

**Fig.1** – Representação gráfica de um modelo de sPLA2 gerado por homologia. Estrutura da proteína representada em NewCartoon (a ciano), com destaque para os resíduos do centro ativo His47 (Carbonos a verde) e Asp48 (Carbonos a cinza) – a coordenar o íão  $\text{Ca}^{2+}$  (a laranja). Átomos de Hidrogénio e Nitrogénio a vermelho e ciano, respetivamente. Interações eletroestáticas com o cofator assinaladas. Molécula de água catalítica a branco. Proteína molde utilizada: 5tfv.1.A. *Software* de visualização molecular: VMD.

## Modelação por Homologia, Ancoragem Molecular e Triagem Virtual, no estudo de um sistema sPLA2 & Aplicação de métodos QM/MM, na simulação do mecanismo reacional da enzima DszB

Maria \*\*\*\*\*

Nº Mec.: up\*\*\*\*\*

M.: Bioinformática e Biologia Computacional

**Maria \*\*\*\*\***

Nº Mec.: up\*\*\*\*\*

**Modelação por Homologia, Ancoragem Molecular e Triagem Virtual, no estudo de um sistema sPLA2; aplicação de métodos QM/MM, na simulação do mecanismo reacional da DszB**

### **Relatório**

Trabalho realizado sob a orientação científica dos Professores Maria João Ramos, Ana Luísa Oliveira e Pedro Alexandrino Fernandes, no âmbito da unidade curricular de Bioinformática Molecular.

### **Palavras-chave**

Modelação por homologia; *molecular docking*; *virtual screening*; mecânica molecular; mecânica quântica; QM/MM; *high layer*; *low layer*; enzima; proteína; DszB; fosfolipase A2 secretada (PLA2); conformação; geometria; complexo; associação; ligando; inibidor; fármaco; *hit*; recetor; *scoring function*; minimização; energia.

### **Agradecimentos**

(Secção omitida por motivos de privacidade)

## Abreviaturas

Å	Angström
AMBER	<i>Assisted Model Building with Energy Refinement</i>
Asp	Aspartato
Arg	Arginina
BLOSUM62	<i>BLOcks SUBstitution Matrix</i>
C	Carbono
Ca <sup>2+</sup>	Cálcio
Cl	Cloro
Cys	Cisteína
DNA	<i>DeoxyRibonucleic Acid</i>
DszB	<i>2'-Hydroxybiphenyl-2-Sulfinate Desulfinate</i>
Gly	Glicina
H	Hidrogénio
His	Histidina
HSO <sub>3</sub> <sup>-</sup>	Hidrogenossulfito
IC <sub>50</sub>	Metade da concentração inibitória máxima
K <sub>i</sub>	Constante(s) de inibição/dissociação
QM	<i>Quantum Mechanics</i>
MM	<i>Molecular Mechanics</i>
O	Oxigénio
OBP	<i>2'-Hydroxybiphenyl-2-Sulfinic Acid</i>
PDB	<i>Protein Data Bank</i>
pH	poder do Hidrogénio
PLA2	<i>Phospholipase(s) A2</i>
sPLA2	<i>secreted Phospholipase A2</i>
RMN	Ressonância Magnética Nuclear
RMSD	<i>Root Mean Square Deviation</i>
RNA	<i>RiboNucleic Acid</i>
S	Enxofre
Ser	Serina
SO <sub>2</sub>	Dióxido de Enxofre
UniProtKB	<i>UniProt Knowledge Base</i>
VMD	<i>Visual Molecular Dynamics</i>
VDW	<i>Van der Waals</i>
3D	Tridimensional

## Índice

<b>Introdução.....</b>	<b>1</b>
sPLA2.....	1
Modulação por homologia.....	2
Ancoragem Molecular.....	3
Pesquisa.....	4
Pontuação das poses.....	5
Triagem Virtual.....	7
Preparação da estrutura do recetor.....	7
Preparação da biblioteca de compostos.....	8
Seriação dos compostos da biblioteca.....	8
Inspeção dos compostos melhor seriados.....	8
Métodos QM/MM.....	9
Mecânica Molecular.....	9
Mecânica Quântica.....	10
DszB.....	10
<b>Materiais e Métodos &amp; Resultados e Discussão.....</b>	<b>12</b>
Modulação por homologia.....	12
Minimização de energia.....	16
Ancoragem Molecular.....	19
Triagem Virtual.....	23
Métodos QM/MM.....	26
<b>Bibliografia.....</b>	<b>30</b>

## Introdução

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

*Steve Ballmer*

O presente relatório resume o trabalho desenvolvido sob: uma PLA2 secretada, básica e cataliticamente ativa, no estudo das técnicas de modelação por homologia, ancoragem molecular (*Molecular Docking*) e triagem virtual (*Virtual Screening*); bem como, o mecanismo de catálise da enzima DszB, por aplicação de métodos híbridos QM/MM (*Quantum Mechanics/Molecular Mechanics*).

### sPLA2

As sPLA2 (em inglês, *secreted Phospholipase A2*) constituem um grupo de enzimas, características do veneno de várias víboras – nomeadamente, as serpentes *Bothrops asper* e *Daboia Russelii*. Quando possuem atividade catalítica ( $\text{Ca}^{2+}$  no centro ativo), são responsáveis pela degradação de ácidos gordos, atuando como hidrolases. A nível não catalítico, estão envolvidas na regulação da coagulação - por ligação a recetores específicos na cascata -, potenciando (tromboses) ou inibindo (hemorragias) a agregação plaquetária (1). Este controlo duplo, aliado à toxicidade e inerente falência de órgãos, motiva o desenvolvimento de fármacos preventivos.

Estruturalmente, as sPLA2 cataliticamente ativas apresentam uma conformação altamente conservada, caracterizada pela presença de uma hélice- $\alpha$  no terminal amina, um *loop* arredor do cofator metálico, duas longas hélices- $\alpha$  e duas folhas- $\beta$  antiparalelas. A coordenar o ião, encontra-se, tipicamente, um resíduo de Asp catalítico, seguido de dois monómeros de Cys e precedido de um de His, na vizinhança de uma molécula de água (mediadora da reação) (2).

## Modelação por homologia

São várias as razões que limitam a determinação experimental, i.e., por cristalografia de raios-X ou Ressonância Magnética Nuclear (RMN), de uma estrutura proteica. O próprio tamanho da molécula, cristais com pouca resolução, impuros e/ou dificuldades na medição da densidade eletrónica são alguns exemplos. Motivada por estes constrangimentos, a técnica de modelação por homologia é capaz de inferir a estrutura tridimensional de uma proteína, a partir da sua estrutura primária (sequência de resíduos de aminoácido) e da conformação tridimensional de proteínas homólogas (que partilhem com ela um ancestral comum).

