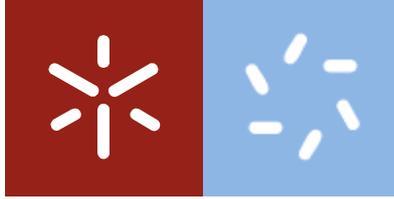


Universidade do Minho
Escola de Ciências

Sara Flor da Silva Queirós

**Modelação da Interação da Glicoproteína
E do Vírus do Dengue com Ligantes de
Interesse Farmacológico.**



Universidade do Minho

Escola de Ciências

Sara Flor da Silva Queirós

**Modelação da Interação da Glicoproteína
E do Vírus do Dengue com Ligantes de
Interesse Farmacológico.**

Dissertação de Mestrado

Mestrado em Biofísica e Bionanossistemas

Trabalho efetuado sob a orientação do

Professor Doutor Luís Silvino Alves Marques

A quem devo tudo. Aos meus pais.

Agradecimentos

Começo por dirigir os meus agradecimentos ao meu orientador, o Professor Dr. Luís Marques. Agradeço-lhe, primeiramente, por me ter acolhido com orientanda. Depois, por ao longo deste ano me ter dado apoio sempre que necessitei e, por me ter transmitido parte do seu conhecimento.

Agradeço, imenso, ao Veniero que foi o meu “salvador” tantas e tantas vezes. Por toda a paciência, por todo o incentivo, por toda a ajuda e, por todo o conhecimento que partilhou comigo. Quando iniciei este trabalho de tese, tinha conhecimentos escassos sobre o assunto e, com a ajuda do Veniero, consegui evoluir imenso e aprender tudo o que necessitava para realizar este trabalho com sucesso. Foi um pilar importantíssimo durante este ano, mais do que um colega do Grupo em que estava inserida, tornou-se um Professor a tempo inteiro e, por isso, um muito obrigada.

À Professora Dra. Marta Ramos, por ter sido sempre incansável comigo. Por me ter apoiado e motivado sempre; por me ter mostrado que trabalhar com um sorriso é sempre o melhor caminho. Muito obrigada!

Às minhas amigas Eduarda, Juliana, Margarida e Marta: agradeço-vos por tudo. Agradeço-vos a Amizade que, para mim, será sempre das melhores coisas que levarei daqui. Obrigada por terem partilhado comigo esta caminhada, por tornarem este ano mais fácil de suportar, pela companhia, pelas gargalhadas e choros, pela força e motivação que nunca deixaram de dar (mesmo quando estavam longe Guida e Marta), por me terem ajudado sempre, sem nunca pedir retorno. Vocês são uma das razões que fez estes dois anos terem valido a pena. Obrigadíssima.

Ao Daniel, que mesmo longe e em momentos complicados, me acompanhou sempre. Foi o meu confidente nas horas de maior aperto, e a melhor pessoa para celebrar comigo as minhas conquistas. Por me ter dito sempre, que tinha orgulho e que acreditaria sempre em mim. Obrigada por me aturares sempre, em qualquer altura.

Por último, aos que tornaram isto tudo possível: os meus pais. Isto é tanto meu como vosso. Obrigada por, mesmo numa situação complicada, me terem dado esta oportunidade. Agradeço-vos o vosso sacrifício diário para eu poder chegar até aqui. Um simples “Obrigada” nunca será suficiente para agradecer tudo o que fazem por mim, mas é um começo: Obrigada.

Resumo

A doença do Dengue tem vindo a afetar parte da população mundial, sem que haja cura ou prevenção possível. O vírus do Dengue (DENV) tem atualmente quatro serotipos distintos, existindo já rumores de um quinto. A doença provocada pelo DENV pode atingir extremos, sendo que o doente pode ter uma recuperação a 100% ou poderá morrer devido a um choque hemorrágico. Desde o século passado que se luta contra a doença tentando encontrar algum tipo de cura ou prevenção, mas sempre sem sucesso. Estas investigações, têm tido diferentes focos, desde o mosquito *Aedes Aegypti* à conceção de fármacos antivirais eficazes contra o DENV.

A Simulação Computacional, tem sido uma preciosa ajuda no estudo da estrutura do vírus DENV, e no desenvolvimento de novos fármacos antivirais. Em combinação com as técnicas laboratoriais, a Simulação Computacional, tem fornecido informações que é impossível observar experimentalmente. Com a ajuda computacional, obtemos informações como valores de energias e orientações moleculares possíveis em mecanismos de ação de determinada biomolécula.

Este trabalho de tese consiste no estudo por simulação computacional das estruturas específicas da Glicoproteína E presente na superfície do vírus do Dengue e a identificação de possíveis alvos para fármacos antivirais. Esta tese foi levada a cabo em paralelo com um estudo experimental realizado pelo Prof. Léo Degreve, na Universidade de São Paulo, visando a identificação e teste de novos fármacos antivirais para o DENGUE. O objetivo desta tese é a elaboração de um protocolo de simulação que permita prever de forma sistemática e precisa, a energia de ligação entre um ligante e um recetor, usando métodos de simulação híbridos combinado simulações de dinâmica molecular e mecânica quântica.

Esta tese está organizada em quatro capítulos. O primeiro capítulo, será a parte introdutória será uma breve explicação do vírus em estudo e, dos métodos computacionais. No segundo capítulo, estarão explicados todos os materiais utilizados e, assim como os parâmetros usados nas simulações computacionais. Nesta secção, será apresentado o protocolo completo para a realização deste estudo. O terceiro capítulo, consistirá na apresentação e discussão dos resultados obtidos de todo o processo anterior. Por último, o quarto capítulo, consistirá na

conclusão final do trabalho analisando todo o processo, evidenciando o que foi favorável e o que deveria ter sido feito para obter melhores resultados.

Abstract

Dengue disease is affecting part of the world population, with no cure or prevention possible. Currently, the virus of dengue (DENV) has four distinct serotypes, and there are rumors of a fifth. The disease caused by DENV can attain extremes, where the patient may have a recovery of 100% or may die due to hemorrhagic shock. Since the last century large efforts have been done to fight the disease, trying to find some kind of cure or prevention, but always without success. These investigations have had different focuses, from the *Aedes Aegypti* mosquito to the development of effective anti-viral drugs against DENV.

Computer Simulation has been a great help in the study of DENV structure, and in the development of new antiviral drugs. In combination with laboratory techniques, Computational Simulation, has provided information that is impossible to observe experimentally. With computer assistance, we obtain information as energy values and possible molecular orientation for a given biomolecule.

This thesis presents a computer simulation study of the specific structures of glycoprotein E present on the surface of Dengue virus and the identification of possible targets for antiviral drugs. This thesis was carried out in parallel with an experimental study by Prof. Léo Degreuve at the University of São Paulo, for the identification and testing of new antiviral drugs for the DENGUE. The aim of this thesis is to develop a simulation protocol to predict systematically and accurately, the binding energy between a ligand and a receptor, using hybrid simulation methods combining molecular dynamics and quantum mechanics simulations.

This thesis is organized in four chapters. The first chapter is the introductory part consisting in a brief explanation of the virus under study and the computational methods used. In the second chapter it is presented all the materials and as well as the parameters used in computer simulations. In this section we will present the full protocol for this study. The third chapter consists in the presentation and discussion of the results. Finally, the fourth chapter presents the conclusions of this works.

Índice

| | |
|--|----|
| 1. Introdução | 1 |
| 1.1 Dengue..... | 2 |
| 1.2 História e Epidemiologia..... | 4 |
| 1.3 Vírus do Dengue | 7 |
| 1.4 Glicoproteína E..... | 8 |
| 1.5 Métodos Computacionais para os estudos da interação proteína-ligante | 12 |
| 1.5.1 Campos de Força Empíricos | 12 |
| 1.5.2 Dinâmica Molecular | 26 |
| 1.5.3. Docking Molecular | 37 |
| 1.5.4. Métodos Quânticos (QM) | 40 |
| 1.5.5. Métodos Híbridos..... | 42 |
| 2. Materiais e Métodos: Protocolo de Simulação da interação proteína-ligante | 50 |
| 2.1 <i>Materiais</i> | 51 |
| 2.2 Métodos | 53 |
| 2.2.1. Estruturas da Glicoproteína E e Ligante | 53 |
| 2.2.2. Simulação da Dinâmica da proteína por Dinâmica Molecular..... | 54 |
| 2.2.3. Docking Proteína-Ligante | 60 |
| 2.2.4. Energias de Ligação..... | 64 |
| 2.2.5. ONIOM 2 | 70 |
| 2.2.6 ONIOM 3 | 75 |
| 3. Resultados..... | 82 |
| 3.1 Dinâmica da Proteína | 83 |
| 3.2 Docking..... | 87 |
| 3.3 Energias de Ligação usando o método ONIOM..... | 90 |
| • Método ONIOM de 2 camadas (ONIOM2) | 90 |
| • Método ONIOM usando 3 camadas (ONIOM3) | 91 |
| 4. Conclusão..... | 93 |
| 5. Bibliografia..... | 95 |

Índice de Figuras

| | |
|--|----|
| Figura 1.1 – Principais sintomas do Dengue. Adaptado de http://www.wikiwand.com/pt/Dengue | 3 |
| Figura 1.2 – Transmissão do vírus de primatas não humanos, para humanos. Adaptado de Whitehead et al. 2007..... | 4 |
| Figura 1.3- Média do número anual de DF/DHF. Adaptado de Tavodova 2012..... | 5 |
| Figura 1.4 - Representação das áreas do globo em risco de ocorrências de casos de Dengue, em 2008..... | 6 |
| Figura 1.5 - Constituição do DENV. Adaptado de FREIRE 2014..... | 7 |
| Figura 1.6- (A) Glicoproteína E com Domínio I representado a vermelho, Domínio II a amarelo e Domínio III a azul-escuro. (B) Conformação dimérica da Glicoproteína E. Adaptado de Alisson et al., 1999..... | 8 |
| Figura 1.7 - Processo de Fusão das membranas, com o auxílio da Glicoproteína E. Adaptado de Mukhopadhyay et al. 2005..... | 10 |
| Figura 1.8 - Potencial de Ligação entre dois átomos usando restrição harmónica. Adaptado de Van der Spoel et al. 2010..... | 14 |
| Figura 1.9 - Potencial do ângulo de ligação entre três átomos com restrição harmónica. Adaptado de van der Spoel et al. 2010..... | 15 |
| Figura 1.10 - Exemplo de um diedro. | 16 |
| Figura 1.11 - Potencial do ângulo diedro. Adaptado de van der Spoel et al. 2010..... | 17 |
| Figura 1.12 - À esquerda, está representado a aplicação de um diedro impróprio para manter a quiralidade, mantendo o ângulo formado pela ligação AB com o plano BCD. À direita, usa-se os diedros impróprios para manter a planaridade, mantendo-se a ligação CD no plano ABC, em que o ângulo de equilíbrio será zero. Adaptado de Soares 1997..... | 17 |
| Figura 1.13 - À esquerda, o ângulo diedro impróprio num tetraedro. À direita, o potencial do ângulo diedro impróprio. Adaptado de van der Spoel et al. 2010..... | 18 |
| Figura 1.14- Representação de um potencial Eletrostático resultante de duas cargas pontuais. Adaptado de Leach 2001..... | 19 |
| Figura 1.15 - Potencial de Lennard-Jones. Adaptado de van der Spoel et al. 2010..... | 20 |
| Figura 1.16 - Minimum image convention num sistema com duas dimensões. A caixa central contém cinco moléculas, assim como a caixa construída contendo a molécula 1 no centro. O círculo representa o potencial cut-off. Adaptado de Cheung et al. 2008..... | 22 |

| | |
|--|----|
| Figura 1.17 - Construção da esfera com caixas de simulação. Adaptado de Cheung et al. 2008 | 23 |
| Figura 1.18 - Processo de Minimização de energia modificando a geometria da molécula até encontrar um mínimo. Adaptado de Standard 2015 | 24 |
| Figura 1.19 - Fluxograma para realização de simulações no AMBER. Adaptado de D.A. Case, et al. 2015..... | 31 |
| Figura 1.20- Conceito das α -esferas aplicado à detecção de cavidades. Adaptado de Schmidtke et al. 2010..... | 35 |
| Figura 1.21 - Ilustração do docking de uma proteína com um ligante, a ligarem-se para formarem um complexo estável. | 37 |
| Figura 1.22 - Princípio do método híbrido QM/MM com divisão do sistema em duas partes... .. | 42 |
| Figura 1.23 - Terminologia do ONIOM, usando o Etano como exemplo. | 44 |
| Figura 1.24 - Representação esquemática da equação usada para o sistema de três camadas do ONIOM. | 46 |
| Figura 1.25 - Representação da abordagem de três camadas do ONIOM. A Camada Baixa é representada pelas linhas finas, a Camada Média por tubos e, a Camada Alta por bolas e traços..... | 47 |
| Figura 2.1 – (A) Montagem biológica contendo a Glicoproteína E representada no centro da estrutura global do DENV. (B) 1THD: Complexa organização da Glicoproteína E do DENV2, contendo três cadeias – A, B e C- e os átomos C-alfa Adaptado de (Fuzo & Degreève 2013) | 53 |
| Figura 2.2 - Metodologia utilizada para a realização das simulações de Dinâmica Molecular... .. | 54 |
| Figura 2.3- (A) Script “mdin1” representando o script criado para o primeiro passo deste processo, dos 0K aos 50K. (B) Script “mdin6” representando o script criado para o sexto passo deste processo, dos 250K aos 300K. | 58 |
| Figura 2.4 - script "mdinres" criado para realizar o relaxamento da proteína. | 59 |
| Figura 2.5 – Cavidade escolhida na Glicoproteína E. | 60 |
| Figura 2.6 - Exemplo de ficheiro de configuração para o ligante zinc_1785. | 62 |
| Figura 2.7 - Criação da Gridbox com centro no Ligante..... | 63 |
| Figura 2.8 – Diferentes tamanhos de esferas testados..... | 66 |
| Figura 2.9 – Representação de “NME” e “ACE” no corte dos resíduos. Adaptado de (OTHMAN 2010) | 68 |
| Figura 2.10 - Ilustração do Complexo, Substrato e Ligante. | 69 |
| Figura 2.11 - Processo de seleção de camadas, selecionando os átomos que se encontram a 3Å de cada átomo do ligante. 1) A “Low Layer” está representado pelas linhas finas e, 2) “High Layer” com bolas e traços. | 70 |
| Figura 2.12 - Esquema de cálculos realizado para ONIOM de duas camadas. | 71 |

| | |
|---|----|
| Figura 2.13 - Representação (1) do Complexo e (2) do Substrato. A amarelo está representado o Ligante que desaparece no Substrato..... | 73 |
| Figura 2.14 - Representação do (1) Complexo e (2) do Ligante..... | 73 |
| Figura 2.15- Representação dos dois erros mais comuns de divisão de camadas. Os que apresentam a circunferência vermelha, indicam o caso em que há a divisão de camadas numa ligação dupla, em que fica o Nitrogénio na HL e o Carbono com o Oxigénio na LL. A circunferência a roxo, indica o caso oposto..... | 76 |
| Figura 2.16 - Esquema de cálculos do método ONIOM com três camadas..... | 77 |
| Figura 2.17 - Estrutura com as três camadas definidas. A Amarelo encontra-se o ligante; representados com bolas e traços os resíduos que pertencem à HL; com tubos a ML e, com linhas finas o LL. | 78 |
| Figura 3.1 - RMSD resultante da DM..... | 83 |
| Figura 3.2 – Cavidade P ₂ ⁰ representada na Glicoproteína E. | 84 |
| Figura 3.3 - Demonstração do comportamento dinâmico da cavidade, através da variação do volume da mesma ao longo da simulação de DM. | 85 |
| Figura 3.4 - Hidrofobicidade da cavidade P ₂ ⁰ | 86 |
| Figura 3.5 - Score resultante do Docking | 88 |
| Figura 3.6 - Ilustração do comportamento do ligante dentro da cavidade de estudo, durante o processo de Docking. | 89 |

Índice de Tabelas

| | |
|--|-----------|
| Tabela 1 – Comparação entre a incorporação mecânica (ME) e incorporação eletrostática (EE). Fonte: (Zhang 2013) | 49 |
| <i>Tabela 2 - Resíduos presentes nas esferas, com os respectivos números de identificação.</i> | <i>67</i> |
| Tabela 3 - Informação sobre o número de átomos por camada, nas esferas de cada conformação estudada. | 78 |
| Tabela 4 - Resultados obtidos nas quatro conformações estudadas com ONIOM2. | 90 |
| Tabela 5 – Resultados obtidos nas quatro conformações estudadas com ONIOM3..... | 91 |

Índice de Ilustrações

| | |
|--|----|
| Ilustração 1 - Esquema de cálculos do sistema..... | 74 |
| Ilustração 2 - Esquema de Cálculos para obter a Energia Total..... | 75 |
| Ilustração 3 - Esquema de cálculos do sistema..... | 80 |
| Ilustração 4 - Esquema de Cálculos para obter a Energia Total..... | 81 |

Abreviaturas

| | |
|-------|--|
| DENV | Vírus do Dengue |
| DC | Dengue Clássico |
| DF | <i>Dengue Fever</i> |
| FHD | Febre Hemorrágica do Dengue |
| DHF | <i>Dengue Haemorrhagic Fever</i> |
| DSS | <i>Dengue Shock Syndrome</i> |
| ARN | Ácido Ribonucleico |
| ADN | Ácido Desoxirribonucleico |
| DM | Dinâmica Molecular |
| PMEMD | <i>Particle Mesh Ewald Molecular Dynamics</i> |
| GAFF | <i>General Amber Force Field</i> |
| AG | Algoritmo Genérico |
| MC | Algoritmo Monte Carlo |
| CI | Método Configuração de Interação |
| MP | Método Møller-Plesset |
| MM | Mecânica Molecular |
| QM | Mecânica Quântica |
| LIE | <i>Linear Interaction Energy</i> |
| HL | <i>High Layer/</i> Camada Alta |
| ML | <i>Middle Layer/</i> Camada Intermédia |
| LL | <i>Low Layer/</i> Camada Baixa |
| BE | <i>Binding Energy /</i> Energia de Ligação |
| LA | <i>Link Atoms</i> |
| LAC | <i>Link Atoms Connection</i> |
| LAH | <i>Link Atoms Host</i> |
| ONIOM | <i>Our own N-layers Integrated molecular Orbital molecular Mechanics</i> |

| | |
|-------|---|
| ME | <i>Mechanical Embedding</i> |
| EE | <i>Electronic Embedding</i> |
| VMD | <i>Visual Molecular Dynamics</i> |
| AMBER | <i>Assisted Model Building with Energy Refinement</i> |
| RMSD | <i>Root-Mean-Square-Deviation</i> |

1. Introdução

1.1 Dengue

O Dengue é uma doença tropical infecciosa transmitida através de um mosquito, o *Aedes Aegypti*. Esta doença é causada pelo vírus DENV, que é proveniente da família de vírus *Flaviviridae*, sendo – tal como o Zika – um dos sessenta e oito géneros de Flavivírus. O DENV apresenta quatro tipos imunológicos distintos (serotipos) sendo eles o DENV1, DENV2 (o mais comum), DENV3 e DENV4, em que a distinção entre eles baseia-se na sua antigenicidade. A infeção causada por um dos quatro serotipos dá imunidade permanente para o serotipo em causa e, imunidade parcial e temporária aos outros três. (Mundial et al. 1940)

Ruralmente, o nome alternativo dado ao Dengue é “febre quebra-ossos” pelo facto de um dos sintomas ser a dor muscular e nas articulações. No entanto, nesta doença os sintomas não são iguais em todos os casos, por essa razão foram criadas categorias onde são inseridos os diferentes níveis de infeção provocados pelo vírus do Dengue: Dengue assintomática, Dengue Clássico (DC) – ou *dengue fever* (DF) - e, Febre Hemorrágica do Dengue (FHD) – ou *dengue haemorrhagic fever* (DHF). A última, foi ainda dividida em quatro secções de gravidade, sendo que a terceira (III) e quarta (IV) são definidas como sendo síndrome de choque do dengue (*dengue shock syndrome* (DSS)). (Who 2009)

A infeção provocada pelo Dengue é considerada uma doença sistémica e dinâmica. Isto ocorre para o caso da manifestação clínica ser do Dengue Clássico (DC) e, pode existir outros sintomas como febres altas, dores de cabeça e erupções cutâneas, tal como demonstrado na Figura 1.1. No entanto, existe outro caso clínico que é a Febre Hemorrágica do Dengue (FHD), que consiste no choque hemorrágico do doente. Normalmente, ocorre entre o terceiro e sétimo dia da doença e, resulta do aumento de permeabilidade vascular que desencadeia a falência circulatória. Este quadro clínico caracteriza-se pela sua curta duração, podendo atingir extremos, num prazo de 24 horas, levando ou à morte ou à recuperação rápida do paciente. Esta manifestação clínica, FHD, atinge cerca de 500 000 pessoas por ano, sendo que há registos de mortalidade com percentagens de 10% para os doentes que receberam assistência médica e, de 30% por os doentes que não tiveram acesso a nenhum tipo de tratamento hospitalar. (Mundo, et al. 1940)

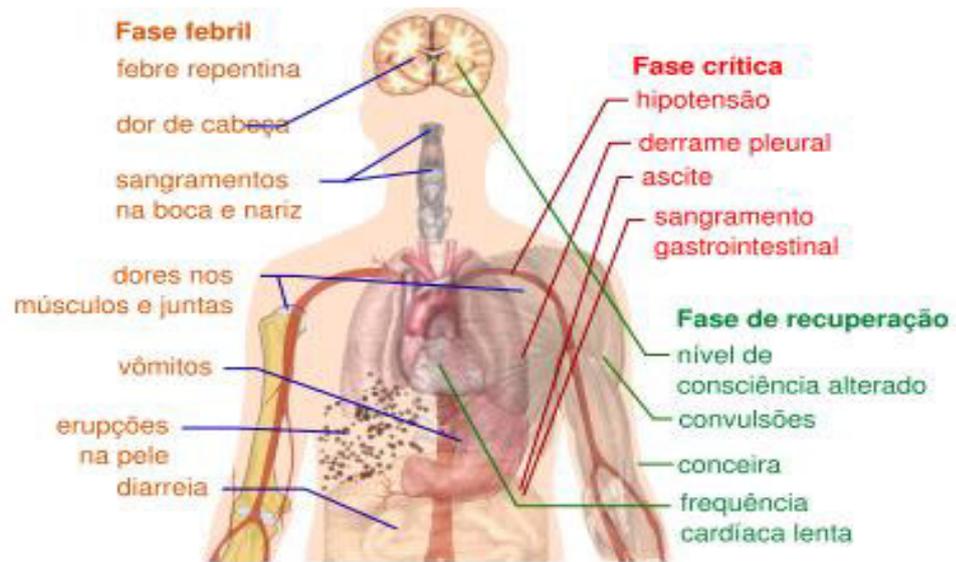


Figura 1.1 – Principais sintomas do Dengue. Adaptado de <http://www.wikiwand.com/pt/Dengue>

1.2 História e Epidemiologia

O vírus do Dengue, teve origem em vírus que circulavam em primatas não-humanos, nas proximidades da península da Malásia. Com o crescimento populacional, as habitações da região aproximaram-se da selva e, desse modo houve a transmissão do vírus dos primatas não humanos, para humanos.

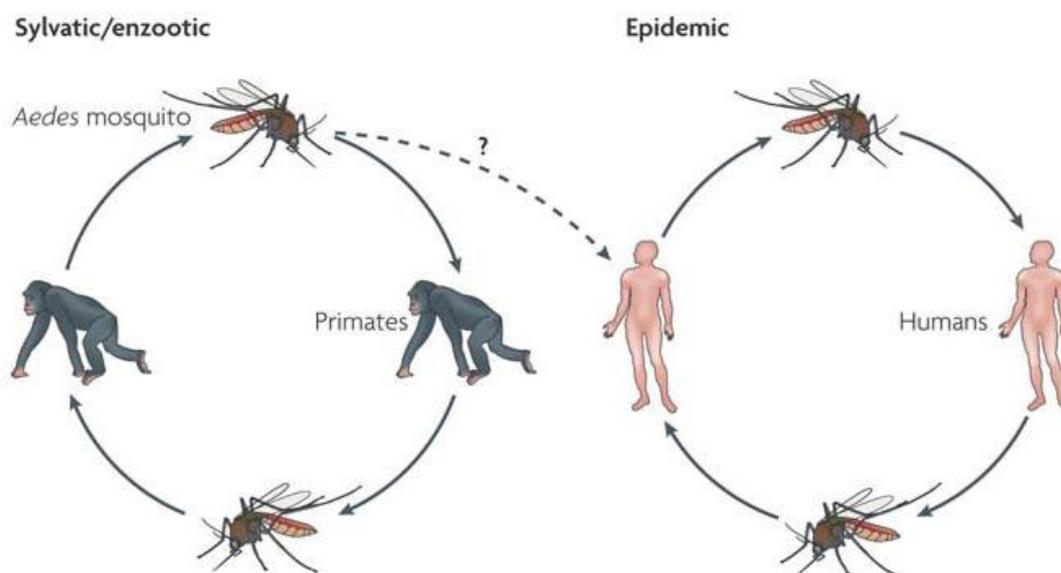


Figura 1.2 – Transmissão do vírus de primatas não humanos, para humanos. Adaptado de Whitehead et al. 2007.

Devido ao comércio de escravos, o mosquito *Aedes Aegypt*, espalhou-se para fora de África, durante os séculos XV e XIX. Os primeiros relatos de grandes epidemias, de supostamente, o Dengue ocorre na Ásia, África e América do Norte em 1779 e 1780. A partir daí, as epidemias de Dengue tornaram-se constantes. Em 1907, o Dengue tornou-se a segunda doença na história (a primeira foi a febre amarela), de carácter viral confirmada. Após a Segunda Guerra Mundial, houve um aumento da propagação do Dengue. Consequência da perturbação ecológica ocorrida no Sudeste da Ásia e do Pacífico que, ofereceu as condições ideais para o aumento da transmissão do Dengue. Este aumento, levou também à dispersão de diferentes sorotipos para novas áreas e, às primeiras ocorrências da Febre Hemorrágica. Este caso clínico, foi relatado

pela primeira vez em 1953, nas Filipinas e, num prazo de vinte anos, a epidemia da doença tinha-se espalhado por todo o Sudoeste Asiático.

