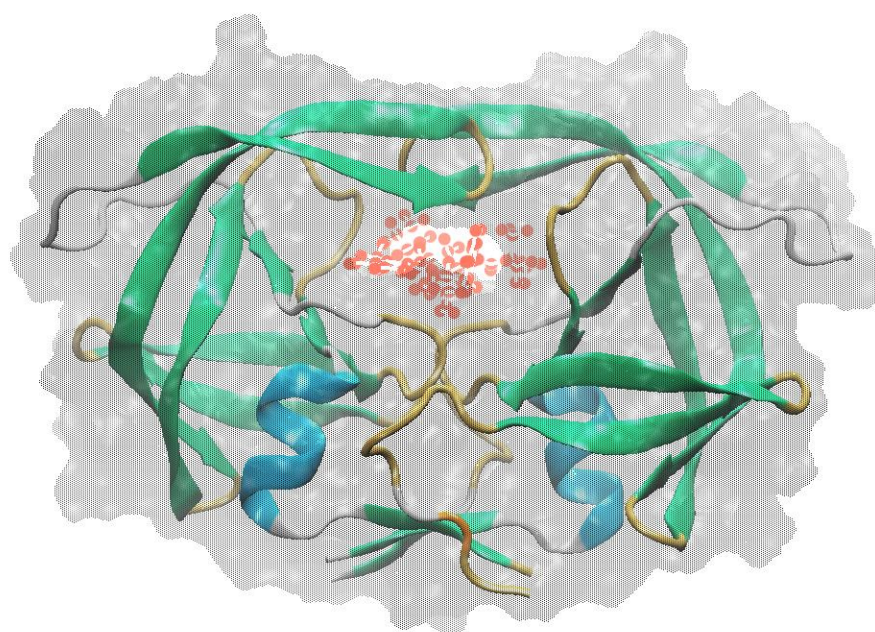


**Maria Ribeiro Vieira**

**Nº Mec.: up201908936**

**Mecânica e Dinâmica Moleculares no estudo do  
sistema Protease HIV-1–Nelfinavir (inibidor)**



**Fig.1** – Representação gráfica da estrutura da Protease do HIV-1, em complexo com o fármaco inibidor Nelfinavir. PDB ID: 1OHR. *Software* de visualização molecular: VMD. Proteína colorida por estrutura secundária (azul pastel –  $\alpha$ -helice; verde pastel –  $\text{extended } \beta$ -sheet; laranja –  $\beta$ -bridge; laranja pastel –  $\text{turn}$ ; branco –  $\text{coil}$ ), em sobreposição com a respetiva representação de superfície; inibidor (a rosa) representado no modelo *ball & stick*.

**Maria Ribeiro Vieira**

**Nº Mec.: up201908936**

**Mecânica e Dinâmica Moleculares no estudo do sistema Protease HIV-1–Nelfinavir (inibidor)**

**Relatório**

Trabalho realizado sob a orientação científica dos Professores Pedro Fernandes e Alexandre Magalhães, no âmbito da unidade curricular de Bioquímica Computacional.

**Palavras-chave**

Mecânica Molecular (MM); Dinâmica Molecular (MD); proteína; protease do HIV-1; inibidor; nelfinavir; estrutura; trajetória; conformação; microestado; energia; potencial; termodinâmica; ponte de hidrogénio; RMSD; distância interatômica; comprimento de ligação; RDFs.

**Agradecimentos**

Ao Sérgio Santos e ao José Santos, orientadores de projeto de licenciatura, por me apresentarem à Bioquímica Computacional; ao Pedro Vieira, meu pai, por lidar, compreensivelmente, com os frutos do meu *stress*; ao Pedro Fernandes e ao Alexandre Magalhães, professores da unidade curricular, por guiarem o trabalho; e a todos os outros “alfaiates de moléculas” cujo nome desconheço, mas que contribuem para o enriquecimento da Bioquímica Computacional.

## Abreviaturas

Å	<i>Angstrom</i>
AMBER	<i>Assisted Model Building with Energy Refinement</i>
Asp	Aspartato
atm	Atmosfera (unidade de pressão)
Cl <sup>-</sup>	Cloreto
DNA	<i>DeoxyRibonucleic Acid</i>
fs	Femtosegundo
Glu	Glutamato
HIV	<i>Human Immunodeficiency Virus</i>
Ile	Isoleucina
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
K	Kelvin
MD	<i>Molecular Dynamics</i>
MM	<i>Molecular Mechanics</i>
ns	Nanosegundo
NPT	Número de partículas, Pressão e Temperatura constantes
OD	Oxigénio ligado ao carbono <i>Delta</i>
OE	Oxigénio ligado ao carbono <i>Épsilon</i>
PDB	<i>Protein Data Bank</i>
PME	<i>Particle Mesh-Ewald</i>
ps	Picosegundo
RDFs	<i>Radial Distribution Functions</i>
RMSD	<i>Root Mean Square Deviation</i>
RNA	<i>RiboNucleic Acid</i>
SIDA	Síndrome da Imunodeficiência Adquirida
s	Segundo
Δt	Passo de integração
t	Tempo
VMD	<i>Visual Molecular Dynamics</i>

## Índice

<b>Introdução.....</b>	<b>1</b>
Protease do HIV-1.....	1
Mecânica Molecular.....	1
Campo de Forças.....	2
Dinâmica Molecular.....	3
Integração das equações de movimento.....	3
Condições periódicas de fronteira.....	4
Cálculo das forças de curto alcance.....	5
Cálculo das forças de longo alcance.....	5
Minimização, Equilibração e Produção.....	6
RDFs ( <i>Radial Distribution Functions</i> ).....	6
RMSD ( <i>Root Mean Square Deviation</i> ).....	7
<b>Materiais e Métodos.....</b>	<b>8</b>
Obtenção da estrutura inicial.....	8
Preparação do ficheiro *.pdb.....	8
Ficheiros necessários às simulações.....	10
Ficheiro de parâmetros, topologias e coordenadas.....	10
Ficheiros de comandos.....	12
Etapas da simulação.....	13
Minimizações de energia.....	13
Dinâmica Molecular.....	15
Equilibração (MD1).....	16
Produção (MD2).....	16
<b>Resultados e Discussão.....</b>	<b>17</b>
Visualização dos resultados.....	17
Propriedades gerais.....	19
RMSD.....	22
Pontes de Hidrogénio.....	23
Distâncias interatómicas e comprimentos de ligação.....	25
RDFs.....	27
<b>Conclusão.....</b>	<b>32</b>
<b>Bibliografia.....</b>	<b>33</b>

## **Introdução**

*“A ciência computacional é o sistema operativo para toda a inovação.”*

*Steve Ballmer*

Com o intuito de reproduzir a realidade física experimental de sistemas bioquímicos e estudar o seu comportamento ao longo do tempo, a aplicação de técnicas de modelação molecular a simulações computacionais emerge como área em otimização.

O presente relatório resume o trabalho desenvolvido sob o complexo Protease do HIV-1–Nelfinavir (um inibidor análogo do substrato), por aplicação de métodos de Mecânica e Dinâmica Moleculares.

### **Protease do HIV-1**

A protease do HIV-1 é um homodímero proteico que catalisa um passo fulcral na maturação do vírus da imunodeficiência humana (HIV): ativação das proteínas virais, para constituição de um vírus maduro competente/funcional. Este, para além de possuir uma elevadíssima capacidade de replicação ( $10^9$  cópias/dia), induz a apoptose celular, levando à morte gradual do organismo infetado. Desta forma, o estudo da protease do HIV-1 constitui um campo de suma importância para o desenvolvimento de fármacos combatentes da SIDA.

Existem várias formas de evitar a propagação do vírus: inibidores da transcriptase reversa – que, ao impedirem o processo de transformação do RNA viral de cadeia simples em DNA de cadeia dupla, inabilitam a sua consequente integração no genoma; inibidores, nomeadamente, péptidos, com maior afinidade para o recetor transmembranar que as proteínas virais; e inibidores da própria protease do HIV (1), como é o caso do Nelfinavir.

### **Mecânica Molecular**

Uma das áreas de suporte mais relevantes da Bioquímica Computacional é a Mecânica Molecular (*Molecular Mechanics* – MM) – uma metodologia computacional que se destina ao cálculo da energia potencial total de um arranjo específico de átomos.

É um método de *simples* e rápida implementação, com potencial aplicação a sistemas moleculares de grandes dimensões. Nesta abordagem, os átomos são descritos como partículas esféricas carregadas e indeformáveis, ligadas covalentemente por “molas” ou em interação (2). Para descrever o seu movimento, a MM assume que a massa nuclear das entidades atômicas é superior à dos elétrons da nuvem. Desta forma, as partículas de carga negativa acompanham os movimentos nucleares e possibilita-se o cálculo de superfícies de energia potencial, em função das suas coordenadas. Intuitivamente, a função de energia total é dada pela soma das funções individuais, descritas por parâmetros experimentais – como constantes de força e valores de equilíbrio – definidos para as várias interações ligantes (energias de ligação, torção e flexão de ângulos) e não ligantes, de curto ou longo alcance (energia de *Van der Waals* e energia eletrostática, respetivamente), específicos do campo de forças selecionado (3).

### **Campo de forças**

Os campos de forças descrevem, então, a interação entre as partículas (neste caso, átomos) de um sistema. São constituídos por um conjunto de equações de base física, necessárias para o cálculo da energia potencial e das forças subjacentes, e pelos respetivos parâmetros. À combinação destes componentes atribui-se o nome de parametrização. Imprecisões na parametrização levam, muitas das vezes, a conformações incorretas e enganosas (4). Notavelmente, parâmetros como o comprimento de ligação, ângulos de ligação, quiralidades, planos e restrições (ajustes na função energética para evitar desvios bruscos do modelo em estudo) de posição, orientação e distância, diedros e diedros impróprios, são calculados com base em valores ideais fixos. Isto resulta numa função de energia mais simples e, portanto, numa eficiência muito maior do algoritmo.

No âmbito do trabalho vigente, utilizou-se o AMBER (*Assisted Model Building with Energy Refinement*) – um pacote de programas para simulações de Dinâmica Molecular, equipado com uma família de campos de força, especialmente direcionados para a simulação de sistemas biológicos, nomeadamente, proteínas e ácidos nucleicos. Para a realização deste estudo, utilizaram-se as bibliotecas gerais FF03, que possui parâmetros para os diversos resíduos presentes, e GAFF (*General Amber Force Field*), com parâmetros para alguns tipos de átomos, ligações, ângulos e diedros presentes em moléculas orgânicas genéricas, bem como parâmetros de van der Waals (3).

Apesar da variedade de programas constantes do AMBER, para executar uma simulação tradicional de Dinâmica Molecular são apenas necessários dois aplicativos: o módulo de cálculo *Xleap* e o *Sander*. O *Sander* é o “pai” de todo o mecanismo da simulação: lê os ficheiros de entrada que descrevem o sistema molecular a ser simulado, i.e., ficheiros de parâmetros e topologia (\*.prmtop) e de coordenadas iniciais (\*.inpcrd). Estes são gerados pelo *Xleap*, a partir de coordenadas predeterminadas no ficheiro \*.pdb; lê um arquivo de comandos, especificando as condições da simulação (\*.in); e executa-a, calculando uma base de trajetória da Dinâmica Molecular clássica nessas informações (\*.mdcrd) e armazenando as respectivas coordenadas finais, num ficheiro de reinicialização (\*.rst - *restart*). Para além dos cálculos associados, a interface gráfica do *Xleap* possibilita a visualização do sistema molecular no espaço tridimensional (ver secção ‘Preparação de ficheiros necessários às simulações’).

Neste trabalho, fez-se também uso do *software* auxiliar *ptraj* – um aplicativo integrado, destinado à análise e processamento de séries temporais de coordenadas (trajetórias) e dos dados correspondentes (5). No caso em questão, o módulo foi utilizado para o cálculo dos desvios padrão de posição (RMSD), ângulos e comprimentos médios de pontes de hidrogénio, determinação das funções de distribuição radial (RDFs) e respectivas correlações temporais.

## **Dinâmica Molecular**

### **Integração das equações de movimento**

A Dinâmica Molecular (*Molecular Dynamics* – MD) é uma extensão da MM, onde o comportamento dinâmico de um sistema atómico é, resumidamente, simulado através da integração das equações Newtonianas de movimento. Utiliza a derivada da função de energia de ligação e calcula, para a posição correspondente a um dado instante  $t_0$ , a força resultante em cada átomo; a partir desta, determina a respetiva aceleração, velocidade e coordenadas da partícula no espaço, no tempo seguinte; as forças são, então, atualizadas, de acordo com as novas posições atómicas e o ciclo de cálculo retoma, o número de vezes definido pelo utilizador (proporcional ao número de estruturas geradas). Assim, de uma simulação de MD, resulta um conjunto de configurações ordenadas no tempo, equivalente a um *ensemble* microestados ou, por outras palavras, uma sequência temporal de trajetórias (de coordenadas cartesianas e momentos atómicos lineares definidos) do sistema macroscópico. No entanto, é de notar que as geometrias obtidas não

passam de aproximações (6). Isto porque a força atuante em cada átomo depende da sua posição em relação aos outros e calculá-la implica assumir que se mantém constante durante o passo de integração definido – o intervalo de tempo  $\Delta t$ , desde a posição inicial, em  $t_0$  até à posição atualizada, em  $t_0 + \Delta t$ . Ora, estando em constante movimento no espaço tridimensional, as entidades colidem entre si; consequentemente, as diferentes velocidades são transferidas “partícula-a-partícula”; e o cálculo das suas posições exatas fica tanto mais comprometido, quanto maior o passo de integração – que, na realidade, equivale a menos de 1fs de estaticidade posicional! Desta forma, e visto que a cada macroestado corresponde um número infinito de microestados e que o tempo prático de cada simulação, apesar de longo, é limitado – restringindo a exploração de todo o espaço conformacional –, as propriedades das trajetórias geradas nunca serão, exatamente, equivalentes às características observáveis. Apesar disso, a Mecânica Estatística assume que: (a) na prática, a média de cada propriedade molecular é *igual* à propriedade termodinâmica correspondente, i.e., função de estado; (b) sendo  $\alpha$  o número de componentes de um sistema, são necessárias  $\alpha+2$  funções de estado independentes para definir o seu estado macroscópico (3). Deste modo, tendo em conta que o modelo em estudo é formado por 4 elementos (proteína, nelfinavir, iões cloreto e água – adicionada explicitamente), seriam necessárias  $4+2 = 6$  variáveis descritoras para definir o seu estado termodinâmico. No entanto, no presente relatório são apenas apresentadas as relações temporais das funções de estado Volume, Pressão, Temperatura e Energia Total.