Apesar da divergência evolutiva provocar alterações nas sequências proteicas, as repercussões “fenotípicas” nas conformações 3D das proteínas correspondentes, em organismos taxonomicamente próximos, são suficientemente insignificativas para gerar modelos por homologia precisos. Isto acontece porque a evolução tende a conservar resíduos relevantes para a funcionalidade das proteínas. Consequentemente, as mutações por resíduos quimicamente distintos são desfavoráveis e pouco prováveis. Neste sentido, o rigor do modelo gerado é dependente da identidade e similaridade, entre as sequências da proteína alvo (*target* – a modular) e da proteína molde (*template* – para construir o modelo). A identidade corresponde à percentagem de resíduos de aminoácidos comuns e deve ser igual ou superior a 30%. Acima deste limiar, a chance de duas sequências produzirem estruturas muito diferentes é praticamente nula. A similaridade é calculada com base em matrizes de pontuação (BLOSUM62, por ex.), em que cada célula corresponde ao *score* atribuído à substituição de um dado resíduo por outro. Esta medida tem em consideração a natureza bioquímica dos resíduos (polares, apolares, aromáticos, carregados), a probabilidade de ocorrência numa determinada posição sequencial e o grau de influência na função/estrutura da proteína. Outro fator relevante é o tamanho das sequências em comparação (molde e alvo): por ex., 30% de identidade sequencial em proteínas de tamanhos muito distintos não é o mesmo que 30% em proteínas do mesmo tamanho. Todos estes aspetos influenciam a seleção do molde adequado. Adicionalmente, a existência de uma estrutura de boa resolução (se cristalográfica) deve ser analisada e o *template* e o *target* devem partilhar o máximo possível de fatores ambientais (pH e temperatura a que a foram determinados, solvente e interação com outras proteínas). Se a estrutura da proteína a modular englobar ligandos ou cofatores metálicos, deve optar-se

por um molde que os contenha, de forma a evitar rearranjos conformacionais indesejados, aquando da adição manual.

Uma vez selecionada a estrutura molde adequada, pode proceder-se à construção da estrutura alvo, em *softwares* específicos – os quais implementam um dos vários métodos de modelação (montagem de corpos rígidos, correspondência de segmentos e modelação por satisfação de restrições espaciais).

Finalizado o processo, deve avaliar-se a estrutura gerada – a qual apresentará uma precisão tanto mais significativa, quanto maior a identidade sequencial com a proteína molde (de onde foi derivada). Existem duas formas de avaliação: interna – que visa averiguar o quão estereoquimicamente preciso é o modelo, comparando as suas características (ângulos, distâncias, pontes dissulfureto) com as que seriam expectáveis encontrar em proteínas nativas; e externa – que avalia a probabilidade do enrolamento gerado ocorrer naturalmente. No âmbito do trabalho atual, ao contrário do que acontece tipicamente, a estrutura cristalográfica do modelo é conhecida e a avaliação assentou num *simples* cálculo de RMSD (*Root Mean Square Deviation*).

## **Ancoragem Molecular**

A Ancoragem Molecular (em inglês, *Molecular Docking*) é uma técnica que prevê as geometrias de associação (poses) de moléculas com estrutura tridimensional conhecida, e estima, do ponto de vista energético, a melhor conformação do complexo intermolecular. Geralmente, esta corresponde à pose com o maior número de interações – à partida, mais estável (i.e., com menor energia potencial) e próxima da real nativa.

Existem dois tipos principais de Ancoragem Molecular: *docking* proteína-ligando, amplamente utilizado na indústria farmacêutica, aquando da descoberta de pequenos fármacos; e *docking* proteína-proteína, aplicado a estas macromoléculas e, portanto, menos preciso e não tão utilizado. No contexto do trabalho desenvolvido, foi estudado o *docking* proteína-ligando, aplicado ao complexo sPLA2-[1-Aminooxalyl-3-benzyl-2-ethyl-indolizin-8-yloxy]-acetic acid (um ligando de natureza anfipática, à semelhança de um fosfolípido).

## Pesquisa

A primeira etapa de qualquer protocolo de *Docking* é a pesquisa. Esta tem como objetivo gerar poses o mais próximas possível da verdadeira (desconhecida). No contexto, o algoritmo faz uma análise de todos os graus de liberdade internos do ligando – relativos às ligações simples rotáveis –, gerando várias geometrias. Depois, a partir dos graus de liberdade rotacional (rotações da molécula, quando o seu centro geométrico está fixo no espaço) e translacional (segundo os eixos ortonormados), é feito outro cálculo matemático independente, que procura encaixar a molécula num espaço limitado, englobante do centro ativo. Isto leva a que o máximo de conformações e orientações do ligando na cavidade (*binding site*) do recetor (proteína) seja explorado.

Posteriormente, num processo denominado *Scoring*, a cada pose encontrada pela pesquisa é atribuída uma pontuação. Assim, é possível criar uma seriação de estruturas e estimar, a partir dela, a geometria de associação mais provável.

Nos casos em que a flexibilidade da proteína não é explorada (fixando-se na sua conformação mais estável) ou, se é, não interfere significativamente com o local de reação (conservado), fala-se de “pesquisa com recetor rígido”; quando se explora o espaço conformacional do recetor, considerando a possibilidade de não se encontrar na sua forma mais estável e/ou que se modifica com a aproximação do ligando, fala-se de “pesquisa com recetor flexível”. Em ambas as vertentes, o ligando é, naturalmente, flexível, possuindo cerca de 5 a 12 ligações rotáveis. Ora, de experiências anteriores, sabe-se que a sPLA2 é muito estável, independentemente, das condições do meio. Isto acontece devido à presença de 7 pontes dissulfureto, que conferem rigidez à conformação. Sendo assim, a implementação da metodologia “com recetor rígido” (apenas o ligando flexível) é quase ideal/pouco prejudicial e foi a selecionada, no contexto do problema.

Existem vários tipos de algoritmos de pesquisa com recetor rígido: métodos sistemáticos – que exploram todos os graus de liberdade (translacional, rotacional e internos) na molécula (ligando), na expectativa de varrer todas as possibilidades conformacionais; computacionalmente desgastantes (basta pensar que cada ângulo diedro possui 3 posições possíveis e cada uma delas outras 3 e assim, sucessivamente...); métodos estocásticos – que procuram fazer uma amostragem, gerando geometrias de forma semi-aleatória (o que maximiza a probabilidade de encontrar a conformação *certa*); e métodos de simulação – que se baseiam no cálculo das equações Newtonianas do



movimento. No âmbito deste trabalho, utilizou-se o *software* AutoDock, o qual recorre a algoritmos de pesquisa genéticos, incluídos nos métodos sistemáticos. Este tipo de algoritmos baseia-se na teoria da evolução biológica: cada conformação é definida por um conjunto de “genes”, equivalente a cada grau de liberdade do ligando (i.e., possíveis valores para ângulos diedros, rotação, translação e orientação); a totalidade de variáveis/graus de liberdade de uma pose denomina-se “cromossoma”; o conjunto de variáveis de estado de um ligando equivale ao “genótipo”; e as coordenadas atómicas correspondentes ao “fenótipo”. A exploração do espaço conformacional tem por base “operadores genéticos”, os quais executam “mutações aleatórias” nos genes e “*cross-overs*” entre as porções mais estáveis das poses.