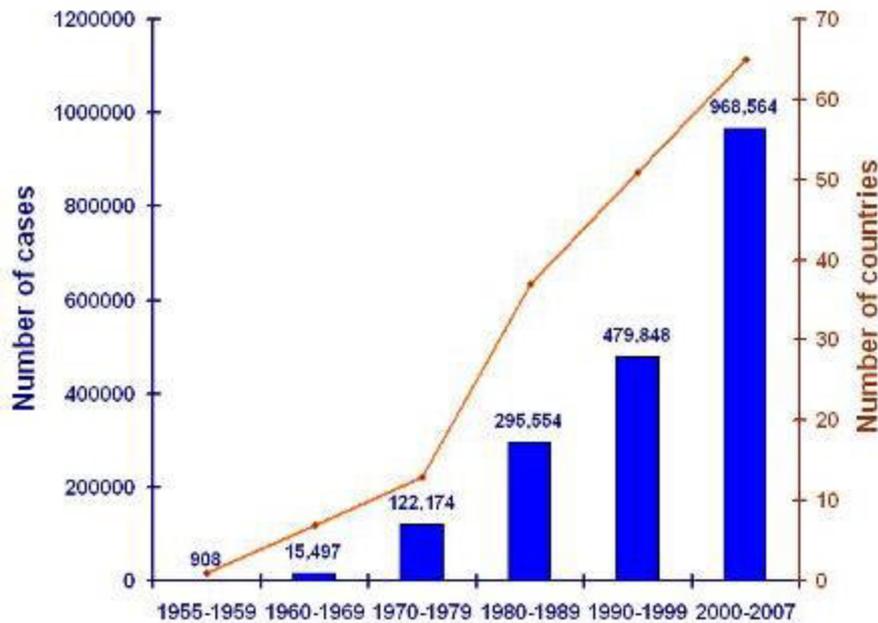
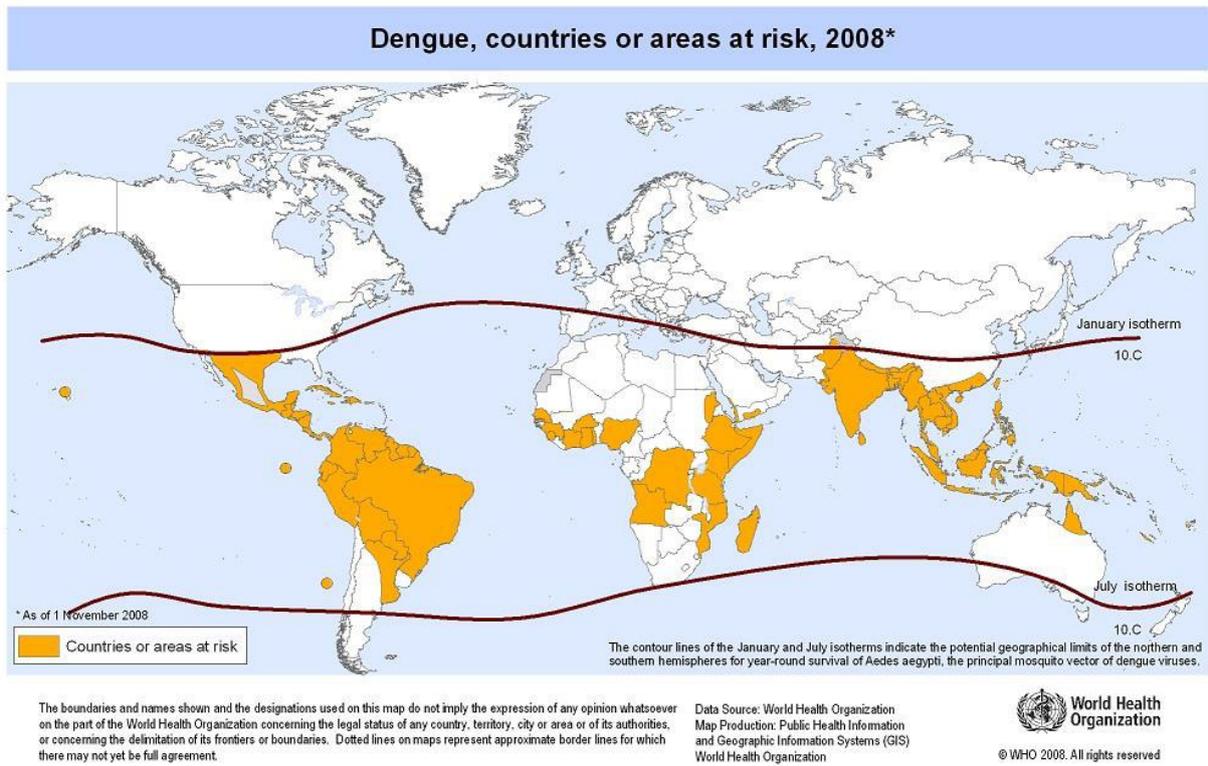


Figura 1.3- Média do número anual de DF/DHF. Adaptado de Tavodova 2012

Mais recentemente, em 2002, a Assembleia Geral da Organização Mundial de Saúde, colocou o Dengue como uma questão de prioridade de saúde pública internacional. Com isto, procura-se reforçar o controlo da doença e a investigação para combater ao máximo a mesma. Relatórios recentes, têm demonstrado que há maior taxa de incidência nas zonas rurais, em comparação com as zonas urbanas. Aponta-se para que isto seja uma consequência do aumento do transporte e mobilidade da urbanização.

O mapa epidemiológico de Dengue, realizado em 2008 onde demonstra os locais do globo que apresentam maior risco de incidência da doença. E, analisando a Figura 1.4, é possível verificar que em 2008, os locais em que começaram as epidemias de Dengue se mantêm como as que apresentam maior risco de incidência.



*Figura 1.4 - Representação das áreas do globo em risco de ocorrências de casos de Dengue, em 2008.
 Adaptado de Tavodova 2012*

1.3 Vírus do Dengue

O vírus do Dengue é composto por RNA positivo que se encontra envolvido por um capsídeo proteico. Por sua vez, este capsídeo proteico está contido por uma bicamada lipídica e, finalmente protegido por um envelope glicoproteico esférico. (Figura 1.5) (Pokidysheva et al. 2006)

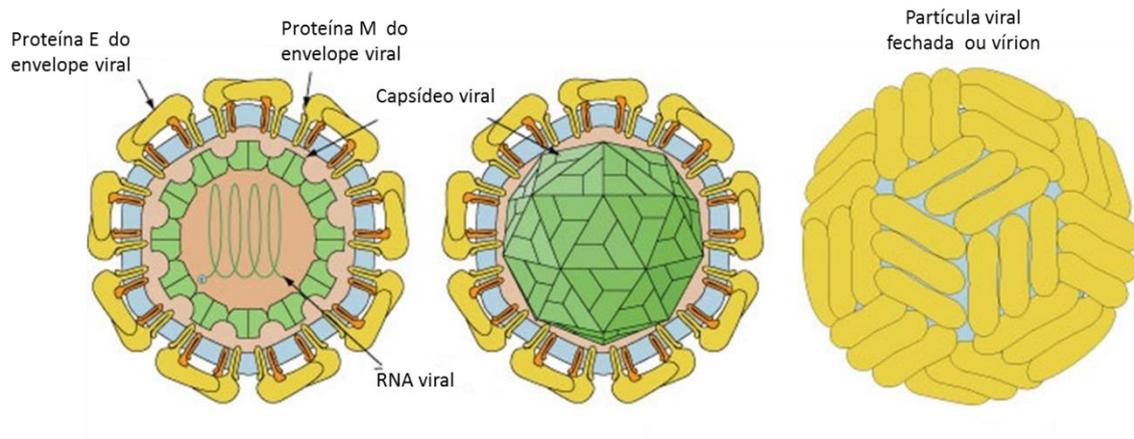


Figura 1.5 - Constituição do DENV. Adaptado de FREIRE 2014

O genoma do vírus é constituído por cerca de 11 000 nucleótidos que codificam três tipos de proteínas (C, M e E) que se encontram na constituição da partícula viral e, sete outros tipos de proteínas (NS1, NS2a, NS2b, NS3, NS4a, NS4B e NS5) que são encontradas nas células hospedeiras infetadas. (Ross 2010) Temos portanto, proteínas estruturais e não estruturais. Nas proteínas estruturais encontra-se três tipos de proteínas, a proteína do capsídeo (C) é extremamente básica e é complexada com o RNA viral de modo a formar o nucleocapsídeo. Esta proteína é encontrada, em células infectadas com DENV, em torno das gotículas lipídicas que derivam do retículo endoplasmático sendo imprescindível para a formação das partículas virais. (Chambers et al. 2003)(Chambers et al. 1990)(Wilschut 2010)

A proteína M faz parte do envelope lipídico do DENV em estado maduro, no entanto, esta é sintetizada pela proteína glicosilada prM; esta forma precursora da proteína M está relacionada com a maturação dos viriões (partículas virais infecciosas) e ocorrem em ambientes com pH ácido e, durante o seu transporte através do Complexo de Golgi. Esta prM também interage com a proteína E, de modo a evitar erros de dobramento nesta aquando da maturação viral. (Barbosa- et al. 2009)

1.4 Glicoproteína E

Os estudos realizados no vírus do Dengue, têm sido feitos nas diversas estruturas deste. Muitos dos estudos incidem nas proteínas não-estruturais e, outros nas proteínas estruturais. Uma das proteínas que tem vindo a ser intensamente estudada é a Glicoproteína E, devido ao papel importantíssimo na fusão da membrana do vírus com a membrana da célula hospedeira.

Esta é a maior componente glicoproteica na superfície do DENV. Esta proteína é importante pois é necessária em diversos processos, participando no processo preliminar da infeção viral, uma vez que esta atua como proteína ligante que irá interagir com recetores das células alvo, mediando assim a endocitose da partícula viral. Também é essencial no processo de fusão do vírus ao endossoma, participando na dissociação do nucleocapsídeo no citoplasma. Esta glicoproteína é considerada o principal alvo para a indução da resposta imunoprotectora baseada em anticorpos neutralizantes. (Santos Soares 2013)

A proteína E é constituída por um dímero, que se estende paralelamente à bicamada lipídica da partícula viral, que é formado por dois monómeros em que cada um deles é composto por um ectodomínio que corresponde à junção de três domínios: I, II e III. (Allison et al. 1999)

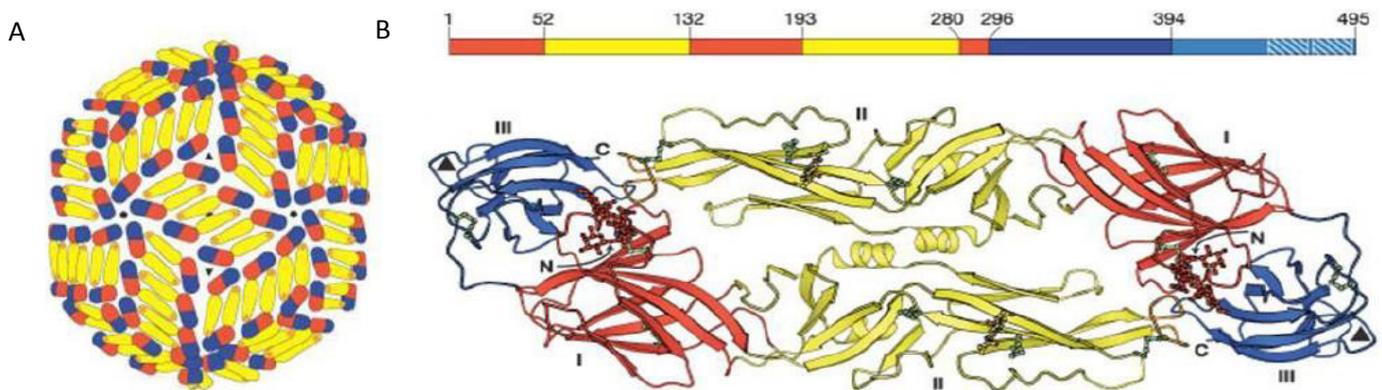


Figura 1.6- (A) Glicoproteína E com Domínio I representado a vermelho, Domínio II a amarelo e Domínio III a azul-escuro. (B) Conformação dimérica da Glicoproteína E. Adaptado de Allison et al., 1999.

O domínio I (DI) contém cerca de 120 aminoácidos e, encontra-se entre o domínio II (DII) e o domínio III (DIII), estando assim localizado no centro da proteína. Este domínio, apresenta vários epítomos (chamados também de determinantes antigénicos) que induzem resposta imune a anticorpos, essencialmente, não-neutralizantes mas também com atividade neutralizante. (Roehrig et al. 1998)

O domínio II contém cerca de 180 aminoácidos e, contém na sua estrutura a região que é responsável pela dimerização dos monómeros da proteína E. DII apresenta-se em forma de dedo e, contém na sua extremidade o peptídeo de fusão (*loop*) com uma sequência de aminoácidos hidrofóbicos, conservada em todos os *Flaviviridae*, e que apenas é exposta na presença de pH ácido. Esta região da proteína E é então responsável pela fixação da glicoproteína à membrana da célula alvo que constitui uma etapa essencial do processo de infeção do vírus. (Zhang et al. 2014)(Modis et al. 2003)(Modis et al. 2004)

O domínio III contém aproximadamente 98 aminoácidos e, tem vindo a ser muito estudado como um importante alvo na produção de anticorpos monoclonais e vacinas, uma vez que este é responsável pela interação entre o vírus e a célula alvo. Este domínio, contém na sua estrutura locais onde se ligam recetores das células humanas, promovendo a endocitose da partícula viral. DIII está localizado na região C-terminal da proteína E.

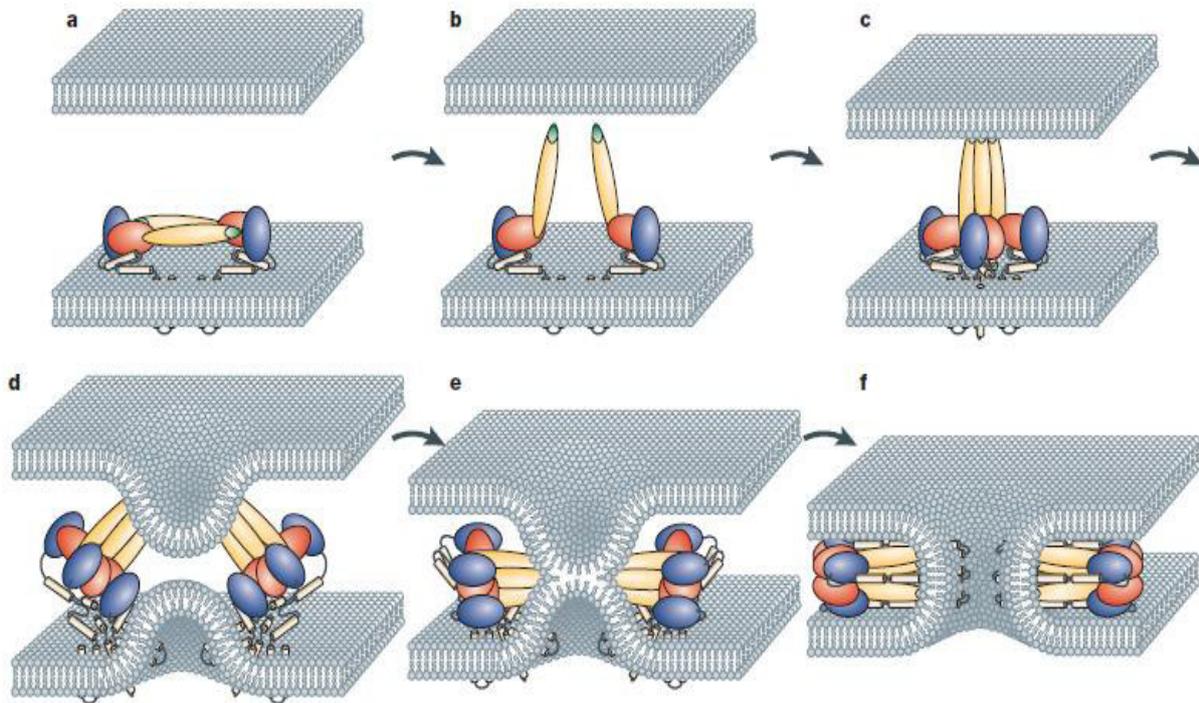


Figura 1.7 - Processo de Fusão das membranas, com o auxílio da Glicoproteína E. Adaptado de Mukhopadhyay et al. 2005

Através da Figura 1.7, é possível analisar o processo de fusão entre a membrana viral e a membrana da célula hospedeira, com o auxílio da Glicoproteína E. O primeiro passo no processo de infecção viral é a ligação a um recetor (por exemplo o sulfato de heparina ou CD209) na superfície da célula hospedeira. Neste ponto, a Glicoproteína é constituída por um dímero, como referido anteriormente, que contém péptidos de fusão (representado a verde). Após a ligação a um recetor, e a conseqüente exposição às condições de pH ácido no endossoma, a glicoproteína forma uma conformação prolongada e o péptido de fusão estende-se através da membrana da célula hospedeira. O processo de inserção dos péptidos de fusão no folheto exterior da membrana, desencadeia uma trimerização da Glicoproteína E. O Domínio III desta, dobra para trás de si própria, trazendo consigo a membrana viral de encontro à membrana da célula hospedeira. Com o movimento do Domínio III no sentido do Domínio II, ocorre a hemifusão das membranas lipídicas, ficando as regiões transmembranares e os péptidos de fusão muito próximos. (Mukhopadhyay et al. 2005)

Esta mudança conformacional da proteína E irá, obviamente, facilitar a libertação do ARN viral para o citoplasma das células hospedeiras, uma vez que a fusão das membranas (virais e

celulares) irá gerar um canal através do qual o material genético irá entrar para o interior da célula, onde irá ocorrer a tradução e replicação do ARN viral. (Faheem et al. 2011)(Rodenhuis-zybert 2010)(Melino & Paci 2007)

A Glicoproteína E constitui assim um alvo preferencial no desenvolvimento de abordagens terapêuticas que possam interferir na fusão do vírus com as células hospedeiras e, conseqüentemente com o ciclo de vida do vírus. Para que haja sucesso neste tipo de investigações é fundamental a existência de informação detalhada acerca da proteína em causa. Neste sentido, as simulações computacionais podem ser uma mais-valia importante no estudo do comportamento da estrutura da proteína E em condições próximas às fisiológicas, trazendo á luz novas informações sobre os mecanismos de ação do DENV.

1.5 Métodos Computacionais para os estudos da interação proteína-ligante

1.5.1 Campos de Força Empíricos

Para o estudo das estruturas e funções de sistemas biológicos é possível usar modelos moleculares baseados na mecânica clássica. Esta apresenta formas funcionais simples o que, permite a execução de um grande número de cálculos num curto espaço de tempo, permitindo por sua vez o estudo de sistemas biológicos com maior grau de complexidade molecular.

Um campo de força empírico é então caracterizado como sendo o conjunto de funções de energia com um conjunto de parâmetros a elas associados. Os parâmetros não são únicos para todos os tipos de sistemas, estes devem ser ajustados para que o campo de força esteja de acordo com resultados obtidos através de cálculos de mecânica quântica, em moléculas simples, e com resultados experimentais realizados para sistemas modelo.

Um campo de forças usado em mecânica molecular é definido como sendo o somatório de vários termos de energia, que são divididos por termos de energia relativamente a átomos ligados quimicamente – onde se incluem os ângulos, diedros e comprimentos de ligações – e, a átomos não-ligados quimicamente – onde se considera as interações de Van Der Waals e de energias de repulsão e atração electrostática. (Namba et al. 2008)

Na Equação 1.1, temos a expressão típica para um campo de forças empírico, definindo a energia potencial total do sistema. Os primeiros três termos descrevem as interações de átomos ligados quimicamente e, o último as interações de átomos não ligados.

Equação 1.1

$$U(r^N) = \sum_{\text{ligação}} \frac{k_b}{2} (l_i - l_{i,0})^2 + \sum_{\text{ângulo}} \frac{k_\theta}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{diedro}} \frac{v_i}{2} (1 + \cos(n\omega_i - \gamma_i)) + \sum_{i=1}^N \sum_{j>i}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

O primeiro termo da equação descreve, com um potencial harmónico, a interação entre dois átomos ligados entre si. O segundo termo, tal como indicado, descreve a modelação dos ângulos

de valência. E, por fim, o terceiro descreve potencial resultante da torção do ângulo de uma ligação. O quarto termo está associado às interações de átomos não ligados. Incluídos estão os átomos da mesma molécula, que se encontram a pelo menos três ligações de distância um do outro ou, de diferentes moléculas. As interações entre dois átomos não ligados, incluem o potencial de Lennard-Jones associado à descrição das interações de Van Der Waals, e o potencial de Coulomb para interações eletrostáticas. Nas secções seguintes vamos descrever em detalhe os vários termos desta equação.

1.5.1.1 Interações Ligantes

- **Ligações**

A Lei de Hooke é responsável por restringir, harmonicamente, o comprimento da ligação covalente entre dois átomos em relação a uma distância de referência. Deste modo, é possível conhecer o potencial de uma ligação entre dois átomos. (Equação 1.2)

Equação 1.2

$$U(l) = \frac{1}{2}k_b(l - l_0)^2$$

O potencial de ligação é conseguido através dos parâmetros k_b que corresponde à constante de elasticidade de ligação dos átomos e, $(l - l_0)$ será a variação do comprimento de ligação, sendo que l_0 , chamado de “comprimento de ligação de referência”, consiste no valor de comprimento que a ligação tem quando todos os termos do campo de força são ajustados para zero. Como se pode verificar na Figura 1.8, no potencial de ligação entre dois átomos, há um perfil simétrico em relação a l_0 , correspondente ao mínimo observado.

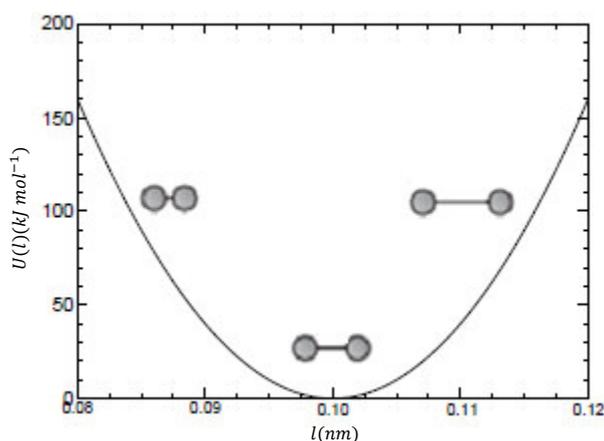


Figura 1.8 - Potencial de Ligação entre dois átomos usando restrição harmônica.
Adaptado de Van der Spoel et al. 2010

O potencial harmónico não descreve corretamente sistemas que apresentem elevadas deformações de comprimento de ligação, levando à sua rotura. No entanto na maior parte dos sistemas biomoleculares estes acontecimentos são muito raros.

- **Ângulos de ligação**

A forma funcional de um ângulo de ligação entre três átomos, também pode ser definida recorrendo à Lei de Hooke, sendo assim um potencial harmónico. (Equação 1.3)

Equação 1.3

$$U(\theta) = \frac{1}{2} k_{\theta} (\theta - \theta_0)^2$$

O potencial harmónico para o ângulo de ligação é então definido por dois parâmetros: k_{θ} correspondente à constante elástica e, o ângulo de referência θ_0 .

Analisando a Figura 1.8 e a Figura 1.9, verifica-se que é necessária uma menor energia para modificar um ângulo para longe do estado de equilíbrio, em comparação com a energia necessária para comprimir ou esticar uma ligação entre dois átomos.

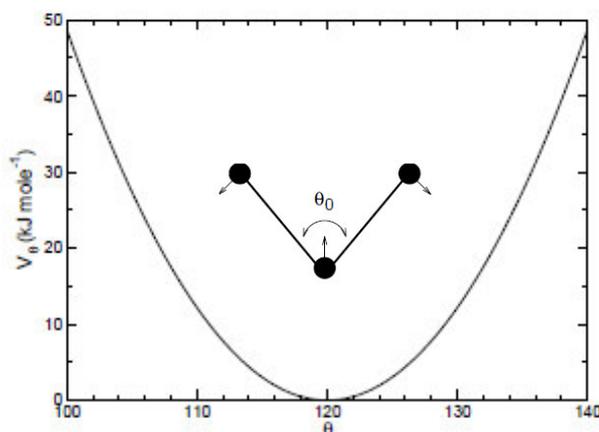


Figura 1.9 - Potencial do ângulo de ligação entre três átomos com restrição harmónica. Adaptado de van der Spoel et al. 2010

- **Diedros**

Um ângulo diedro é caracterizado por ser formado entre dois planos, sendo que são constituídos por quatro átomos ligados ao longo de três ligações.

Como exemplo de diedro, temos a Figura 1.10, onde estão representados os quatro átomos (**i**, **j**, **k** e **l**) ligados entre si por três ligações. Os dois planos, neste caso concreto, seriam o plano **i, j** e **k** e, o segundo, **j, k** e **l**.

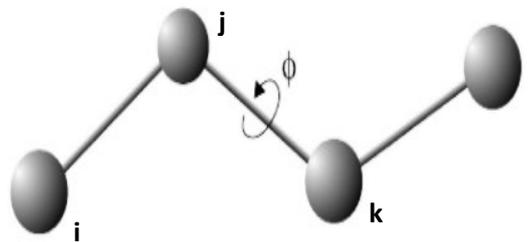


Figura 1.10 - Exemplo de um diedro.

A rotação de uma ligação tem uma grande influência nas propriedades estruturais de uma molécula, podendo por exemplo, gerar alterações locais no arranjo espacial de uma cadeia de um polímero. Estas rotações das ligações ocorrem com mais facilidade que a variação de um comprimento de ligação ou da alteração de um ângulo entre três átomos, uma vez que a rotação necessita de menos energia e, apresenta uma grande flexibilidade. (Rapaport 2004)

O perfil de energia associado à rotação de uma ligação é definido pela forma funcional do diedro, especificada na Equação 1.4.

Equação 1.4

$$U(\omega) = \sum_{n=0}^N \frac{v_n}{2} [1 + \cos(n\Phi - \gamma)]$$

A função do diedro (ω), é conseguida através de vários parâmetros. O parâmetro v_n define a altura da barreira de energia existente em relação à rotação. O parâmetro N é a multiplicidade, responsável por determinar o número de mínimos na função em questão. Por fim, o parâmetro γ , é o fator de fase responsável por determinar a posição do mínimo de energia.

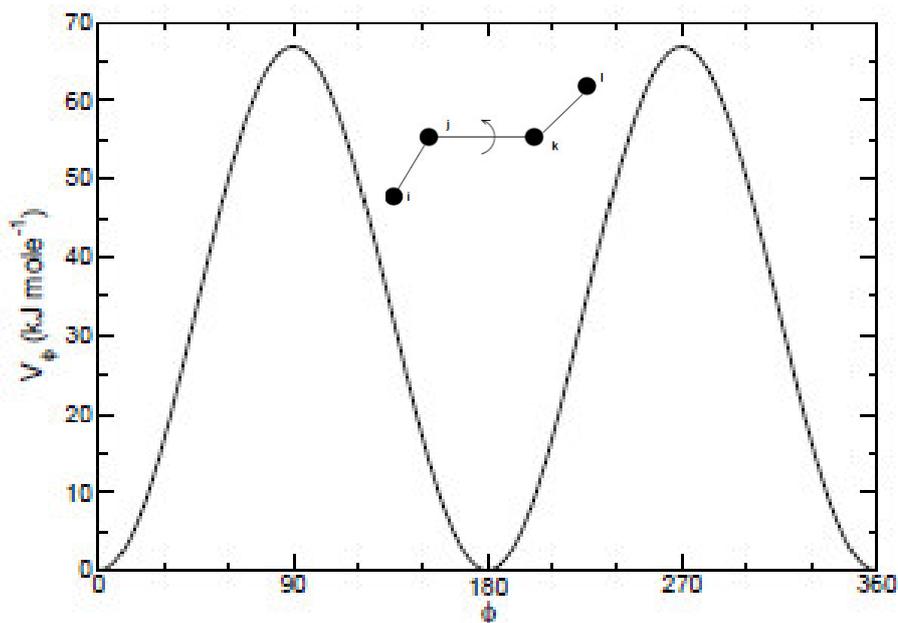


Figura 1.11 - Potencial do ângulo diedro. Adaptado de van der Spoel et al. 2010

Até aqui, falou-se de diedros próprios que são os que geralmente estão presentes como um termo do campo de forças empírico. No entanto, existem os diedros impróprios que permitem restringir a geometria das moléculas, sendo responsáveis por manter quiralidade em tetraedros ou a planaridade. (Figura 1.12)

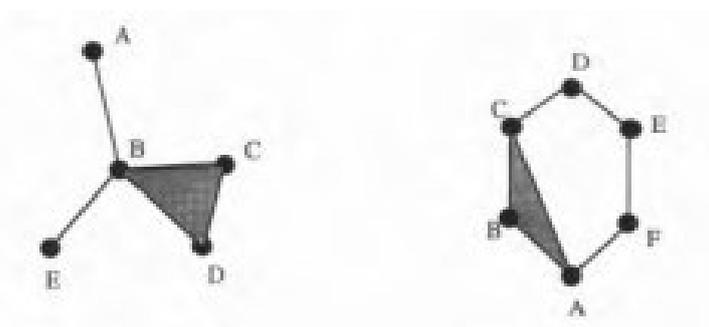


Figura 1.12 - À esquerda, está representado a aplicação de um diedro impróprio para manter a quiralidade, mantendo o ângulo formado pela ligação AB com o plano BCD. À direita, usa-se os diedros impróprios para manter a planaridade, mantendo-se a ligação CD no plano ABC, em que o ângulo de equilíbrio será zero. Adaptado de Soares 1997

O potencial dos ângulos diedros impróprios (ξ) podem ser definidos recorrendo a potenciais harmônicos. (Equação 1.5)

Equação 1.5

$$U(\xi_{ijkl}) = \frac{1}{2} k_{\xi} (\xi_{ijkl} - \xi_0)^2$$

Uma vez que estes são definidos harmonicamente, o seu potencial é semelhante aos restantes termos que contêm restrições harmônicas, como se pode verificar pela Figura 1.13. O ângulo diedro impróprio é definido como o ângulo entre os planos (i,j,k) e (j,k,l) .