### **Condições periódicas de fronteira**

Antes de executar uma simulação de MD é necessário delimitar o espaço de coordenadas. Para tal, o sistema deve ser pré-introduzido numa caixa de simulação, de dimensão e geometria específicas para o modelo em estudo. Esta equivale a uma rede 3D infinita de réplicas da célula central (célula de simulação), a única na qual são efetuados os cálculos necessários para a dinâmica. Os movimentos das estruturas nas células envolventes são, rigorosamente, mimetizados a partir dos movimentos correspondentes na célula de simulação. Assim, as células-réplica são constituídas por moléculas-imagem das moléculas da célula central, proporcionando o nivelamento das interações ao longo de todo o espaço conformacional, i.e., ultrapassando os constrangimentos inerentes aos efeitos de superfície. Para além disso, as condições periódicas de fronteira garantem a conservação do número total de partículas (N) na caixa central e na globalidade do sistema (se um átomo sai da célula original, a sua imagem entra pela face oposta, com a



mesma velocidade). No entanto, é, natural e computacionalmente, impossível calcular infinitas interações. Sendo assim, para evitar que uma dada partícula interaja com a sua imagem ou consigo própria, e permitir um cálculo aproximado das interações não ligantes – energia de *Van der Waals* e energia eletrostática – para todo o sistema, é necessário delimitar um espaço de cálculo, de acordo com uma dada distância interatômica (7).

### **Cálculo das forças de curto alcance**

No caso de potenciais de curto alcance, recorre-se, habitualmente a uma de duas metodologias: convenção da imagem mínima ou corte esférico. Na primeira, os átomos são centrados numa região de forma e dimensão semelhantes à da célula de simulação e apenas são contabilizadas, para o cálculo das interações do átomo central, as partículas nela contidas; na segunda, os átomos são centrados numa esfera de raio pré-definido (*cutoff*) e procede-se analogamente (8).

### **Cálculo das forças de longo alcance**

No caso das energias de longo alcance, como o potencial de interação não decai tão rapidamente com o aumento da distância interatômica, truncamentos abruptos nas interações calculadas podem comprometer, significativamente, a exatidão das trajetórias. Para ultrapassar este problema, utilizam-se, habitualmente, métodos baseados nas somas de Ewald, como o PME (*Particle Mesh-Ewald*). Nesta metodologia, as cargas pontuais das partículas interatuantes (a longa distância) são neutralizadas pela adição das nuvens eletrónicas de carga oposta. Consequentemente, ocorre o fenómeno de blindagem dos núcleos e a distância de interação diminui o suficiente para que o potencial passe a ser considerado uma força de curto alcance. Deste modo, as metodologias anteriormente enunciadas podem ser aplicadas para o cálculo aproximado das interações. Posteriormente, de modo a anular o efeito de cancelamento do potencial eletrostático e restaurar as condições “originais”, nuvens de carga igual às das partículas são somadas ao sistema. Assim, tendo em conta que a densidade eletrónica de cada nuvem segue uma distribuição gaussiana centrada no ião, a energia total pode ser, analiticamente, calculada. Esta resulta da soma das energias de interação núcleo-núcleo, do potencial atrativo entre as nuvens dos iões blindados e da interação entre as nuvens somadas para contrabalançar as cargas totais (3).

Para executar as simulações de Dinâmica Molecular relativas ao modelo Protease do HIV-1–Nelfinavir, considerou-se um  $\Delta t = 1$  fs e simulou-se o sistema, com um *cutoff*

= 10 Å, numa caixa paralelepipedica de solvatação explícita (incorporação de moléculas de água no modelo, como componente adicional), cujo limite da célula central distava 12 Å da extremidade do complexo biomolecular. Desta forma, impossibilitou-se a interação entre réplicas conformacionais de células diferentes.

### **Minimização, Equilibração e Produção**

Notavelmente, de forma a eliminar maus contactos entre os átomos e aumentar a estabilidade do sistema inicial, antes de iniciar uma simulação de MD, deve proceder-se à minimização da energia potencial do sistema. Apesar de informação mais detalhada sobre a técnica de minimização poder ser encontrada na secção ‘Materiais e Métodos\ Etapas da Simulação\ Minimizações’, adianta-se que, após a minimização, o sistema, ainda está longe de atingir o equilíbrio termodinâmico. Desta forma, inicia-se a Equilibração – a primeira etapa de MD. Nesta fase, o sistema é, gradualmente, aquecido até à Temperatura de interesse (neste caso, de 0 K a 310.15 K) e as velocidades iniciais ( $\approx$  quantidade de movimento), aleatoriamente, atribuídas a cada partícula, de forma a que no conjunto sigam uma distribuição de Maxwell-Boltzmann (e possibilitem a integração das equações de movimento). Precisamente por corresponder a uma fase de adaptação – do sistema e das suas propriedades – este período de Inicialização é variável, dando-se por finalizado quando se alcança o equilíbrio termodinâmico. Inicia-se, portanto, a Produção – a segunda fase da MD, na qual se obtém as configurações, aproximadamente, equivalentes aos microestados do sistema e se calcula as respetivas propriedades termodinâmicas (7). No trabalho em questão, as trajetórias da etapa de produção foram geradas num *ensemble* NPT (isotérmico-isobárico), i.e., mantendo constantes as funções de estado Número de partículas, Pressão e Temperatura.

### **RDFs (*Radial Distribution Functions*)**

A ponte entre as propriedades termodinâmicas macroscópicas e a realidade de interações intermoleculares tem por base funções de distribuição radial (*Radial Distribution Functions* – RDFs). Tal como o nome indica, as RDFs representam a densidade de partículas (neste caso, átomos) que se encontram a uma determinada distância radial de uma partícula central, de natureza igual ou diferente das entidades circundantes. Usualmente e no caso em questão, as RDFs são utilizadas em sistemas biomoleculares em solução, para estudar a distribuição de moléculas de solvente, em torno de um determinado átomo de referência (3).

### **RMSD (*Root Mean Square Deviation*)**

Claramente, após a execução de várias simulações, é necessário averiguar a compatibilidade entre as estruturas tridimensionais previstas (geradas pelas dinâmicas), que mimetizam a molécula em solução, e a estrutura globular nativa, na forma de cristal. Uma maneira comum de estudar a similaridade entre conformações proteicas é o cálculo do RMSD (*Root Mean Square Deviation*), cuja fórmula se encontra representada na equação 1, onde  $\delta_i$  é a diferença entre as posições do átomo  $i$  (trajetória) e a entidade correspondente na estrutura de referência ou a posição média dos  $N$  átomos equivalentes.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (Eq. 1)$$

Assim, quanto maior for o RMSD, maiores são as diferenças entre as estruturas em comparação; quanto menor, mais similares são as conformações. Se o RMSD é igual a zero, então foi encontrado um “emparelhamento perfeito”, *i.e.*, os átomos da estrutura prevista e de referência ocupam exatamente as mesmas coordenadas no espaço tridimensional (9). Mais especificamente: apesar de se preferirem desvios de posição atômica inferiores a 1 Å, é comum a obtenção de confórmeros semelhantes numa gama de RMSD entre 1 a 2 Å; se as estruturas diferem vagamente nas cadeias laterais, então desviam-se entre 2 a 3 Å; se as diferenças nos grupos R dos resíduos são consideráveis e afetam a cadeia principal, as coordenadas de posição são significativamente distintas e distam, em média, 3 a 4 Å; RMSDs superiores a 4 Å apontam para indícios de desnaturação; e, a partir dos 6 Å, a molécula encontra-se claramente desnaturada.

## **Materiais e Métodos**

*“O nosso cérebro simula a realidade.”*

*Stephen LaBerge*

*“A boa notícia é que os computadores fazem o que lhes dizemos.”*

*Theodor Nelson*

### **Obtenção da estrutura inicial**

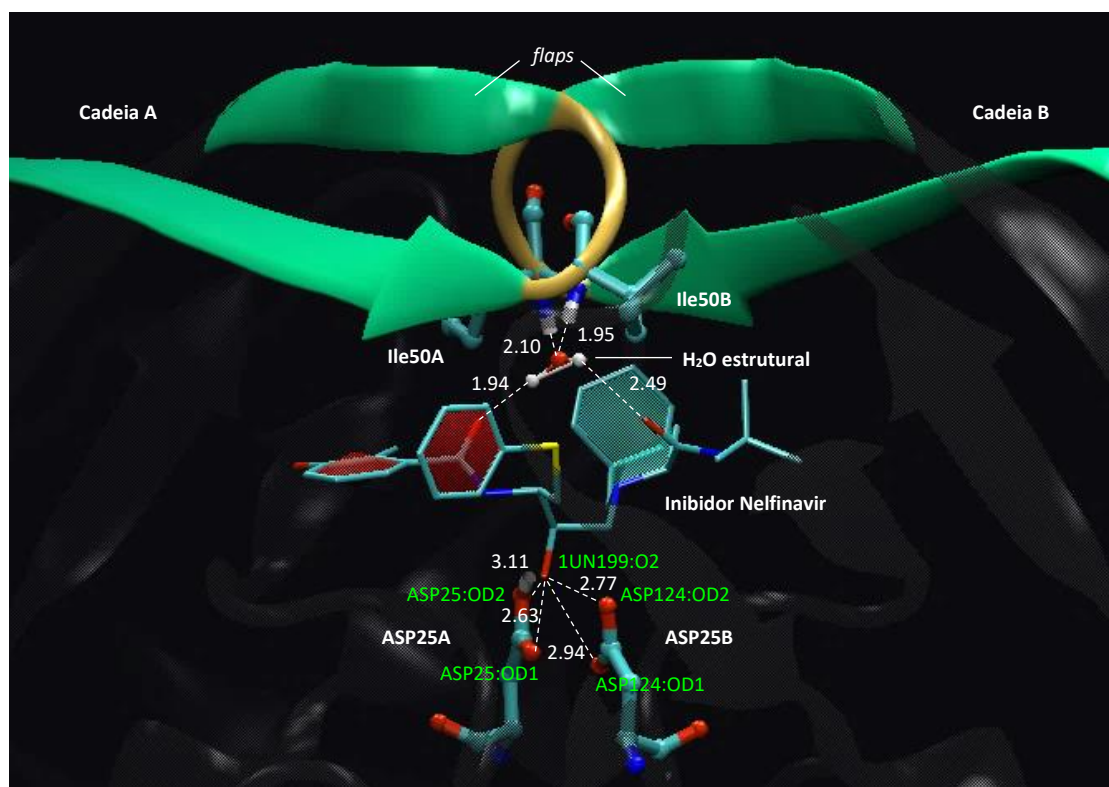
Um ficheiro \*.pdb é um tipo de ficheiro utilizado pelo PDB (*Protein Data Bank*), para representar tridimensionalmente a estrutura de uma proteína nativa (referência). No âmbito do projeto vigente, obteve-se, a partir da base de dados supracitada, o ficheiro \*.pdb inicial (1OHR.pdb), correspondente ao complexo Protease do HIV-1 – Nelfinavir. Assegurou-se que o arquivo selecionado correspondia à estrutura isenta de mutações com menor resolução. Contudo, aos ficheiros \*.pdb de origem estão, tipicamente, associados problemas resultantes de limitações experimentais: por ex., é praticamente impossível obter um modelo cristalográfico enzima-substrato não modificado, visto que tal exigiria uma continua inalteração do centro ativo da proteína e, apesar de rápidas ( $< 1$  s), as reações de consumo de substrato levam a alteração dos seus resíduos de aminoácidos; em outros casos, dada a sua volatilidade, verifica-se a falta de moléculas de água (muitas vezes, imprescindíveis para a reação de catálise); e, para além disso, uma vez que os raios-X só difratam na nuvem eletrónica, é habitual, em estruturas obtidas por cristalografia, a ausência dos átomos de hidrogénio e/ou a incorreta distribuição dos mesmos, quando posteriormente colocados no espaço tridimensional. Por este motivo, o trabalho foi orientado a partir de uma configuração inicial previamente fornecida, sem hidrogénios (1OHR\_withoutHs.pdb) e sem camada de solvatação (moléculas de água até  $\approx 2$  Å de distância), ultrapassando o problema associado à conformação de origem (1OHR.pdb).

### **Preparação do ficheiro \*.pdb**

No entanto, é de notar ainda que estruturas obtidas a partir do PDB raramente são compatíveis com os programas utilizados e precisam de ser ajustadas aos modelos pretendidos. Desta forma, para garantir que o arquivo \*.pdb contém a informação necessária para os cálculos de Mecânica e Dinâmica Moleculares e que as linhas que não conseguem ser modeladas pelo *software* estão ausentes, o novo ficheiro foi limpo: a

informação irrelevante foi removida e, de forma a evitar a repulsão no centro ativo, um dos resíduos de aspartato catalítico (Asp25, na cadeia A) renomeado para a designação equivalente à forma protonada (ASH) no AMBER. Para além disso, teve-se em atenção os requisitos de preparação dos ficheiros de parâmetros, topologia e coordenadas, no módulo *Xleap* (ver secção ‘Preparação dos ficheiros necessários às simulações’). Este aplicativo assume que a ordem pela qual os resíduos aparecem listados no ficheiro \*.pdb corresponde à ordem de ligação, a não ser que estejam separados por uma linha com a designação TER – identificador de terminação de cadeia. Deste modo, certificou-se que as duas subunidades peptídicas (cadeia A e cadeia B) e inibidor se encontravam assim delimitadas, no arquivo 1OHR\_withoutHs.pdb. Assim, no final do processo, o ficheiro \*.pdb continha apenas a informação compatível com o *software* e a necessária para as simulações: coordenadas atómicas, elemento químico, resíduo a que cada átomo pertence e a sua posição. Assegurou-se que todas as entidades estavam corretamente nomeadas e numeradas. Este arquivo foi útil para avaliar visualmente a qualidade do alinhamento com as estruturas geradas pela dinâmica (ver secção ‘Propriedades gerais\ RMSD’).

Para visualização da estrutura “1OHR\_withoutHs” (fig.1) e análise da conformação dos resíduos de aspartato catalíticos, bem como da localização espacial da molécula de água estrutural (resíduo 205), do fármaco Nelfinavir e dos *flaps* da enzima (resíduos 45 a 55, na cadeia A e 144 a 154, na cadeia B), recorreu-se ao *software* de visualização molecular VMD (*Visual Molecular Dynamics*) (fig.2). Todavia, para que o inibidor seja reconhecido pelo AMBER, de forma a poder-lhe atribuir os parâmetros correspondentes, a nomenclatura dos elementos estruturais equivalentes (1UN) deve ser reconhecida nas bibliotecas do *software*. Uma vez que isso não acontece, parâmetros específicos para o inibidor e consistentes com o campo de forças utilizado foram, previamente, determinados e gravados: no ficheiro 1UN.fcmold ficou guardada a informação relativa aos tipos de átomos, ligações, ângulos, diedros e parâmetros de van der Waals; no ficheiro 1UN.prepc, os dados relativos às conectividades e cargas do Nelfinavir. Assim, a adição da biblioteca ao programa é suficiente para o reconhecimento da molécula e possibilita a sua visualização no espaço tridimensional (fig.2).