### **Pontuação das poses**

Ora, como referido anteriormente, a pontuação da pose é uma medida da qualidade do encaixe (*fitness*) do ligando, no centro ativo. Determiná-la, geralmente, envolve cálculos simples de energia, os quais são dependentes de uma função de pontuação (em inglês, *scoring function*) previamente definida. De entre os métodos de *scoring*, existem os baseados em campos de força, no conhecimento e empíricos.

Nos primeiros, campos de força *standard* quantificam a energia de interação proteína-ligando, através da combinação dos termos Energia Interna do ligando, Energia Eletroestática e de *Van der Waals*. Este tipo de função peca pela ausência dos termos de solvatação e entropia, que são tratados empiricamente. Quando a conformação do ligando é menos favorável e se criam tensões estereoquímicas para que este se adapte ao centro ativo, são introduzidas penalizações na função de energia total.

Nas funções de pontuação com base no conhecimento, os potenciais são calculados a partir da frequência de ocorrência de diferentes contactos átomo-átomo e outras interações típicas, presentes em grandes conjuntos de dados de complexos moleculares. Ora, comparando o número relativo de vezes que um determinado arranjo atómico ocorre (“valor real”), com a probabilidade aleatória correspondente (“valor esperado”), cria-se um sistema de pontuação em que esta será tanto maior, quanto mais estável a associação. As geometrias mais frequentes refletem as estruturas com interações mais favoráveis. A desvantagem desta função é ser dependente da informação disponível em conjuntos limitados de estruturas.

Nas funções empíricas, a energia livre total também é dada pela combinação linear das interações individuais (provenientes do campo de forças), contudo, a contribuição (peso/coeficiente) de cada termo para o potencial energético resulta de um ajuste multilinear contra dados experimentais (empíricos). Basicamente, a partir de energias livres, laboratorialmente, determinadas e/ou de *training sets* com informação sobre complexos moleculares é possível, por meio de regressão linear, estimar os pesos de cada componente. O constrangimento desta *scoring function* está no facto de, ao estar dependente de dados experimentais, não ser versátil, nem transferível, i.e., a mesma função não pode ser aplicada a diferentes sistemas.

Notavelmente, nenhuma função de *score* é perfeita e nem todas as forças estabelecidas contribuem significativamente para a Energia de Associação calculada. Por exemplo, as pontes de hidrogénio não fixam o ligando ao recetor, *apenas* o obrigam a ancorar numa dada posição, relevante para as reações químicas que ocorrem. Quando o ligando está em solução, estabelece pontes de hidrogénio com a água; quando está num complexo, liga-se, da mesma forma, à proteína. Por isso, o balanço é, praticamente, neutro. O mesmo acontece com as interações iónicas: são essenciais para manter o Cálcio do centro ativo ligado ao resíduo de Aspartato, mas não são determinantes para a associação em si. Por outro lado, as forças de *Van der Waals* – hidrofóbicas, que se estabelecem entre grupos não polares da proteína e do ligando –, apesar de fracas, são numerosas e muito mais fortes e favoráveis na proteína do que em água. Conferem, portanto, a grande parte da energia de associação.

Apesar de, atualmente, já ser possível prever a geometria de associação proteína-ligando, com uma exatidão de cerca de 70 a 80%, fatores como: a identificação da geometria mais favorável, por parte das *scoring functions*; o efeito do solvente e a participação direta das moléculas de água nas interações proteína-ligando; e as resoluções limitadas da maior parte dos *targets* cristalográficos, são ainda limitantes.

## Triagem Virtual

Os fármacos são, por norma, pequenas moléculas que se associam a recetores biológicos (maiores), com elevada afinidade. Estes correspondem, geralmente, a enzimas ou proteínas transmembranares, e modificam a sua estrutura, aquando do fenómeno de associação. São estas alterações na estrutura do recetor que determinam a atividade biológica subsequente, inibindo ou estimulando a sua função. Tendo isto em mente, a ideia de criar pequenos ligandos, controladores de funções biológicas de interesse, conquistou a indústria farmacêutica. Atualmente, a técnica *in silico* de Triagem Virtual (em inglês, *Virtual Screening*) possibilita o *design* racional de fármacos, a partir da estrutura das proteínas recetoras, emergindo, assim, como método complementar da descoberta experimental. O presente relatório foca-se na Triagem Virtual baseada na estrutura do recetor (sPLA2) – uma técnica que se propõe distinguir ligandos de elevada afinidade para uma molécula alvo (HITs), de compostos com os quais isso não se sucede (não-ligandos). Esta metodologia é, na verdade, uma das principais aplicações da Ancoragem Molecular e, no caso em questão, foi executada com recetor (proteína) fixo.

### Preparação da estrutura do recetor

O primeiro passo de um protocolo de *Screening* é a preparação da estrutura do recetor. Esta deve estar completa, principalmente, ao nível do centro ativo (local de ancoragem), conter cofatores necessários à reação, e, no caso de estruturas determinadas por cristalografia, ter uma resolução, preferencialmente, inferior a 2 Å. Se existirem complexos moleculares pré-determinados, também devem ser tidos em consideração. Por exemplo, saber, antecipadamente, que a conformação de um recetor sofre alterações significativas, aquando da associação é uma motivação para não executar ancoragens com essa molécula estática (e vice-versa). Adicionalmente, deve ter-se em atenção diferentes formas tautoméricas, resultantes da má distinção entre átomos de H e N (comum em estruturas cristalográficas), e estados de ionização – quer do recetor, quer do composto a ancorar. Informação adicional sobre a importância deste passo pode ser encontrada na secção ‘Materiais e Métodos & Resultados e Discussão\ *Virtual Screening*’). Moléculas de água mediadoras da associação devem ser incluídas, de forma a prever geometrias de associação tanto mais exatas quanto possível. No entanto, no âmbito do trabalho atual, optou-se por não o fazer (ver secção ‘Materiais e Métodos & Resultados e Discussão\ *Molecular Docking*’).

## **Preparação da biblioteca de compostos**

Para além da qualidade da estrutura do recetor, o sucesso de uma campanha de triagem virtual é altamente dependente dos compostos a ancorar. Por exemplo, a procura de moléculas com potencial farmacêutico e estrutura diferente de um *hit* já conhecido para o mesmo recetor, pode ser uma mais-valia para o desenvolvimento de fármacos alternativos (com outros efeitos secundários). O conjunto de compostos químicos desenhados *in silico* e/ou naturalmente sintetizados formam uma biblioteca. Por norma, as bibliotecas utilizadas em *Virtual Screening* são derivadas de outros conjuntos de dados e, antecipadamente, filtradas segundo o potencial farmacocinético das moléculas que incluem (i.e., forma e velocidade de biodistribuição, no organismo). Para além desta seleção, há a opção de considerar filtros adicionais, baseados nas propriedades da estrutura do recetor e de ligandos já conhecidos. Por ex., se a cavidade de ligação tem uma dimensão limitada, então, se calhar, faz sentido excluir do conjunto de compostos todos as moléculas que a ultrapassem. No entanto, isto reduz o número de soluções do *docking* e pode, facilmente, eliminar bons ligandos. No trabalho desenvolvido, utilizou-se a base de dados pública BindingDB, da qual foram selecionados (por limitações de tempo) 6 dos 512 compostos presentes, sem a aplicação de filtros adicionais.