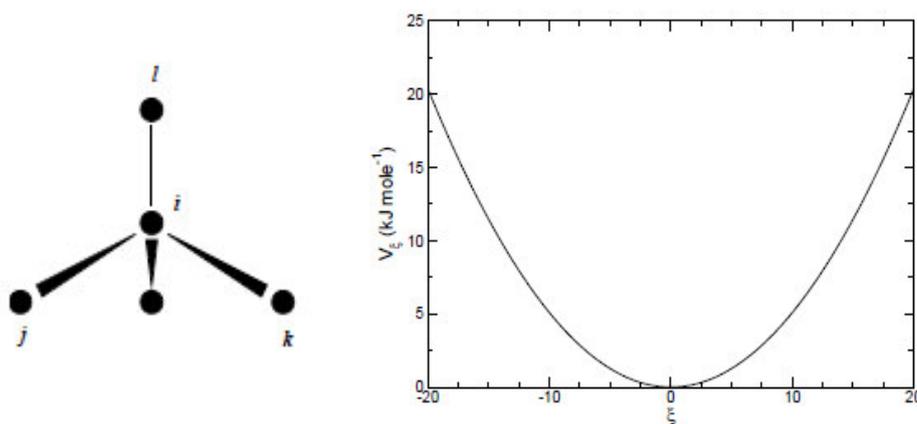


Figura 1.13 - À esquerda, o ângulo diedro impróprio num tetraedro. À direita, o potencial do ângulo diedro impróprio. Adaptado de van der Spoel et al. 2010

1.5.1.2 Interações não Ligantes

- **Interações Eletrostáticas**

Os fenômenos eletrostáticos surgem a partir das forças que as cargas elétricas exercem entre si. Estas cargas são concebidas para reproduzir as propriedades eletrostáticas de uma molécula e, a distribuição destas podem ser representadas de variadas formas. Geralmente, a abordagem consiste no fracionamento da carga da molécula resultando num conjunto de cargas pontuais distribuídas pelos vários átomos ao longo da molécula. Assim, a interação eletrostática entre duas moléculas, ou entre diferentes partes da mesma molécula, é calculado como a soma das interações entre pares de cargas pontuais, usando a Lei de Coulomb. (Equação 1.6)

Equação 1.6

$$U = \sum_{i=1}^{N_A} \sum_{\substack{j=1 \\ j>i}}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

O cálculo do potencial eletrostático, depende dos parâmetros N_A e N_B que são os números de cargas pontuais em duas moléculas (A e B). Os parâmetros q_i e q_j correspondem às cargas em cada átomo; ϵ_0 é a constante elétrica que tem um valor já definido e, por fim o parâmetro r_{ij} que corresponde à distância entre as duas cargas.

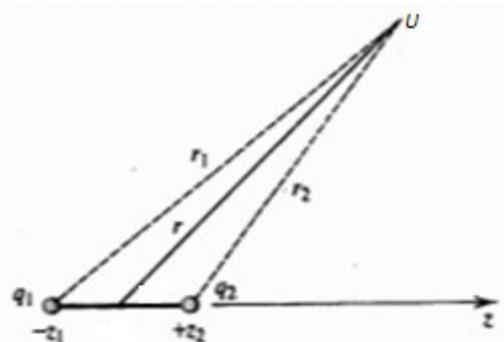


Figura 1.14- Representação de um potencial Eletrostático resultante resultante de duas cargas pontuais. Adaptado de Leach 2001

- **Interações de Van Der Waals**

As interações eletrostáticas não conseguem englobar todas as interações de átomos não ligados num sistema. Certas interações eletrostáticas tipo dipolo-dipolo poderão existir momentaneamente entre átomos, devido à interação das respetivas nuvens eletrónicas. Estas interações podem ser descritas por forças de Van Der Waals, que são caracterizadas como sendo uma soma de várias contribuições de forças repulsivas e atrativas.

A função mais conhecida por representar o potencial das interações de Van Der Waals, é a função de Lennard-Jones, descrevendo a interação entre dois átomos. (Lennard-Jones 1924)

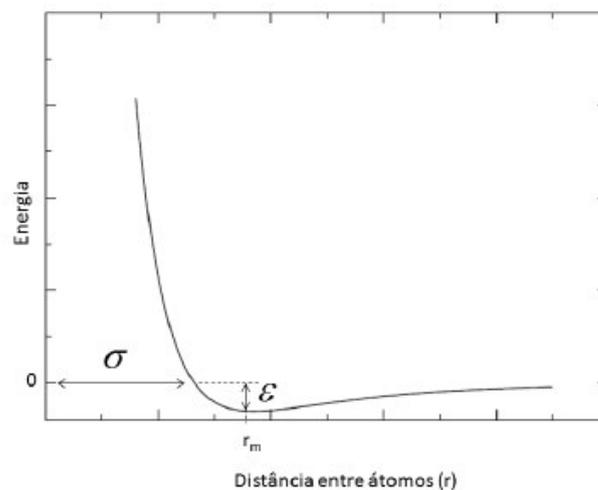


Figura 1.15 - Potencial de Lennard-Jones. Adaptado de van der Spoel et al. 2010

O potencial de Lennard-Jones é caracterizado por descrever a variação das forças atrativas com r^{-6} e as forças repulsivas com r^{-12} e, por essa razão é conhecido como a função Lennard-Jones 12-6. (Equação 1.7)

Equação 1.7

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

Neste potencial, apenas dois parâmetros são ajustados, sendo eles o parâmetro σ que corresponde à distância na qual a energia é nula. O parâmetro ϵ corresponde à profundidade do “poço” do potencial.

- **Tratamento da Interações não Ligantes**

Nas simulações de Mecânica Molecular, o cálculo das energias e/ou forças das interações não ligantes é a parte mais dispendiosa do processo. O custo de avaliação das interações ligantes é proporcional ao número de átomos do sistema, enquanto para os termos não ligantes varia com N^2 . Este tipo de interações é calculado entre todos os pares de átomos do sistema, o que nem sempre se justifica. De modo a limitar o número de interações a ser avaliadas, é usado um raio de *cut-off* que define a distância a partir da qual as interações não ligantes entre pares de átomos deixa de ser considerada, o que faz sentido dado que estas interações variam rapidamente com o inverso da distância.

A definição da distância do *cut-off* está relacionada com a dimensão da caixa em que está inserido o sistema em estudo, usando o método da *minimum image convention*, como representado na Figura 1.16. A implementação deste método pressupõe a definição das condições fronteira periódicas em que se supõe que o sistema original é envolvido por um conjunto de réplicas do sistema original. Deste modo, a distância de *cut-off* deve ser sempre inferior a metade do comprimento da dimensão da célula menor, para garantir que os átomos não interagem com as suas imagens. Em *minimum image convention*, a energia é calculada com o átomo mais próximo ou imagem. Analisando a Figura 1.16, o círculo representa o *cut-off*, e neste caso, as moléculas que se encontram dentro do círculo terão uma contribuição no átomo 1. No entanto, os átomos 4, 5 e 6, por exemplo, não irão ter qualquer contribuição no átomo 1, uma vez que os seus centros não se encontram no interior do círculo. (Cheung et al. 2008)

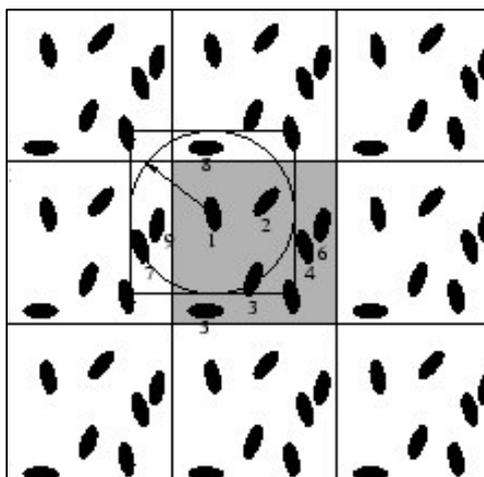


Figura 1.16 - Minimum image convention num sistema com duas dimensões. A caixa central contém cinco moléculas, assim como a caixa construída contendo a molécula 1 no centro. O círculo representa o potencial *cut-off*. Adaptado de Cheung et al. 2008

Quando o *cut-off* é utilizado, as interações entre todos os pares de átomos que estão mais distantes do que a distância de *cut-off*, são postos a zero, tendo em conta a imagem mais próxima. (Figura 1.16) (Fadrná et al. 2005)(Leach 2001)

Este método, no entanto, não é muito eficaz no caso das interações de longo alcance, como é o caso das interações eletrostáticas, que decaem gradualmente com a distância. Isto ocorre, porque este método não permite uma descrição rigorosa das interações na zona junto ao limite de *cut-off*. Para estes casos, em que o *cut-off* não é suficientemente rigoroso, existe o tratamento de interações eletrostáticas de longo alcance: *Particle Mesh Ewald* (PME).

Este método é conhecido por conseguir considerar todas as interações entre cargas dentro do mesmo sistema, e as interações entre as cargas e as imagens respetivas. Ewald (1921), propôs este método em que uma partícula interage com todas as outras partículas da caixa de simulação e com todas as suas imagens numa infinita variedade de células periódicas. A Figura 1.17 ilustra como a matriz de células de simulação é construída. A estratégia é criar caixas por ordem de proximidade à caixa central, normalmente as células unitárias são adicionadas em sequência, sendo que: o primeiro termo tem $|n| = (0,0,0)$; o segundo termo será $|n| = L$ ("L" corresponde ao comprimento de lado da caixa) que contém as seis caixas centradas em $n = (\pm L, 0, 0)$, $(0, \pm L, 0)$ e $(0, 0, \pm L)$.

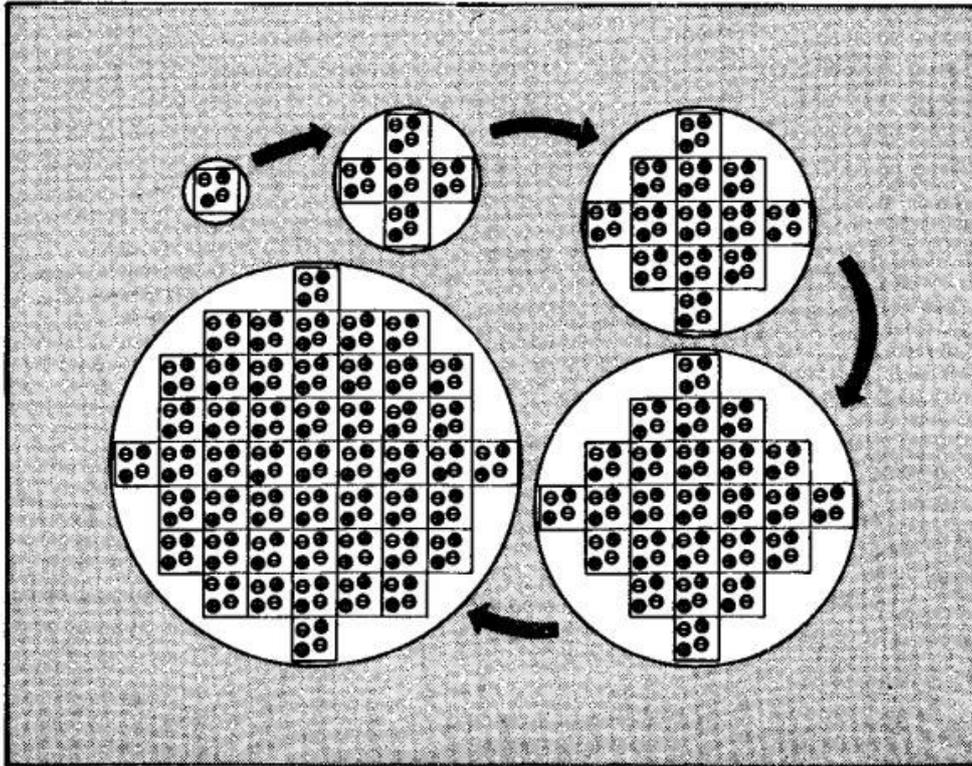


Figura 1.17 - Construção da esfera com caixas de simulação. Adaptado de Cheung et al. 2008

A interação entre pares de cargas na caixa de simulação central (célula unitária $n=0$) é descrita baseando-se na Lei de Coulomb. (Equação 1.8)

Equação 1.8

$$U = \frac{1}{2} \sum'_{n=0} \sum_{i=1}^N \sum_{\substack{j=1 \\ j>i}}^N \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij} + n|}$$

Neste caso, as interações entre as cargas pontuais, q_i e q_j , que se encontram dentro da célula unitária são tidas em consideração, no entanto as interações em que $i=j$ são ignoradas, assim como as imagens presentes a uma distância $|r_{ij} + n|$.

1.5.1.3 Minimização de energia

Na modelação molecular têm-se especial interesse em encontrar pontos mínimos na superfície da energia potencial, uma vez que estes pontos mínimos correspondem a estados estáveis do sistema. O processo de Minimização de energia é um procedimento numérico criado para encontrar um mínimo da superfície de energia potencial, a partir de uma estrutura inicial de energia mais elevada, variando a posição dos átomos até encontra-lo. (Figura 1.18) No final deste processo, poderá existir vários mínimos de energia, no entanto, o que apresentar menor energia será considerado o mínimo de energia global.

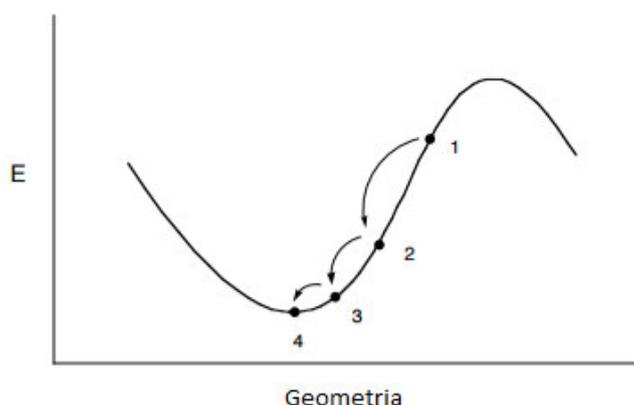


Figura 1.18 - Processo de Minimização de energia modificando a geometria da molécula até encontrar um mínimo. Adaptado de Standard 2015

Para ter acesso às diferentes conformações associadas a pontos mínimos da superfície de energia, é possível recorrer a alguns programas de minimização. Um programa capaz de realizar minimização de energia, é o SANDER (Simulated Annealing with NMR-Derived Energy Restraints) presente no *software* AMBER. Este é um programa básico de minimização de energia e dinâmica molecular, que suporta vários tipos de algoritmos de minimização entre eles – os mais comuns – método *steepest descent* e o gradiente conjugado. (Case et al. 2011)

Geralmente, a minimização de energia recorrendo ao SANDER é feita com a combinação dos algoritmos referidos anteriormente. Os primeiros passos são feitos usando o *steepest descent*, pelo facto de este ser um método robusto, que é eficiente na procura dos mínimos mesmo que o

ponto inicial de minimização esteja bastante afastado. Os passos seguintes são geralmente feitos recorrendo ao método do gradiente conjugado, que consegue uma convergência mais rápida que o método anterior.

No primeiro método, a direção pela qual segue a primeira minimização é a do gradiente mais íngreme, em relação ao ponto inicial. Encontrado um mínimo, é a partir deste que começa uma segunda minimização e, do mesmo modo que a primeira minimização, a sua direção será a do gradiente mais íngreme encontrado. Este processo é repetido até atingir um mínimo em todas as direções ou, até atingir o número de iterações definido. O *steepest descent* é usado na parte inicial da minimização, quando a estrutura está longe de ser a conformação com menor energia.

O método de gradiente conjugado inicia-se da mesma maneira que o método anterior, no entanto, difere nos passos seguintes. Este método tem sempre em consideração as direções anteriores, evitando assim grandes oscilações da superfície de potencial e, chegando mais rapidamente ao mínimo. Assim, consegue-se atingir a convergência mais rapidamente, uma vez que não é necessário passos adicionais até chegar ao mínimo, seguindo um caminho “direto” até lá. (Standard 2015)

1.5.2 Dinâmica Molecular

A Dinâmica Molecular é caracterizada como sendo uma técnica de simulação computacional de sistemas complexos que são modelados a nível molecular. Neste processo, as equações de movimento (Equações de Newton) são resolvidos numericamente de modo a existir uma evolução no tempo do sistema, permitindo a derivação de variadas propriedades de interesse, como as termodinâmicas e cinéticas.

Esta técnica de simulação é fundamentada nos princípios da Mecânica Clássica. Deste modo, para que se consiga informações sobre o comportamento dinâmico de determinado sistema, é necessário a aplicação da mecânica estática e da mecânica molecular. A primeira, terá a responsabilidade de realizar os cálculos das propriedades observáveis a nível macroscópico a partir das propriedades microscópicas, conseguindo-se assim valores de pressão, energia interna, volume, temperatura, entropia, etc. A mecânica molecular irá permitir que as moléculas sejam tratadas como um conjunto de átomos, sujeitos a uma série de potenciais de interação que os mantêm unidos, e definidos como sendo campos de forças empíricos.

Para que seja possível gerar DM a partir de modelos de potencial contínuos, como é o caso dos campos de forças, são utilizados alguns métodos de integração de diferenças finitas, descritos em seguida.

- **Algoritmo de Verlet**

Técnicas de diferenças finitas são usadas para criar trajetórias de dinâmica molecular com modelos de potenciais contínuos. Para qualquer arranjo dos átomos no sistema, a força que atua sobre cada átomo devido às interações com outros átomos pode ser calculada por diferenciação da função de energia. A partir da força em cada átomo é possível determinar a sua aceleração através da segunda lei de Newton. A integração das equações do movimento deve então produzir uma trajetória capaz de descrever como as posições, velocidades e acelerações das partículas variam com o tempo.

Matematicamente falando, a trajetória de átomos em função de tempo será dada por:

$$\mathbf{r}^N(t)$$

Em seguida, encontrar-se-á as posições para intervalos de tempo discretos, em que δt é o *time step*. Quanto menor for o valor de δt , mais precisa será a solução das equações de movimento.

$$\mathbf{r}^N(0), \mathbf{r}^N(\delta t), \mathbf{r}^N(2\delta t), \mathbf{r}^N(3\delta t), \dots$$

Existem vários algoritmos para a integração das equações de movimento usando métodos de diferença finita, que são implementados para os cálculos de dinâmica molecular. Todos os algoritmos assumem que as posições e as propriedades dinâmicas – velocidades, acelerações – podem ser aproximadas como séries de Taylor:

Equação 1.9

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \frac{1}{6} \delta t^3 \mathbf{b}(t) + \frac{1}{24} \delta t^4 \mathbf{c}(t) + \dots$$

Equação 1.10

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \delta t \mathbf{a}(t) + \frac{1}{2} \delta t^2 \mathbf{b}(t) + \frac{1}{6} \delta t^3 \mathbf{c}(t) + \dots$$

Equação 1.11

$$\mathbf{a}(t + \delta t) = \mathbf{a}(t) + \delta t \mathbf{b}(t) + \frac{1}{2} \delta t^2 \mathbf{c}(t) + \dots$$

Como se pode verificar através das equações anteriores, a velocidade (\mathbf{v}) resulta da primeira derivada das posições (\mathbf{r}) em função do tempo; a aceleração (\mathbf{a}) é dada pela segunda derivada.

O algoritmo Verlet é o método de integração das equações de movimento mais usado em simulações de dinâmica molecular. Este algoritmo usa as posições e acelerações no instante t , e as posições resultantes do passo anterior, $\mathbf{r}(t - \delta t)$, para calcular novas posições no instante $(t + \delta t)$, ou seja, $\mathbf{r}(t + \delta t)$. (Equação 1.12)

Equação 1.12

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t)$$

No algoritmo de integração de Verlet, as velocidades não aparecem explicitamente, neste caso uma das possibilidades é estimar as velocidades para metade do *time step*, $t + \frac{1}{2}\delta t$. (Equação 1.13)

Equação 1.13

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t)] / \delta t$$

O facto de as velocidades terem de ser calculadas “à parte”, torna a obtenção das velocidades mais complicada e, é necessário que as posições do *time step* seguinte sejam calculadas para que seja possível obter as velocidades. De modo a minimizar essa desvantagem, foi criado o método da velocidade de Verlet (Swope 1982) que é capaz de calcular as posições, velocidades e acelerações ao mesmo tempo, sem que a precisão seja comprometida.

Equação 1.14

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t)$$

Equação 1.15

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2} \delta t [\mathbf{a}(t) + \mathbf{a}(t + \delta t)]$$

Através da Equação 1.14, verifica-se que para o cálculo de novas velocidades é necessário as acelerações no tempos t e $(t + \delta t)$. Assim, na primeira etapa, as posições no tempo $(t + \delta t)$ são calculadas de acordo com a Equação 1.15, usando as velocidades e acelerações do tempo t . As velocidades para metade do *time step*, $t + \frac{1}{2}\delta t$, são determinadas pela seguinte equação:

Equação 1.16

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = \mathbf{v}(t) + \frac{1}{2}\delta t\mathbf{a}(t)$$

Por fim, novas forças são calculadas a partir das últimas posições, tendo assim $\mathbf{a}(t + \delta t)$. No último passo, as velocidades são por fim calculadas para o tempo $(t + \delta t)$. (Equação 1.17)

Equação 1.17

$$\mathbf{v}(t + \delta t) = \mathbf{v}\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t\mathbf{a}(t + \delta t)$$

- **Termóstatos**

A segunda lei de Newton preserva a energia total do sistema, e uma integração simples da segunda lei de Newton, por conseguinte, conduz a simulações que preservam a energia total do sistema (E), o número de átomos (N) e o volume do sistema (V). No entanto se pretendermos realizar simulações a temperatura constante, em vez de energia constante, o sistema em estudo é acoplado a um reservatório térmico e é definido um termostato para manter constante a temperatura do sistema. Os termostatos mais utilizados são Langevin, Nosé-Hoover e Berendsen.

- **Solvatação**

Grande parte dos processos em sistemas biológicos ocorrem num meio aquoso, e por isso é importante perceber de que modo é que o solvente afeta o comportamento de determinado sistema. Existem sistemas em que as moléculas de solvente estão diretamente envolvidas em reações ou então, encontram-se tão fortemente ligadas que passam a ser parte integrante do soluto. Nesses casos, as moléculas de solvente devem ser modeladas explicitamente. Noutros sistemas, o solvente não interage diretamente com o soluto mas, fornece um ambiente que afeta fortemente o comportamento do soluto. Nestes casos, o solvente atua como uma perturbação do

comportamento da fase gasosa do sistema e, para isso existem os modelos de solvente contínuo. (Leach 2001)

Os modelos de solvente contínuo têm uma grande economia em termos computacionais e podem fornecer uma razoável descrição do comportamento do solvente. Os modelos de solvente explícito, por sua vez, são mais dispendiosos em termos computacionais, mas fornecem uma boa descrição física do solvente. Dos modelos de solvente explícito mais comuns, existe o modelo *Transferable Intermolecular Potential 3P* (TIP3P) (Jorgensen et al. 1983), este tem três sítios de interação eletrostática, sendo que as duas cargas parciais positivas são nos dois átomos de hidrogénio e, a negativa no átomo de oxigénio. As interações de Van Der Waals têm o seu ponto de interação no átomo de oxigénio.

AMBER

O software AMBER é caracterizado por ser um conjunto de programas que permite aos utilizadores realizar simulações de dinâmica molecular, particularmente em biomoléculas. Os programas incluídos no AMBER, trabalham em conjunto de modo a fornecer uma estrutura que será sujeita a vários cálculos.

O termo AMBER é também usado para fazer referência a um campo de força, no entanto, é importante referir que este não é exclusivo para este software e, que este software não contém apenas este campo de forças.

O AMBER está dividido em dois programas principais: O AmberTools15 e o Amber14. O primeiro, consiste num conjunto de programas de simulação e análise biomolecular; o segundo é baseado no programa pmemd (que será descrito à frente). É possível realizar várias simulações recorrendo apenas ao AmberTools15, no entanto, a combinação de ambos, permite a realização de simulações mais extensas e complexas. (D.A. Case, et al. 2015)

Todos os programas de simulação têm o seu método de atuação. Neste caso, o AMBER tem alguns passos estabelecidos, que são realizados recorrendo aos diferentes programas que este disponibiliza. (Figura 1.19)

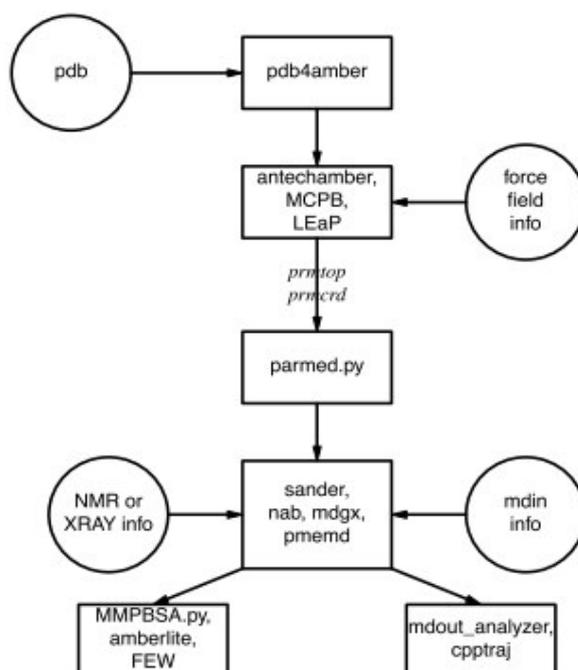


Figura 1.19 - Fluxograma para realização de simulações no AMBER. Adaptado de D.A. Case, et al. 2015

Os passos iniciais, que constituem o processo de preparação do sistema para as simulações, podem estar associados a erros de simulações e, por essa razão têm de ser feitos com especial atenção.

Para a realização das simulações, o sistema que se pretende estudar deve conter as coordenadas cartesianas de cada átomo (geralmente, os ficheiros que contêm as coordenadas têm o nome de “prmcdr”). Por norma, os sistemas vêm num formato Protein Data Bank (PDB), que podem não ser compatíveis com os programas AMBER, por essa razão há ficheiros que necessitam de algumas modificações. O programa LEaP permite fazer algumas edições, no entanto, também o *script pbd4amber* é capaz de o fazer.

O segundo passo para a realização das simulações é relacionado com a topologia, onde está incluído as conectividades, nomes e tipos de átomos e nome de resíduos (por norma, este ficheiro tem o nome “prmtop”). O software AMBER contém programas que têm bases de dados que usam para retirar informações acerca de aminoácidos padrão, ADN, ARN, açúcares e lípidos comuns. O *antechamber* é um programa do AMBER que permite o desenvolvimento de campos de força para moléculas “não-padrão” usando o *General Amber force field* (GAFF). O uso deste programa pode ser feito diretamente recorrendo ao LEaP.