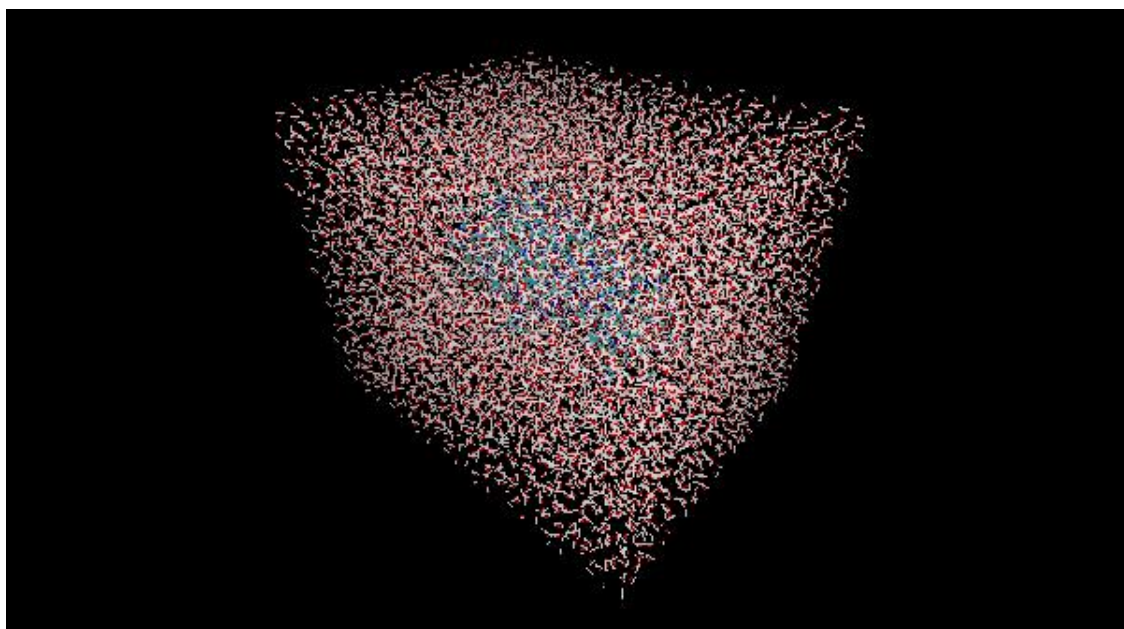
**Fig.2** – Representação gráfica do centro ativo da Protease do HIV-1, ilustrando as principais interações enzima-inibidor, bem como a posição dos resíduos de aspartato catalítico (Asp25, na cadeia A e Asp124, na cadeia B) e da molécula de água estrutural, mediadora das ligações por pontes de hidrogénio assinaladas. PDB ID: 1OHR. *Software* de visualização molecular: VMD. Proteína colorida por estrutura secundária, com *flaps* destacados; inibidor, molécula de água e restantes elementos estruturais coloridos por tipo de átomo e representados num modelo *spacefill* (CDK).

## Preparação dos ficheiros necessários às simulações

### Preparação dos ficheiros de parâmetros, topologia e coordenadas (*Xleap*)

Para preparar os ficheiros necessários ao início do processo de cálculo de uma simulação de Dinâmica Molecular no AMBER, recorreu-se ao módulo *Xleap* – um *software* de cálculo e visualização molecular integrado, originalmente, equipado com um conjunto de parâmetros e conectividades, definidos para os vários resíduos. Tal como explicado anteriormente, o AMBER (e, consequentemente, o *Xleap*) não reconhece elementos estruturais diferentes dos constantes nas suas bibliotecas. Deste modo, para além das bibliotecas gerais FF03 e GAFF, para que o inibidor fosse reconhecido no aplicativo, as bibliotecas de parâmetros correspondentes, anteriormente criadas pelo utilizador, foram também carregadas. O ficheiro \*.pdb modificado foi adicionado e os átomos de hidrogénio em falta, automaticamente, recolocados pelo *software*, em posições iniciais aleatórias/não otimizadas. Posteriormente, com recurso ao comando “*charge*”,

determinou-se a carga global do sistema biomolecular ( $5^+$ ). Assim, constatou-se que havia um excesso de carga positiva que necessitava de ser contrabalançado, antes da inclusão do modelo em solvente explícito: adicionaram-se contra-íões  $\text{Cl}^-$  (resíduos 200 a 204), na proporção necessária para anular a carga total. Para tal, o programa constrói uma função de potencial de *Coulomb* sob uma grelha de dimensões = 1 Å e adicionou os contra-íões, um a um, nos pontos de coordenadas correspondentes aos mínimos ou, neste caso, máximos de potencial eletrostático. No final da etapa, o complexo enzima-inibidor e o sistema envolvente ficaram eletricamente neutros. Posteriormente, de forma a mimetizar as condições experimentais, delimitar o espaço de trajetórias e garantir um cálculo lógico das funções energéticas do campo para todas as corridas independentes, o sistema foi introduzido numa caixa de água paralelepípedica pré-equilibrada (pré-parametrizada), cujas faces distam 12 Å das extremidades da proteína (fig.3). No decorrer deste passo, constrangimentos alostéricos entre o solvente e contra-íões são evitados por eliminação automática de moléculas de água específicas, num raio pré-definido. Como explicado anteriormente, a informação gerada pelo *Xleap* foi guardada nos ficheiros de terminologia \*.prmtop e \*.inpcrd – arquivos de entrada para o módulo *Sander*, utilizado nas minimizações.



**Fig.3** – Representação gráfica da caixa de simulação de solvente explícito (moléculas de água) para o modelo proteína-inibidor considerado. *Software* de visualização molecular: *Xleap*.

### **Preparação dos ficheiros de comandos (Editor de texto)**

Finalmente, obtiveram-se os ficheiros necessários para a otimização geométrica, i.e., recolocação dos átomos no espaço tridimensional, em posições compatíveis com um estado de menor energia (mais estável) da estrutura – aquele na qual a resultante de forças atuante em cada átomo é zero. Os ficheiros de comandos com as instruções relativas às três etapas de minimização subsequentes foram criados manualmente, com recurso a um editor de texto e de acordo com especificações conhecidas: volume constante; 500 ciclos de minimização por etapa – 250 dos quais, com recurso à meta-heurística *Steepest Descent* e os 250 seguintes, com recurso ao algoritmo *Conjugate Gradient*, mais eficaz na otimização perto dos mínimos da função.

Em ambos os casos, o sistema é minimizado de acordo com a topologia (disposição e propriedades dos átomos no espaço): a energia potencial da conformação atual é avaliada; as forças resultantes são calculadas e aplicadas a cada átomo; os quais se movem nas direções e sentidos determinados, com uma certa intensidade; “salta-se” assim para um novo ponto na função energética; e a conformação de partida é atualizada para a estrutura resultante das novas coordenadas espaciais. O ciclo repete-se o número de vezes definidas pelo utilizador.

No algoritmo de minimização *Steepest Descent*, as novas coordenadas (de cada átomo) dependem: da posição anterior; do sinal e amplitude da derivada calculada no ponto da função energética equivalente à conformação atual da proteína; e de um *learning rate*, correspondente à magnitude de salto necessária para atingir as posições seguintes (equivalentes à estrutura atualizada e, portanto, a um mínimo da função de energia potencial). No algoritmo de otimização iterativa *Conjugate Gradient*, a pesquisa no espaço de coordenadas não só é orientada na direção do gradiente (derivada), como usa o histórico de saltos para acelerar a convergência para o mínimo local (10).

Notavelmente, as forças calculadas são aplicadas a todos os átomos e, consequentemente, os diedros das estruturas secundárias podem mover-se. No entanto, a estabilização das conformações locais é praticamente garantida (fig.6) pela função de energia potencial, a qual incorporou todas as restrições impostas nos ficheiros de entrada (distâncias entre átomos, coordenadas de posição, etc) e obriga os diedros das estruturas secundárias a confinarem-se aos valores de equilíbrio pré-definidos. Assim, por cada força calculada que obrigue os átomos a executar um movimento que potencie a



destabilização de uma estrutura local, dado esse potencial, uma força oposta, obriga-os a permanecer na conformação. Para além das restrições de origem (\*.inpcrd), garantiu-se que os novos constrangimentos espaciais aplicados à parte do sistema que se desejou fixar (primeira e segunda otimizações) constavam dos arquivos de comandos.

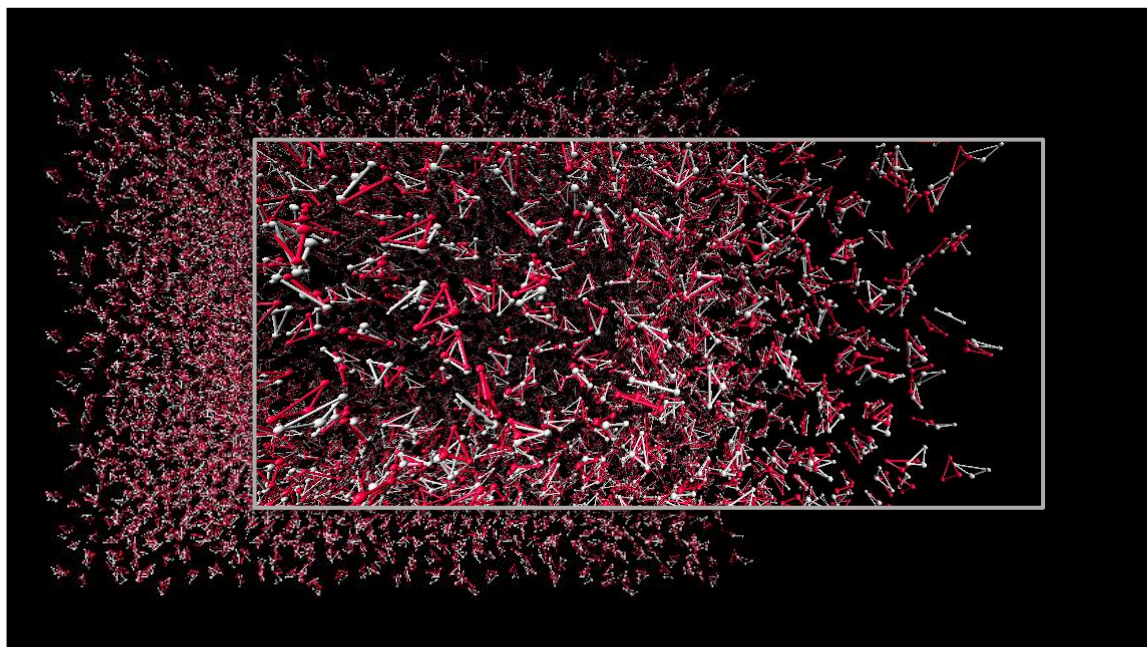
## Etapas da simulação

### Minimizações de energia

*“A vida não se encontra em átomos ou moléculas (...), mas sim na organização.”*

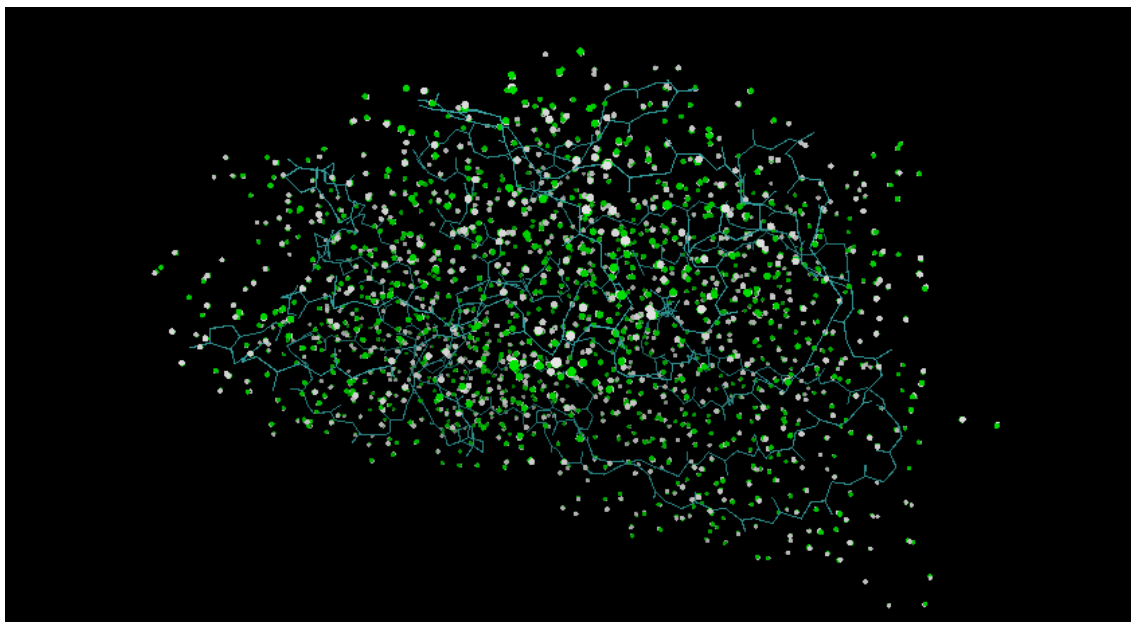
*Edwin Conklin*

Preparados os ficheiros, invocou-se o módulo *Sander* e executou-se cada uma das três minimizações supracitadas: na primeira (min1), as moléculas de água praticamente em equilíbrio, foram otimizadas e o resto do sistema manteve-se fixo. Este passo é importante para evitar que águas mal colocadas induzam a desnaturação. Daqui resultou um modelo não muito diferente do inicial (fig.4), cujas coordenadas serviram de entrada para a minimização seguinte;

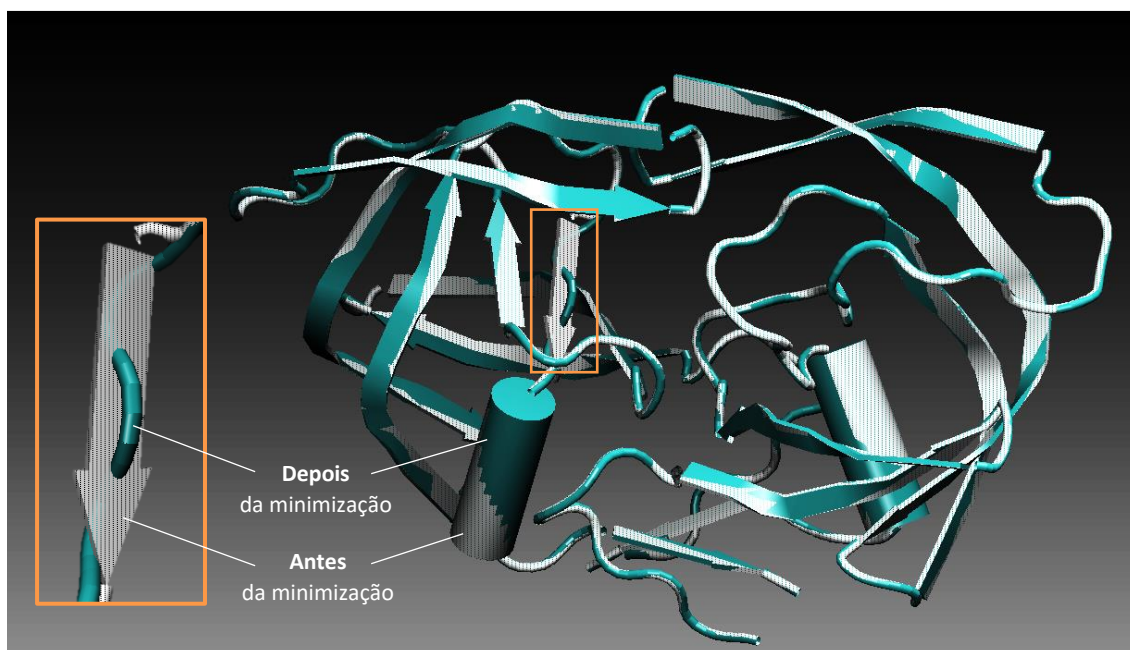


**Fig.4** – Representação gráfica da caixa de simulação de solvente explícito (moléculas de água) para o modelo proteína-inibidor considerado, antes (a branco) e depois (a rosa) da primeira minimização energética. *Software* de visualização molecular: VMD.

na segunda (min2), os átomos de hidrogénio, aleatoriamente posicionados pelo *Xleap*, foram recolocados no espaço tridimensional e o sistema gerado ficou, significativamente, mais refinado. Para uma melhor exposição das alterações resultantes desta otimização, os átomos de hidrogénio foram tratados como partículas independentes, na fig.5a.