## **Seriação dos compostos da biblioteca**

De forma a identificar e seriar ligandos e não-ligandos, procede-se à ancoragem molecular de cada composto da biblioteca. Idealmente, se possível, as poses previstas devem ser comparadas com associações proteína-ligando conhecidas (no PDB), de forma a avaliar a qualidade do *docking* executado (informação adicional na secção ‘Materiais e Métodos & Resultados e Discussão\ Ancoragem Molecular’). Posteriormente, avalia-se a qualidade da seriação, i.e., a capacidade do método de ancoragem classificar como ligandos e não-ligandos, as moléculas que, de facto, os constituem.

## **Inspeção dos compostos melhor seriados**

Porque finalizá-la é motivo de comemoração, a última etapa da campanha de triagem é a *HIT Picking Party* – uma análise visual dos ligandos selecionados (Nesta fase, todos com grande e similar qualidade de associação).

## Métodos QM/MM

### Mecânica Molecular

Uma das áreas de suporte mais relevantes da Química Computacional é a Mecânica Molecular (*Molecular Mechanics* – MM) – uma metodologia computacional, baseada na física clássica, 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 (como DNA, RNA e macromoléculas). 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 os elétrons acompanham a deslocação dos núcleos atômicos (de maior massa), o que possibilita 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). Por não descrever explicitamente os elétrons, fenómenos como a quebra e formação de ligações químicas ou a atividade enzimática, em reações de catálise, não podem ser estudados por MM.

### Mecânica Quântica

Como metodologia complementar da Mecânica Clássica supracitada, surge a Mecânica Quântica (*Quantum Mechanics* – QM) – uma técnica rigorosa, que visa descrever o movimento de qualquer tipo de sistema microscópico (como é o caso de elétrons, proteínas/enzimas, átomos e núcleos). É sustentada por ciências exatas (física e matemática) e baseada numa função extraordinária, que encerra em si toda a informação do sistema químico (função de onda). Esta recebe, como argumentos, as coordenadas espaciais dos elétrons e núcleos do sistema e retorna a probabilidade de os encontrar, numa dada posição do espaço, independentemente do tempo (estado estacionário). Ora, operando o hamiltoniano molecular sob a função de onda, é possível calcular a energia global do sistema molecular e ter conhecimento sobre todas as suas propriedades. A equação que relaciona estas três variáveis denomina-se equação de Schrödinger e é função da QM resolvê-la o mais exatamente possível, ainda que requeira aproximações –

caso dos sistemas químicos com mais do que um elétron. Uma simplificação típica é a proposta por Born e Oppenheimer, que, à semelhança do que acontece na MM, assenta no facto da massa nuclear ser muito maior que a dos eletrões. Assim, é plausível assumir-se que os núcleos estão fixos em relação às partículas da nuvem e o termo de energia cinética do hamiltoniano é reduzido ao movimento eletrónico. Os métodos de QM que admitem este tipo de simplificações matemáticas, mas que não consideram nenhum parâmetro empírico/experimental para as nivelar, denominam-se *ab initio*; quando há recurso a algum tipo de parametrização, fala-se de cálculos semi-empíricos. Em ambos os casos, a elevada exatidão das abordagens de Mecânica Quântica permite determinar estados de transição e de excitação.

### **Métodos QM/MM**

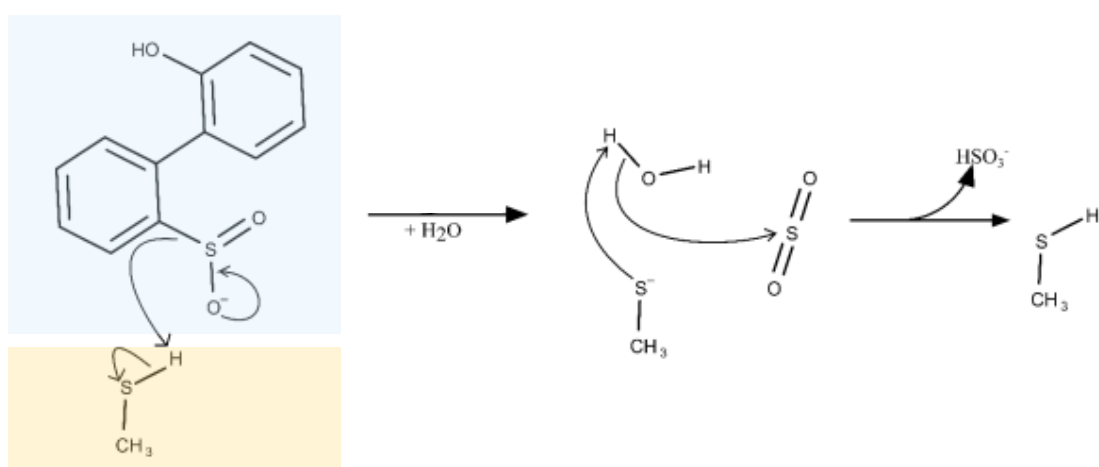
A rapidez e capacidade de aplicação a sistemas grandes da MM, aliadas à elevada exatidão e descrição explícita de eletrões da QM hibridam-se na metodologia única QM/MM. Desta forma, “o melhor dos dois mundos” pode unir-se no estudo de sistemas biomoleculares, dividindo-os em: *high layer* (ou região QM) – pequena, eletronicamente relevante, onde são executados cálculos quânticos; *low layer* (ou região MM) – com um maior número de átomos, passível de tratamento clássico; e *medium layer*, se aplicável – onde, por norma, são executados cálculos menos rigorosos de Mecânica Quântica.

Ora, precisamente, por incluírem um pequeno local de reação química (centro ativo, substrato e cofatores) e um amplo ambiente circundante, cujas interações eletroestáticas com o cerne reacional não são significativas para provocar rearranjos eletrónicos (solvente e restante porção da macromolécula), as enzimas constituem o sistema biológico ideal para empregar a abordagem QM/MM. No trabalho vigente, a proteína DszB foi alvo de estudo.

### **DszB**

A DszB (do inglês, *2'-Hydroxybiphenyl-2-Sulfinase Desulfinase*) é uma enzima de suma importância no processo de desfosforilação de compostos orgânicos, nomeadamente, o petróleo. Por participar na remoção de enxofre, S, a temperatura e pressão ambientes, o estudo do seu mecanismo de catálise acarreta um interesse biotecnológico acrescido. Na verdade, já são várias as formas de aumentar a dependência de compostos de enxofre, em bactérias sintetizantes de DszB. Os métodos híbridos QM/MM são úteis para descrever a reação metabólica em causa, a qual pode seguir duas

vias: adição nucleofílica ou substituição eletrofílica. Por ter associada uma menor barreira energética e, conseqüentemente, uma maior probabilidade de ocorrência, no contexto do trabalho, optou-se por estudar o segundo mecanismo. De forma geral, a substituição inicia-se com a quebra da ligação SH do resíduo Cys8 da enzima, que liberta o próton para ligação ao anel aromático do substrato (no caso, OBP – *2'-hydroxybiphenyl-2-sulfinic acid*). O SO<sub>2</sub> é, então, eliminado, e atacado pela molécula de água catalítica, formando HSO<sub>3</sub><sup>-</sup>. Ora, aproximando, gradualmente, o H do grupo tiol da DszB, ao C–SO<sub>2</sub> do OBP, é possível simular a reação de protonação em causa. No trabalho em questão, a aproximação foi feita em incrementos de 0.1 Å.