De seguida, é necessário definir os parâmetros dos campos de força para todas as ligações, ângulos e tipos de átomos do sistema. No entanto, também nas bases de dados do AMBER, existem parâmetros padrão para vários campos de força.

Uma vez terminados os ficheiros de topologia, o AMBER dispõe de um script – o *parmed.py* – que examina se os parâmetros definidos na topologia são válidos para o sistema.

Posto isto, a parte de preparação é finalizada. Seguindo-se a parte de simulação, que é feita recorrendo a programas de simulação incluídos no AmberTools, como o programa SANDER. Posteriormente, procede-se à análise das simulações, recorrendo a programas de análise do AMBER, como o CPPTRAJ.

Programas integrados no AMBER

- **LEaP – Programa de Preparação**

LEaP é o nome genérico dado aos programas `tleap` e `xleap`, presentes no software AMBER. Através destes, é possível ler *inputs* contendo o conjunto de parâmetros, ler e editar ficheiros com formato PDB, interligar resíduos e criar complexos de moléculas não-ligados, modificar as coordenadas internas de uma molécula e criar ficheiros de topologia, coordenadas e parâmetros compatíveis com o software AMBER.

Assim, este programa consegue combinar e executar quatro módulos (*prep*, *link*, *edit* e *parm*) necessários para a preparação do sistema. Deste modo, o LEaP torna-se uma ferramenta muito utilizada no processo de simulações de dinâmica molecular, pela sua capacidade de realização de diversas tarefas. (D.A. Case, et al. 2015)

- **SANDER – Programa de Simulação**

Como já referido anteriormente, SANDER é o programa de minimização de energia e de dinâmica molecular do software AMBER.

Este programa relaxa a estrutura por iterações do movimento dos átomos, na procura de um mínimo de energia. As estruturas, por norma, são sujeitas a minimizações antes das simulações de DM. E por essa razão, este programa é uma boa opção, uma vez que este é capaz de realizar uma minimização de energia antes de iniciar os cálculos de DM.

Uma versão otimizada do SANDER que é usada normalmente, é a Particle Mesh Ewald Molecular Dynamics (`pmemd`) que foi criado com o objetivo de melhorar o desempenho dos métodos utilizados no SANDER. Esta versão suporta simulações de PME, e é compatível com campos de força AMBER e CHARMM. No `pmemd` é possível ainda, adicionar um acelerador de GPU, criando a versão `pmemd.cuda`. (Case et al. 2011)

- **CPPTRAJ – Programa de Análise**

O cpptraj é um programa versátil, incluído no AMBER, utilizado no pós-processamento da trajetória obtida com DM e para análise de dados. O processamento usando o cpptraj, é mais rápido do que as anteriores versões que o AMBER continha, essa melhoria é, particularmente, notória no processamento de trajetórias com formato NetCDF.

Este programa veio trazer várias ferramentas adicionais ao tratamento de dados obtidos da simulação de DM. Para além de ser capaz de processar ficheiros de trajetórias com diferentes topologias, ao mesmo tempo, é também capaz de extrair coordenadas, calcular valores de ligações, ângulos e diedros, variação de posições atômicas, etc. (Case et al. 2014)

MDPOCKET

Nas investigações feitas para novos fármacos, recorrendo a métodos computacionais, a maioria das vezes não se conhecem nem os ligantes, nem cavidades da macromolécula onde estes possam ser inseridos. Para que se conheça todas as cavidades existentes em determinada macromolécula, foram criados ao longo do tempo vários algoritmos de deteção de cavidades. Estes algoritmos têm vindo a melhorar, sendo agora capazes de detetar e caracterizar as cavidades, tendo em consideração o comportamento dinâmico da macromolécula.

MDpocket é um programa que foi criado com o objetivo de identificar e caracterizar locais de ligação e canais, que possam ser gerados durante simulações de DM. Este programa é baseado no Fpocket, que é um algoritmo “*geometry-based*”. Este tipo de algoritmos procura as cavidades por medições de tamanho e de outras propriedades na superfície da proteína. De modo geral, este algoritmo assume como possíveis sítios de ligação na proteína, as cavidades que apresentarem um volume considerável e, outras propriedades definidas pelo utilizador. (Erlanson Daniel A. et al. 2016)

Este programa baseia-se no conceito de α -esferas, citada por Liang et al. (1998). α -esferas são esferas que se encontram em contacto com quatro átomos, não contendo nenhum outro átomo dentro dela. Este conceito é utilizado para a deteção de cavidades, uma vez que assumem que

as esferas com raios pequenos encontram-se dentro da proteína, e as grandes, no exterior; As esferas com raios intermédios são interpretadas como sendo possíveis cavidades.

Por norma, a deteção das cavidades é feita depois da DM, permitindo assim o uso das diferentes conformações resultantes da simulação. O MDpocket irá atuar sobre cada conformação retirada da DM. Analisando a Figura 1.20, é fácil de entender como através do conceito das α -esferas é possível detetar cavidades na proteína. Assumindo que temos duas conformações diferentes (uma representada pela cor laranja e, outra a vermelho), cada uma terá as suas α -esferas associadas a um ponto i , na grade com espaçamento de 1 Angstrom. Este ponto é aplicado o mais próximo possível do centro da α -esfera. Cada ponto i , de cada conformação, está associado a um parâmetro de ocupação (δ) que apenas pode ser 0 ou 1, sendo que 1 indica que já há uma α -esfera associada ao ponto i em questão, caso contrário é 0. A normalização do número de α -esferas, por cada ponto de grade i , pelo número de conformações gera um mapa de densidade, que irá dar informações sobre o que rodeia as cavidades.

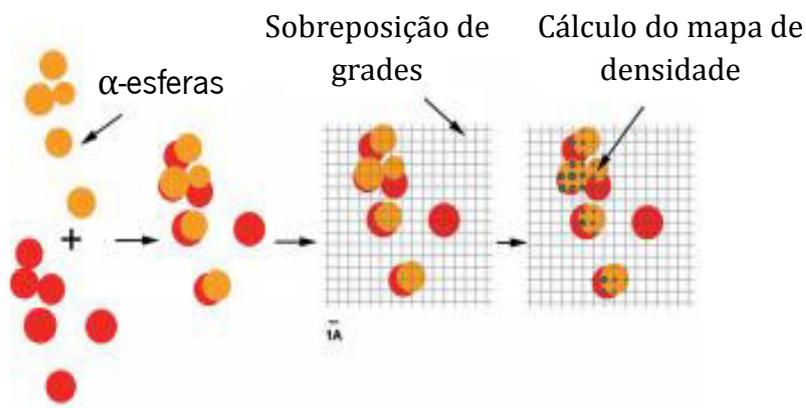


Figura 1.20- Conceito das α -esferas aplicado à deteção de cavidades. Adaptado de Schmidtke et al. 2010

Uma vez atribuídos os valores de δ_i , é possível gerar o mapa de frequências (φ) normalizando a soma de todos os δ_i , de todas as conformações (n). (Equação 1.18)

Equação 1.18

$$\varphi_i = \frac{1}{n} \sum_{i=1}^n \delta_i$$

Esta equação, permite que se conheça a frequência de abertura de cavidades durante as trajetórias provenientes da DM. Através do resultado da equação é possível saber se um determinado ponto da grade se encontra permanentemente acessível ($\varphi_i=1$) ou inacessível ($\varphi_i=0$). (Schmidtke et al. 2011)

O MDpocket permite assim fornecer inúmeras informações importantes. Através dos mapas – de densidade e frequência – é possível explorar a dinâmica (abertura e fecho) da cavidade. No entanto, no final do processo tem-se um ficheiro que contém os descritores, que incluem a área acessível de superfície e volume da cavidade, o número de α -esferas, a hidrofobicidade, entre outras informações que poderão ser úteis na caracterização da cavidade.

1.5.3. Docking Molecular

O Docking Molecular é um método computacional que prevê a ligação não-covalente de macromoléculas, ou entre uma macromolécula (receptor) e uma molécula pequena (ligante), começando com as estruturas não ligadas, obtidas através da Dinâmica Molecular, por exemplo. Enquanto que a Dinâmica Molecular lida diretamente com as energias (campo de forças), o Docking tem o objetivo de reproduzir potenciais químicos que determinem a conformação de ligação preferencial e a energia livre de ligação. O facto de este prever a orientação preferencial de uma molécula em ligação com outra, para formar um complexo estável, faz com que seja uma mais-valia na procura de novos fármacos. Normalmente, é usado para a modelação da interação entre uma pequena molécula (chamada também de ligante) e uma proteína, permitindo a caracterização do comportamento do ligante no local de ligação da proteína em estudo.

Fischer E. (1894) para explicar a ligação ligante-proteína, recorreu à teoria do sistema chave-fechadura. (Fischer 1894) Hoje, o mesmo se pode usar para explicar a atuação do Docking Molecular. Num sistema chave-fechadura, é necessário encontrar a orientação correta da chave, para a inserir no buraco da fechadura. Quando inserida na fechadura, a chave tem de sofrer uma rotação para abrir a fechadura. O mesmo ocorre num sistema ligante-proteína, em que o ligante é a chave e a proteína é a fechadura. (Figura 1.21)

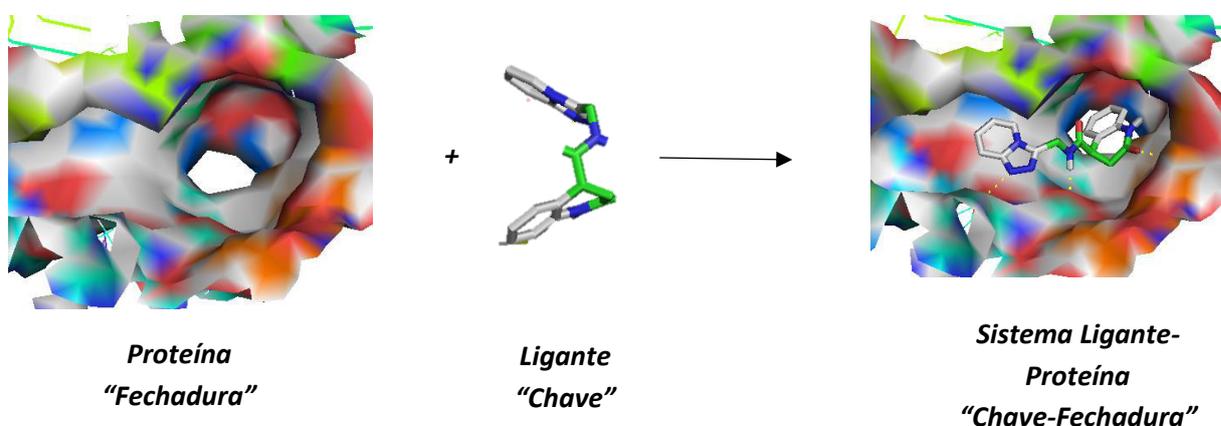


Figura 1.21 - Ilustração do docking de uma proteína com um ligante, a ligarem-se para formarem um complexo estável.

Para a realização do Docking, facilita que haja já um conhecimento prévio de possíveis cavidades onde se possa inserir um (ou mais) ligante(s), um método utilizado para esse fim, é o MDpocket. Uma vez conhecida uma cavidade, e inserido o ligante na mesma, é estudada a melhor orientação do ligante dentro da proteína e, depois avaliada a afinidade do ligante com a proteína. Assim, para realização do processo de Docking existem dois passos principais, sendo eles a previsão da conformação do ligante que inclui a posição e orientação do ligante na proteína, e a avaliação da afinidade da ligação entre a proteína e ligante. Para reproduzir as conformações do ligante é usado algoritmos de amostragem e, para as classificar em termos de energia, recorre-se a funções de *score*. (Meng et al. 2011)

Metodologias do Docking

A diferença entre as metodologias utilizadas no Docking estão relacionadas com a flexibilidade ou rigidez do ligante e proteína/recetor em estudo. Existem três metodologias principais:

- *Docking com Ligante Rígido e Recetor Rígido:* Nesta metodologia, quer o ligante quer o recetor são tratados do mesmo modo, considerando-se como estruturas rígidas. Isto deve-se à menor liberdade de rotação, assim como de espaço translacional.
- *Docking com Ligante Flexível e Recetor Rígido:* Esta é a metodologia mais utilizada nos softwares de Docking. O conceito é baseado no “ajuste induzido” (“*induced fit*”) da interação entre o ligante e o recetor, que consiste na alteração de conformações (flexibilidade) que são sujeitos, o ligante e o recetor, para que haja um encaixe perfeito, com um mínimo de energia. O que acontece é que, quando o recetor é flexível também, torna-se computacionalmente muito dispendioso. Com esta metodologia, procura-se atingir um balanço entre a precisão do encaixe e a exigência computacional. Deste modo, apenas o ligante é sujeito a mudanças de conformação, enquanto a proteína é mantida rígida durante o encaixe.
- *Docking com Ligante Flexível e Recetor Flexível:* Tal como referido anteriormente, esta seria a metodologia ideal para uma elevada precisão de encaixe, no entanto

este torna-se computacionalmente muito dispendioso e, por essa razão raros são os softwares que optam pela sua utilização. (Dastmalchi 2016)

AutoDock Vina

São vários os programas que podem ser utilizados para realizar o Docking molecular. O AutoDock Vina é um deles, que foi desenvolvido mais recentemente pela mesma equipa do AutoDock. (Trott & Olson 2010). Em comparação, o AutoDock Vina atinge duas vezes mais velocidade do que a versão mais recente do AutoDock – AutoDock 4 -, ao mesmo tempo apresenta uma melhoria na precisão das previsões dos modos de ligação.

Tal como referido anteriormente, o processo de Docking para ser realizado necessita de algoritmos de amostragem e de funções de *score*. Os algoritmos de amostragem são responsáveis pela procura de conformações do ligante, modificando aleatoriamente o ligante. Os mais típicos denominam-se algoritmos Genético (AG) e de Monte Carlo (MC). O AutoDock Vina usa o algoritmo Genético. Este consiste em codificar os graus de liberdade do ligante em *strings* binárias chamadas genes. O conjunto destes genes, compõe o chamado cromossoma que nada mais é do que a conformação que o ligante adquire. Existem dois tipos de operadores neste algoritmo, as mutações que são responsáveis por fazer mudanças aleatórias nos genes; o cruzamento é responsável pela permuta de dois genes no cromossoma. De cada vez que os operadores afetam os genes, há uma nova conformação do ligante, que serão depois analisadas pelas funções de *score*. (Jones et al. 1997)

A função de *score* utilizada pelo AutoDock Vina é a combinação de dois tipos de funções diferentes, aproveitando as vantagens de cada uma delas. As duas funções utilizadas, são as funções baseadas no conhecimento ("*knowledge-based*") e as empíricas.

As funções de *score* empíricas obtêm energias de ligação através da decomposição em vários componentes de energia, como ligações de hidrogénio, interações iónicas, etc. Cada componente de energia é multiplicado por um coeficiente e depois, tudo é somado de modo a dar um resultado final. O coeficiente é obtido através das análises de regressão feita, em conjuntos de complexos proteína-ligante, com afinidades de ligação já conhecidas.

As funções de *score* baseadas no conhecimento ("*knowledge-based*") fundamentam-se em análises estatísticas das estruturas cristalinas dos complexos em estudo, obtendo a frequência

com que existe contacto interatômico. Isto é importante, uma vez que esta função baseia-se no pressuposto de que quanto mais favorável for a interação, mais frequente ela será. A pontuação (o *score*) é calculada favorecendo os contactos que são mais frequentes e, desfavorecendo as repulsões entre cada átomo no ligante e proteína dentro de uma determinada distância de *cut-off*.

Estas duas funções combinadas tornam-se mais eficientes, uma vez que se tornam complementares. Um exemplo disso, é que a função "*knowledge-based*" consegue atuar em algumas interações incomuns, às quais a função empírica não é eficiente. Este método de combinação destas duas funções de *score*, permite que obtenha informação empírica das preferências conformacionais do complexo e valores de afinidade experimentais. (Trott & Olson 2010)(Meng et al. 2011)

1.5.4. Métodos Quânticos (QM)

Os métodos quânticos derivam em última análise, da equação de Schrödinger trazida à luz no final de 1920. Trata moléculas como coleções de núcleos e eletrões, sem qualquer assunção prévia de existência de "ligações químicas". A solução para a equação de Schrödinger é em termos do movimento dos eletrões, que por sua vez leva diretamente a estrutura molecular e energia do sistema entre outros observáveis, bem como a informação sobre a ligação entre os átomos. No entanto, a equação de Schrödinger não pode ser resolvida de forma exata para qualquer sistema, apenas para sistemas com um único eletrão (O átomo de hidrogénio), e aproximações têm de ser feitas. Os vários métodos quânticos diferem na natureza dessas aproximações, e abrangem uma vasta gama, em termos de capacidade e fiabilidade e custo computacional.

Uns dos primeiros métodos a ser desenvolvido foi a aproximação Hartree-Fock aplicada à equação de Schrödinger para sistemas de muitos eletrões. Esta aproximação substitui a descrição "correta" do movimento dos eletrões por um quadro no qual os eletrões se comportam essencialmente como partículas independentes. Os modelos de Hartree-Fock oferecem boas descrições de geometrias de equilíbrio e conformações, bem como alguns tipos de comparações termoquímicas. No entanto, os modelos de Hartree-Fock descrevem mal a termoquímica de reações que envolvem a criação ou quebra de ligações. A falha destes modelos pode ser

atribuída a uma descrição incompleta da "correlação de elétrons" ou, simplesmente, da maneira pela qual o movimento de um elétron afeta os movimentos de todos os outros elétrons. Duas abordagens fundamentalmente diferentes são usadas para obviar às limitações dos modelos de Hartree-Fock.

Uma abordagem é construir uma descrição mais flexível do movimento dos elétrons em termos de uma combinação de descrições Hartree-Fock para o estado fundamental e estados excitados. Os modelos deste tipo mais utilizados são: configuração de interação (CI) e Møller-Plesset (MP). O modelo Møller-Plesset (MP2) é o mais prático e geralmente oferece excelentes descrições de geometrias de equilíbrio e conformações, bem como, da termoquímica de reações em que há quebra ou formação de ligações.

Uma abordagem alternativa para melhorar os modelos Hartree-Fock são os modelos de funcionais da densidade e envolvem a inclusão de um termo explícito para descrever a "correlação de elétrons". Este termo é obtido a partir da solução "exata" para um sistema de idealizado, sendo introduzido usando parâmetros empíricos. Os modelos de funcionais de densidade provaram ser bem-sucedidos para determinação de geometrias de equilíbrio e conformações, e na termoquímica de reações.

A aproximação de Hartree-Fock é também o modelo base para os métodos semi-empíricos. Estes introduzem aproximações adicionais, bem como parâmetros empíricos para simplificar os enormemente os cálculos, com um efeito adverso mínimo sobre os resultados. Os modelos mais populares são AM1, PM3 e PM6, os quais têm provado ser bem-sucedidos no cálculo de geometrias de equilíbrio, incluindo as geometrias de compostos de metais de transição.

A principal limitação dos métodos quânticos é a custo computacional elevado para descrever sistemas de grandes dimensões com milhares de átomos frequentemente encontrados em sistemas biomoleculares.

1.5.5. Métodos Híbridos

Nos últimos anos, os métodos computacionais têm vindo a ser muito utilizados no estudo de grandes sistemas biomoleculares, principalmente com recurso a simulações de Dinâmica Molecular, com campos de forças de mecânica molecular. No entanto, a mecânica molecular (MM), não é capaz de descrever alguns processos que ocorrem no sistema, por exemplo as quebras e formações de ligações e transferência de cargas. Para uma descrição correcta desses processos é necessário recorrer-se à mecânica quântica (QM), que se torna mais dispendioso computacionalmente para sistemas de grandes dimensões. Com o objetivo de usufruir da precisão da mecânica quântica e da velocidade de computação da mecânica molecular, Warshel & Levitt (1976) apresentaram o método híbrido QM/MM, Técnicas que combinam dois ou mais métodos computacionais em apenas um cálculo, permite uma exploração mais precisa de sistemas de grandes dimensões. Métodos QM/MM consistem na divisão dos sistemas, em duas ou mais partes, sendo cada uma das partes, tratada a diferentes níveis de teoria. Tal como representado na Figura 1.22, a região de interesse é tratada usando métodos de QM que permite a investigação de todas as reações químicas. O resto do sistema é tratado com mecânica molecular (MM), que permite obter algum conhecimento sobre a contribuição deste, na região de interesse. (Frisch 2003)

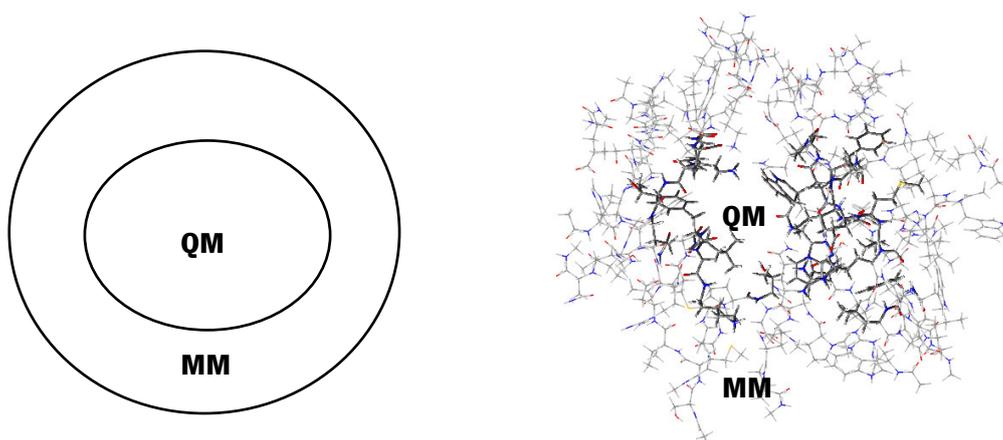


Figura 1.22 - Princípio do método híbrido QM/MM com divisão do sistema em duas partes.

Este método pode ser implementado em diversos tipos de sistemas, levando a que não haja um modelo único que o investigador possa seguir. Para cada sistema, haverá diferentes divisões do sistema, podendo este ser dividido em duas (mais comum) ou três partes. Também os métodos que irão tratar cada parte serão diferentes de sistema para sistema, uma vez que estes são escolhidos de acordo com o que se está a estudar e com o que se pretende obter.

Para ajudar o utilizador, foram criadas algumas abordagens para obter energias finais dos sistemas. As duas abordagens principais são a aditiva e a subtrativa. Na primeira, a energia total é obtida através da soma da energia da camada tratada com QM (E^{QM}), com a energia da camada tratada com MM (E^{MM}) e a interação entre ambas as camadas ($E^{QM/MM}$), como representado na Equação 1.19.

Equação 1.19

$$E_{total} = E^{QM} + E^{MM} + E^{QM/MM}$$

A segunda abordagem, a subtrativa, é a mais usada. Esta abordagem obtém a energia total do sistema, através da Equação 1.20. A $E^{real,MM}$ representa a energia de todo o sistema tratada com o nível de teoria mais baixo (métodos de MM); A $E^{modelo,QM}$ é a energia resultante da camada mais reduzida, que é tratada com um nível de teoria superior (métodos de QM). Por fim a $E^{modelo,MM}$, é a energia do núcleo QM calculado com um baixo nível de teoria (MM). (Banáš et al. 2009)

Equação 1.20

$$E_{total} = E^{real,MM} + E^{modelo,QM} - E^{modelo,MM}$$

ONIOM (Our own N-layers Integrated molecular Orbital molecular Mehanics)

Esta técnica computacional foi desenvolvida por Morokuma et al. (1996) e é capaz de calcular energias, fazer otimizações de geometria, prever frequências de vibração e propriedades elétricas e magnéticas, entre outras coisas. ONIOM é um método híbrido, que se baseia na aproximação subtrativa, anteriormente referida. Este permite que seja usado qualquer método computacional, podendo ser combinado QM com QM assim como QM com MM, em determinado número de camadas definidas.

Tal como observado na Equação 1.20, para que se obtenha a energia total do sistema no seu todo, é necessário que este seja dividido em sistema real e sistema modelo. O primeiro, contém todos os átomos e é calculado com MM. O último, por sua vez, contém apenas os átomos que irão ser tratados com QM. No entanto, o sistema modelo terá de ser sujeito a cálculos com QM e MM. Na Figura 1.23 tem-se o exemplo do Etano, em que foi incluído na região tratada com QM um grupo Metil e, o outro na região a ser tratada com MM. Tal como representado, as duas regiões encontram-se ligadas no sistema real. No sistema modelo, é necessário que se inclua um átomo de ligação (Hidrogénio) para saturar o Carbono (Figura 1.23). Em qualquer sistema, é importante incluir átomos de ligação de modo a saturar as ligações oscilantes do sistema. Os átomos de ligação (LA “*Link Atoms*”), são colocados na linha em que há a ligação das duas regiões (LAC “*Link Atoms Connection*”), substituindo o átomo, da região MM, que se encontrava nessa ligação (LAH “*Link Atoms Host*”).

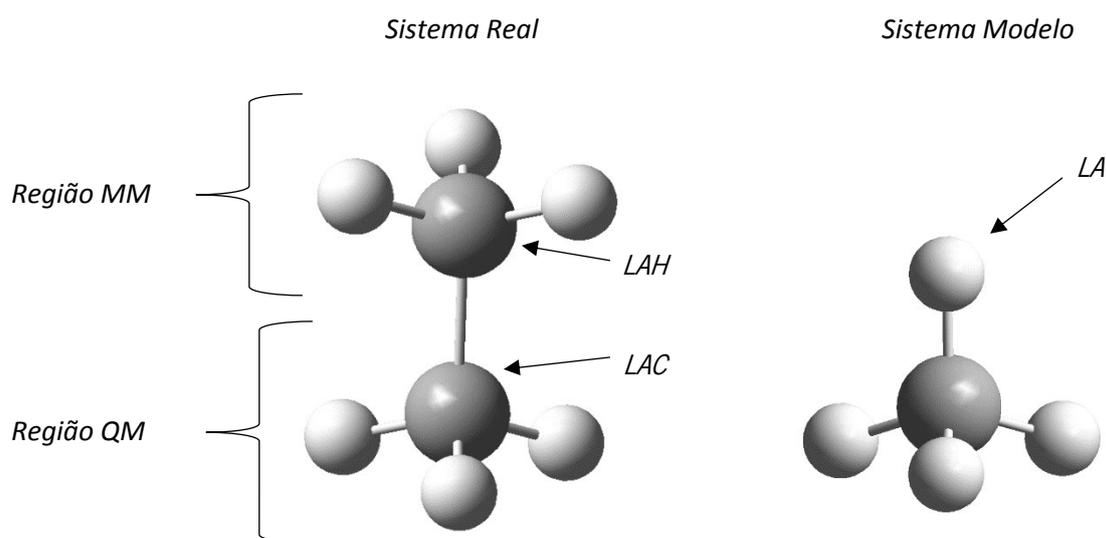


Figura 1.23 - Terminologia do ONIOM, usando o Etano como exemplo.

Para um sistema de duas camadas, a energia é obtida através da seguinte equação:

Equação 1.21

$$E = E_{MM,real} + E_{QM,modelo} - E_{MM,modelo}$$

Tal como explicado anteriormente, o *real* corresponde ao sistema no seu todo em que se inclui as moléculas de água, enzimas, ligante, resíduos catalíticos, etc. A componente *modelo* é a região de interesse, onde está incluído o ligante e alguns resíduos. Neste caso, é fácil de perceber que o ONIOM se baseia na aproximação de subtração e que, através desta é possível descrever as duas regiões (QM e MM), os átomos incluídos nelas e, a interação entre ambas. (Vreven & Morokuma 2006)

Uma das grandes vantagens do ONIOM é que permite misturar métodos de qualquer nível de teoria. Devido à facilidade de formulação é possível dividir o sistema em “n-camadas”. Quando há a divisão do sistema em três camadas, a energia é obtida recorrendo-se à Equação 1.22.