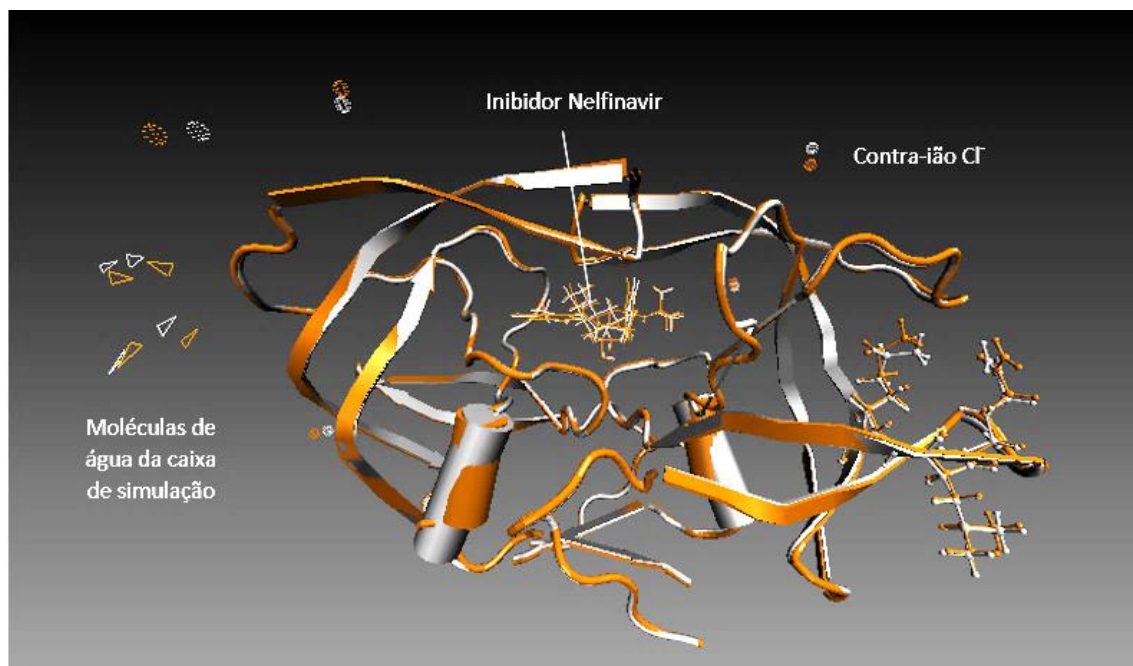
**Fig.5a** – Representação gráfica da distribuição dos átomos de hidrogénio, do modelo proteína-inibidor considerado, no espaço tridimensional, antes (a branco) e depois (a verde) da segunda minimização energética. *Backbone* da proteína em ambos os sistemas (sobreposição perfeita), representada a azul. *Software* de visualização molecular: VMD.



**Fig.5b** – Representação gráfica da influência da recolocação dos átomos de hidrogénio na estrutura secundária da proteína em estudo. Estrutura inicial, representada a branco; estrutura com hidrogénios otimizados, a azul ciano. *Software* de visualização molecular: VMD.

No entanto, na prática, influenciam diretamente a estabilidade termodinâmica das conformações locais. Como tal, na fig.5b, pode observar-se, precisamente, a desnaturação (perda de estrutura) de uma folha- $\beta$  para um *loop*;

na terceira (min3), partiu-se das coordenadas do modelo minimizado anteriormente e executou-se a otimização geométrica de todos os átomos. Obteve-se um sistema biomolecular muito mais próximo do nativo correspondente (fig.6).



**Fig.6** – Representação gráfica do modelo proteína-nelfinavir em estudo, ilustrando as mudanças de distribuição espacial resultantes da minimização global do sistema biomolecular (a laranja), face ao modelo inicial (a branco). Os cinco pares de contra-íons  $\text{Cl}^-$  foram representados na forma de *dotted spheres*. Software de visualização molecular: VMD.

## Dinâmica Molecular

Feitas as otimizações geométricas, a energia potencial do sistema encontra-se minimizada e este está, convenientemente, preparado para ser simulado. Nesta etapa da metodologia, a Dinâmica Molecular foi dividida em duas fases: equilíbrio (MD1) e produção (MD2), realizadas separadamente. Para a sua execução utilizou-se o módulo *Sander* e, como tal, ficheiros de comandos com as instruções de simulação foram previamente preparados. Seguem as especificações mais relevantes:

**Errata:** A fig.6 deveria ter sido ilustrada em conformidade com a representação anterior, i.e., exibindo a cadeia com as 2 folhas- $\beta$  (estrutura inicial) do lado esquerdo (do ponto de vista do leitor) da imagem.

### Equilibração (MD1)

Aquecimento gradual do sistema de  $T_0 = 0$  K até  $T_{eq.} = 310.15$  K;  $V =$  constante (definição das condições periódicas de fronteira); frequência de colisão =  $1 \text{ ps}^{-1}$ ; *cutoff* =  $10 \text{ Å}$ ; trajetórias calculadas de 100 em 100 passos de simulação, num total de 20000;  $\Delta t = 1 \text{ fs}$ ;  $t_{\text{simulação}} = 20 \text{ ps} \rightarrow$  *Ensemble* de 200 microestados.

A partir do ficheiro de coordenadas finais da fase de adaptação, criou-se o ficheiro de entrada para a segunda etapa da simulação.

### Produção (MD2)

$P =$  constante ( $1 \text{ atm}$ ) (relaxação da densidade da água para o estado de equilíbrio);  $T =$  constante ( $310.15 \text{ K}$ ); frequência de colisão =  $1 \text{ ps}^{-1}$ ; *cutoff* =  $10 \text{ Å}$ ; trajetórias calculadas de 100 em 100 passos de simulação, num total de 100000;  $\Delta t = 1 \text{ fs}$ ;  $t_{\text{simulação}} = 100 \text{ ps} \rightarrow$  *Ensemble* de 1000 microestados.

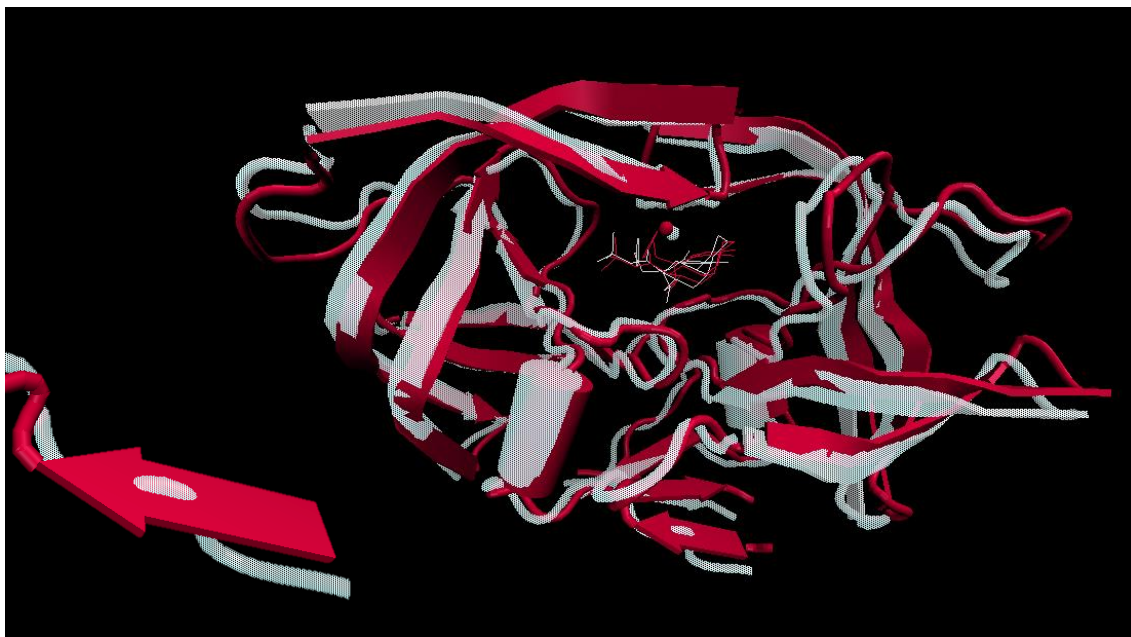
O aumento e a manutenção da temperatura (nas etapas de equilibração e produção, respetivamente) tiveram por base a implementação do algoritmo do “termóstato” de Langevin. Considerando que o sistema biomolecular se encontrava, explicitamente, “embebido” num solvente, é compreensível que a sua dinâmica seja afetada pelas colisões aleatórias entre as moléculas de água que o constituem e as forças de atrito daí resultantes. Notavelmente, as partículas de solvente deslocam-se com uma dada energia cinética, dependente da temperatura. Ora, a finalidade do “termóstato” de Langevin é, precisamente, mimetizar esse banho térmico e incorporar os seus efeitos na energia do sistema. Como? O algoritmo corresponde a uma variação da segunda lei de Newton, a qual descreve o movimento de uma partícula (de soluto) sob ação de uma força determinística e das tais forças – entre elas, a de fricção – resultantes da interação com o banho de calor (solvente). Assim, a cada integração das equações de movimento, o algoritmo de Langevin – dependente da temperatura vigente – é implementado e as coordenadas posicionais são atualizadas na direção da energia cinética pretendida, i.e., o sistema converge para a temperatura termodinâmica alvo (definida pelo utilizador) (11). Para que isto seja possível, é imprescindível a definição da variável “frequência de colisão” – a qual visa simular a forma como o banho absorve ou transmite energia ao sistema. De uma maior frequência de colisão, depreende-se uma maior velocidade e, portanto, um menor o tempo de deslocamento das partículas. Consequentemente, menores serão as oscilações de temperatura do sistema e vice-versa (3).



## Resultados e discussão

### Visualização dos resultados

As trajetórias (\*.mdcrd) relativas a cada uma das fases da dinâmica foram visualizadas no VMD. Na fig. 7, encontra-se representado o último microestado de MD1, em sobreposição com a configuração minimizada de coordenadas inicial (min3.rst).

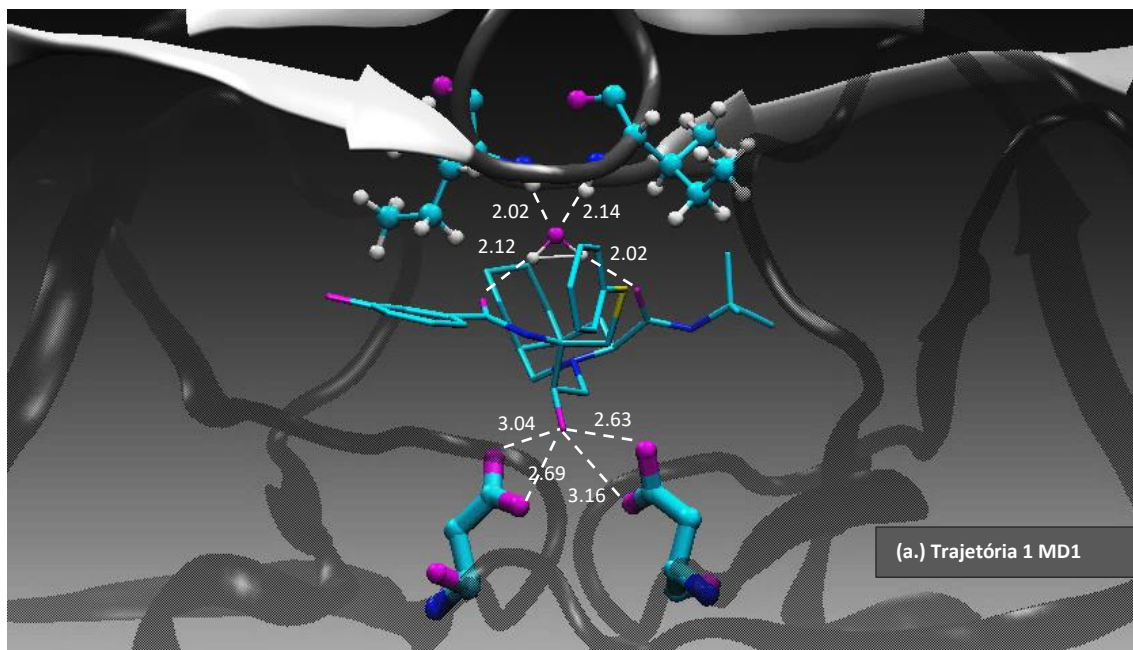


**Fig.7** – Representação gráfica do modelo proteína-inibidor em estudo, ilustrando as mudanças conformacionais entre o sistema biomolecular simulado em MD1 (última trajetória representada a rosa) e o modelo inicial, previamente minimizado (a branco), com ênfase numa nova folha- $\beta$ . Esfera de água estrutural, no centro ativo. *Software* de visualização molecular: VMD.

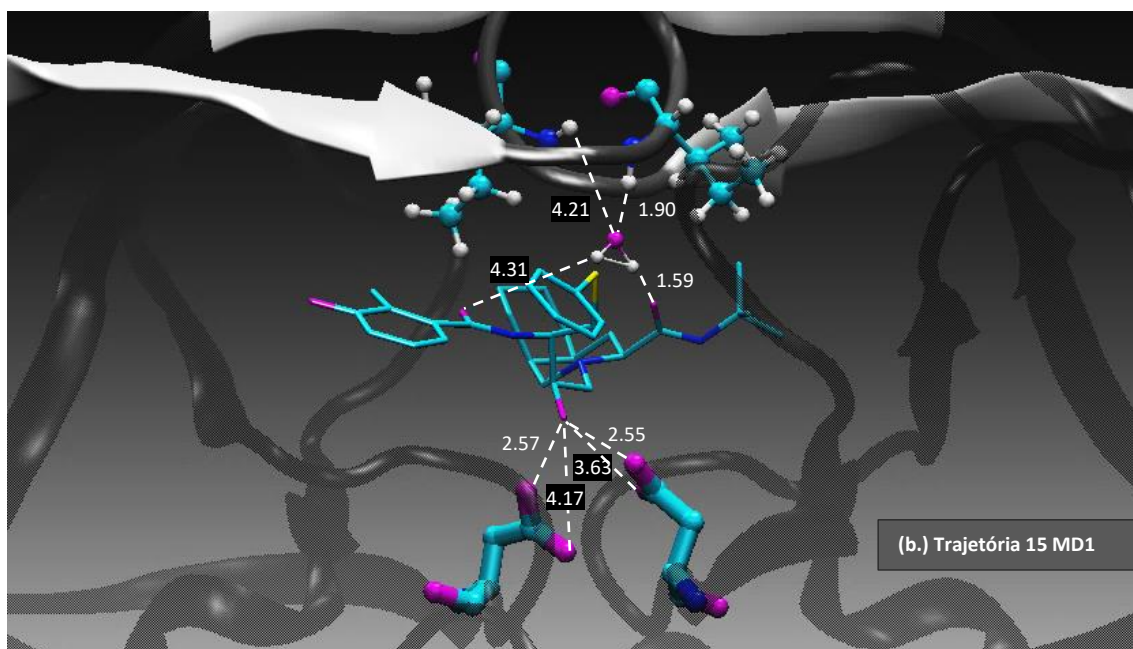
Com a MD1, o sistema sofreu alguns rearranjos conformacionais significativos, nomeadamente, a nível da formação de uma nova folha- $\beta$  (fig.7).

