | DNA | FOXO3 | FOXO3 | FOXO3 +DNA |
| 9 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 | FOXO3 +DNA |
| 10 | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 11 | FOXO3 +DNA | FOXO3 +DNA | DNA | DNA | FOXO3 +DNA | FOXO3 | FOXO3 |
| 12 | FOXO3 +DNA | FOXO3 | DNA | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA |
| 13 | FOXO3 +DNA | DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA |
| 14 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | |
| 15 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | FOXO3 +DNA | |
| 16 | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 | FOXO3 +DNA | |
| 17 | | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | FOXO3 +DNA | |
| 18 | | FOXO3 +DNA | | | FOXO3 +DNA | | |
| 19 | | FOXO3 | | | | | |
| 20 | | FOXO3 +DNA | | | | | |