Equação 1.22

$$E = E_{baixo,real} + E_{médio,intermédia} + E_{alto,modelo} - E_{baixo,intermédia} - E_{médio,modelo}$$

Nesta equação, é acrescentado mais um termo que anteriormente não existia. Os termos baixo, médio e alto relacionam-se com os níveis de teoria utilizados. Os termos *real* e *modelo* têm significados iguais aos que têm no sistema de duas camadas. O termo *intermédia* corresponde à camada entre a região de interesse e o restante sistema. Analisando a Figura 1.24, ficará mais fácil de entender como as três camadas são tratadas com diferentes níveis de teoria.

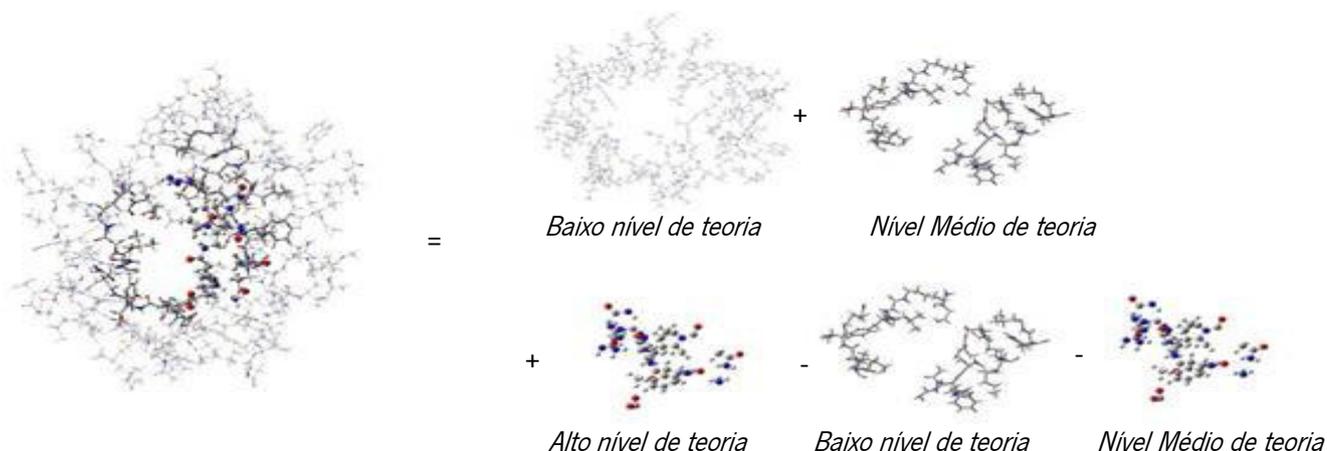


Figura 1.24 - Representação esquemática da equação usada para o sistema de três camadas do ONIOM.

Definição de Camadas

No ONIOM, tal como já referido, é possível realizar o estudo de um sistema recorrendo a “n-camadas” que, serão tratadas utilizando métodos sucessivamente mais precisos. Normalmente, os estudos realizados nunca passam as três camadas, sendo elas denominadas como Camada Alta (“*High Layer*”), Média (“*Middle Layer*”) e a Baixa (“*Low Layer*”).

- A Camada Alta (HL) é a menor de todas as camadas do sistema e, é tratada com o método mais preciso (QM). É nesta camada que ocorrem as quebras e formações de ligações que se pretende estudar. Quando é usado a abordagem de apenas duas camadas do ONIOM, esta camada é simplesmente chamada de Sistema de Modelo.
- A Camada Baixa (LL), na abordagem de apenas duas camadas do ONIOM, é o que resta do sistema sem a Camada Alta. O cálculo desta camada irá corresponder aos efeitos do ambiente molecular que estão à volta do ponto de interesse (Camada Alta). Sendo uma camada de importância inferior, esta é tratada usando métodos não tão precisos (MM).
- A Camada Média (ML), é uma camada que apenas é utilizada na abordagem de três camadas do ONIOM. Tal como o nome indica, esta camada deve estar numa zona intermédia entre as duas restantes camadas e, deve ser tratada com métodos de precisão, também, intermédios.

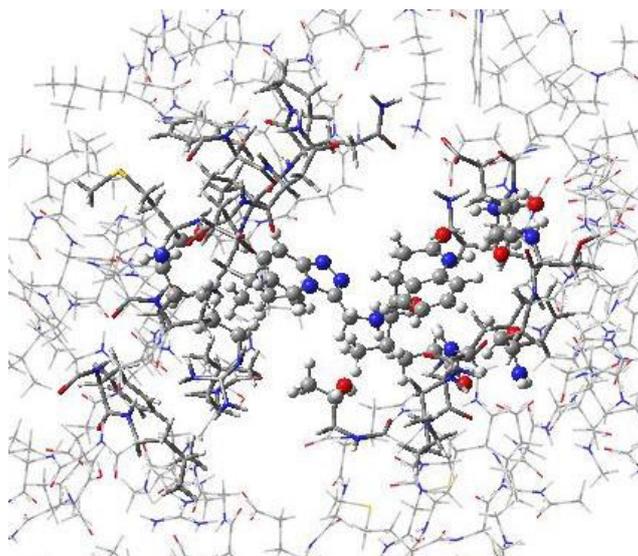


Figura 1.25 - Representação da abordagem de três camadas do ONIOM. A Camada Baixa é representada pelas linhas finas, a Camada Média por tubos e, a Camada Alta por bolas e traços.

A definição das camadas é um passo essencial no sucesso do estudo de um sistema, utilizando o ONIOM. O princípio básico que se deve ter em mente na definição das camadas é que, a parte de interesse (por exemplo, a cavidade e o ligante) deve estar sempre incluída no sistema de modelo, sendo tratada com mecânica quântica. A parte restante do sistema, deve fazer parte do sistema real que será tratado com níveis baixos de teoria (MM). No entanto, para que a definição de camadas seja definida de um modo mais correto possível, deve-se ter sempre em consideração mais alguns pontos:

- Ligações que sofrem quebras ou formam novas ligações, não devem estar incluídas na região de MM;
- O limite entre as regiões deve ser definido a, pelo menos, três ligações de distância das ligações que se quebram ou que se formam. Isto é importante pelo facto de nestes casos, existir transferência de cargas e, seria incorreto cortar essas ligações.
- Átomos ligados por ligações duplas ou triplas, devem ser incluídos sempre dentro da mesma camada.
- Nos anéis aromáticos, sempre que houver um átomo de interesse que esteja incluído no anel, todos os restantes átomos deste devem ser incluídos na mesma camada.

Incorporação (“Embedding”)

No acoplamento entre o ligante e uma proteína, a interação eletrostática entre ambos, é um elemento chave. Dependendo do modo de tratamento deste parâmetro, o sistema pode ser tratado de duas diferentes formas: com incorporação mecânica (“*Mechanical Embedding*”, ME) ou com incorporação eletrostática/eletrônica (“*Electronic Embedding*”, EE).

A incorporação mecânica realiza os cálculos de QM na região de interesse sem a presença do restante sistema e, trata as interações entre a região de interesse e o restante sistema com MM. Neste tipo de incorporação, a distribuição de cargas da região de MM irá interagir com a distribuição de cargas da região QM. Neste caso, as cargas parciais da região MM estão incluídas na região QM, que irá permitir uma maior precisão na descrição das interações eletrostáticas. Assim, a incorporação mecânica trata as interações eletrostáticas dos limites das duas regiões com MM.

Na incorporação eletrostática as cargas pontuais nos átomos presentes na região de MM, estão envolvidas em operadores de QM. Isto significa que a interação eletrostática entre as duas regiões é calculado com níveis de teoria superiores (QM), sendo que a polarização da parte principal do sistema, pela distribuição de carga da região de menor importância é realizada automaticamente. Teoricamente, este tipo de incorporação torna-se mais precisa que a incorporação mecânica. (Zhang 2013)

Através das breves explicações apresentadas anteriormente, é fácil de perceber que existem diferenças entre as duas incorporações. Obviamente que, estes tipos de incorporação devem ser adequados ao que se pretende estudar.

Tabela 1 – Comparação entre a incorporação mecânica (ME) e incorporação eletrostática (EE). Fonte: (Zhang 2013)

| | ME | EE |
|---|---|---|
| Electrostatic interaction between the primary system (PS) and the secondary system (SS) | Handled in the standard MM way | Treated by including certain one-electron terms in the QM Hamiltonian |
| Advantage | Simple | <ol style="list-style-type: none"> 1. Do not need electrostatic MM parameters for PS atoms, which may change their character during the simulation 2. The electronic structure of the PS adjusts to the charge distribution in the SS |
| Disadvantage | <ol style="list-style-type: none"> 1. An accurate set of electrostatic MM parameters is often not available for PS atoms 2. Ignores the potential perturbation of the electronic structure of the PS by the charge distribution of the SS | <ol style="list-style-type: none"> 1. More computational effort 2. Need to construct an appropriate representation for the charge distribution in the SS |

2. Materiais e Métodos: Protocolo de Simulação da interação proteína-ligante

2.1 **Materiais**

- **Hardware**

As simulações de Docking e alguns processos realizados no software VMD foram realizadas num computador com o sistema operativo Linux. Todos os outros cálculos foram realizados noutra computador com sistema operativo Windows 8. Ambos se encontravam no Grupo de Física Teórica e Computacional da Universidade do Minho (GFCT).

Os cálculos quer de DM, quer de energias de ligação foram realizados usando os clusters computacionais Higgs do Grupo GFCT e Search6 do departamento de Informática da Universidade do Minho.

- **Software**

As simulações de Dinâmica Molecular da proteína foram realizadas usando *Amber Molecular Dynamics Package* versão 2015. AMBER (*Assisted Model Building with Energy Refinement*) corresponde a uma família de campos de forças utilizados para a dinâmica molecular de biomoléculas. O *software* AMBER é constituído então por um conjunto de programas que permitem a simulação desses campos de forças, ou seja, que permite a realização de simulações de dinâmica molecular em biomoléculas. (D.A. Case, et al. 2015)(Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW 1995)

Para a realização dos estudos de *Docking* foi usado o *software* AutoDock Vina como *plugin* no *software* Pymol. AutoDock Vina é um “*open-source*” que é usado para realizar trabalhos de *docking molecular*, tendo como vantagem ser mais preciso na previsão do modo de ligação em comparação com os restantes *softwares* capazes de efetuar trabalhos de docking. (<http://vina.scripps.edu/index>) (Steffen et al. 2010) Pymol fornece funcionalidades excepcionais de visualização 3D que podem ser muito úteis na conceção de medicamentos à base de estrutura. Além de ser de fácil utilização, o *plugin* AutoDock Vina do Pymol, abrange uma grande quantidade de funcionalidades que permitem a preparação do recetor e do(s) ligante(s) automaticamente, gerar e visualizar mapas de rede de afinidade, visualização de posições de

encaixe e, análise e exportação dos resultados do *docking*. (<http://wwwuser.gwdg.de/~dseelig/adplugin.html>) (Seeliger & De Groot 2010)

A construção dos complexos (proteína e ligante) a serem estudados, foi feita usando o software VMD – *Visual Molecular Dynamics*. Este programa foi projetado para ser capaz de modelar, visualizar e analisar sistemas biológicos. (<http://www.ks.uiuc.edu>) No presente caso, foi usado para analisar a trajetória das simulações de dinâmica molecular e, posteriormente construir as esferas que seriam estudadas pelo método híbrido QM/MM.

Os cálculos de energia de ligação foram feitos com o auxílio dos programas Gaussian versão 09 e GaussianView versão 5. Gaussian09 fornece-nos previsões de energias, estruturas moleculares, frequências vibratórias e propriedades moleculares de moléculas, recorrendo às leis fundamentais da mecânica quântica, de sistemas complexos. Os resultados obtidos podem ser visualizados, monitorizados e controlados, através do GaussView. (<http://www.gaussian.com>)

2.2 Métodos

2.2.1. Estruturas da Glicoproteína E e Ligante

A complexa estrutura da Glicoproteína E do DENV-2 foi conseguida através da plataforma online *Protein Data Bank* (PDB), com o código de identificação 1THR, como se pode verificar na Figura 2.1. Esta plataforma é a única, a nível mundial, que fornece informações acerca das estruturas 3D das moléculas biológicas. (Berman et al. 2000)

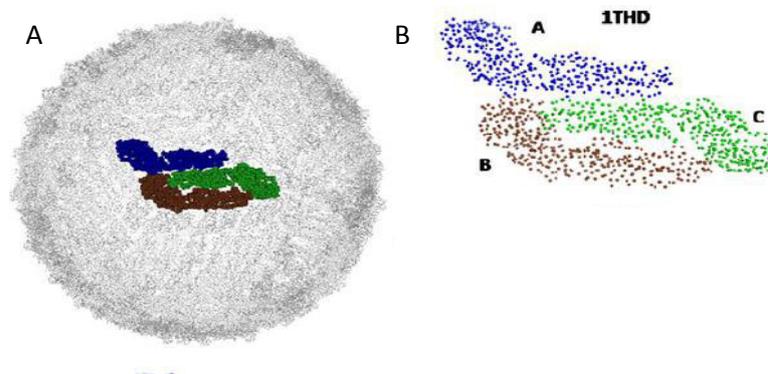


Figura 2.1 – (A) Montagem biológica contendo a Glicoproteína E representada no centro da estrutura global do DENV. (B) 1THR: Complexa organização da Glicoproteína E do DENV2, contendo três cadeias – A, B e C- e os átomos C-alfa Adaptado de (Fuzo & Degreuve 2013)

Esta estrutura foi usada para o estudo da Dinâmica Molecular, resultando em doze imagens que posteriormente irão ser usados para os restantes estudos.

O Ligante que foi estudado foi recolhido da plataforma ZINC, com o código de identificação ZINC22232179. A base de dados ZINC é uma plataforma online que disponibiliza as estruturas tridimensionais de compostos químicos, sendo que cerca de 35 milhões podem ser comprados, podendo ser usados para realizar trabalhos como *docking*. (Irwin & Shoichet 2005)

2.2.2. Simulação da Dinâmica da proteína por Dinâmica Molecular

Com o objetivo de analisar o comportamento dinâmico da proteína ao longo do tempo, foram realizadas simulações de Dinâmica Molecular através do *software* AMBER15 (D.A. Case, et al. 2015), no sistema operativo Linux.

Tudo isto foi tido em consideração ao longo do nosso trabalho de simulação de DM, a Figura 2.2 descreve o processo que foi executado, para chegar às simulações de DM. O primeiro passo foi o *download* da estrutura 1THD da plataforma *Protein Data Bank* (PDB).

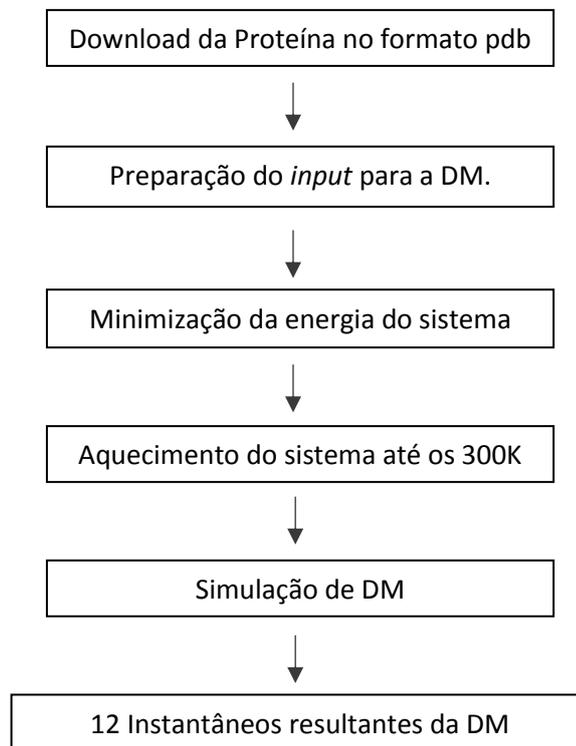


Figura 2.2 - Metodologia utilizada para a realização das simulações de Dinâmica Molecular.

Os requisitos para a utilização de ficheiros, com formato pdb, no AMBER são: nome dos átomos, nome dos resíduos, identificação das cadeias – caso haja mais que uma- e, as coordenadas dos átomos. Deste modo, é crucial que se analise o ficheiro pdb e se adeque o mesmo – através de um ficheiro de texto, por exemplo- para que não haja qualquer incompatibilidade com o AMBER. Se o processo falhar, o programa responsável pela preparação do *input* - LEaP- irá analisar o ficheiro pdb, verificando os nomes dos átomos e resíduos. Se as denominações não forem reconhecidas pelo LEaP, este irá criar uns novos, de modo a que não haja perda de nenhum

átomo na proteína. No entanto, se esses novos resíduos não constarem na base de dados do LEaP, as cargas e tipologia dos mesmos serão tratados como desconhecidos.

Para que não nos deparássemos com esse tipo de problemas, e outros, promovendo a ocorrência de instabilidade das simulações, fez-se a “limpeza” da proteína e, posteriormente a verificação do ficheiro pdb. A “limpeza” da proteína inclui a remoção de todos os Hidrogénios, contra-íões, todas as moléculas de água e, ligantes caso existam. Fez-se também a verificação do ficheiro e alterou-se a denominação de alguns resíduos como a Histidina (HIS), Cisteína (CYS), Treonina (THR) e Fenilalanina (PHE).

A preparação do *input* é feito com o auxílio do programa LEaP, incluído no software AMBER15. Tal como já foi referido, esta é uma ferramenta básica de construção de ficheiros de campos de forças. Um campo de forças AMBER normalmente refere-se à forma funcional usada pela família de campos de forças AMBER. Esta forma funcional, inclui um número de parâmetros e, cada campo de forças AMBER fornece valores a esses parâmetros, tendo assim cada um, o seu próprio nome. (Cornell WD, et al. 1995)

AMBER utiliza vários campos de forças simples, apesar de que o que é usado geralmente é o que contém parametrizações desenvolvidas por Peter Kollman, o criador do software AMBER. No LEaP, o mais usado, hoje em dia, é o campo de forças ff14SB, que é o mais recomendado para ser aplicado em proteínas e ácidos nucleicos. Este campo de forças foi também o escolhido para a elaboração deste trabalho. O campo de forças ff14SB é uma continuação do ff99SB que, desde 2006, ficou evidente que apresentava limitações a nível da parametrização. O mais recente, foi desenvolvido com o intuito de minimizar a dependência dos parâmetros das cadeias laterais em relação à cadeia principal, o facto de limitarem um diedro por cadeia secundária, permitiu uma maior redução da dependência à cadeira principal. Foi também ajustado o processo de geração de dados, sendo que houve a minimização da dependência da estrutura principal da proteína, em que se inclui as limitações dos diedros da estrutura principal e a otimização das ligações e ângulos de cada conformação. (D.A. Case, et al., 2015)

A preparação da molécula para o AMBER, tal como já referido anteriormente, é feita usando o programa LEaP e, existem três passos essenciais:

1. Carregar a estrutura com formato pdb;
2. Construção/ Edição da molécula através do LEaP;

3. Deve-se obter ficheiros de saída (*outputs*) um com a topologia da estrutura e outro com as coordenadas.

No presente trabalho, a estrutura que se carregou foi o 1THD, proveniente da plataforma PDB. De seguida, procedeu-se à edição da estrutura. Iniciou-se o processo recorrendo ao comando *list* que lista todas as variáveis, definidas pelo LEaP, da estrutura. Assim, é possível corrigir alguns erros de compatibilidade. Posteriormente a este processo de verificação, criou-se uma caixa de solvatação usando o comando *solvatebox*. Este passo irá criar uma caixa periódica de solvente à volta da estrutura, esta será tratada como sendo o soluto. Assim, usou-se uma caixa pré-equilibrada com o modelo de água TIP3P e, especificou-se também que a distância entre a estrutura e a aresta da caixa seria de 11 Angstroms. O modelo TIP3P (*Transferable Intermolecular Potential 3P*) tem três sítios de interação e pode ser descrito como potenciais compostos por termos de Lennard-Jones – modelo matemático simples que estuda a interação entre um par de átomos neutros ou de moléculas - e Coulomb – termo que descreve a força da interação entre partículas eletricamente carregadas. (Naeem 2013)(Browne et al. 2010)(Mark & Nilsson 2001)

De modo a neutralizar a carga da proteína, adicionou-se alguns iões. Terminou-se a preparação da proteína gerando ficheiros com a sua topologia e as suas coordenadas, recorrendo ao comando “SaveAmberParm”, que é o utilizado em campos de forças aditivos. (D.A. Case, et al. 2015)

Uma vez com o ficheiro prmtop criado – onde está incluído a topologia molecular, campo de força, tipo de caixa periódica e nome dos átomos e resíduos – estamos em condições de iniciar a minimização da energia usando o programa Sander. Sander (*Simulated Annealing with NMR-Derived Energy Restraints*) é um módulo do AMBER capaz de calcular energias de minimização e de fazer simulações de dinâmica molecular, sendo responsável por dar suporte direto aos campos de forças das proteínas e aos diferentes modelos de água.

A minimização de energia é efetuada com o objetivo de eliminar pontos de contacto indesejáveis, ou seja, más interações estéricas (interações que ocorrem entre átomos da mesma molécula, havendo sobreposição das nuvens eletrónicas, podendo afetar a conformação final da molécula e a sua reatividade). (Stephan 2008)(Kerrigan, E. 2011) No presente caso, fez-se uma minimização com 2000 ciclos.

Após a fase de energia de minimização ter finalizado, usou-se o ficheiro *restart* - resultante da minimização - para iniciar a simulação de Dinâmica Molecular. Começou-se com um processo de aquecimento do sistema. Este processo foi realizado até chegar aos 300K (kelvins), através de seis passos, sendo que de 50 em 50 fentosegundos havia o aumento de 50K.

Para realizar este processo, de seis passos, recorreu-se a um *script* - para cada passo (mdin1,2,3,4,5 e mdin6). Nestes scripts especificou-se, tal como representado na Figura 2.3, o tipo de minimização (“imin”) que no presente caso não seria nenhuma, uma vez que já se tinha feito anteriormente; o número de iterações (“nstlim”); o modo de leitura do ficheiro que contém as coordenadas, velocidades e tamanhos da caixa de solvatação (“ntx”), neste caso o número 5, indica que o ficheiro será lido ou com o formato NetCDF ou ASCII; o tempo de cada passo (“time step”) em fentosegundos (“dt”); para a temperatura será usado um termóstato Berendsen (“ntt=1”) tendo sempre especificado a temperatura mínima (“tempi”) e máxima (“temp0”); a pressão constante (“ntp”) que será usada no sistema, neste caso “0” implica que se use um baróstato Berendsen que irá manter a pressão constante; “ntb” é a variável que controla se são impostas, ou não, barreiras periódicas durante os cálculos das interações não-covalentes. Esta variável depende dos outros parâmetros, isto é, se o parâmetro “ntp” for 0, “ntb” será 1 indicando que o sistema manterá um volume constante ao longo dos cálculos; “ntc” e “ntf” são também parâmetros dependentes, sendo que “ntc” é o parâmetro relacionado com o algoritmo SHAKE que deve ser usado em todos os cálculos de DM. SHAKE remove a liberdade de alongamento das ligações, que é o movimento que apresenta a maior frequência de oscilação, permitindo assim um maior tempo de simulação, uma vez que é a maior frequência de oscilação que determina o tamanho máximo do “time step”. Regra geral, sempre que é usado o modelo de água TIP3P, “ntc” é igual a “ntf” e ambos têm o número 2 que corresponde à restrição das ligações de hidrogénio. É importante que o movimento dos hidrogénios presentes nas moléculas de água estejam fixos, caso contrário, poderá promover erros de precisão nos cálculos de densidade, por exemplo. O “ntf” está relacionado com o cálculo de forças de restrição das ligações, que não é necessário quando é usado o SHAKE. O número 2 associado ao “ntf” está também relacionado com os hidrogénios. “cut” é usado para definir o ponto de corte das ligações não-covalentes, em Angstroms. (D.A. Case, et al. 2015)(Mark & Nilsson 2001)

A

```
initial heatup phase
&cntrl
imin=0, nstlim=10000,
ntx=5,
dt=0.002
ntt=1, tempi=0 temp0=50.0
ntp=0,
ntb=1, ntf=2, ntc=2,
cut=10,
/
```

B

```
initial heatup phase
&cntrl
imin=0, nstlim=10000,
ntx=5,
dt=0.002
ntt=1, tempi=250.0 temp0=300.0
ntp=0,
ntb=1, ntf=2, ntc=2,
cut=10,
/
```

Figura 2.3- (A) Script “mdin1” representando o script criado para o primeiro passo deste processo, dos 0K aos 50K. (B) Script “mdin6” representando o script criado para o sexto passo deste processo, dos 250K aos 300K.

Depois do processo de aquecimento do sistema, iniciou-se o procedimento para a realização da Dinâmica Molecular. Este, será feito ao longo de 500 nanosegundos, sendo que os primeiros 80 nanosegundos serão de relaxamento da proteína. Para o relaxamento da proteína, fez-se um *script* com os parâmetros, representados na Figura 2.4. O parâmetro “ig” é um gerador de números aleatórios de que depende a velocidade de partida das simulações de DM, o “-1” é baseado na data e hora sendo que será diferente para cada passo de simulação, para este caso era vantajoso permitindo que fizéssemos cada passo de simulação em 24h. “irest=1” permitirá reiniciar a simulação, tendo acesso às coordenadas e velocidades do ficheiro anterior. O parâmetro “ntxo” permite escolher o formato em que serão escritas as coordenadas, velocidades e tamanho de caixa de solvatação nos ficheiros para reiniciar as simulações, neste caso, o 2 escolhe o formato binário (NetCDF) que é mais leve e menos problemático. Em comparação com os scripts anteriores, mudou-se os parâmetros “ntp” e “ntb” colocando o sistema com uma pressão constante em todas as direções.

```
&cntrl
lg=-1,
lmin=0, nstlim=4000000,
lrest=1, ntx=5,
ntwx=25000,
ntwr=100000,
ioutfm=1,
ntxo=2,
dt=0.002,
ntt=1, temp0=300.0,
ntp=1,
ntb=2, ntf=2, ntc=2,
cut=10,
/
```

Figura 2.4 - script "mdinres" criado para realizar o relaxamento da proteína.

Recorreu-se a um subconjunto de funcionalidades do Sander, para fazer a realização da DM usando o pmemd.cuda. Para isso, apenas é necessário a especificação do *input*, *output* (nome desejado), ficheiro contendo a topologia, ficheiros de reinicialização (o anterior e o que irá ser criado). Os resultados foram analisados com o programa cpptraj, também incluído no AMBER, permitindo a obtenção de 12 imagens instantâneas do sistema (*snapshot*), obtidas em intervalos de tempo regulares (após os 80 nanosegundos iniciais de relaxamento da proteína) e o cálculo do desvio médio quadrático (RMSD) da distância média entre os átomos das diferentes conformações da proteína, resultantes da DM. (Michael Hardy, et al., 2016)

Uma vez finalizada a DM, realizou-se a procura de cavidades na proteína através do programa MDPOCKET, explicado anteriormente. Para a realização deste passo, "apenas" foi necessário a utilização das trajetórias dos átomos resultantes da DM. Através do programa MDpocket é possível descobrir todas as cavidades que aparecem em mais de 50% das vezes. Daqui, é possível visualizá-los e analisar a sua estrutura na proteína.

Deste modo, é possível selecionar uma cavidade de interesse, tendo em consideração fatores como a sua localização e estabilidade. Após escolhida uma determinada cavidade, o MDpocket retorna ficheiros com a descrição da mesma. No presente caso, foi escolhida a cavidade que se encontrava entre as três cadeias da proteína, tal como demonstrado na Figura 2.5. Esta cavidade foi a escolhida devido à sua localização, uma vez que poderá ser um local favorável à

interrupção da mudança de conformação da Glicoproteína, na altura da fusão e, devido à sua estabilidade ao longo de toda a simulação.