**Fig.2** – Representação esquemática do mecanismo de substituição eletrofílica da enzima DszB. Substrato destacado a azul. C-Grupo R da cisteína catalítica, a laranja.

## **Materiais e Métodos & Resultados e Discussão**

*“O nosso cérebro simula a realidade.”*

*Stephen LaBerge*

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

*Theodor Nelson*

### **Modelação por homologia**

#### **Recolha da sequência da proteína a modelar (*target*)**

Para dar início ao protocolo de modelação por homologia, acedeu-se ao UniProtKB (*UniProt Knowledge Base*) – uma base de dados de sequências de proteínas. Dos resultados obtidos para a pesquisa “organism: *Bothrops asper* (Terciopelo) [8722] AND name: *“basic phospholipase a2”* AND”, seleccionou-se o correspondente à molécula de interesse (Código UniProt: P20474) e descarregou-se, em formato FASTA, a sua sequência. Esta é composta por um péptido sinal (resíduos 1-16) e pela sequência da proteína alvo, com 122 resíduos (17-138). O Asp64 é precedido por uma His63, ambos com atividade catalítica.

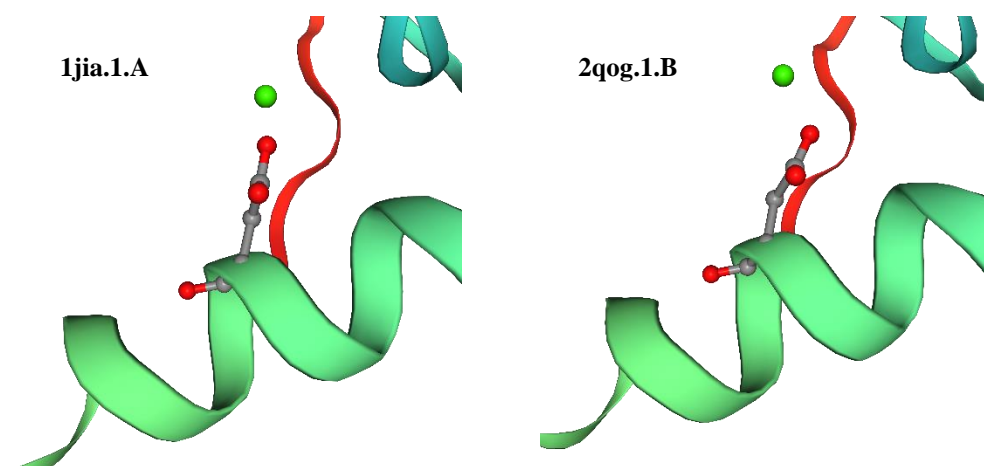
#### **Pesquisa de proteínas homólogas e alinhamento sequencial**

Uma vez obtida a estrutura primária da molécula, recorreu-se ao Swiss-Model – um servidor de procura de proteínas homólogas. Socorrendo-se de uma biblioteca pessoal, pesquisa, internamente, por sequências candidatas a *template* e executa o alinhamento entre cada uma delas e a sequência de entrada (a modelar). Aos *matches* encontrados associa medidas de qualidade, as quais oferecem uma previsão sobre o quão preciso seria o modelo gerado. Com base neste fator, é retornada uma lista ordenada de potenciais sequências molde, e as suas respetivas características – percentagem de identidade, método experimental de determinação estrutural (e, se aplicável, resolução), estado oligomérico, ligandos e percentagem de sequência alvo presente no *template*.



## Escolha da sequência molde

Da seriação obtida, considerando os fatores enunciados na secção ‘Introdução\ Modelação por homologia’, destacaram-se duas potenciais proteínas molde: a 1jia.1.A, com identidade sequencial 63.93%  $\geq$  30% (limiar necessário para um modelo preciso), resolução 2.1 Å, cofator  $\text{Ca}^{2+}$  e 122 resíduos (tamanho igual ao da sequência a modular); e a 2qog.1.B, com 63.11% de identidade, 2.3 Å de resolução, cofator e mesmo tamanho. Apesar de possuírem medidas de qualidade bastante similares, o resíduo de Asp da conformação 1jia.1.A coordena o Cálcio de forma muito mais alinhada (fig.3) e, portanto, foi a estrutura escolhida para *template*.