Ao longo de toda a simulação de Dinâmica Molecular (MD1 + MD2): verificou-se, visualmente, uma maior flexibilidade dos *loops* – qualquer cadeia “solta” entre duas extremidades de estrutura local e, portanto, com uma maior mobilidade inerente às poucas restrições; hélices- $\alpha$  e folhas- $\beta$  superiormente conservadas; o inibidor permaneceu no centro ativo da enzima, preservando a sua conformação (mesmo que com variações na geometria) e sem se deslocar, significativamente, da posição inicial; a molécula de água estrutural não manteve a posição relativa prevista na estrutura cristalográfica. Verificaram-se ainda alterações nas pontes de hidrogénio (fig.8). Estas correspondem a ligações não covalentes fortes e estabelecem-se entre pares de átomos eletronegativos (dador e aceitador de eletrões), como o Nitrogénio (a azul escuro) e o Oxigénio (a violeta).

Para que uma interação aconteça, os átomos pesados devem distar entre si 2.5 a 3.5 Å, o equivalente a 1.5 a 2.5 Å entre o próton e o seu aceitador.



**Fig.8** – Representação gráfica do modelo proteína-inibidor em estudo, ilustrando as principais mudanças, ao nível das pontes de hidrogénio estabelecidas, entre o (a.) primeiro e (b.) o décimo quinto microestados da equilibração (MD1). Representação Distâncias assinaladas medidas em Å. *Software* de visualização molecular: VMD.

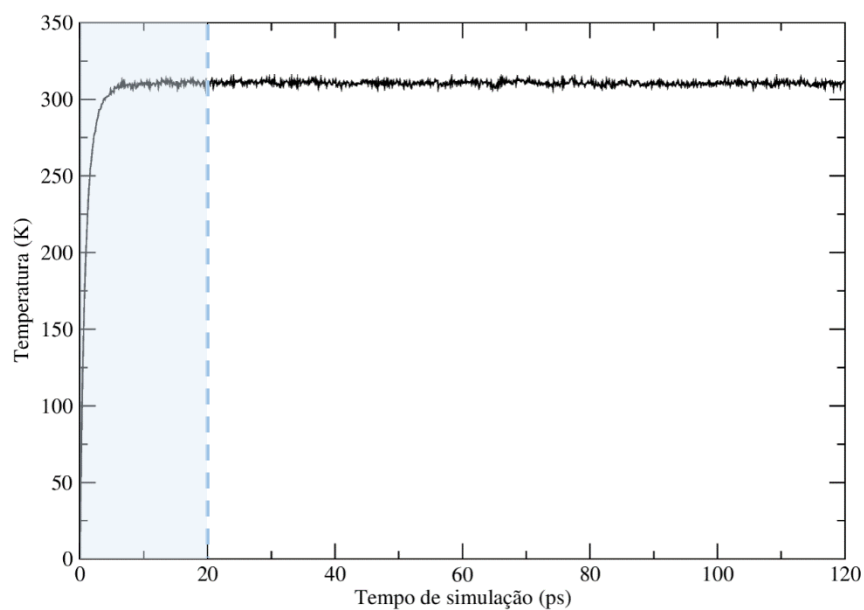


A título exemplificativo, encontra-se ilustrado na figura 8, os comprimentos das pontes de hidrogénio mais prováveis. Das representações, conclui-se que as distâncias não só variaram com a trajetória, como resultaram em quebras pontuais de ligação (distâncias assinaladas a preto, na figura exemplo).

## Propriedades gerais

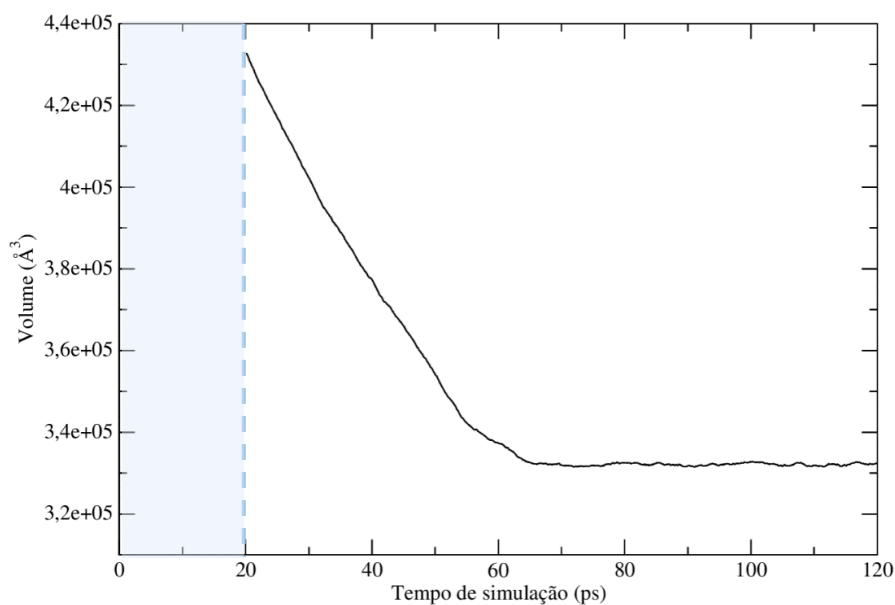
Para análise das propriedades gerais, recorreu-se a um *script*, previamente, fornecido (\*.perl) e compararam-se os ficheiros *output*, com recurso ao *xmgrace* – um *software* de visualização gráfica. Os gráficos obtidos encontram-se abaixo listados, nos quais a área a azul representa o período de adaptação.

### Temperatura



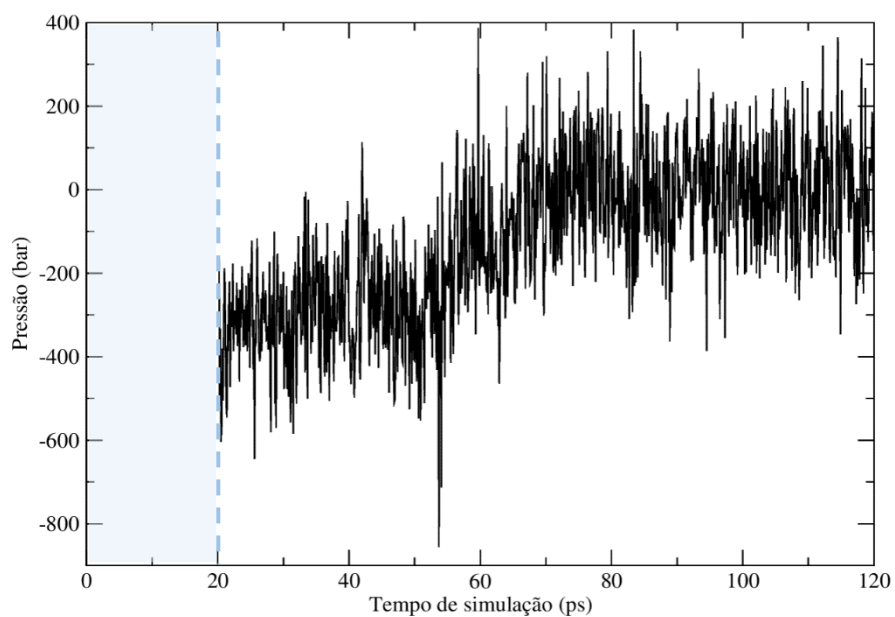
**Fig.9**– Representação gráfica da variação da Temperatura do sistema, em função do tempo de simulação.

### Volume



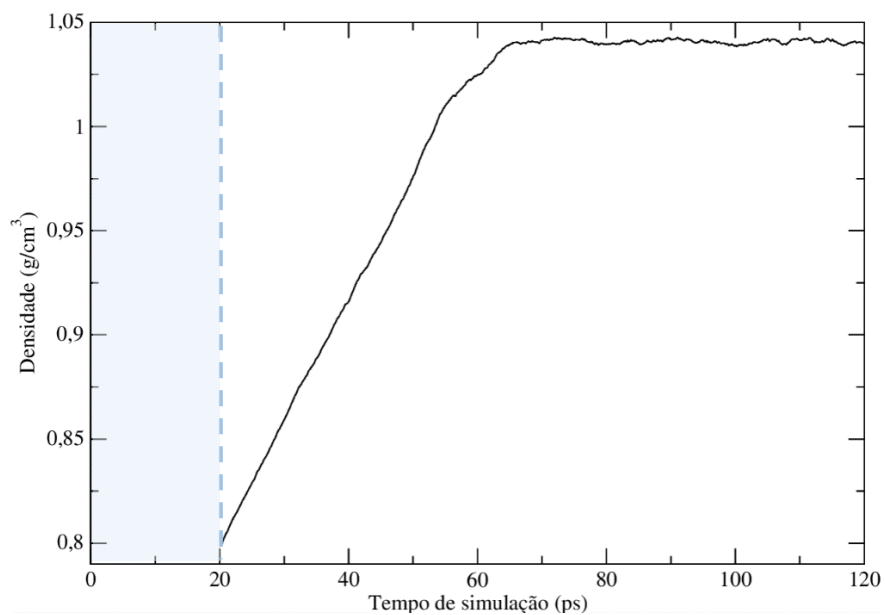
**Fig.10** – Representação gráfica da variação do Volume do sistema, em função do tempo de simulação.

## Pressão



**Fig.11** – Representação gráfica da variação da Pressão do sistema, em função do tempo de simulação.

## Densidade



**Fig.12** – Representação gráfica da variação da Densidade do sistema, em função do tempo de simulação.

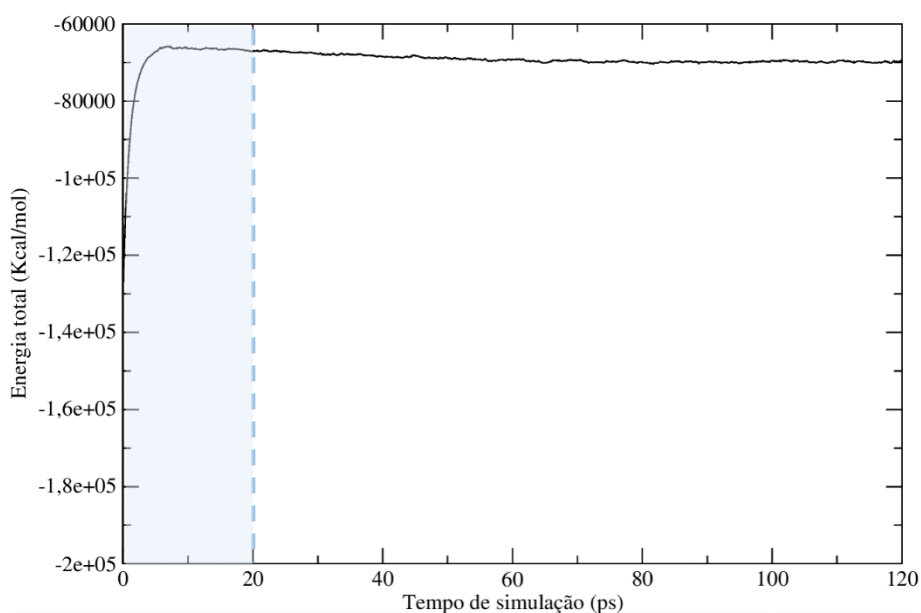


Das representações gráficas da evolução da temperatura (fig. 9) e RMSD (fig. 14, na secção seguinte) – variável dependente – com o tempo de simulação, pode concluir-se que, tal como previsto, convergem para um valor médio a partir de  $t_{\text{simulação}} \approx 20$  ps, apontando para a estabilização/equilíbrio do sistema biomolecular considerado.

É de relembrar que, no decorrer do período de inicialização, algumas propriedades termodinâmicas, como o volume e, consequentemente, pressão e densidade se mantêm fixas. No entanto, por defeito, o *software* atribui a essas variáveis valores unitários que não correspondem à realidade e que foram, por isso, removidos. Assim, nas representações gráficas relativas a essas características (figs. 10, 11 e 12), verifica-se uma ausência de dados até  $t = 20$  ps, correspondente ao final do período de adaptação.

Da observação da variação do volume com o tempo de simulação (fig.10), verifica-se que (na fase de estabilização) o sistema contraiu, ao longo do tempo, refletindo a aproximação das águas envolventes, para preenchimento das cavidades da proteína. Ora, diminuindo o volume de fluído, diminui o volume do sistema global e a densidade (fig.12) – inversamente proporcional – aumenta, até  $t \approx 60$  ps  $\gg 20$  ps, a partir do qual as propriedades atingem um limiar praticamente constante, corroborando a ideia de que quanto maior o tempo de simulação, maior a predisposição para o equilíbrio.

### Energia total

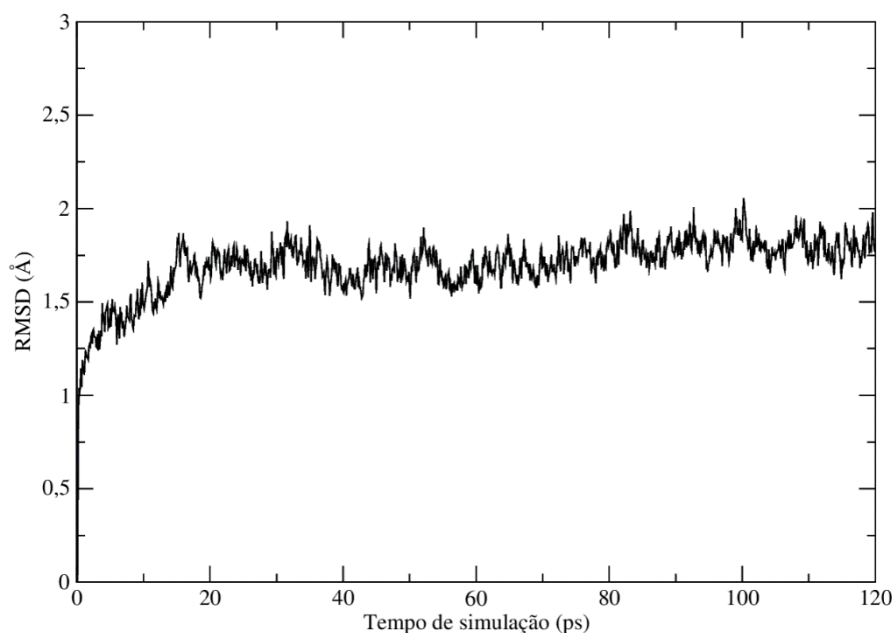


**Fig.13** – Representação gráfica da variação da Energia total do sistema, em função do tempo de simulação.

## RMSD

O cálculo dos diferentes valores de RMSD foi efetuado internamente, após sobreposição de cada um dos conformeros previstos com a configuração inicial, no VMD. As estruturas foram alinhadas pelos Carbonos- $\alpha$  da *backbone*, cujas coordenadas atômicas permitiram o cálculo dos desvios correspondentes. Os dados relativos às duas fases da Dinâmica Molecular foram então listados e exportados, para posterior análise. No final de MD2 – correspondente a  $t = 120$  ps de simulação (fig.3a) –, obteve-se um  $\text{RMSD}_{\text{MD2}} = 1.69026 \text{ \AA} < 2 \text{ \AA}$ , indicativo de uma elevada similaridade entre o sistema cristalográfico de referência e a última estrutura gerada por Dinâmica Molecular.