Figura 2.5 – Cavidade escolhida na Glicoproteína E.

No final de todo o processo de DM, temos doze conformações diferentes da Glicoproteína E e, a escolha da cavidade de interesse. Posto isto, é possível avançar para o processo seguinte que consistirá no primeiro estudo de interação entre ligante(s) e proteína.

2.2.3. Docking Proteína-Ligante

Docking, tal como já referido anteriormente, é um método capaz de prever a melhor conformação de um ligante num local de ligação alvo de uma proteína, quando estes estão ligados para formarem um complexo estável. O conhecimento da orientação da molécula pode por sua vez, ser utilizado para prever a força de associação ou a afinidade de ligação entre a molécula e a proteína. (Lengauer & Rareyt 1996)

Esta etapa do trabalho foi realizada com o intuito de melhor compreender o processo de procura de novos fármacos e assim selecionar um bom inibidor da mudança de conformação da Glicoproteína E que, conseqüentemente, inibiria a fusão das membranas celulares e virais. À partida, sabíamos que o ligante ZINC 22232179 seria uma boa opção, devido a alguns testes preliminares desenvolvidos pelo grupo do Professor Léo Degrève. No entanto, quisemos testar o acoplamento (em termos de energia) de vários ligantes – inclusive o ZINC 22232179 – com a

Glicoproteína E. Foram recolhidos cerca de 1000 ligantes diferentes, através da plataforma ZINC, selecionando todos os que tinham 70%, ou mais, de analogia com o ligante inicial. Para aumentar o número de ligantes a serem testados, usou-se também ligantes que tinham sido testados noutras partes do vírus (como as proteínas NS1) e seus análogos. No total testamos 1016 ligantes na Glicoproteína E e usou-se a metodologia de Docking de Ligante Flexível e Recetor Rígido.

Esta etapa é feita primeiramente usando o sistema operativo Linux e, *softwares* como AutoDock Vina. Inicialmente, ponderou-se realizar o estudo do Docking usando o AutoDock4, mas segundo a literatura com o AutoDock Vina obteríamos uma maior precisão na previsão do modo de ligação, em menos tempo, em comparação com o AutoDock4, por essa razão, optou-se pelo AutoDock Vina. (Trott & Olson 2010)

Posteriormente, uma vez já familiarizados com os *softwares*, desenvolveu-se um *script* que usando o AutoDock Vina nos permitiu estudar o acoplamento de 1016 ligantes e, depois visualizar os resultados no Pymol.

O *script* realizado inclui diversos passos que eram necessários para a realização do estudo do Acoplamento. Inicialmente, o primeiro passo seria modificar o formato dos ligantes (que foram obtidos com o formato mol2 e da proteína para o formato PDBQT. Este formato, armazena as coordenadas atómicas, as cargas e os tipos de átomos reconhecidos pelo AutoDock Vina. O formato PDBQT no AutoDock tem como principal característica a capacidade de analisar os átomos detalhadamente; por exemplo, é capaz de distinguir nitrogénios que podem ter ligações com hidrogénios dos que não podem. Este primeiro passo é conseguido através do *software* MGL Tools que é usado especialmente para criar *inputs* com formato PDBQT, no AutoDock/AutoDock Vina. Quando todos os ligantes e proteína, estiverem no formato desejado, é criado uma pasta para cada ligante. Em cada pasta, estaria incluído os ficheiros PDBQT do ligante em questão e da proteína, e um ficheiro de configuração onde incluía os parâmetros necessários para realizar o acoplamento do ligante à proteína, tal como representado na Figura 2.6.

```
receptor=protein.pdbqt
ligand=zinc_1785.pdbqt

size_x=20.25
size_y=19.875
size_z=22.5
center_x=76.47
center_y=60.56
center_z=30.19

out=final_zinc_1785.pdbqt
log=results_zinc_1785.txt
num_modes=10
energy_range=4
```

Figura 2.6 - Exemplo de ficheiro de configuração para o ligante zinc_1785.

Neste ficheiro de configuração são especificados todos os parâmetros essenciais para a realização do estudo de acoplamento; começa-se por identificar o recetor que é a proteína retirada da primeira conformação resultante da dinâmica molecular (deu-se o nome de “*protein*” com o formato PDBQT) e, é usada a mesma para todos os ligantes. O ligante é identificado automaticamente, sendo que irá variar de pasta em pasta (de ligante em ligante).

Para o estudo do Docking é necessário a criação de uma caixa, que será a área que o ligante poderá percorrer dentro da proteína, testando a sua melhor conformação na zona da proteína desejada, como representado na Figura 2.7.

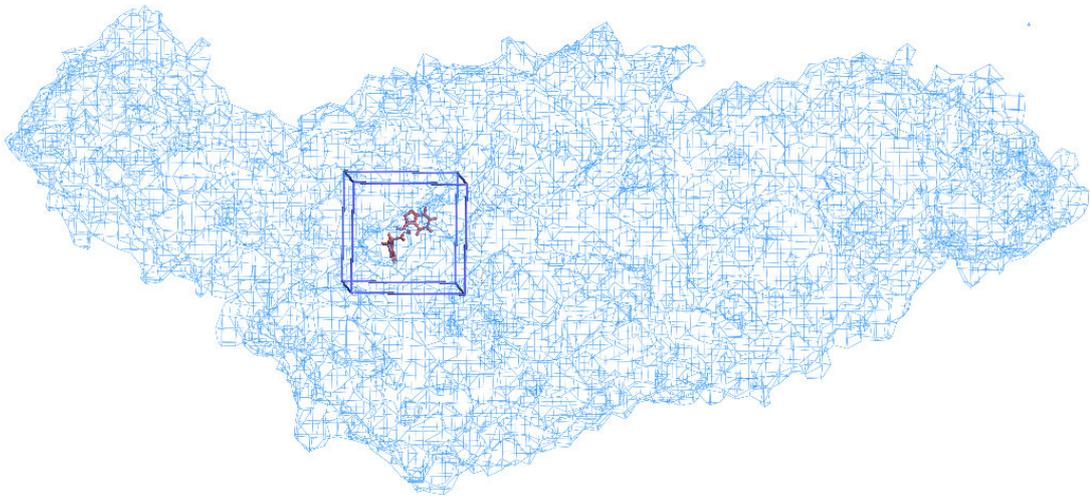


Figura 2.7 - Criação da Gridbox com centro no Ligante.

Assim, para a criação desta caixa, era necessário determinar o seu tamanho, em Angstroms, nas três direções. Os tamanhos foram especificados em “size_x”, “size_y” e “size_z”, e foram definidos tendo em consideração a cavidade que estava em estudo, o objetivo seria que a caixa ficasse o mais próxima possível do tamanho da cavidade para que, o deslocamento dos ligantes para fora da cavidade fosse evitada. O centro da caixa foi definido em função do ligante ZINC22232179, uma vez que a disposição inicial deste percorria o interior da cavidade, seleccionando-se as coordenadas de um átomo central do ligante, ter-se-ia as coordenadas aproximadas do centro da cavidade.

As últimas especificações correspondem aos nomes que teriam os outputs, com o formato PDBQT para que fossem visualizados e, em formato de texto contendo os valores de energias. O parâmetro “num_modes” corresponde ao número de posições de ligação que o ligante pode adquirir ao longo do processo; por fim, o parâmetro “energy_range” é a diferença máxima de energia entre o melhor modo de ligação e o pior, o número 4 é o valor padrão estabelecido. (Trott & Olson 2010)

Com todos os parâmetros definidos, a terceira parte do *script* passa pela execução do AutoDock Vina para todos os ligantes. Cada ligante, terá uma energia para cada conformação de ligação, isto é, no final, cada ligante apresentará 10 conformações de ligação diferentes no interior da

cavidade, cada uma associada a uma energia. Essas energias serão descritas no *output* em formato de texto criado anteriormente, para cada ligante.

De modo a obter uma comparação de todos os ligantes, todos os resultados obtidos, foram reunidos num só ficheiro de texto e, ordenados os valores de energia por ordem crescente. Os resultados que apresentassem valores de energia menores, seriam os mais estáveis e, conseqüentemente, que teriam maior probabilidade de ligação com a cavidade.

O Docking é um processo interessante pelo facto de ser capaz de reproduzir potenciais químicos, que determinam a preferência de conformação de ligação e as energias livres de ligação.

2.2.4. Energias de Ligação

O método de Docking anteriormente descrito, fornece uma lista contendo os ligantes com maiores probabilidades de acoplamento com a Glicoproteína. Este passo é importante na medida em que, é-nos possível escolher um ligante para testar ou, ter acesso a uma rápida análise sobre a possível ligação de um determinado ligante a uma proteína de interesse. No entanto, variadas são as vezes em que se procura mais métodos, que comprovem a fiabilidade de um determinado ligante, numa proteína, obtendo energias de ligação. Para além do processo de Docking, tivemos acesso a resultados, sobre o ligante ZINC22232179, de outro método por Dinâmica Molecular – LIE. No entanto, estes métodos dão-nos maioritariamente informações acerca dos movimentos dos átomos, sem que haja informação sobre a quebra e formação de ligações e reações químicas. Por essa razão, realizamos o método ONIOM para obter energias de ligação finais.

- **Energia Interação Linear (LIE – Linear Interaction Energy)**

Este método foi desenvolvido, originalmente, para prever a afinidade de ligação entre um grupo de inibidores estruturalmente relacionados com uma proteína da família de aspartil proteases. (Bharatam et al. 2008)

LIE é baseado na combinação linear das variações entre os valores médios de energia, resultantes das simulações de DM da interação do ligante com a proteína/recetor, permitindo o cálculo da energia livre. Por outras palavras, este é um método empírico que é baseado na interação do ligante com o meio que o rodeia. Para chegar à energia livre de ligação (ΔG), este método, divide as contribuições em dois termos: o eletrostático (contribuições polares) e de Lennard-Jones (contribuições apolares). (Hammes 2012)

O resultado do cálculo de energia livre do ligante ZINC22232179, usando o método LIE, foi-nos fornecido, pelo Professor Léo Dégreve, para que pudéssemos ter um valor para comparação com os restantes métodos.

- **Por Método Híbrido QM/MM**

Dos métodos anteriores, escolheu-se um ligante que apresentasse um *score* e energia razoável. Para que pudéssemos fazer um estudo mais profundo da interação do ligante na cavidade da proteína, incluindo as quebras e novas formações de ligações, recorreremos aos métodos híbridos de QM/MM.

Tal como anteriormente descrito, o ONIOM pode ser usado com a definição de duas ou três camadas, no presente trabalho, foram testadas ambas as metodologias.

Dos resultados de dinâmica molecular, usou-se apenas quatro das doze *snapshots* resultantes: conformação 3, 6, 9 e 12, devido a limitações de tempo para efetuar os cálculos para todos os *snapshots*.

Todo o processo foi efetuado utilizando uma esfera da proteína, com 10 Å de raio sendo que o centro seria o centro da cavidade. Foram testados vários tamanhos de esferas (Figura 2.8), e também cálculos com toda a proteína.

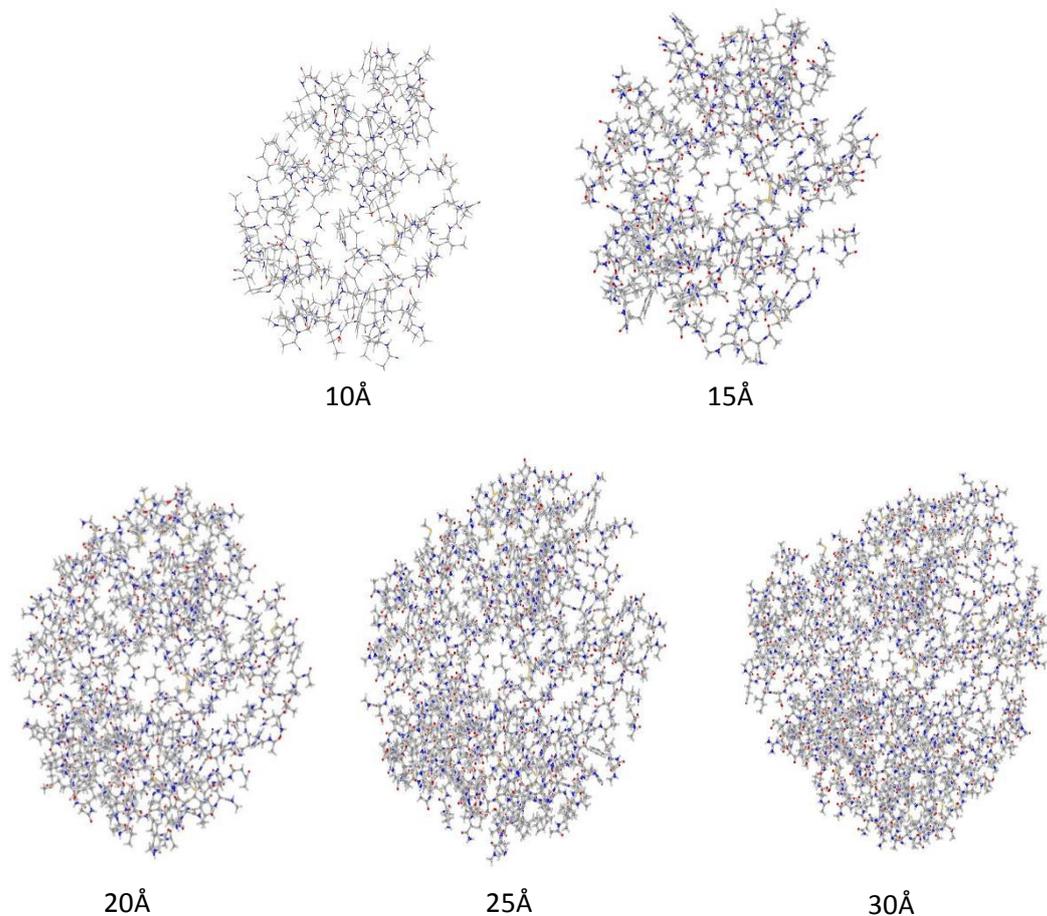


Figura 2.8 – Diferentes tamanhos de esferas testados.

Após efetuar alguns cálculos em todas as esferas, conclui-se que seria melhor utilizar a esfera de 10 Å, pelo facto de ser mais rápido e apresentar resultados que não diferiam muito das restantes esferas. As esferas foram feitas recorrendo a um *script* capaz de definir as esferas apenas com determinados resíduos/aminoácidos, com um raio escolhido pelo utilizador, a partir de um ponto também definido pelo mesmo.

Para saber quais os resíduos que deveriam estar incluídos na esfera, usou-se o VMD para analisar todas as *snapshots* selecionadas da DM. Para cada uma, escolhíamos o ligante como centro – uma vez que este se encontrava no centro da cavidade – e, considerando uma esfera de 10Å, selecionamos todos os resíduos presentes na esfera. O objetivo desta etapa, era garantir que todas as esferas teriam o mesmo número de átomos, contendo os resíduos mais importantes de todas as conformações. Os resíduos retirados, eram então colocados no *script*, sendo assim capaz de construir esferas com 10Å em todas as *snapshots* resultantes da DM.

Tabela 2 - Resíduos presentes nas esferas, com os respectivos números de identificação.

| <i>“Resname”</i> | <i>“Resid”</i> |
|---------------------------------|---|
| <i>Denominação dos Resíduos</i> | <i>Número de Identificação dos Resíduos</i> |
| LEU | 56 |
| THR | 76 |
| LEU | 107 |
| PHE | 108 |
| GLY | 109 |
| LYS | 110 |
| ILE | 129 |
| VAL | 130 |
| GLN | 131 |
| PRO | 132 |
| PHE | 193 |
| ASN | 194 |
| GLU | 195 |
| MET | 196 |
| VAL | 197 |
| ARG | 210 |
| PHE | 213 |
| LEU | 214 |
| LYS | 310 |
| ALA | 313 |
| GLU | 314 |
| THR | 315 |
| GLN | 316 |

Com a lista de resíduos feita, o *script* irá incluir todos os que se encontram num raio de 10Å. Para que isto aconteça, é necessário que haja o corte de algumas ligações que existem entre resíduos e posteriormente, que as terminações N- e C- sejam tapadas. Para as terminações N- o *script* acrescenta o grupo acetil proveniente do aminoácido seguinte, dando o nome de “NME”; para as terminações C- é acrescentado grupos metilamina, com o nome “ACE”, como representado na Figura 2.9. É então realizado um passo de minimização de energia, restringindo as duas posições dos restantes átomos.

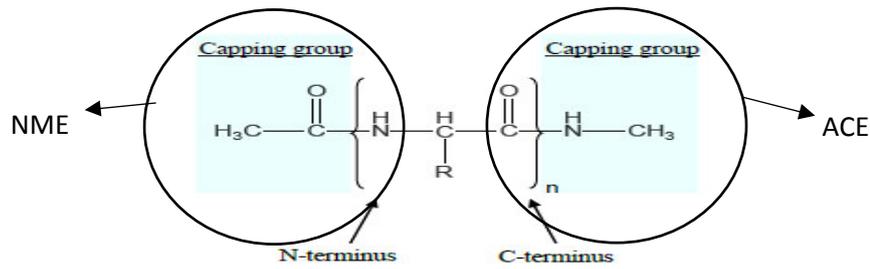


Figura 2.9 – Representação de “NME” e “ACE” no corte dos resíduos.
Adaptado de (OTHMAN 2010)

Contendo todas as esferas de todas as doze conformações, passa-se ao seguinte passo que é realizado no GAUSSIAN09. Todas as conformações vêm com as cargas AMBER provenientes da DM, o que por vezes acontece é que, existam alguns resíduos que não estejam a ser reconhecidos no GAUSSIAN e, nesses casos é necessário que sejam adicionadas as cargas. Quando todos os resíduos estiverem devidamente identificados e, com as respetivas cargas, inicia-se o método ONIOM.

- **Complexo, substrato e ligante**

Para a realização dos cálculos no método ONIOM, é necessário que o sistema (esfera de 10Å com o ligante) sejam tratados de diferentes modos, ao longo de todo o processo de cálculos. O sistema será tratado como complexo, em que terá o ligante e a porção de proteína; como substrato, em que terá apenas a porção de proteína; e, apenas o ligante.

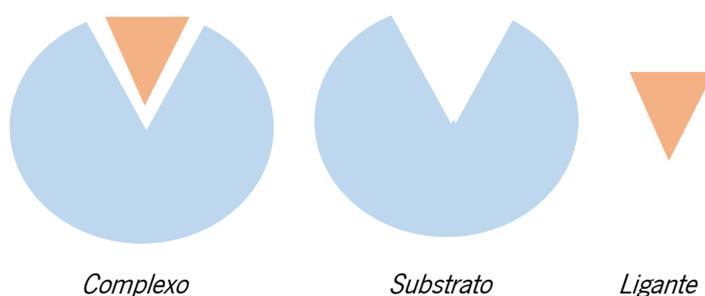


Figura 2.10 - Ilustração do Complexo, Substrato e Ligante.

Os cálculos serão realizados por etapas, começando os cálculos do sistema enquanto Complexo (Substrato com o Ligante) e, mais tarde é que o sistema é tratado como Substrato e Ligante. Este é um procedimento importante para obter as energias totais. Tal como é possível verificar através da Equação 2.1, para se obter a energia total do sistema, é necessário que à energia do Complexo (E_c), se subtraia as energias individuais, a energia do Ligante (E_L) e a energia do Substrato (E_s).

Equação 2.1

$$E_T = E_c - (E_s + E_L)$$

2.2.5. ONIOM 2

Inicialmente, a metodologia escolhida foi o ONIOM usando duas camadas: “*High Layer*” e “*Low Layer*”. O processo é feito no GAUSSIAN, começando com o reconhecimento de alguns resíduos nas esferas, tal como já referido. A fase seguinte, é então a definição das camadas. A metodologia que se utilizou – para além das regras já apresentadas anteriormente na secção de Definição de Camadas -, foi a seleção dos resíduos que se encontravam a 3Å de cada átomo do ligante, conforme apresentado na Figura 2.11.

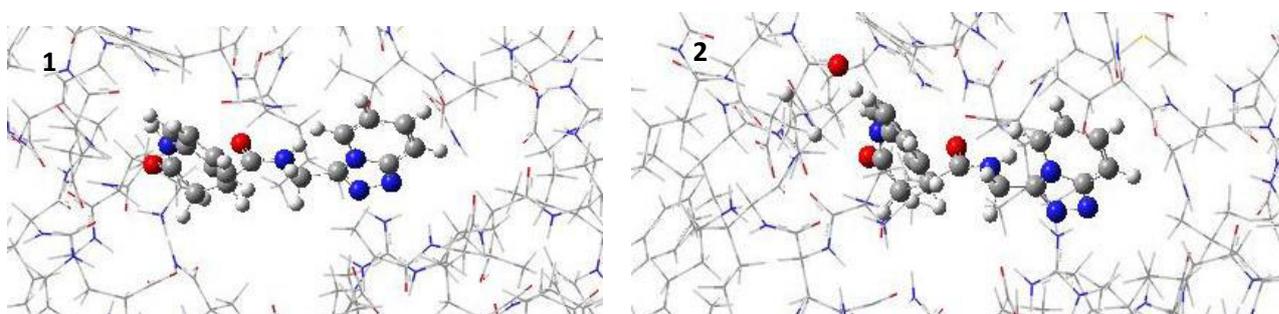


Figura 2.11 - Processo de seleção de camadas, selecionando os átomos que se encontram a 3Å de cada átomo do ligante. 1) A “*Low Layer*” está representado pelas linhas finas e, 2) “*High Layer*” com bolas e traços.

Após a definição das camadas, cada conformação apresentava um diferente número de átomos na HL, mas mantendo-se um total de 1764 átomos. A partir deste ponto, é possível iniciar o protocolo de cálculos, que é iniciado apenas com o complexo, como demonstrado na Figura 2.12. Todo o processo de cálculos foi feito usando a incorporação eletrónica.

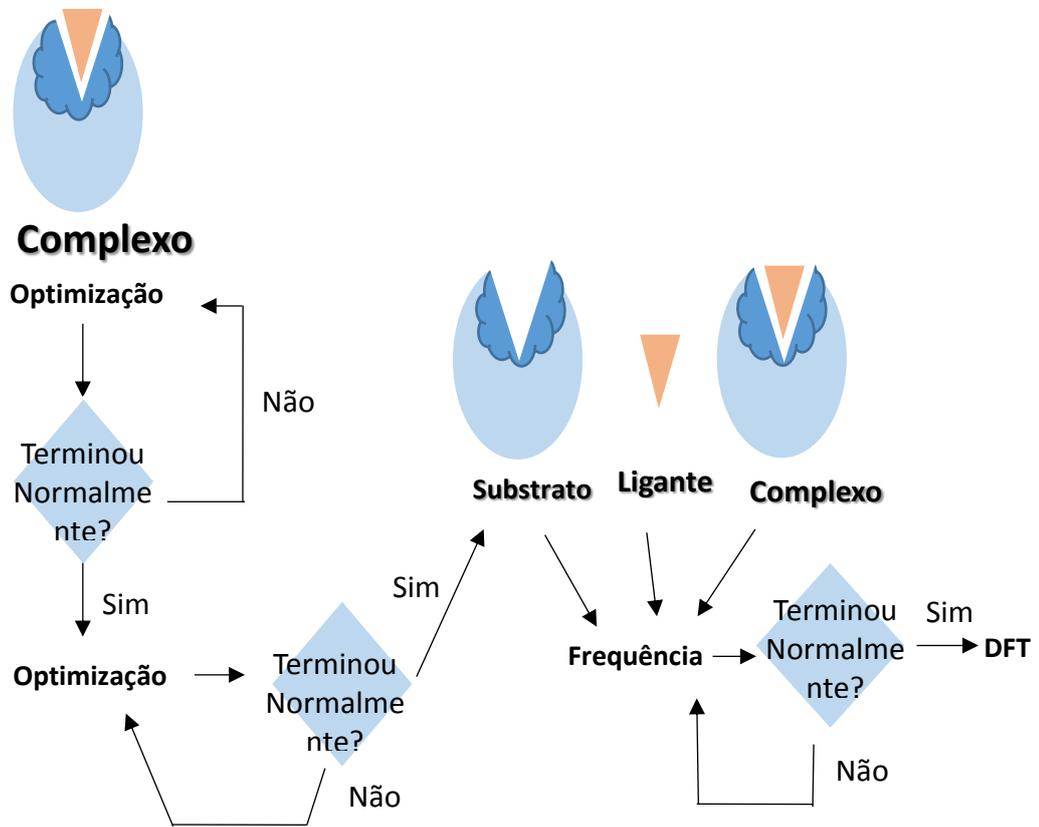


Figura 2.12 - Esquema de cálculos realizado para ONIOM de duas camadas.

Cálculos aplicados ao Complexo

Inicialmente é necessário obter uma geometria relaxada para a estrutura do complexo. A otimização de geometria do complexo iniciou-se com a otimização “*loose*” que é a indicada para otimizações iniciais de grandes sistemas, tendo a característica de definir o critério de convergência. Os cálculos são feitos usando o método quântico semi-empírico PM6 e o método AMBER (parte de MM). Após esta otimização, o complexo é sujeito a outro tipo de otimização: “*calcfc*”. Este tipo de otimização faz com que as constantes de forças sejam calculadas em primeiro lugar, usando os métodos determinados anteriormente.

Quando estas otimizações são finalizadas, são realizados cálculos de frequência para a geometria relaxada do complexo. Isto permite calcular vibrações do sistema, que fornece informações sobre a estabilidade do complexo, por exemplo. Através dos cálculos da frequência é possível saber se existe problemas com a geometria do sistema em estudo, caso se obtenha frequências negativas (frequências imaginárias), isso indica que há instabilidades do sistema. Aqui, são mantidos os mesmos métodos de cálculo que os passos anteriores (PM6:amber). Quando os resultados dos cálculos de frequência demonstram estabilidade, segue-se para o passo seguinte: cálculo da energia total do sistema usando um cálculo de “*Single point energy*”.

O último passo do processo de cálculos consiste no cálculo do “*Single point energy*”, que fornece um valor numérico para a superfície de energia potencial (*potential energy surface*). Este valor numérico descreve a energia do sistema, em função de alguns parâmetros, como por exemplo a posição dos átomos. Para este último passo, usou-se um método de Teoria dos funcionais da Densidade (Density *Functional Theory* – DFT), usando o conjunto de bases 6-31G+(d,p) e o funcional híbrido B97D3, o qual contém um termo de dispersão empírico. Para a parte de MM manteve-se o método Amber.

Cálculos aplicados ao Substrato

Tal como representado na Figura 2.12, o substrato é gerado a partir da estrutura otimizada do complexo. Este consiste, simplesmente, no complexo sem o ligante.

O processo de cálculos do substrato inicia-se com os cálculos de frequência, sendo estes e, os cálculos seguintes iguais aos que foram explicados anteriormente, para o complexo.