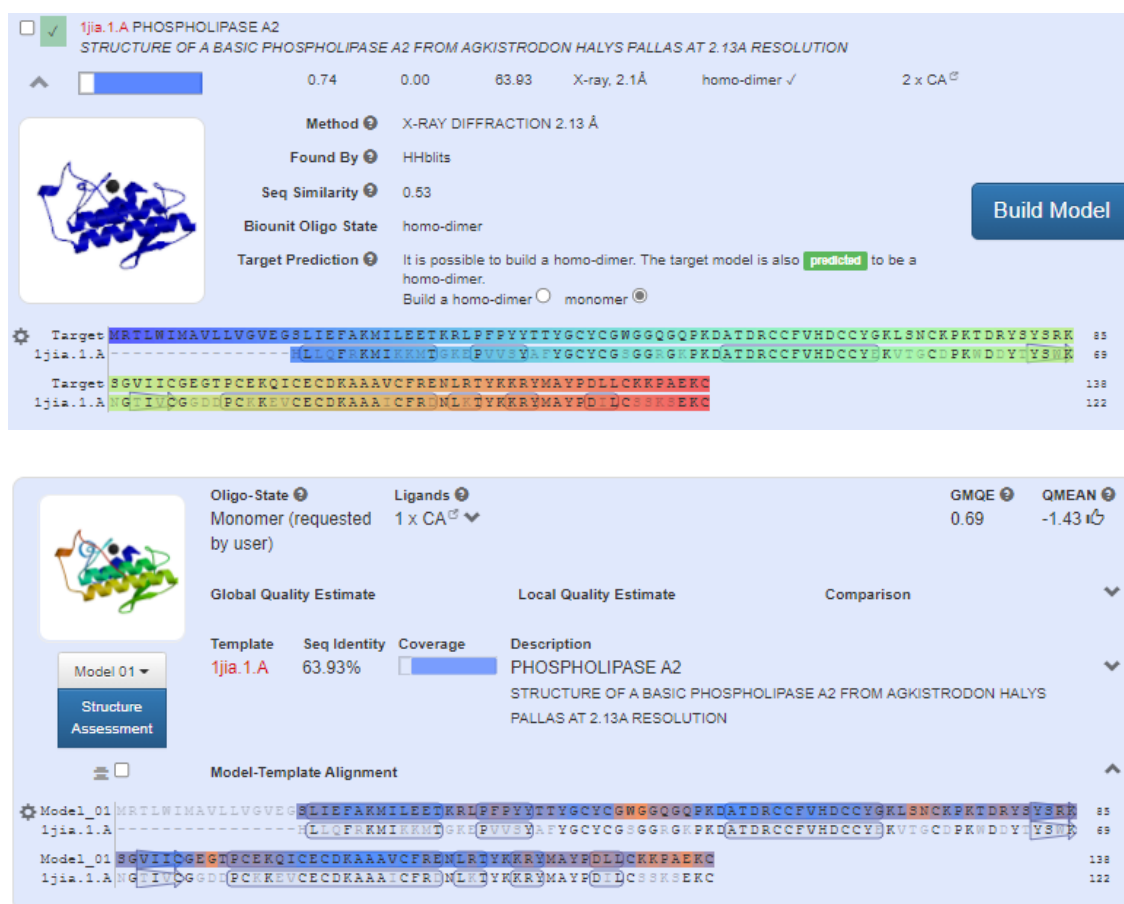


**Fig.3** – Representação gráfica do centro ativo das proteínas 1jia.1.A e 2qog.1.B – candidatas a molde –, com ênfase para a interação eletroestática estabelecida entre o resíduo de Asp catalítico e o íon  $\text{Ca}^{2+}$ . Fonte: Swiss-Model.

A estrutura proteica 5tfv.1.A, com resolução 2.5 Å, íon  $\text{Ca}^{2+}$  e 122 resíduos (sem péptido inicial) foi destacada para posterior comparação com o modelo criado. Por ter associada uma percentagem de identidade sequencial muito próxima de 100% (99.18%,) é passível assumir-se que corresponde à estrutura nativa da proteína a modular (habitualmente, não disponível) e, portanto, útil para avaliação da qualidade do modelo gerado por homologia. A estrutura cristalográfica 5tfv foi descarregada a partir do PDB.

## Construção do modelo tridimensional da proteína alvo

Escolhida a proteína molde, instruiu-se o servidor a construir o modelo, no estado oligomérico apropriado (monómero). Na fig.4, encontra-se a informação resultante do Swiss-Model. A estrutura foi guardada em formato \*.pdb (1\_Homology\_Modelling.pdb).



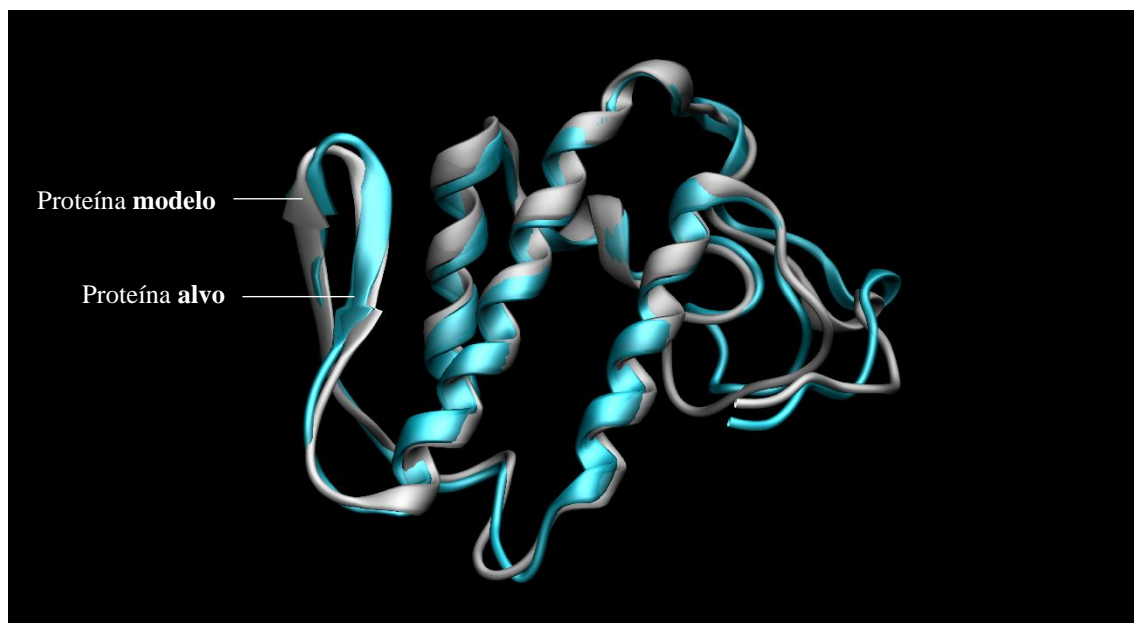
**Fig.4** – (a) Informação sobre a proteína molde 1jia.1.A. (b) Informação sobre o modelo criado por homologia. Fonte: Swiss-Model.

Notavelmente, o modelo gerado por homologia não engloba a secção correspondente ao péptido inicial, originalmente presente na sequência do alvo. Apesar do objetivo não ser modular essa porção, é importante perceber o motivo pelo qual isto acontece: as sequências da proteína molde (sem sinal) e do alvo têm tamanhos diferentes. Desta forma, o alinhamento só acontece a partir do resíduo 17 (fig.4).

### Avaliação da qualidade da estrutura modelada

Para avaliar a qualidade da estrutura modelada, comparou-se, por sobreposição, a estrutura tridimensional do modelo gerado (1\_Homology\_Modelling.pdb) e a estrutura nativa do alvo correspondente (cadeia A da proteína 5tfv.pdb) (fig.5). Calculou-se o RMSD – uma medida de similaridade entre conformações, correspondente ao desvio padrão posicional. Quanto menor (mais próximo de 0), mais coincidentes são as coordenadas atômicas e mais próximas as estruturas em comparação. No caso, obteve-se um  $\text{RMSD} \approx 1.33 \text{ \AA} \approx 1 \text{ \AA}$ , indicativo de uma alta semelhança estrutural e reflexo da elevada qualidade estereoquímica com que o modelo foi gerado.

Este foi ainda comparado com a estrutura 3D molde (1jia.1.A) e obteve-se um  $\text{RMSD} \approx 0.38 \text{ \AA}$ , muito próximo de  $0 \text{ \AA}$  – o esperado, dada a elevada identidade entre as sequências.



**Fig.5** – Representação gráfica da estrutura da sPLA2 obtida por homologia (modelo) – a branco –, e da conformação nativa correspondente 5tfv.1.A (alvo) – a ciano –, em sobreposição.

### Modelação da molécula de água mediadora da catálise

A molécula de água catalítica foi identificada na estrutura 1jia.1.A e as suas coordenadas transferidas para o ficheiro 1\_Homollogy\_Modelling.pdb.

## Minimizações de energia

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

*Edwin Conklin*

De forma a eliminar maus contactos entre os átomos, diminuir as tensões estereoquímicas e torsões desfavoráveis, resultantes de erros na modulação, assim como, preparar o sistema para o protocolo de *Docking*, procedeu-se à minimização da sua energia potencial (por outras palavras, à otimização da geometria do sistema).