Notavelmente, apesar das trajetórias terem, garantidamente, associados valores de energia progressivamente mais próximos do equilíbrio (fig.13), nem sempre correspondem a valores de RMSD (fig.14) praticamente constantes. Hipóteses de explicação para o grau de alinhamento considerado nem sempre refletir a qualidade da dinâmica estão inerentes a defeitos na parametrização do campo de forças e/ou à presença de regiões móveis “soltas”, i.e., *loops/ coils*, cuja diferença nas estruturas em comparação pode ser suficiente para distorcer o valor do desvio (12).



**Fig.14** – Representação gráfica da variação do RMSD do sistema, em função do tempo de simulação.

## Pontes de hidrogénio

Tal como para o RMSD, para a análise das pontes de Hidrogénio, recorreu-se ao módulo *ptraj*. No entanto, os resultados apresentados são agora relativos a uma terceira simulação de MD (MD3), mais extensa ( $t_{\text{simulação}} = 12 \text{ ns}$ ), a qual foi executada pelo docente. Começou então por criar-se um ficheiro de instruções para o aplicativo, com a indicação do arquivo de trajetórias em análise (relativo, portanto, à fase de produção de MD3) e dos átomos dadores e aceitadores, nos resíduos existentes. Estes foram, notavelmente, definidos de acordo com as designações compatíveis com o *software*, i.e., contrariamente ao convencionado pela IUPAC. Consideravelmente, tendo em conta a geometria dos aminoácidos de prolina, instruiu-se o programa a desprezar os hidrogénios do grupo básico do resíduo. Para remover os movimentos de difusão, as estruturas foram centradas no centro de geometria formada por todos os Carbonos- $\alpha$  e executou-se a análise, considerando apenas as pontes de hidrogénio prevalentes em mais de 5% do tempo de simulação. Na figura 15, encontra-se representada a informação relativa a algumas das pontes de hidrogénio mais importantes, i.e., com maior tempo de existência.

HBOND SUMMARY:

data was sorted, intra-residue interactions are NOT included,  
Distance cutoff is 3.50 angstroms, angle cutoff is 140.00 degrees  
Hydrogen bond information dumped for occupancies > 0.05

Aceitador de H DONOR		H doado p/ ligação ACCEPTORH		Dador de H ACCEPTOR		%occupied	distance	angle
atom#	:res@atom	atom#	:res@atom	atom#	:res@atom			
729	:45@O	875	:56@H	874	:56@N	94.38	2.813 ( 0.08)	18.46 ( 9.90)
2293	:144@O	2439	:155@H	2438	:155@N	94.13	2.807 ( 0.09)	20.54 (11.48)
1935	:121@O	1750	:110@H	1749	:110@N	92.15	2.828 ( 0.08)	15.64 ( 8.00)
2477	:156@O	2779	:176@H	2778	:176@N	91.82	2.813 ( 0.08)	21.35 ( 9.99)
417	:25@OD1	444	:28@H	443	:28@N	91.65	2.803 ( 0.08)	24.70 (12.97)
370	:22@O	185	:11@H	184	:11@N	90.83	2.828 ( 0.08)	16.16 ( 8.16)
765	:47@O	834	:54@H	833	:54@N	89.83	2.833 ( 0.08)	16.93 ( 7.76)

$\Theta = 180^\circ$  - angle

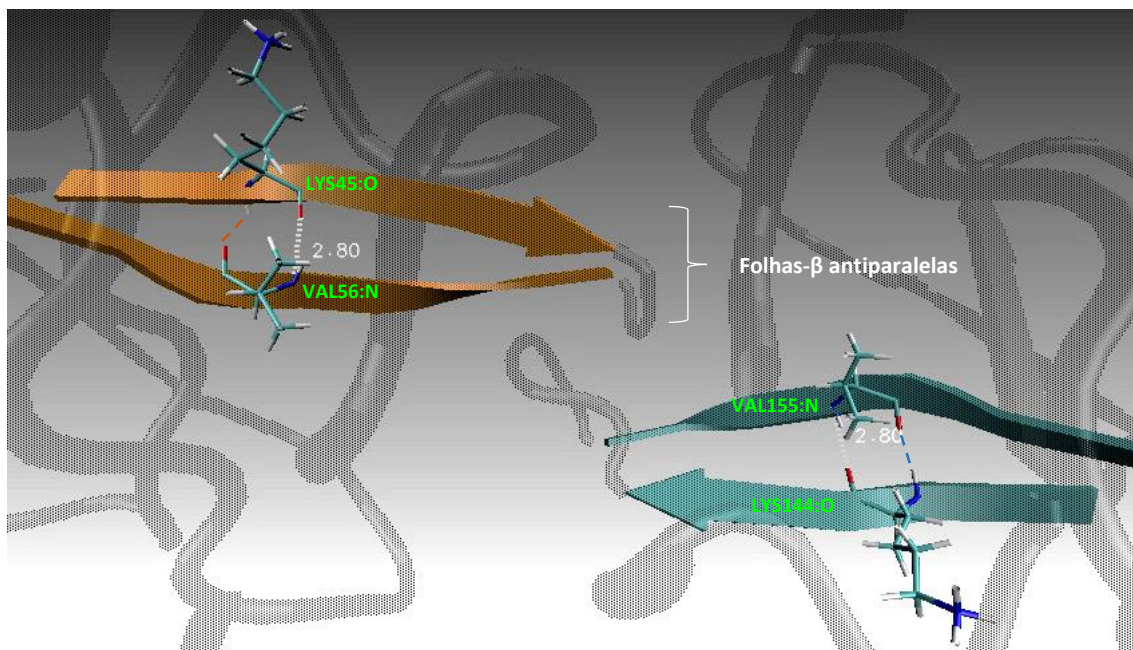
**Fig.15** – Excerto do ficheiro de resultados da análise das pontes de hidrogénio, ilustrando os átomos e resíduos intervenientes nas várias ligações (colunas DONOR\*, ACCEPTORH, ACCEPTOR\*\*), a percentagem de tempo de simulação em que se mantêm formadas (%occupied), bem como o comprimento (distance) medido entre átomos pesados e ângulo médio (angle\*\*\*). Desvios respetivos entre parêntesis. \*DONOR = Aceitador de H; \*\*ACCEPTOR = Dador de H; \*\*\*angle  $\neq \Theta$  entre o H partilhado e o eixo da ligação estabelecida.  $\Theta = 180^\circ$  - angle.

Da análise do ficheiro de resultados, conclui-se que, para 5% do tempo de simulação, aparecem referenciadas 185 pontes de hidrogénio. Parte delas corresponde, certamente, a ligações bifurcadas, visto que o número de átomos aceitadores de H difere e é maior que o número de átomos dadores. As percentagens de prevalência das 5 pontes mais estáveis foram 94.38%, 94.13%, 92.15%, 91.82% e 91.65% e os ângulos de ligação  $\Theta$  correspondentes  $161.54^\circ$ ,  $159.46^\circ$ ,  $164.36^\circ$ ,  $158.65^\circ$  e  $155.3^\circ$ , respetivamente. Deste

modo, depreende-se que uma ligação de H é tanto mais estável, quanto maior (mais próximo de  $180^\circ$ ) o ângulo  $\theta$  (que o seu eixo faz com o H partilhado). Por outro lado, conclui-se que o comprimento destas ligações não covalentes – equivalente a  $2.813\text{\AA}$ ,  $2.807\text{\AA}$ ,  $2.828\text{\AA}$ ,  $2.813\text{\AA}$  e  $2.803\text{\AA}$ , para as pontes consideradas – praticamente, não varia com a percentagem de “ocupação” e, portanto, não influencia a estabilidade da interação.

Tendo em atenção os átomos envolvidos em cada uma destas ligações, conclui-se que constituem pontes de hidrogénio estruturantes, i.e., responsáveis pela estabilidade termodinâmica das estruturas secundárias. Dado que a sua geometria é baseada em valores de equilíbrio, apresentam uma menor flexibilidade. Consequentemente, as pontes de hidrogénio associadas são superiormente conservadas no tempo. Na figura 16 está representado, a título exemplificativo, a localização das duas ligações mais prevalentes, as quais fazem parte de pares de folhas- $\beta$  antiparalelas.

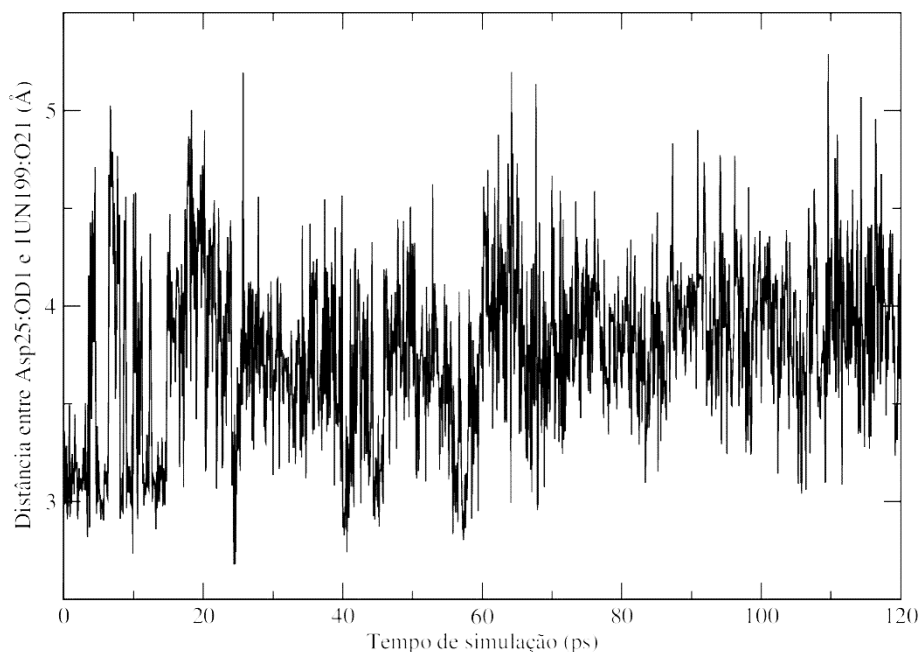
Por fim, como confirmado por representações anteriores, a protease do HIV-1 é constituída por duas subunidades praticamente simétricas, constituindo, por isso, um dímero proteico. Ora, sabendo que o resíduo 1 na cadeia A é equivalente ao resíduo 100 na cadeia B, é possível analisar as pontes de H entre os mesmos resíduos, em subunidades diferentes. Estas têm, previsivelmente, estabilidades similares, as quais estão assinaladas para duas ligações exemplo, na figura 15 (áreas a laranja – subunidade A – e a azul - B).



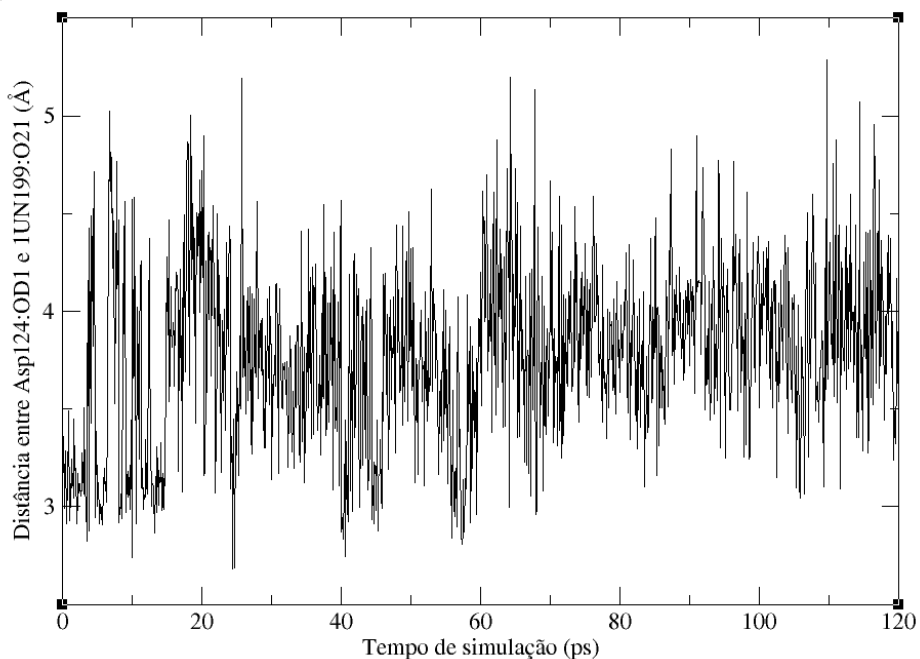
**Fig.16** – Representação gráfica do último microestado de MD3, enfatizando a localização conformacional (folhas- $\beta$ ) das pontes de Hidrogénio mais prevalentes, i.e., entre o Oxigénio do resíduo 45 e o Nitrogénio do resíduo 56, na cadeia A e a ligação correspondente na subunidade B, i.e., entre os átomos eletronegativos dos resíduos 144 e 155.

## Distâncias interatômicas e comprimentos de ligação

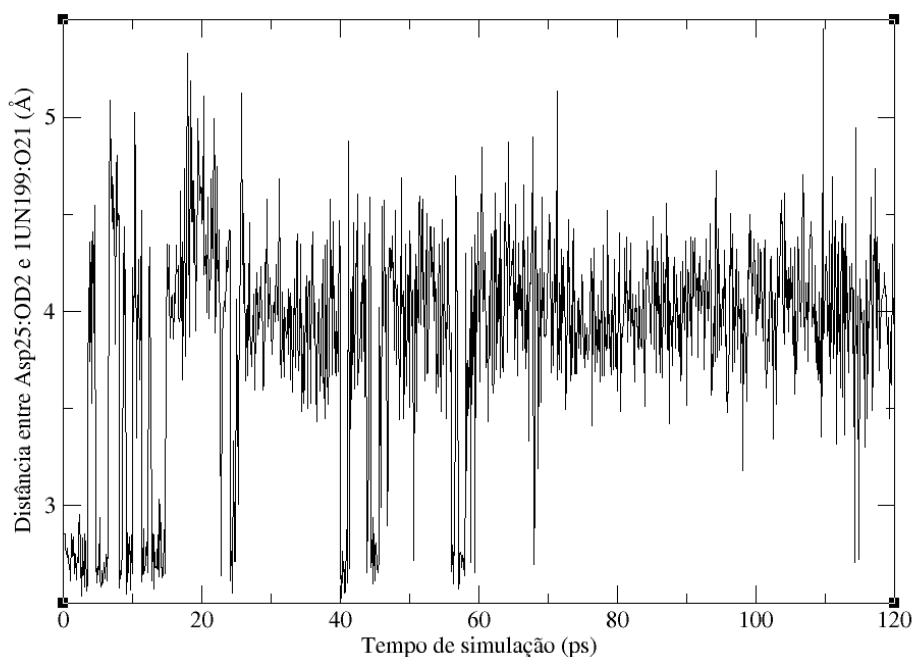
Posteriormente, recorreu-se ao *software ptraj* para determinação das distâncias entre o átomo de oxigénio do grupo carboxilo dos resíduos de aspartato catalíticos e o átomo de oxigénio do grupo hidroxilo do inibidor (fig.2). Para tal, criou-se um ficheiro de instruções, com recurso a um editor de texto, no qual se especificou os átomos intervenientes nas interações. Utilizaram-se as trajetórias das simulações MD1 e MD2.