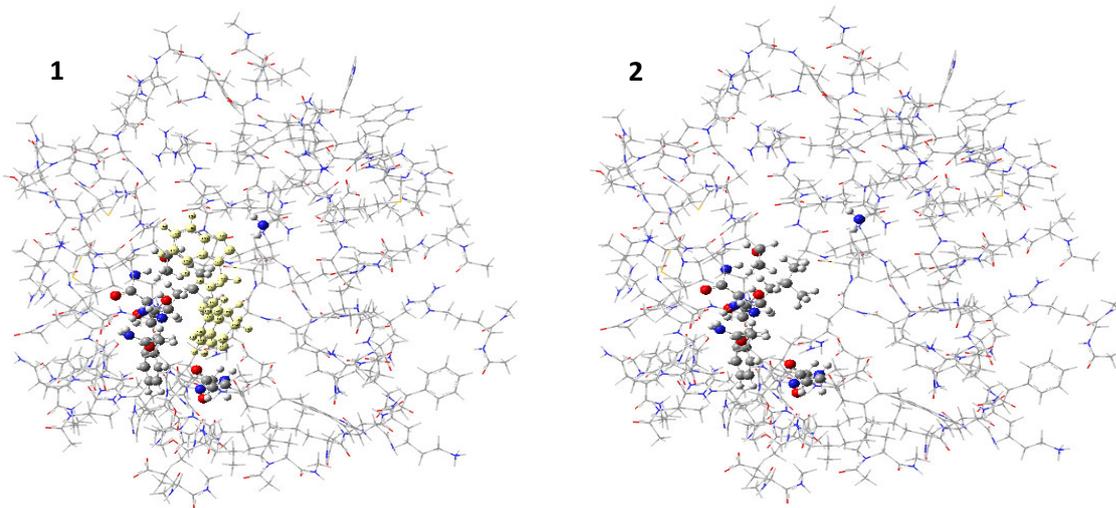


Figura 2.13 - Representação (1) do Complexo e (2) do Substrato. A amarelo está representado o Ligante que desaparece no Substrato.

Cálculos aplicados ao Ligante

O ligante, resulta igualmente do complexo, sendo que do complexo eliminou-se tudo, à exceção do ligante. Este sofreu o mesmo processo que o substrato.

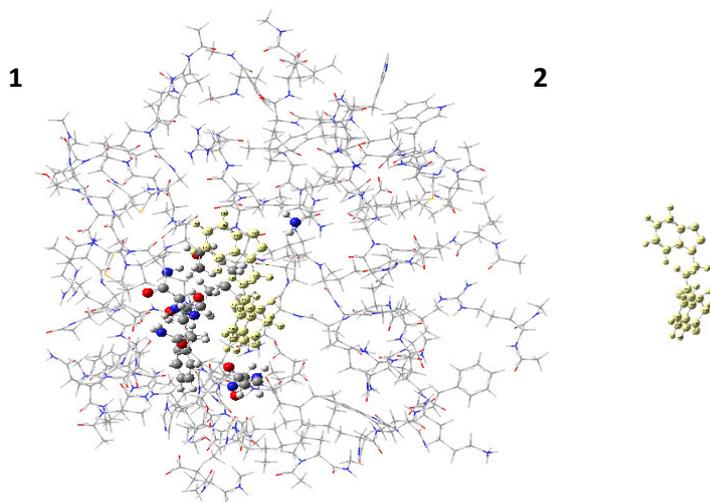


Figura 2.14 - Representação do (1) Complexo e (2) do Ligante.

Como obter as energias de ligação do complexo

Os valores das energias de ligação, em Kcal/mol, foram obtidos através do seguinte processo:

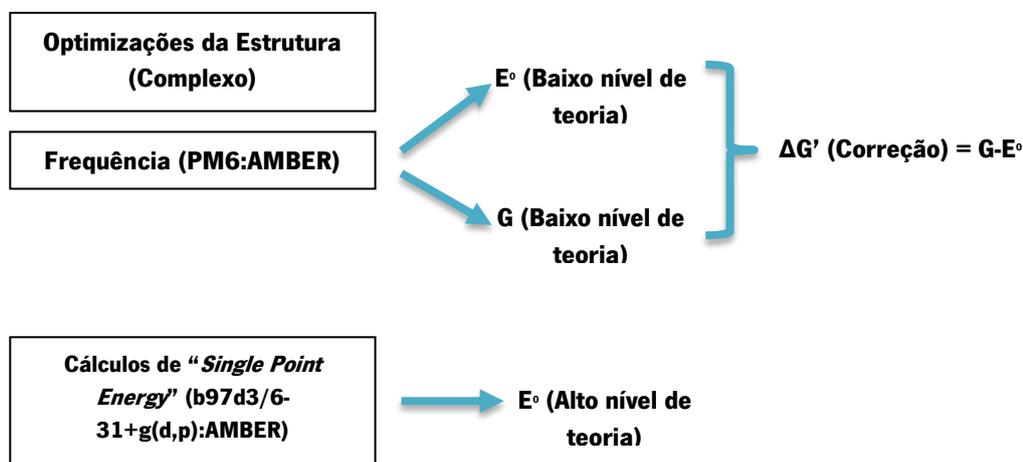


Ilustração 1 - Esquema de cálculos do sistema.

Os primeiros valores usados para chegar à energia final, são os valores obtidos dos cálculos de frequência. Daqui, retiramos a energia eletrônica (E°) e, a energia de *Gibbs* (G), resultantes dos cálculos com baixos níveis de teoria (LT) e, com duas camadas definidas no sistema. A E° é obtida pela subtração das energias das três camadas, e o mesmo ocorre com a energia de *Gibbs*. Deste ponto, e subtraindo a energia de *Gibbs* pela E° , obtém-se a Energia de correção ($\Delta G'$).

Após o cálculo do "Single Point Energy", com um nível de teoria superior, é possível obter a energia de ligação final, calculando a uma nova E° resultante deste último passo. A soma do novo valor de E° com a energia de correção dará o resultado final, em Hartrees, que quando multiplicado por 627.504 fica em Kcal/mol.

$$\begin{aligned}
 &^1 E_{LT}^{\circ} = E_{c_{LT}} - E_{s_{LT}} - E_{lig_{LT}} \\
 &^2 G_{LT} = G_{c_{LT}} - G_{s_{LT}} - G_{lig_{LT}} \\
 &^3 \Delta G' = G_{LT} - E_{LT}^{\circ} \\
 &^4 \text{Energia Total (Kcal/mol)} = (E_{HT}^{\circ} + \Delta G'_{LT}) * 627.504
 \end{aligned}$$

Ilustração 2 - Esquema de Cálculos para obter a Energia Total

2.2.6 ONIOM 3

O método ONIOM com duas camadas por vezes pode falhar, uma vez que temos uma transição abrupta entre as camadas HL e LL. No método ONIOM com três camadas é introduzida uma camada intermédia suavizando a transição entre a HL e LL.

Todo o processo inicial de preparação das esferas do sistema, foi feito da mesma forma, usando o *script* anteriormente referido. Isto, permitiu que no método anterior se tivesse sistemas com o mesmo número de átomos mas, com diferente número de átomos na HL. Para aumentar a precisão dos cálculos, procurou-se que a HL tivesse também o mesmo número de átomos em todas as conformações. Para isso, recorreu-se ao VMD e às esferas já obtidas. De todas as esferas, recolheu-se os resíduos que se encontravam a 3Å do ligante – tal como se fez para as esferas de 10Å. A lista de resíduos obtida, foi colocada no GAUSSIAN de modo a que estes constituíssem a HL. Portanto, depois deste processo, todas as esferas resultantes das quatro conformações de estudo, apresentavam um total de 1764 átomos e, com 429 átomos na HL. Ao definir a HL usando uma lista de resíduos, existiram alguns pormenores que tinham de ser verificados, pois iam contra as regras da definição de camadas. Tal como é possível ver através da Figura 2.15, os erros mais comuns era a divisão das duas camadas em ligações duplas, sendo que algumas vezes tínhamos o Nitrogénio na HL e, o Carbono e Oxigénio na LL e, vice-versa. O que se decidiu fazer foi, incluir sempre o Carbono e Oxigénio na camada em que se encontrava o Nitrogénio.

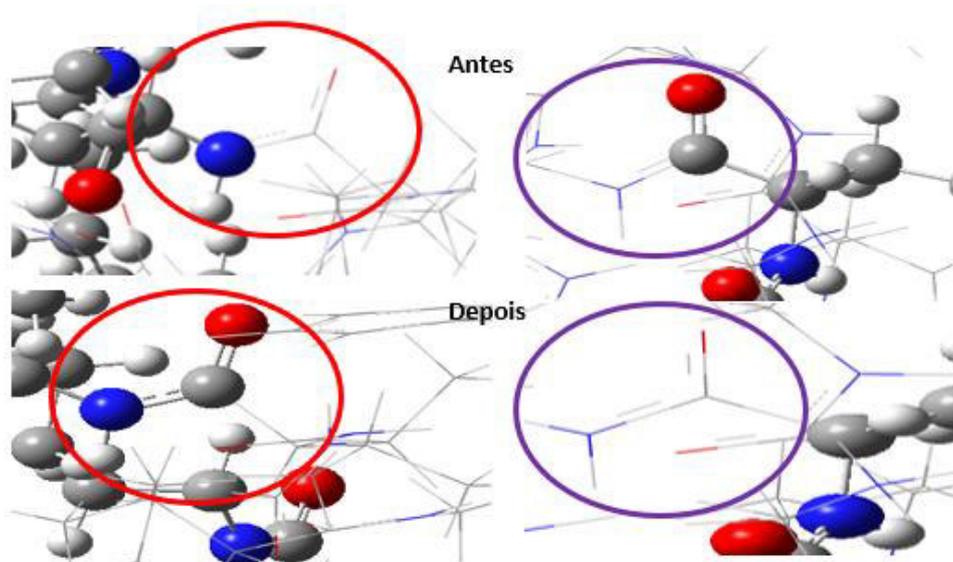


Figura 2.15- Representação dos dois erros mais comuns de divisão de camadas. Os que apresentam a circunferência vermelha, indicam o caso em que há a divisão de camadas numa ligação dupla, em que fica o Nitrogênio na HL e o Carbono com o Oxigênio na LL. A circunferência a roxo, indica o caso oposto.

Terminada esta fase de correção de pequenos erros, era então possível iniciar o processo de cálculos. Este inicia-se do mesmo modo que o método de ONIOM com duas camadas, tal como é possível verificar na Figura 2.16. A seleção da terceira camada é apenas realizada, no complexo e substrato, depois dos cálculos de frequência.

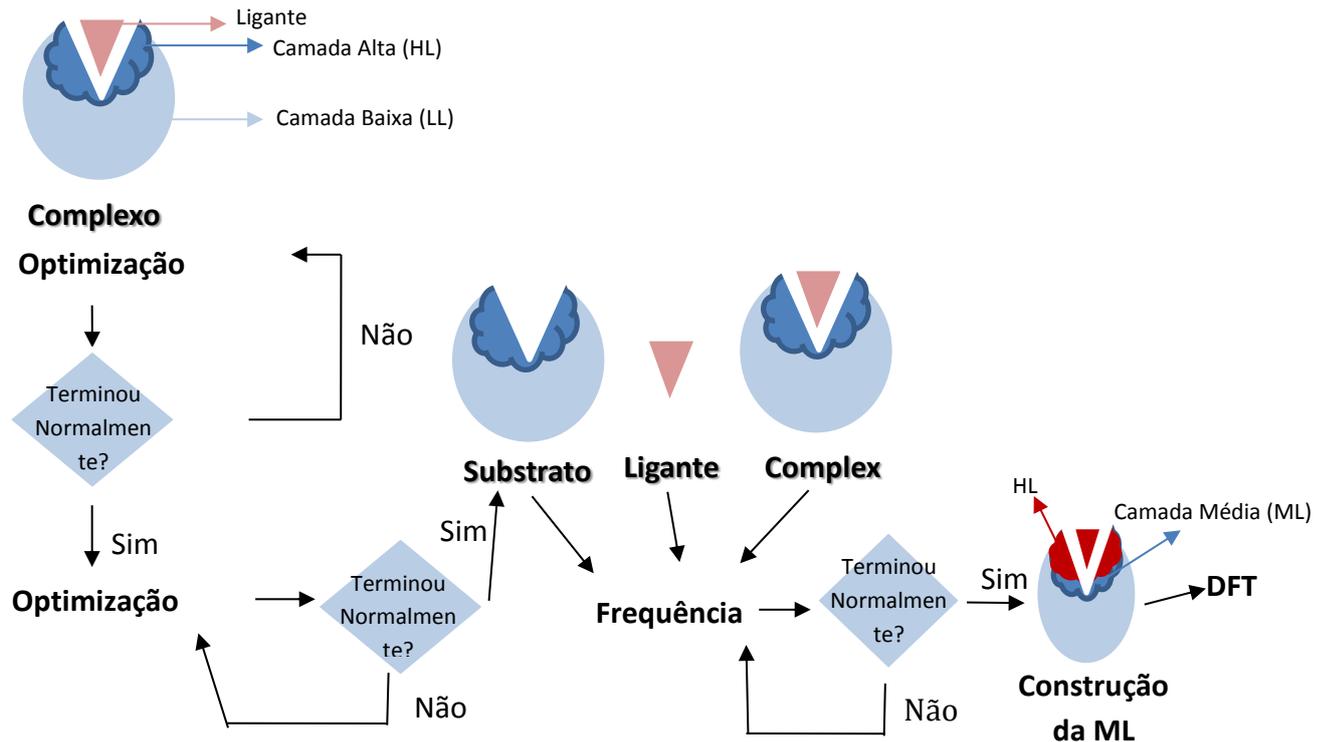


Figura 2.16 - Esquema de cálculos do método ONIOM com três camadas.

A seleção da ML foi realizada tendo como base a metodologia usada na seleção das duas camadas. O primeiro passo, seria passar toda a HL para ML. O segundo, seria colocar o ligante na HL. A partir deste passo, a HL seria definida a partir de cada átomo do ligante, com uma distância de 3 Å. O objetivo seria que a HL não excedesse muito os 100 átomos, sendo que se descartou as interações H-H (Hidrogénio - Hidrogénio) e, as interações entre HL e LL foram ignoradas. As primeiras, foram mantidas na ML e, as últimas sofreram uma pequena alteração, no sentido em que o que pertencia à LL manteve-se e, o que se encontrava na HL passou a ML. No final da seleção, obteve-se um diferente número de átomos na HL de cada conformação estudada, conforme a Tabela 3.

Tabela 3 - Informação sobre o número de átomos por camada, nas esferas de cada conformação estudada.

| SNAPSHOT/CONFORMAÇÃO | 3 | 6 | 9 | 12 |
|-------------------------------------|----------|----------|----------|-----------|
| ÁTOMOS POR ESFERA | 1764 | 1764 | 1764 | 1764 |
| ÁTOMOS POR HL (ANTES DA ML) | 429 | 429 | 429 | 429 |
| ÁTOMOS POR HL (DEPOIS DA ML) | 116 | 123 | 109 | 107 |
| ÁTOMOS POR ML | 313 | 306 | 320 | 322 |
| ÁTOMOS POR LL | 1335 | 1335 | 1335 | 1335 |

Estas “estratégias” de seleção de camadas, faz com que tenhamos um número menor de átomos na HL, sendo tratados com o nível mais alto de teoria o ligante e apenas os resíduos mais próximos deste. (Figura 2.17) Assim, obter-se-á uma maior precisão de cálculos.

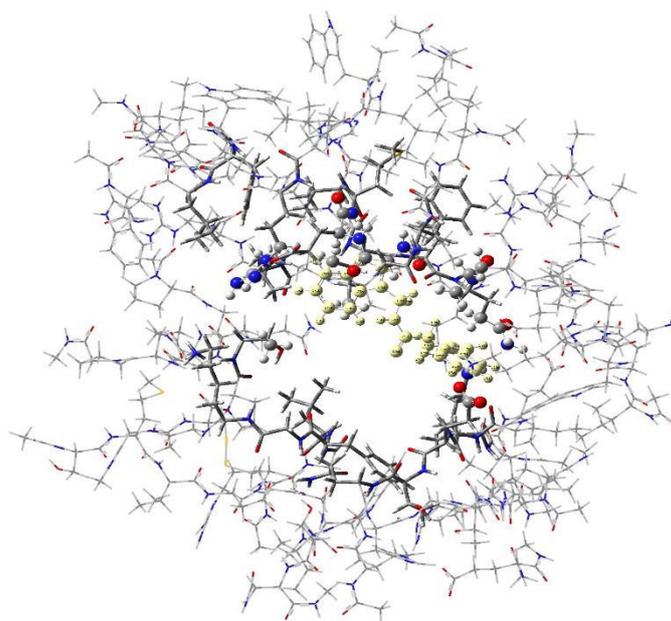


Figura 2.17 - Estrutura com as três camadas definidas. A Amarelo encontra-se o ligante; representados com bolas e traços os resíduos que pertencem à HL; com tubos a ML e, com linhas finas o LL.

Cálculos aplicados ao Complexo

A fase inicial de otimização do complexo, ocorre quase da mesma forma que foi feita com o ONIOM2. A única diferença é que os processos de otimização e frequência do complexo são realizados recorrendo à incorporação mecânica, a qual permite realizar estes processos de forma mais eficiente.

Após a fase de cálculos da frequência do complexo e, depois de comprovada a não existência de frequências imaginárias, realizava-se o cálculo de “*single point energy*”, usando métodos de DFT com incorporação eletrônica. Decidiu-se usar a incorporação eletrônica apenas neste ponto, uma vez que se constatou que quando aplicada em todo o processo, condicionava a convergência das estruturas, atrasando todo o procedimento de cálculos.

Para os cálculos DFT efetuados na fase final, recorreu-se aos mesmos métodos usados anteriormente no ONIOM2, apenas foi adicionado o método PM6 para tratar a ML. Foi importante manter os mesmos níveis de teoria e os mesmos métodos utilizados, para que fosse possível fazer a comparação entre os resultados de ONIOM2 e ONIOM3.

Cálculos aplicados ao Substrato

Os cálculos para o substrato, tal como para o complexo, não diferenciaram muito dos utilizados em ONIOM2. As três camadas do substrato foram definidas após os cálculos de frequência, tal como representados na Figura 2.16 e, a ML foi tratada também com PM6.

Cálculos aplicados ao Ligante

Os cálculos aplicados ao ligante, foram realizados usando o mesmo procedimento que no ONIOM2.

Como obter as energias de ligação do complexo

Os cálculos realizados nesta metodologia, vão muito de encontro ao que foi feito com a abordagem anterior, de duas camadas. A diferença que é mais perceptível, à primeira vista, é o facto de os últimos cálculos necessitarem de três metodologias, ao invés de apenas duas, como anteriormente.

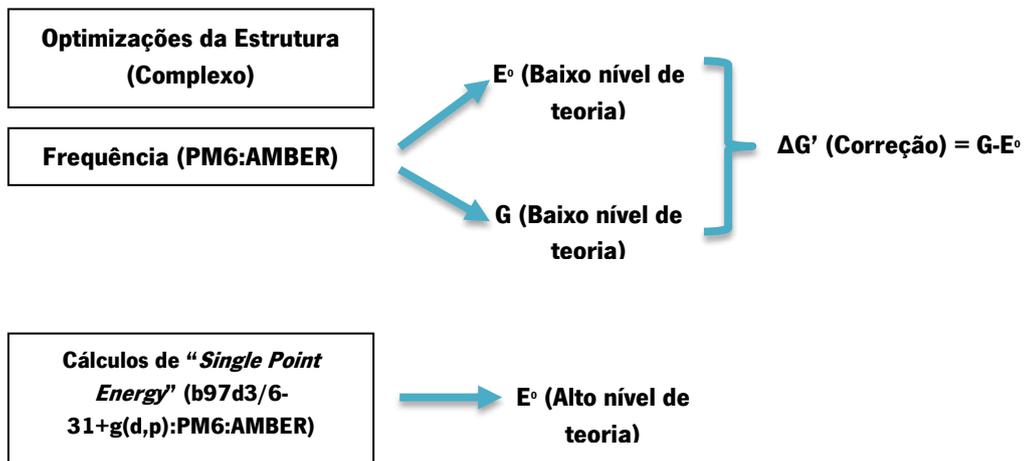


Ilustração 3 - Esquema de cálculos do sistema.

Tal como todo o processo de cálculos é idêntico ao utilizado anteriormente, também o modo de encontrar a energia final é idêntico. No entanto, e uma vez que se tem mais uma camada, é necessário recorrer às equações em diferentes níveis de teoria, do seguinte modo:

$$1 \quad E^{\circ}_{LT} = Ec_{LT} - Es_{LT} - Elig_{LT}$$

$$2 \quad G_{LT} = Gc_{LT} - Gs_{LT} - Glig_{LT}$$

$$3 \quad \Delta G' = G_{LT} - BE_{LT}$$

$$4 \quad E^{\circ}_{HT} = Ec_{HT} - Es_{HT} - Elig_{HT}$$

$$5 \quad G'_{HT} = E^{\circ}_{HT} + \Delta G'_{LT}$$

$$6 \quad \text{Energia Total}_{(Kcal/mol)} = (G'_{HT}) * 627.504$$

Ilustração 4 - Esquema de Cálculos para obter a Energia Total.

Tudo o que resulta dos cálculos de frequência são tratados como baixo nível de teoria (LT) e, o resultante dos cálculos de “*single point energy*” tratados como alto nível de teoria (HT).

3. Resultados

3.1. Dinâmica da Proteína

Apesar do estudo da interação de ligante-proteína poder ser realizado partindo da estrutura de raios -X da proteína armazenada nas bases de dados, dado o carácter dinâmico da proteína, as cavidades interessantes podem pura simplesmente não se encontrar ou estar mal representadas nessas estruturas. Assim, procedeu-se ao estudo prévio da dinâmica da Glicoproteína E por Dinâmica Molecular. Com este processo, foi possível verificar a hidrofobicidade da proteína, a sua dinâmica, as cavidades presentes nela e os seus volumes.

Desta etapa, foi possível obter o RMSD (*root-mean-squared displacement*), que é capaz de monitorizar quantitativamente o comportamento da proteína, ao longo da simulação, em relação à estrutura inicial.

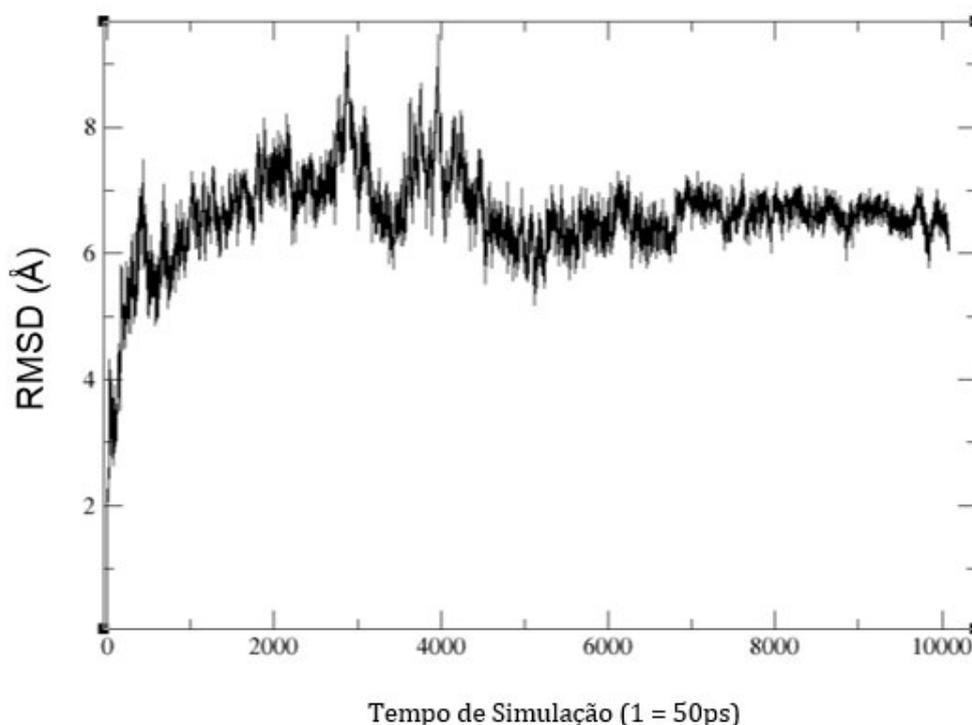


Figura 3.1 - RMSD resultante da DM.

A partir da Figura 3.1, podemos observar que o comportamento da proteína vai estabilizando com o decorrer da simulação. Nos momentos finais da simulação, é possível verificar que há uma menor flutuação do RMSD, conferindo assim uma maior estabilidade da estrutura.

Tal como já foi referido anteriormente, não é apenas o RMSD que é possível obter da DM. Neste caso em concreto, procurávamos uma cavidade que apresentasse as melhores características para o estudar.

Através do MDpocket, foi possível estudar algumas características de uma das cavidades encontrada na estrutura. Uma vez encontrada uma cavidade de interesse, era interessante estudar a sua hidrofobicidade que iria ser crucial na altura de definir o protocolo de estudo, podendo inserir-se a água, ou não. Neste caso, estudou-se a cavidade P_2^0 , representada na Figura 3.2. Esta cavidade tem sido pouco explorada até ao momento, sendo uma candidata promissora para efetuar estudo de inibição da Glicoproteína E com fármacos antivirais.



Figura 3.2 – Cavidade P_2^0 representada na Glicoproteína E.

Na Figura 3.3, está representado o volume da cavidade ao longo da simulação, a qual apresenta algumas oscilações de volume, revelando o seu comportamento dinâmico. Como comportamento dinâmico, poderá estar implícito a abertura e fecho do canal dessa cavidade, onde poderá ser inserido um ligante.

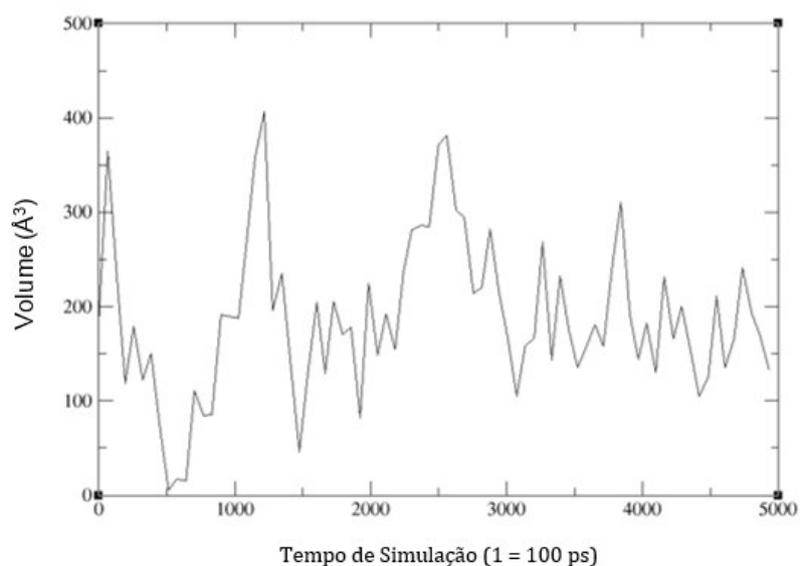


Figura 3.3 - Demonstração do comportamento dinâmico da cavidade, através da variação do volume da mesma ao longo da simulação de DM.

Também desta etapa, é possível verificar um aspeto importantíssimo da cavidade: a Hidrofobicidade. Este é um fator que é extremamente importante conhecer, pois vai implicar (ou não) o uso de métodos de solvatação nos cálculos seguintes.

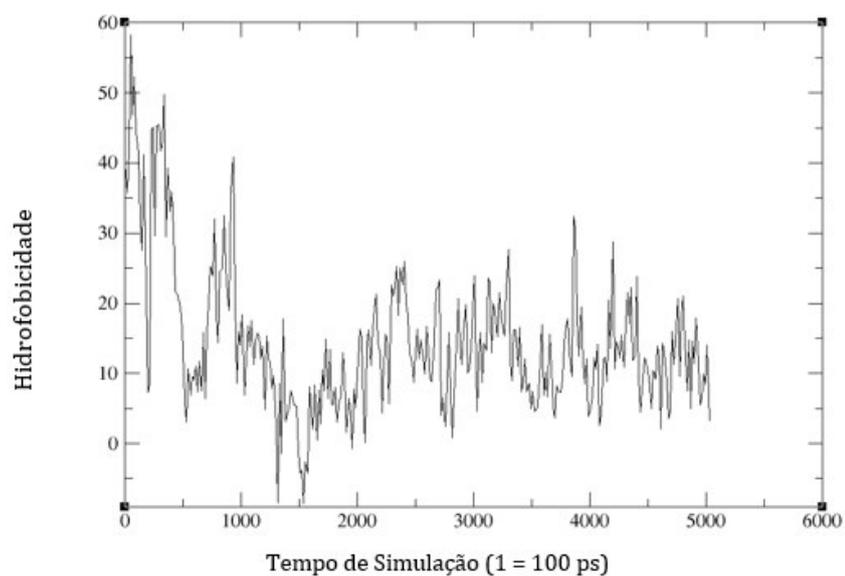


Figura 3.4 - Hidrofobicidade da cavidade P_2^0 .