### Preparação do ficheiro \*.pdb

Apesar de conter a informação necessária para representar tridimensionalmente a estrutura de uma proteína, aos arquivos \*.pdb estão, tipicamente, associados problemas resultantes de limitações experimentais. Os raios-X, por ex., só difratam nas nuvens eletrónicas dos átomos. Isto limita a deteção da densidade eletrónica de partículas muito leves, como o H – o que, habitualmente, resulta na sua ausência (ou incorreta distribuição, quando posteriormente distribuídos) em estruturas obtidas por cristalografia. O mesmo acontece com moléculas de água, por serem muito voláteis. Ora, o modelo de homologia foi gerado a partir de um ficheiro \*.pdb e, portanto, a adição posterior destes elementos é necessária para que fique completo (ver secção ‘Geração dos ficheiros de parâmetros e de topologia’). Para além disso, a presença de informação irrelevante, numeração/nomenclaturas incorretas, ou até incompatíveis com os *softwares* de cálculo a utilizar, tornam a modificação destes arquivos indispensável.

Neste sentido, recorreu-se a um script de conversão (*pdb4amber*), o qual originou quatro ficheiros: *\*nonprot.pdb* – destinado a guardar informação sobre a parte não proteica de um sistema (no caso, vazio); *\*renum.txt* – com a antiga e nova numeração dos resíduos de aminoácidos (Asp64 e His63, por ex., passaram a ocupar as posições 48 e 47, respetivamente); *\*ss\_link.txt* – com a identificação dos resíduos de cisteína entre os quais se estabelecem pontes dissulfureto; e *\*\_amber.pdb* – equivalente ao ficheiro \*.pdb processado. Deste último, removeram-se as linhas relativas às conectividades entre átomos, permitindo que novas pudessem ser adicionadas pelo campo de forças do AMBER (*Assisted Model Building with Energy Refinement*).

## **Geração dos ficheiros de parâmetros e de topologia**

Para gerar um sistema biomolecular que mimetize as condições experimentais e criar os ficheiros necessários ao início de um processo de minimização no AMBER, recorreu-se ao módulo *Xleap* – um *software* de cálculo e visualização molecular integrado. De forma a que todos os elementos estruturais fossem reconhecidos pelo aplicativo, a biblioteca de parâmetros para os vários resíduos da proteína (FF14SB) e a biblioteca geral para moléculas orgânicas (GAFF) foram carregadas. Informação sobre o modelo de água, respetiva parametrização (frcmod.ionsjc\_tip3p) e pontes de dissulfureto foi fornecida ao programa. De forma a contrabalançar a carga global do sistema, simulando a realidade biológica e preparando-a para a adição do solvente explícito, introduziram-se 9 iões Cl<sup>-</sup>. O sistema, agora eletricamente neutro, foi introduzido numa caixa de água, cujas faces distam 12 Å das extremidades da proteína. Os átomos de hidrogénio em falta foram, automaticamente, colocados pelo *software*, em posições iniciais aleatórias/não otimizadas. A informação gerada foi guardada nos ficheiros de topologia e parâmetros \*.inpcrd e \*.prmtop – arquivos de entrada para o módulo *Sander*.

### **Visualização do sistema**

O sistema molecular foi visualizado, com recurso ao VMD, assegurando-se que todas as modificações foram corretamente executadas.

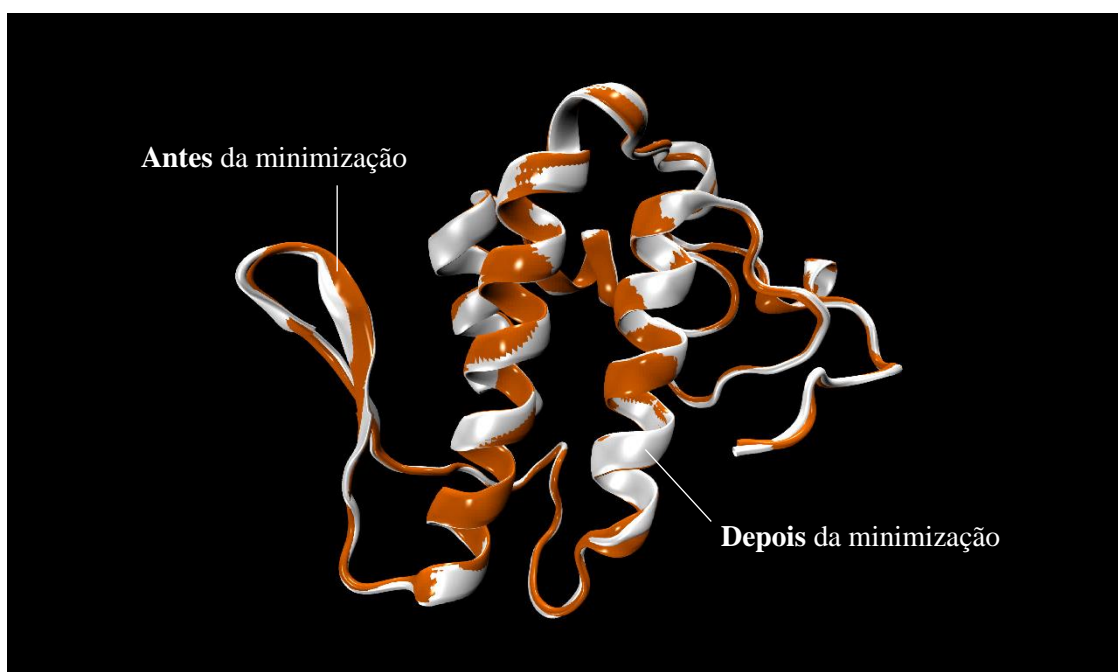
### **Minimizações de energia**

Invocou-se o módulo *Sander* e executou-se cada uma das quatro minimizações energéticas: na primeira (min1), as moléculas de água, praticamente em equilíbrio (pré-parametrizadas), foram otimizadas. O resto do sistema manteve-se fixo. Este passo é importante para evitar que águas mal colocadas induzam a desnaturação; na segunda (min2), os átomos de hidrogénio foram recolocados no espaço tridimensional; na terceira (min3), otimizaram-se as cadeias laterais; e na quarta (min4), todos os átomos. Obteve-se um sistema biomolecular significativamente mais refinado e, estima-se, mais próximo do nativo correspondente.

### Visualização da estrutura minimizada

O centro ativo do sistema minimizado encontra-se representado na fig.1 (capa). A posição da água catalítica (junto ao resíduo de His48) é altamente conservada e a coordenação do cofator  $\text{Ca}^{2+}$ , pelo Asp49, prevalece.

O enrolamento global do sistema inicial e da versão energeticamente minimizada demonstra-se, praticamente, inalterado (fig.6). Isto acontece porque a função de energia potencial obriga os diedros das estruturas secundárias a confinarem-se aos valores de equilíbrio pré-definidos. 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.



**Fig.6** – Representação gráfica da estrutura modelada por homologia antes (a laranja) e depois (a branco) da minimização energética.  $\text{RMSD} \approx 0.38$ . *Software* de visualização molecular: VMD.

## Ancoragem Molecular

Para evitar potenciais impedimentos estereoquímicos e não comprometer a disposição espacial do ligando no centro ativo, garantiu-se que todas as moléculas de água do sistema anteriormente minimizado (sPLA2\_H\_wat.pdb) fossem removidas. O íon Cálcio manteve-se, dada a sua importância para a estabilidade da proteína. A nova informação foi guardada no ficheiro sPLA2.pdb.