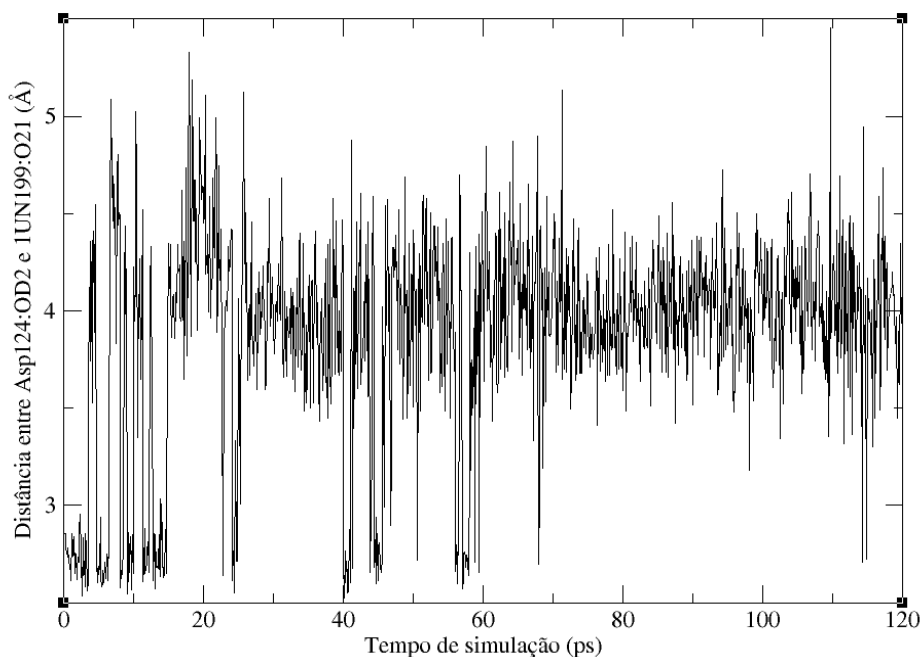
**Fig.17** – Representação gráfica da variação da distância entre Asp25:OD1 e IUN199:O21, em função do tempo de simulação.



**Fig.18** – Representação gráfica da variação da distância entre Asp124:OD1 e IUN199:O21, em função do tempo de simulação.



**Fig.19** – Representação gráfica da variação da distância entre Asp25:OD2 e 1UN199:O21, em função do tempo de simulação.



**Fig.20** – Representação gráfica da variação da distância entre Asp124:OD2 e 1UN199:O21, em função do tempo de simulação.

Das representações gráficas, conclui-se que, tal como supracitado, as interações não se mantêm constantes ao longo da simulação. A título corroborativo, foram calculadas as distâncias das mesmas interações na estrutura PDB original (fig.2), bem como os valores médios e desvios padrão, considerando todos os microestados (tabela 1).

<u>Interação</u>	<u>Distância aprox. (Å)</u> <u>Estrutura PDB inicial</u>	<u>Distância média (Å)</u> <u>Trajетórias MD1+MD2</u>
Asp25:OD1 – 1UN199:O21	2.63	<b>3.77093 ± 0.456811</b>
Asp124:OD1 – 1UN199:O21	2.94	3.87514 ± 0.543337
Asp25:OD2 – 1UN199:O21	3.11	<b>3.77093 ± 0.456811</b>
Asp124:OD2 – 1UN199:O21	2.77	3.87514 ± 0.543337

**Tabela 1** – Distâncias das interações do centro catalítico da enzima Protease do HIV-1, na estrutura PDB original e respetivos valores médios e desvios padrão, para as trajetórias obtidas ao longo de 120 ps de simulação de MD.

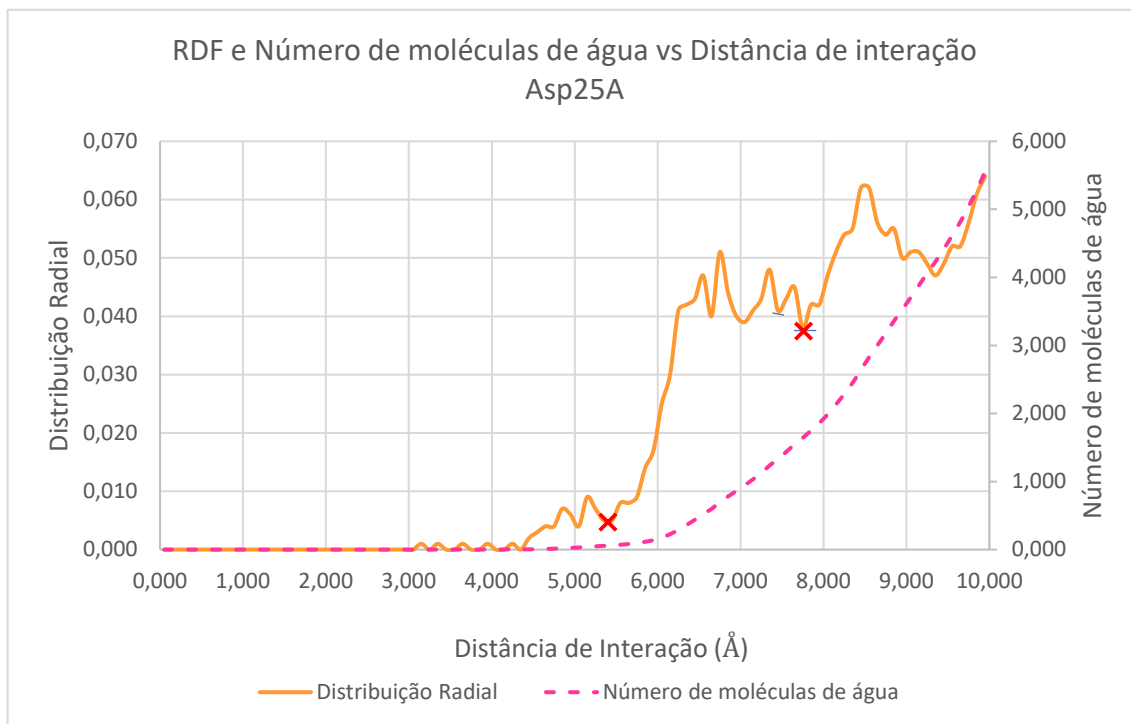
Como previsto, as mesmas interações em cadeias diferentes têm aproximadamente o mesmo comprimento (tabela 1), inerente à simetria da proteína. Sabe-se que se estabelecem por pontes de hidrogénio e, como tal, as distâncias relativas não devem ultrapassar os 3.5 Å – limiar de comprimento, entre átomos pesados, para a formação da ligação. Contudo, pela análise gráfica e agora dos valores médios de interação, isto não se verifica continuamente, ao longo da simulação. Desta forma, corrobora-se o facto de que o Nelfinavir, para além de competitivo, constitui um inibidor reversível, i.e., apesar de se poder encontrar constantemente no centro catalítico, o rompimento pontual das ligações não covalentes, impede-o de desempenhar a sua função: bloquear a reação de catálise.

## RDFs

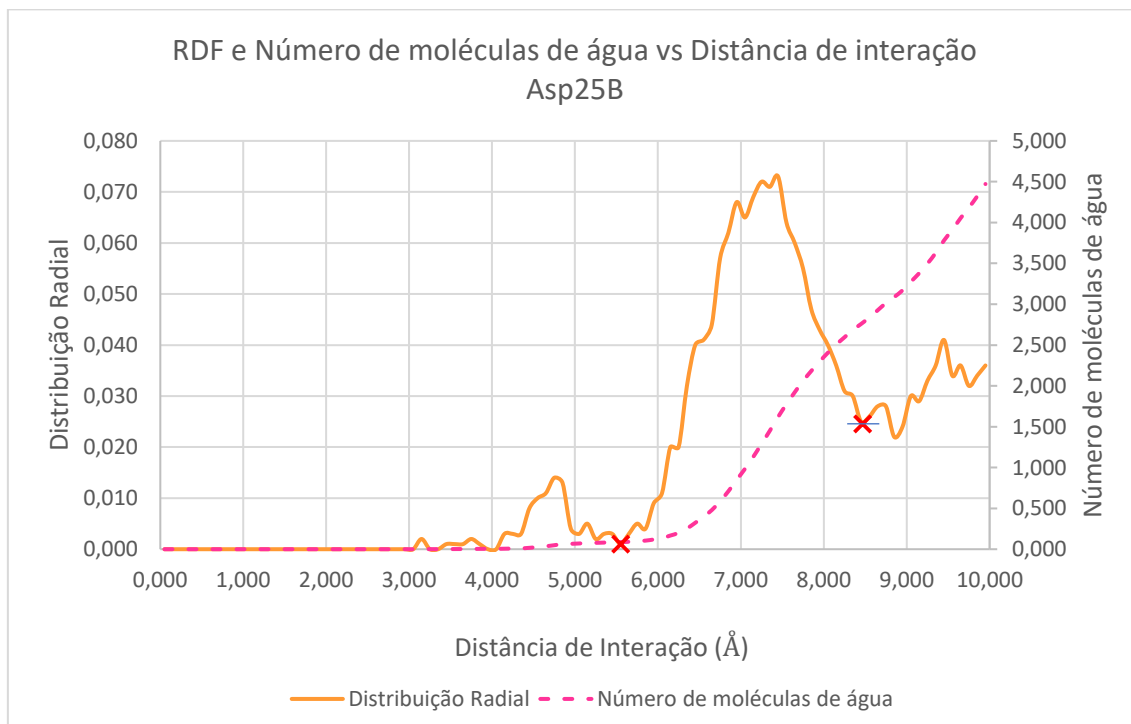
Finalmente, para estudar a variação de interações específicas solvente-proteína e a distribuição de moléculas de água nas esferas de hidratação, calcularam-se as RDFs para as interações Asp25-COO<sup>-</sup> – H<sub>2</sub>O, Asp29-COO<sup>-</sup> – H<sub>2</sub>O e Glu21-COO<sup>-</sup> – H<sub>2</sub>O da cadeia A e os pares correspondentes na cadeia B. Utilizou-se a simulação MD3. A informação foi representada graficamente, desta vez, com recurso ao *Excel*.

Admitindo que a cada pico predominantemente identificável na função de distribuição radial corresponde uma maior probabilidade de partículas na vizinhança (neste caso, moléculas de água), é possível prever o número de esferas de solvatação, apenas por visualização gráfica. Para além disso, considerando que a primeira esfera de hidratação engloba as moléculas de água contidas até ao primeiro mínimo da função; a segunda, o número de águas entre o primeiro e o segundo mínimos; e a terceira, entre o segundo e o

terceiro, determinou-se também o número de partículas em cada esfera. As conclusões mais relevantes foram registradas na tabela 2, com base na informação das figs. 21 a 26.