No caso da cavidade P_2^0 , estamos perante uma cavidade hidrofóbica, como é possível analisar através da Figura 3.4. Assim não é de esperar um papel relevante das moléculas de água na interação do ligante com a proteína, pelo que não será necessário usar qualquer tipo de solvatação no estudo da interação entre ligante e proteína.

3.2. Docking

Após a detecção da cavidade com melhor estabilidade e melhor localização, foi possível verificar quais os ligantes que poderiam ser uma boa aposta para travar a atividade da Glicoproteína E, através de estudo de Docking do ligante com proteína, de modo a obter uma estimativa da probabilidade de ligação dos ligantes à proteína em estudo. Para este estudo foram usados vários *snapshots* da cavidade (P_2^o) obtidos durante a simulação da dinâmica da proteína.

Tal como já foi explicado anteriormente, foram reunidos vários ligantes e testados na cavidade P_2 . Como já se esperava, tendo em conta os valores aos quais se teve acesso por método LIE, o ligante 22232179 poderia ser uma boa molécula para constituir o novo inibidor da atividade da proteína em questão.

O princípio básico de análise dos resultados do Docking é que, quanto mais negativa for a energia obtida, mais estável e provável, será a ligação entre o ligante em causa e a proteína. Através da Figura 3.5, é possível verificar que os primeiros lugares do *score* são os que apresentam menores energias. Isto, implica que o ligante zinc_32524060 seria o que apresentava uma maior probabilidade de interação com a proteína.

Relativamente ao ligante que foi usado neste trabalho, embora este não fosse o melhor, estando no início da lista de *score* final com uma energia de -8.7 Kcal/mol, em conformidade com a energia obtida com o método LIE, mostrou ser uma boa possibilidade de estudo.

| | E (Kcal/mol) | | |
|---|--------------|-------|-------|
| 1.final_zinc_32524060.pdbqt-REMARK VINA RESULT: | -9.4 | 0.000 | 0.000 |
| 2.final_zinc_29077578.pdbqt-REMARK VINA RESULT: | -9.4 | 0.000 | 0.000 |
| 3.final_zinc_16052277.pdbqt-REMARK VINA RESULT: | -9.3 | 0.000 | 0.000 |
| 4.final_zinc_16052277(1).pdbqt-REMARK VINA RESULT: | -9.3 | 0.000 | 0.000 |
| 5.final_zinc_20978604.pdbqt-REMARK VINA RESULT: | -9.2 | 0.000 | 0.000 |
| 6.final_zinc_20610842.pdbqt-REMARK VINA RESULT: | -9.2 | 0.000 | 0.000 |
| 7.final_zinc_18202167.pdbqt-REMARK VINA RESULT: | -9.1 | 0.000 | 0.000 |
| 8.final_zinc_16248871.pdbqt-REMARK VINA RESULT: | -9.1 | 0.000 | 0.000 |
| 9.final_zinc_16248871(1).pdbqt-REMARK VINA RESULT: | -9.1 | 0.000 | 0.000 |
| 10.final_zinc_16248873.pdbqt-REMARK VINA RESULT: | -9.0 | 0.000 | 0.000 |
| 11.final_zinc_16248873(1).pdbqt-REMARK VINA RESULT: | -9.0 | 0.000 | 0.000 |
| 12.final_zinc_20611001.pdbqt-REMARK VINA RESULT: | -8.9 | 0.000 | 0.000 |
| 13.final_zinc_15449300.pdbqt-REMARK VINA RESULT: | -8.8 | 0.000 | 0.000 |
| 14.final_zinc_29077583.pdbqt-REMARK VINA RESULT: | -8.7 | 0.000 | 0.000 |
| 15.final_zinc_22232179.pdbqt-REMARK VINA RESULT: | -8.7 | 0.000 | 0.000 |
| 16.final_zinc_21290170.pdbqt-REMARK VINA RESULT: | -8.7 | 0.000 | 0.000 |
| 17.final_zinc_20744643.pdbqt-REMARK VINA RESULT: | -8.6 | 0.000 | 0.000 |
| 18.final_zinc_20410403.pdbqt-REMARK VINA RESULT: | -8.6 | 0.000 | 0.000 |
| 19.final_zinc_32524059(1).pdbqt-REMARK VINA RESULT: | -8.5 | 0.000 | 0.000 |
| 20.final_zinc_28697056.pdbqt-REMARK VINA RESULT: | -8.5 | 0.000 | 0.000 |
| 21.final_zinc_20744644.pdbqt-REMARK VINA RESULT: | -8.5 | 0.000 | 0.000 |
| 22.final_zinc_20610844.pdbqt-REMARK VINA RESULT: | -8.4 | 0.000 | 0.000 |
| 23.final_zinc_16248897.pdbqt-REMARK VINA RESULT: | -8.4 | 0.000 | 0.000 |
| 24.final_zinc_28697054.pdbqt-REMARK VINA RESULT: | -8.2 | 0.000 | 0.000 |
| 25.final_zinc_27650753(1).pdbqt-REMARK VINA RESULT: | -8.2 | 0.000 | 0.000 |
| 26.final_zinc_18210358.pdbqt-REMARK VINA RESULT: | -8.2 | 0.000 | 0.000 |
| 27.final_zinc_32524059.pdbqt-REMARK VINA RESULT: | -8.1 | 0.000 | 0.000 |
| 28.final_zinc_27650753.pdbqt-REMARK VINA RESULT: | -8.1 | 0.000 | 0.000 |
| 29.final_zinc_18152146.pdbqt-REMARK VINA RESULT: | -8.0 | 0.000 | 0.000 |
| 30.final_zinc_16248899.pdbqt-REMARK VINA RESULT: | -8.0 | 0.000 | 0.000 |
| 31.final_zinc_28436586.pdbqt-REMARK VINA RESULT: | -7.9 | 0.000 | 0.000 |
| 32.final_zinc_18098320(2).pdbqt-REMARK VINA RESULT: | -7.9 | 0.000 | 0.000 |
| 33.final_zinc_17003731.pdbqt-REMARK VINA RESULT: | -7.9 | 0.000 | 0.000 |
| 34.final_zinc_20611041.pdbqt-REMARK VINA RESULT: | -7.8 | 0.000 | 0.000 |
| 35.final_zinc_18847034.pdbqt-REMARK VINA RESULT: | -7.8 | 0.000 | 0.000 |

Figura 3.5 - Energias resultantes do Docking

Também através do Docking foi possível observar um pouco do comportamento do ligante na proteína. Tal como já referido anteriormente, definiu-se um número de “tentativas” de conformações do ligante na proteína, no presente caso, obteve-se 10 conformações diferentes do ligante a interagir com a proteína. (Figura 3.6) Através do *score*, verifica-se que a conformação mais estável, em termos de energia, é preferencialmente a primeira, em grande parte dos ligantes. Isto verifica-se pelo facto de as duas últimas colunas do *score*, estarem a zero, implicando que não houve deslocamentos em relação à posição inicial. Analisando os resultados de *score*, comprova-se que a energia de interação entre ligante e proteína é maior para outros ligantes que não o ZINC22232179, no entanto, este encontra-se com um bom valor de energia (-8.7 Kcal/mol), indicando que poderá ser um bom candidato. Além do mais, resultados experimentais preliminares indicam que este ligante revela atividade inibitória. O Docking é capaz de fornecer alguma informação adicional, em termos visuais, possibilitando o conhecimento do comportamento do ligante, na cavidade, é fácil de verificar que este tem maior preferência de localização no lado direito da cavidade. No entanto, este varia o seu comportamento, não se podendo assim, indicar a posição preferencial deste na cavidade.

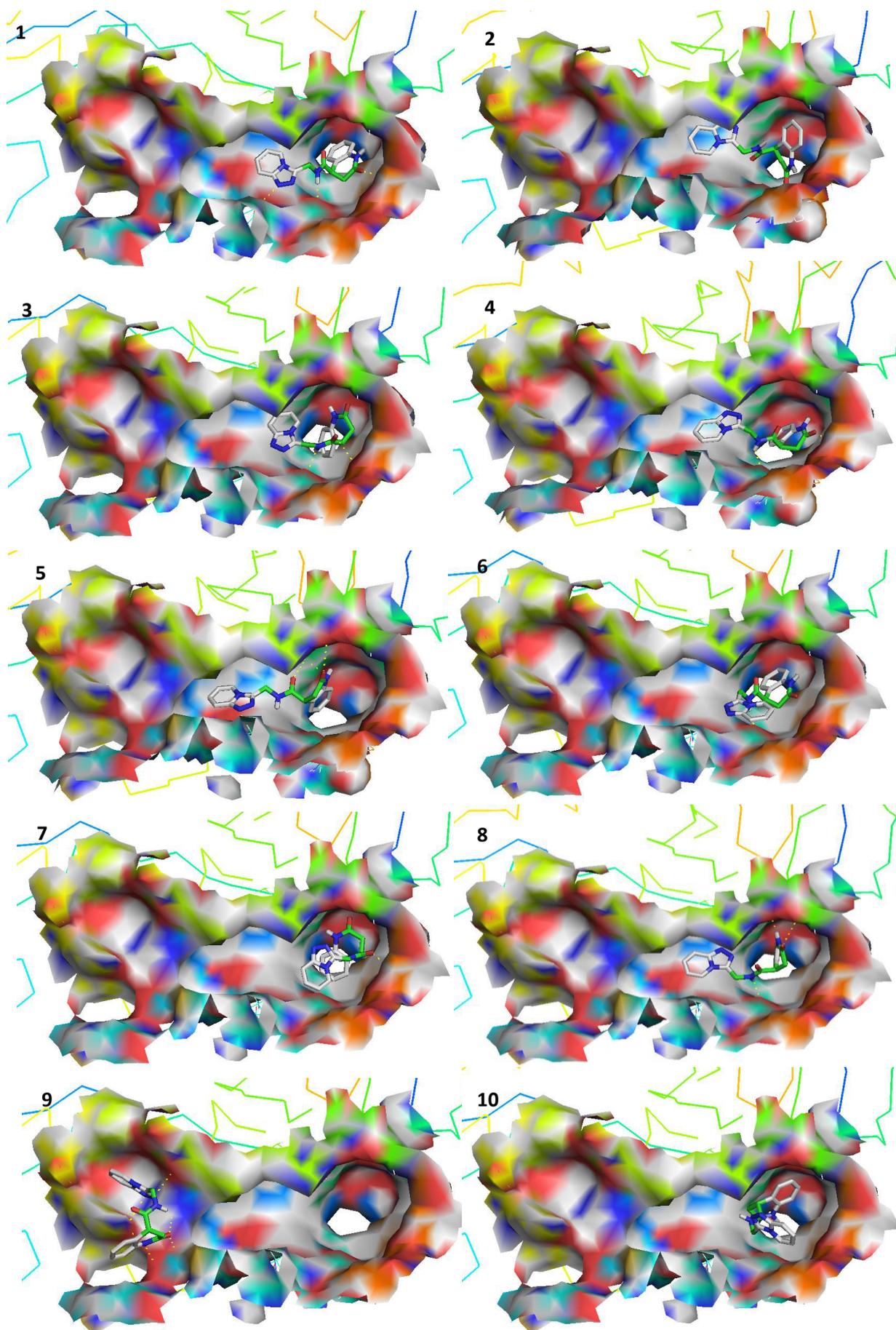


Figura 3.6 - Ilustração do comportamento do ligante dentro da cavidade de estudo, durante o processo de Docking.

3.3. Energias de Ligação usando o método ONIOM

- **Método ONIOM de 2 camadas (ONIOM2)**

A metodologia utilizada de duas camadas é semelhante à reportada na literatura. Como se trata de um sistema de grandes dimensões, mesmo usando uma esfera de 10 Å, obteve-se HL's com dimensões maiores do que o que seria aconselhado.

As energias resultantes da metodologia ONIOM2, tal como é possível verificar Tabela 4, foram um pouco discrepantes.

Tabela 4 - Resultados obtidos nas quatro conformações estudadas com ONIOM2.

| CONFORMAÇÃO | 3 | 6 | 9 | 12 |
|-------------------------------------|----------|----------|----------|-----------|
| Energia Final (Kcal/mol) | -94,3834 | 14,0139 | -46,0567 | -33,6776 |

Os resultados revelam uma grande dispersão nos valores da energia de ligação. A conformação 6 foi a que se revelou mais problemática, uma vez que apresentou valores positivos, enquanto para as restantes obteve-se valores negativos tal como esperado.

Estes resultados são também diferentes dos obtidos por Docking (energias a rondar os -9 Kcal/mol), o que não foi de todo estranho pelo facto de se ter adicionado componentes de incorporação (mecânica e/ou eletrónica), camadas a níveis diferentes e, termos de dispersão. Todos estes componentes contribuem para uma diminuição da energia, como seria de esperar.

As razões para a dispersão de resultados observadas podem ser variadas, desde as regiões de fronteira, entre as camadas, mal realizadas; o facto de se ter um elevado número de átomos na HL, levando à diminuição da precisão na região de interesse; a má escolha do método de incorporação; entre muitas outras possibilidades. Após a análise das possíveis razões que poderão ter levado aos resultados em questão, chegou-se à conclusão que o cerne da questão

está relacionado com a definição das camadas e, as suas grandes dimensões. O que nos levou à adoção da metodologia ONIOM usando três camadas.

- **Método ONIOM usando 3 camadas (ONIOM3)**

Esta abordagem foi realizada com o intuito de se obter uma maior precisão na região de interesse e, conseqüentemente, energias que confirmassem que de facto, o ligante teria grande afinidade de ligação com a Glicoproteína E.

Neste estudo apenas foi possível calcular as energias de ligação para as *snapshots* 3, 9 e 12. Podemos no entanto observar uma menor dispersão nos resultados para as energias de ligação comparados com os do método ONIOM com 2 camadas. Relativamente à *snapshot* 6, era a que apresentava o maior número de átomos na HL depois de definida a ML o que provocou uma maior dificuldade de convergência no processo de otimização, não tendo sido possível obter resultados fiáveis em tempo útil.

Tabela 5 – Resultados obtidos nas quatro conformações estudadas com ONIOM3.

| CONFORMAÇÃO | 3 | 9 | 12 |
|-------------------------------------|--------------|--------------|--------------|
| Energia Final (Kcal/mol) | -17,41347762 | -25,65179579 | -22,60956321 |

De modo geral, é fácil de verificar e de perceber que, com esta abordagem de três camadas, é possível obter energias mais próximas do que era esperado. Para além de obtermos energias mais próximas umas das outras, é um processo idêntico ao utilizado na abordagem de duas camadas, utilizando-se, praticamente, o mesmo tempo e os mesmos recursos computacionais. Neste caso em concreto, observamos também que o facto a presença de água não afeta a interação do ligante com a proteína, devido à elevada hidrofobicidade da cavidade estudada, mas noutros casos poderia ser indispensável, constituindo mais um componente a ter em conta e, que certamente, afetaria o resultado final. No entanto para obter valores realistas da energia de ligação, deveria adicionar-se aos resultados finais uma estimativa para a energia de

solvatação do sistema obtida por exemplo usando MM/PBSA (Molecular Mechanics Poisson Boltzmann Surface Area).

Em conclusão, tendo em conta os recursos disponíveis, os resultados obtidos são bastantes promissores, confirmando que o ligante ZINC22232179 é um fármaco promissor para inibição do Vírus do Dengue.

4. Conclusão

A doença do Dengue, transmitida pelo mosquito *Aedes Aegypti*, tem vindo a afetar parte da população mundial, sem que haja cura ou prevenção possível. O Dengue, têm sido alvo de estudos intensivos usando várias metodologias, sejam elas químicas e biológicas ou computacionais.

Nesta tese, pretendia-se elaborar um protocolo detalhado com os várias metodologias a usar, que permite-se a realização de estudos sistemáticos da interação da Glicoproteína E com novos ligantes e estimar com precisão a energia de ligação dessa interação.

Através da Dinâmica Molecular, foi possível obter uma visão do comportamento dinâmico da proteína e das suas cavidades. O Docking foi possível selecionar um leque de possíveis ligantes com possível atividade inibidora quando inseridos na cavidade (P₂^o). A interação do ligante com a proteína, quando inserido na cavidade (P₂^o), foi estudada usando o método híbrido ONIOM de 2 e 3 camadas, obtendo-se estimativas para energias de ligação entre o ligante e Glicoproteína E.

Uma das conclusões mais evidentes a que se chegou foi que, efetivamente, com a abordagem de 3 camadas do ONIOM é possível obter uma menor dispersão nos resultados da energia de ligação, quando comparado com método ONIOM de 2 camadas. O ligante não apresenta nenhuma ligação preferencial com nenhum dos resíduos da proteína presentes na cavidade, pelo que a interação aparenta ser do tipo de Van Der Waals.

Os resultados das simulações vêm assim de encontro aos resultados experimentais preliminares, realizados pelo Prof. Léo Degréve da Univ. de S. Paulo – Brasil, que confirmam a atividade inibitória do Vírus do Dengue para este ligante.

5. Bibliografia

- Allison, S.L. et al., 1999. Mapping of Functional Elements in the Stem-Anchor Region of Tick-Borne Encephalitis Virus Envelope Protein E. , 73(7), pp.5605–5612.
- Banáš, P. et al., 2009. Theoretical studies of RNA catalysis: Hybrid QM/MM methods and their comparison with MD and QM. *Methods*, 49(2), pp.202–216.
- Barbosa-, G. et al., 2009. Dengue Virus Capsid Protein Usurps Lipid Droplets for Viral Particle Formation. , 5(10).
- Berman, H.M. et al., 2000. The Protein Data Bank. *Nucleic acids research*, 28(1), pp.235–242.
- Bharatam, P. V, Khanna, S. & Francis, S.M., 2008. Drug Design. , 7(3), pp.1–46.
- Browne, T., Coulomb, C.A. De & Volta, A., 2010. Coulomb ' s law The law. , pp.1–8.
- Case, D. a et al., 2011. AMBER 11 Users' Manual. , pp.1–304.
- Case, D. a et al., 2014. Amber 14. *University of California, San Francisco*. Available at: <http://ambermd.org/>.
- Chambers, T.J. et al., 1990. AND REPLICATION ORGJ \ NIZATION , EXPRESSION , I.
- Chambers, T.J. et al., 2003. Yellow Fever Virus / Dengue-2 Virus and Yellow Fever Virus / Dengue-4 Virus Chimeras : Biological Characterization , Immunogenicity , and Protection against Dengue Encephalitis in the Mouse Model. , 77(6), pp.3655–3668.
- Cheatham, T.E., I. et al., 1995. Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal of the American Chemical Society*, 117(14), pp.4193–4194. Available at: <http://pubs.acs.org/doi/abs/10.1021/ja00119a045>.
- Cheung, D.L. et al., 2008. Computer simulation of liquids and liquid crystals. *Computer Physics Communications*, 179(1-3), pp.61–65.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, K.P. (1995)., 1995. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *J. Am. Chem. Soc.* 117: 5179–5197., pp.1–4.
- D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, D.M.Y. and P. a. K., 2015. Amber2015.
- Dastmalchi, S., 2016. *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery*,
- Erlanson Daniel A. et al., 2016. *Fragment-based Drug Discovery: Lessons and Outlook*, Volume 67 de *Methods and Principles in Medicinal Chemistry*.

- Fadrná, E., Hladecková, K. & Koca, J., 2005. Long-range electrostatic interactions in molecular dynamics: an endothelin-1 case study. *Journal of biomolecular structure & dynamics*, 23(2), pp.151–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16060689>.
- Faheem, M., Raheel, U. & Nasir, M., 2011. A molecular evaluation of dengue virus pathogenesis and its latest vaccine strategies. , pp.3731–3740.
- Fischer, E., 1894. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.*, 27, pp.2985–2993.
- FREIRE, C.C.D.M., 2014. *Caracterização de processos evolutivos de vírus de RNA a partir de padrões deixados nas filogenias virais Tese.*
- Frisch, M.J., 2003. Geometry Optimization with QM / MM , ONIOM , and Other Combined Methods . I . Microiterations. *Journal of Computational Chemistry*, 9627775.
- Fuzo, C. a. & Degrève, L., 2013. New pockets in dengue virus 2 surface identified by molecular dynamics simulation. *Journal of Molecular Modeling*, 19(3), pp.1369–1377.
- Hammes, a. S.. O., 2012. MODELAGEM MOLECULAR DE INIBIDORES DE ASPARTIL PROTEASE : POTENCIAIS NOVOS COMPOSTOS ANTIMALARIAIS AMANDA SUTTER DE OLIVEIRA HAMMES Rio de Janeiro Julho de 2012 INSTITUTO OSWALDO CRUZ. *Dissertação de mestrado - FIOCRUZ, Instituto Oswaldo Cruz.*
- Irwin, J.J. & Shoichet, B.K., 2005. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1), pp.177–182.
- Jones, G. et al., 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267(3), pp.727–48. Available at: <http://www.sciencedirect.com/science/article/pii/S0022283696908979>
<http://www.ncbi.nlm.nih.gov/pubmed/9126849>.
- Jorgensen, W.L. et al., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), p.926. Available at: <http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869>
<http://link.aip.org/link/?JCPSA6/79/926/1>.
- Kerrigan, E., J., 2011. AMBER 11.0 Introductory Tutorial. , (732), pp.1–24.
- Leach, A., 2001. *Molecaular Modelling: Principles and Applications*,
- Lengauer, T. & Rareyt, M., 1996. Computational methods for biomolecular docking. , pp.402–406.
- Liang, J., Edelsbrunner, H. & Woodward, C., 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein science : a publication of the Protein Society*, 7, pp.1884–1897.

- Mark, P. & Nilsson, L., 2001. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105(43), pp.9954–9960. Available at: <http://pubs.acs.org/doi/abs/10.1021/jp003020w>.
- Melino, S. & Paci, M., 2007. Progress for dengue virus diseases Towards the NS2B – NS3pro inhibition for a therapeutic-based approach. , 274, pp.2986–3002.
- Meng, X.-Y. et al., 2011. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2), pp.146–57. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3151162&tool=pmcentrez&rendertype=abstract>.
- Michael Hardy, Den fjättrade ankan~enwiki, Thorwald, Saintswithin, Unquietwiki, S. & Gtdp, Kiefek, Blastwizard, Thunderboltz, DanielPenfield, QTCaptain, Dicklyon, Geo.per, Quevvy, B., 2016. Root-mean-square deviation of atomic positions. *Wikipedia*, pp.1–3. Available at: https://en.wikipedia.org/wiki/Root-mean-square_deviation_of_atomic_positions.
- Modis, Y. et al., 2003. A ligand-binding pocket in the dengue virus envelope glycoprotein.
- Modis, Y. et al., 2004. Structure of the dengue virus envelope protein after membrane fusion. , (February).
- Morokuma, K. et al., 1996. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)₃)₂ + H₂ Oxidative Addition. *Journal of Physical Chemistry*, 100(50), pp.19357–19363.
- Mukhopadhyay, S., Kuhn, R.J. & Rossmann, M.G., 2005. A structural perspective of the. , 3(January).
- Mundial, S.G. et al., 1940. Dengue. , pp.1–15.
- Naeem, R., 2013. Lennard-Jones Potential. *Chemwiki*, pp.3–6. Available at: http://chemwiki.ucdavis.edu/Core/Physical_Chemistry/Physical_Properties_of_Matter/Atomic_and_Molecular_Properties/Intermolecular_Forces/Specific_Interactions/Lennard-Jones_Potential.
- Namba, A.M., Silva, V.B. & Silva, C.H.T.P., 2008. Dinâmica molecular : teoria e aplicações em planejamento de fármacos. , 33, pp.13–23.
- OTHMAN, R., 2010. *Ligand-Protein Interactions of Inhibitors With Dengue Virus Type 2 Serine Protease : a Structural Study*.
- Pasquini, C., 2003. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), pp.198–219.
- Pokidysheva, E. et al., 2006. Cryo-EM Reconstruction of Dengue Virus in Complex with the Carbohydrate Recognition Domain of DC-SIGN. , pp.485–493.
- Rapaport, D.C.C., 2004. *The Art of Molecular Dynamics Simulation*,

- Rodenhuis-zybert, I.A., 2010. Dengue Virus Life Cycle : Viral and Host Factors Modulating Infectivity. , (June 2016).
- Roehrig, J.T., Bolin, R.A. & Kelly, R.G., 1998. Monoclonal Antibody Mapping of the Envelope Glycoprotein of the Dengue 2 Virus , Jamaica. , 328.
- Ross, T.M., 2010. Dengue Virus. , 30, p.15261.
- Santos Soares, R.O., 2013. Processos preliminares da infecção viral : estudo estereoquímico da proteína E do Dengue Processos preliminares da infecção viral : estudo estereoquímico da proteína E do Dengue.
- Schmidtke, P. et al., 2011. MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, 27(23), pp.3276–3285.
- Seeliger, D. & De Groot, B.L., 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-Aided Molecular Design*, 24(5), pp.417–422.
- Soares, C.M., 1997. Simulação de proteínas usando métodos de mecânica / dinâmica molecular. , pp.33–44.
- Van der Spoel, D. et al., 2010. 4.3.1 Position Restraints. *Gromacs User Manual version 4.5.6*, pp.77–78. Available at: www.gromacs.org.
- Standard, J.M., 2015. *Energy Minimization Methods Other Minimization Algorithms*,
- Steffen, C. et al., 2010. TmoleX--a graphical user interface for TURBOMOLE. *Journal of computational chemistry*, 31(16), pp.2967–2970.
- Stephan, D.W., 2008. "Frustrated Lewis Pairs": a Concept for New Reactivity and Catalysis. *Organic & Biomolecular Chemistry*, 6(9), pp.1535–1539.
- Swope, W.C., 1982. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76(1), p.637. Available at: <http://link.aip.org/link/?JCP/A6/76/637/1>.
- Tavodova, M., 2012. Dengue fever. *South Sudan Medical Journal*, 5(1), pp.13–16.
- Trott, O. & Olson, A., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 31(2), pp.455–461.
- Vreven, T. & Morokuma, K., 2006. Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Annual Reports in Computational Chemistry*, 2(C), pp.35–51.
- Warshel, a. & Levitt, M., 1976. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2), pp.227–249.

- Whitehead, S.S. et al., 2007. Prospects for a dengue virus vaccine. *Nat Rev Microbiol*, 5(July), pp.518–528. Available at: [internal-pdf://8b-2014368512/8b.pdf\nhttp://www.ncbi.nlm.nih.gov/pubmed/17558424](http://www.ncbi.nlm.nih.gov/pubmed/17558424).
- Who, 2009. Dengue: guidelines for diagnosis, treatment, prevention, and control. *Special Programme for Research and Training in Tropical Diseases*, p.147.
- Wilschut, I.A.R.J., 2010. Dengue virus life cycle : viral and host factors modulating infectivity. , pp.2773–2786.
- Zhang, Y. et al., 2014. Conformational Changes of the Flavivirus E Glycoprotein. , 12(9), pp.1607–1618.
- Zhang, Y., 2013. Electrostatic Interaction of the Electrostatic-Embedding and Mechanical-Embedding Schemes for QM/MM Calculations. *Communications in Computational Chemistry*, 1(2), pp.109–117. Available at: <http://www.global-sci.org/cicc/volumes/v1n2/pdf/12-109.pdf>.