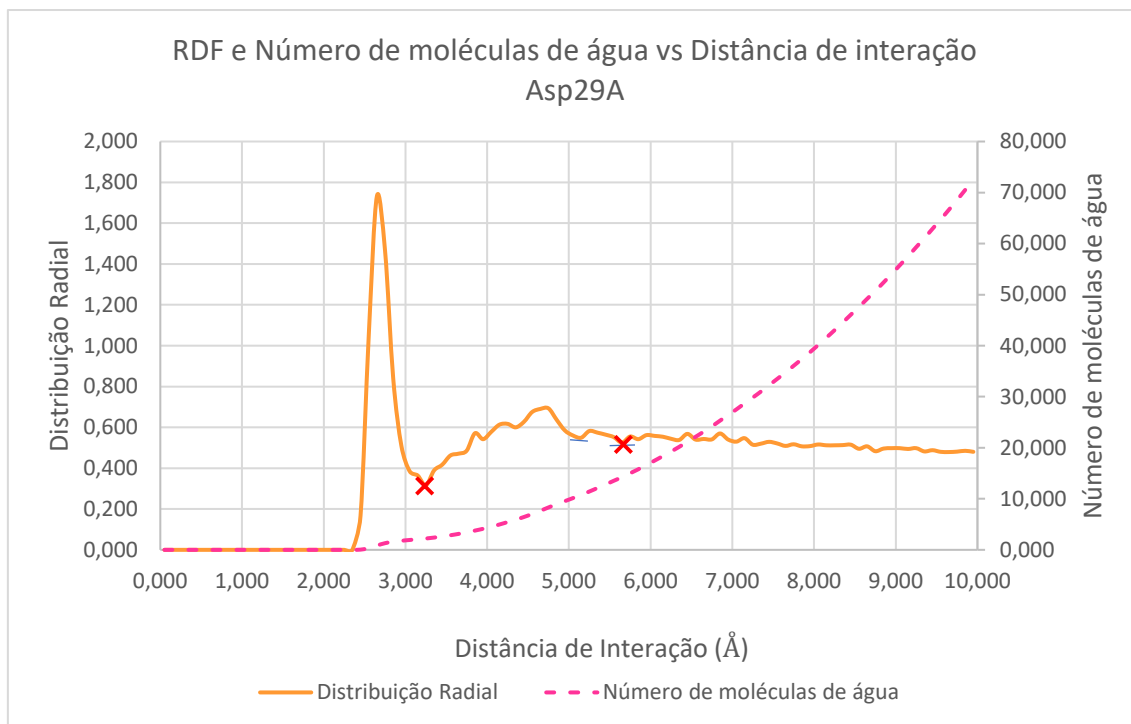


**Fig.21** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Asp25A.

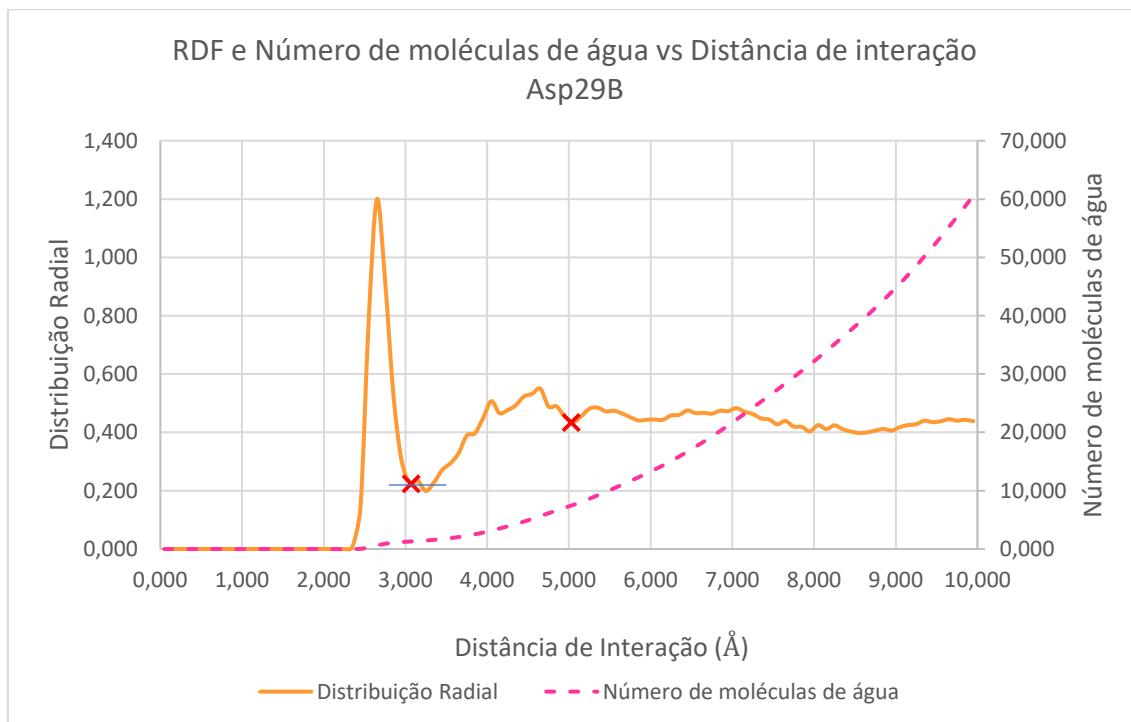


**Fig.22** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Asp25B.



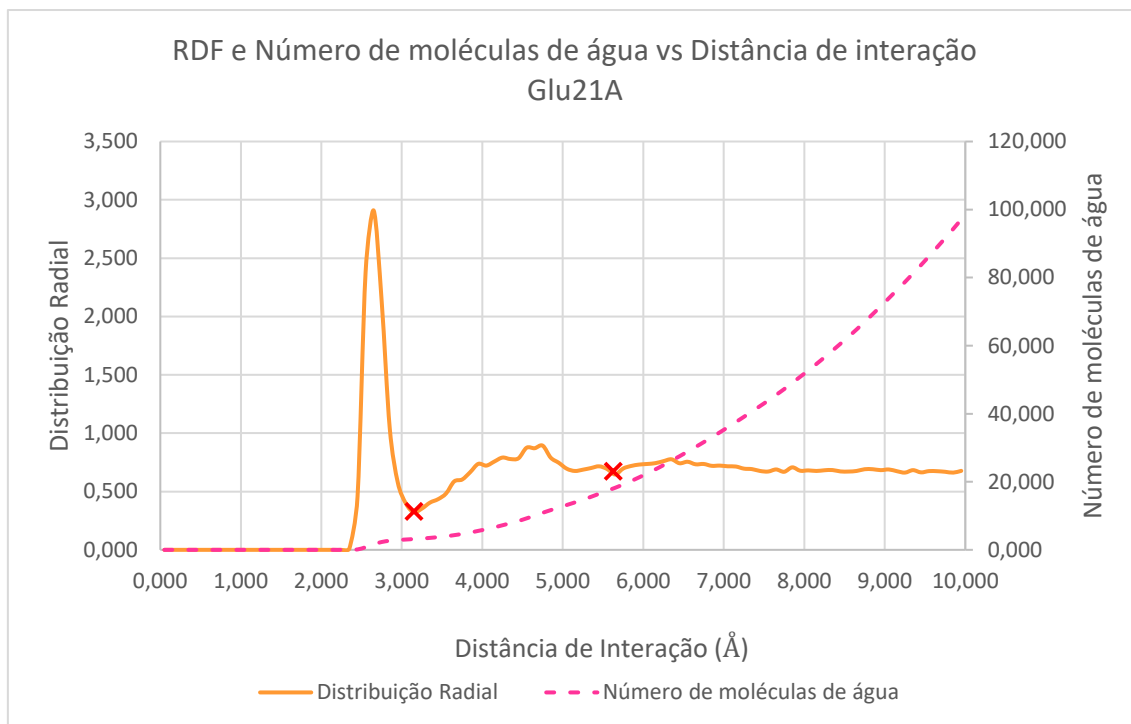


**Fig.23** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Asp29A.

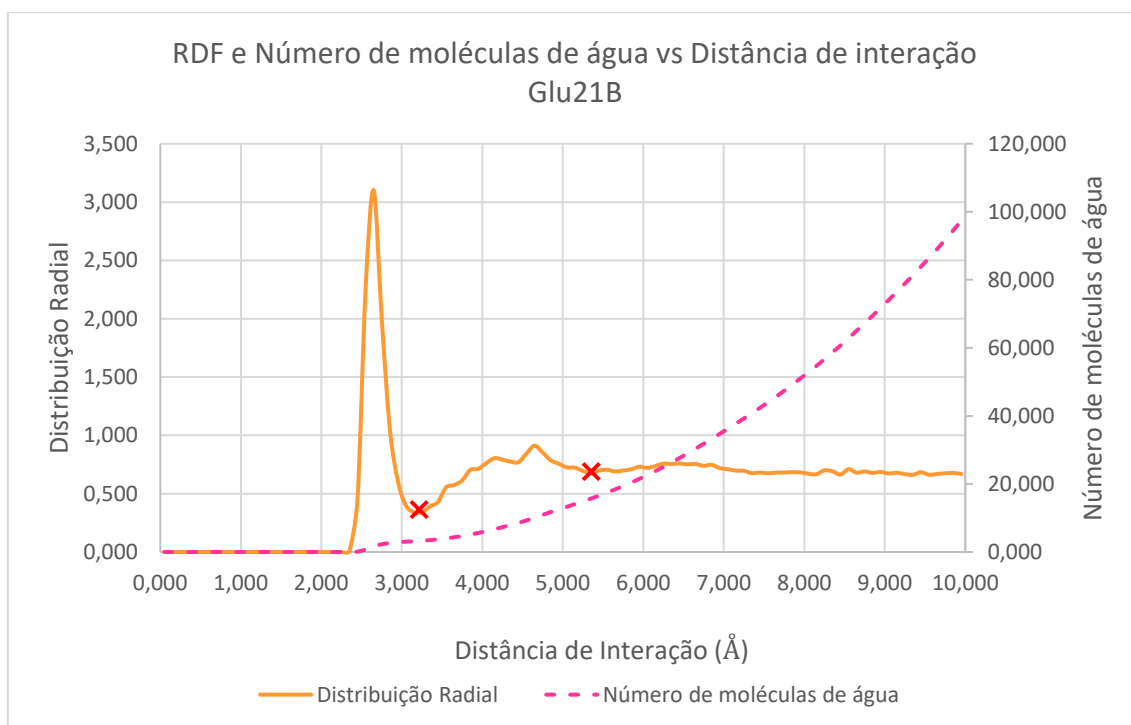


**Fig.24** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Asp29B.

**Nota:** Teórica e notavelmente, o final de uma coroa de solvatação nem sempre coincide com o início da seguinte. No entanto, por limitações de tempo, não foi possível calcular analiticamente os mínimos de cada RDF (derivada no ponto) e universalizou-se, para todos os casos presentes, o parecer inicial.



**Fig.25** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Glu21A.



**Fig.26** – Representação gráfica da função de distribuição radial e do número de moléculas de água, em função da distância de interação com o Oxigênio do Glu21B.

	Esfera de hidratação	1ª	2ª	C/ 1 molécula de H <sub>2</sub> O	C/ 2 moléculas de H <sub>2</sub> O
	Raio aproximado (Å)	r <sub>final</sub>	r <sub>inicial</sub> e r <sub>final</sub>	r <sub>médio</sub>	r <sub>médio</sub>
Interação	Asp25A-COO <sup>-</sup> – H <sub>2</sub> O	5.450	5.450 – 7.750	r <sub>médio</sub> 2ª esfera ≈ 6.6	r <sub>médio</sub> 3ª esfera ≈ 8.5
	Asp25B-COO <sup>-</sup> – H <sub>2</sub> O	5.550	5.550 – 8.450	r <sub>médio</sub> 2ª esfera ≈ 7.0	r <sub>médio</sub> 2ª esfera ≈ 7.0
	Asp29A-COO <sup>-</sup> – H <sub>2</sub> O	3.250	3.250 – 5.750	r <sub>médio</sub> 1ª esfera	r <sub>médio</sub> 1ª esfera
	Asp29B-COO <sup>-</sup> – H <sub>2</sub> O	3.050	3.050 – 5.550	r <sub>médio</sub> 1ª esfera	r <sub>médio</sub> 2ª esfera ≈ 4.3
	Glu21A-COO <sup>-</sup> – H <sub>2</sub> O	3.150	3.150 – 5.750	r <sub>médio</sub> 1ª esfera	r <sub>médio</sub> 1ª esfera
	Glu21B-COO <sup>-</sup> – H <sub>2</sub> O	3.250	3.250 – 5.350	r <sub>médio</sub> 1ª esfera	r <sub>médio</sub> 1ª esfera

**Tabela 2** – Raios de interação para diferentes esferas de hidratação. Considerar r<sub>médio</sub> a “mediana” aproximada (média entre o r<sub>inicial</sub> e o r<sub>final</sub>) de uma coroa de solvatação.

Notavelmente, o OD2 possuiu cerca de 3 coroas de hidratação, nos resíduos Asp25, e 2, nos Asp29. O OE do Glu21 está rodeado por 2 esferas de solvatação.

Visualizando a disposição da molécula de água estrutural, em todas as trajetórias da simulação, constata-se que se encontra, alternadamente, a distâncias na ordem dos 6, 7 e (pontualmente) 8 Å, dos OD2 dos Asp25. Ora, uma vez que estas estão contidas no intervalo de r<sub>inicial</sub> – r<sub>final</sub> relativo à segunda coroa de solvatação para os átomos OD2 (tabela 2) dos Asp25 e que, de acordo com as figs.1 e 2, esta contem 1 e 1 só molécula, conclui-se que a molécula de água estrutural faz parte dessa esfera de hidratação.

Por outro lado, o raio limiar da 1ª esfera de solvatação para os átomos OD2 dos resíduos catalíticos Asp25A e Asp25B equivale, aproximadamente, a 5.450 e 5.550 Å. No entanto, por confrontação gráfica, sabe-se que não engloba nenhuma molécula de solvente, na sua integridade. Isto reflete a ausência da molécula de água catalítica, a qual é compatível com a falta de substrato no sistema. *Normalmente*, encontra-se a uma distância de interação adequada, relativamente aos Asp25 do complexo.

À distância de 9 Å dos átomos de referência encontram-se, em média, 3.615, 3.223, 54.91, 44.78, 72.76 e 72.95 moléculas de água (figs.21 a 26), para cada um dos resíduos Asp25A, Asp25B, Asp29A, Asp29B, Glu21A e Glu21B, respetivamente. A similaridade entre valores de resíduos correspondentes indicia que a disposição espacial de moléculas de solvente acompanha a simetria da proteína, i.e., as moléculas de água num raio de x Å de um resíduo da cadeia A são (praticamente) as mesmas – em número e em geometria – que estão presentes à distância x do mesmo resíduo, na cadeia B. Isto reflete também o sucesso das etapas de minimização e de introdução do sistema no solvente explícito.

## Conclusão

No âmbito deste trabalho, aplicaram-se técnicas de MM e DM à exploração do sistema Protease do HIV-1–Nelfinavir. Dos resultados das simulações, concluiu-se que: propriedades como a Temperatura, Volume, Pressão, Densidade e RMSD convergem, com o tempo de simulação, para um valor médio, indicativo do equilíbrio termodinâmico; o valor de RMSD obtido para a última conformação reflete positivamente a qualidade do alinhamento entre a estrutura gerada por dinâmica e a nativa; as pontes de hidrogénio mais estáveis compreendem estruturas secundárias locais, nomeadamente, folhas- $\beta$ , cuja geometria é baseada em valores de equilíbrio; o comprimento de todas as interações não covalentes é variável, ao longo das trajetórias, evidenciando a dinâmica do sistema; a molécula de água estrutural conservada faz parte da primeira esfera de solvatação, medida entre o OD2 de cada Asp25; os resultados obtidos para os resíduos da cadeia A são, praticamente, equivalentes aos da cadeia B, evidenciando o homodímero em estudo; pelo mesmo motivo, a disposição das moléculas de solvente é, aproximadamente, simétrica.

*“A Dinâmica Molecular não é, primariamente, sobre fazer vídeos de moléculas. Mais habitualmente, é sobre desenvolver previsões quantitativas do tamanho e forma moleculares, flexibilidades, interações com outras moléculas, comportamento sob pressão e a frequência relativa de um estado comparado com outro.”*

*Bernard Leimkuhler*

O *design* racional de enzimas, péptidos e fármacos inibidores para síntese exige um estudo de previsão estrutural precisa e constitui, por isso, um dos desafios para a bioinformática. Apesar disso, sabe-se que a otimização de técnicas de modelação molecular evolui progressivamente – não fosse este trabalho realizado e o Nelfinavir, atualmente, comercializado.

## Bibliografia

1. Weiss J, Haefeli WE. Impact of atp-binding cassette transporters on human immunodeficiency virus therapy. *Int Rev Cell Mol Biol*. 2010;280(C):219–79.
2. HUBBARD RE. Molecular Graphics and Modeling: Tools of the Trade. In: *Guidebook on Molecular Modeling in Drug Design*. Elsevier; 1996. p. 19–54.
3. Cerqueira N, Sousa S, Fernandes P, Ramos M. *Bioquímica Computacional - sebenta essencial*. 2015.
4. Abraham M, Hess B, Spoel D van der, Lindahl E. *GROMACS User Manual- 5.0.7*. WwwwGromacsOrg. 2015;
5. Amber tutorial [Internet]. [cited 2020 Dec 27]. Available from: <https://ambermd.org/tutorials/advanced/tutorial8/loop3.htm>
6. Hollingsworth SA, Dror RO. *Molecular Dynamics Simulation for All*. Vol. 99, Neuron. Cell Press; 2018.
7. Namba AM, da Silva VB, da Silva CHTP. *Dinâmica molecular: teoria e aplicações em planejamento de fármacos* [Internet]. São Paulo, Brasil; 2008. Available from: <http://www.scielo.br/pdf/eq/v33n4/v33n4a02.pdf>
8. Attig N, Binder K, Grubmuller, Helmut; Kremer K. *Computational Soft Matter: From Synthetic Polymers to Proteins*. In: *Computational Soft Matter: From Synthetic Polymers to Proteins* [Internet]. Vol.23. Julich; 2004. p. 1–28. Available from: <https://udel.edu/~arthij/MD.pdf>
9. Maier VN, Crippen GM. Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins. *J Mol Biol* [Internet]. 1994 Jan [cited 2019 Jun 10];235(2):625–34. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022283684710175>
10. Parker T. *Computacional Chemistry 3.4 - Conjugate Gradient* [Internet]. 2017. Available from: <https://youtu.be/AyFXfAhwVBQ>
11. Farago O. Langevin thermostat for robust configurational and kinetic sampling. *Phys A Stat Mech its Appl*. 2019 Nov 15;534.
12. Damm KL, Carlson HA. Gaussian-Weighted RMSD Superposition of Proteins: A Structural Comparison for Flexible Proteins and Predicted Protein Structures. *Biophys J* [Internet]. 2006 Jun [cited 2019 Jun 10];90(12):4558–73. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0006349506726315>