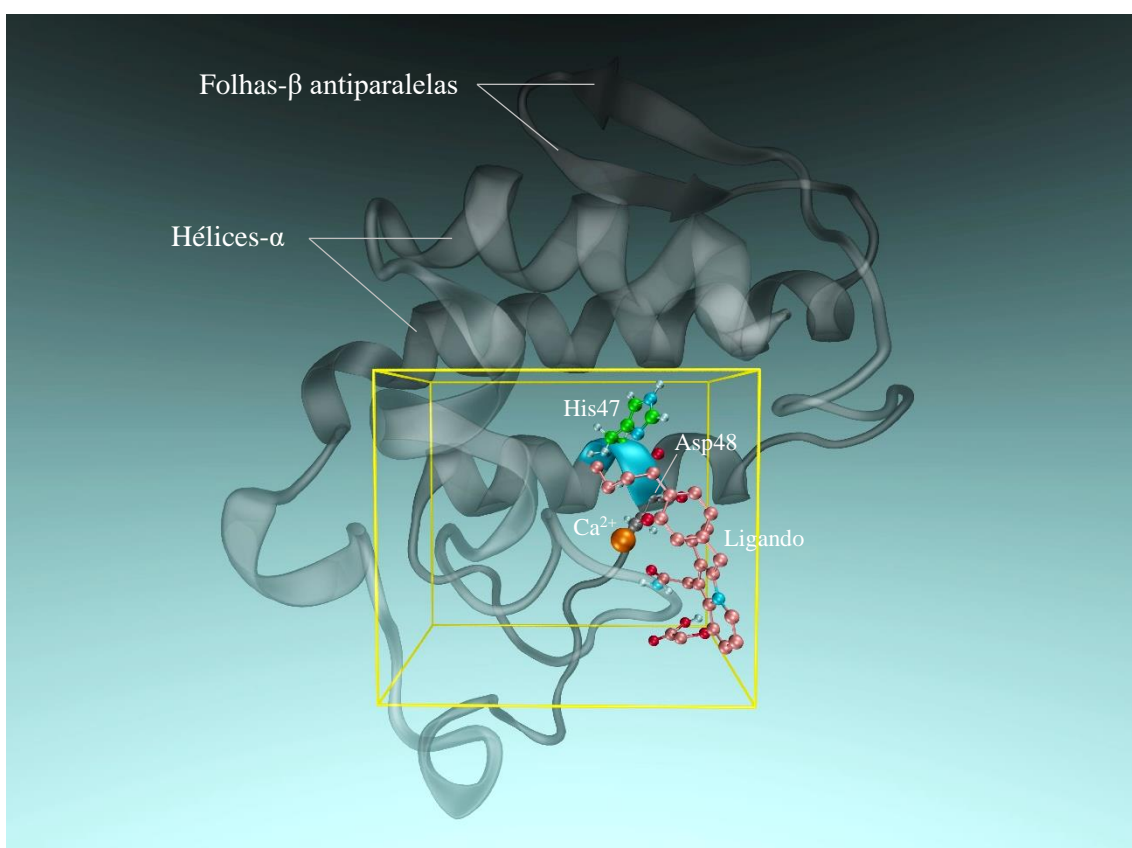
Posteriormente, a partir do PubChem – um repositório *online* de moléculas e respetivas atividades biológicas – obteve-se a estrutura 3D do ligando (*[1-Aminooxalyl-3-benzyl-2-ethyl-indolizin-8-yloxy]-acetic acid*), a qual foi descarregada em formato \*.sdf. Por ser incompatível com o programa de *docking*, procedeu-se à conversão para o formato \*.mol2 (ligand1.mol2), com recurso ao Open Babel.

Uma vez obtidas as estruturas da proteína e do ligando a ancorar, carregou-se a estrutura sPLA2.pdb no VMD e, recorrendo à extensão VsLab, adicionou-se o ligando ao sistema. Acedeu-se, de seguida, ao separador AutoGrid – um programa responsável pela construção de grelhas de energias de interação, entre vários tipos de átomos (como carbonos aromáticos, alifáticos, oxigénios, etc, do ligando) e uma macromolécula (no caso, proteína). As grelhas são geradas numa região tridimensional limitada, equivalente ao *binding site* de interesse, e posteriormente, utilizadas para os cálculos de energia de associação, no AutoDock.

Notavelmente, apesar de, em solução, tenderem a formar dímeros, as sPLA2 formam complexos com as membranas celulares (ricas em fosfolípidos), no organismo. Assim, para além dos resíduos de aminoácido catalíticos (His47 e Asp48), a orientação da proteína na membrana é outro fator a considerar aquando do posicionamento espacial e dimensionamento da caixa de *docking* – a qual, não deve, nomeadamente, incluir as duas longas hélices- $\alpha$ , nem as duas folhas- $\beta$  da enzima. Com isto em mente, a área de ancoragem inicial foi definida segundo as seguintes especificações, em Å: altura:16.43, largura:17.97, profundidade:10.48, posição segundo x:44.76, y:37.90 e z:30.05. As diretrizes foram guardadas e o *docking* executado. Os 10 resultados obtidos encontram-se registados na fig.8, com ênfase para a pose de menor energia ( $\Delta G_{\text{ligação}} = \text{score} = -7.48$ ), representada na fig.7.

Para além do *score* – calculado pela função de pontuação, equivalente à energia da associação entre as moléculas –, a constante de inibição,  $K_i$ , pode (e deve) ser usada como medida de afinidade complementar. Corresponde à concentração de inibidor necessária para produzir metade da inibição máxima e, portanto, quanto menor o  $K_i$ , mais forte a inibição, i.e., mais estável é o complexo. No caso dos resultados enunciados, a seriação das poses, *por acaso*, está em concordância com ambas as variáveis e, portanto, pode assumir-se que a função de pontuação teve um bom desempenho.

Em contrapartida, à semelhança do que está representado na fig.7, pôde verificar-se visualmente, para cada uma das poses, que grande parte da área de pesquisa (à esquerda) não gerou geometrias. Deste modo, reproduziu-se a ancoragem, deslocando a caixa de *docking* para a direita. As conformações obtidas tiveram pontuações e  $K_i$  significativamente melhores – fig.8.



**Fig.7** – Representação gráfica do complexo sPLA2–([1-Aminooxalyl-3-benzyl-2-ethyl-indolizin-8-yloxy]-acetic acid), com ênfase para o centro ativo. Estrutura da proteína representada em NewCartoon (a ciano), com destaque para os resíduos catalíticos His47 (Carbonos a verde) e Asp48 (Carbonos a cinza) – a coordenar o ião  $Ca^{2+}$  (a laranja). Geometria do ligando com maior pontuação representada em CDK (a rosa). Geometria do ligando com maior pontuação representada em CDK (Carbonos a rosa). Átomos de Hidrogénio e Nitrogénio a vermelho e ciano, respetivamente. Área de *docking* a amarelo. *Software* de visualização molecular: VMD.



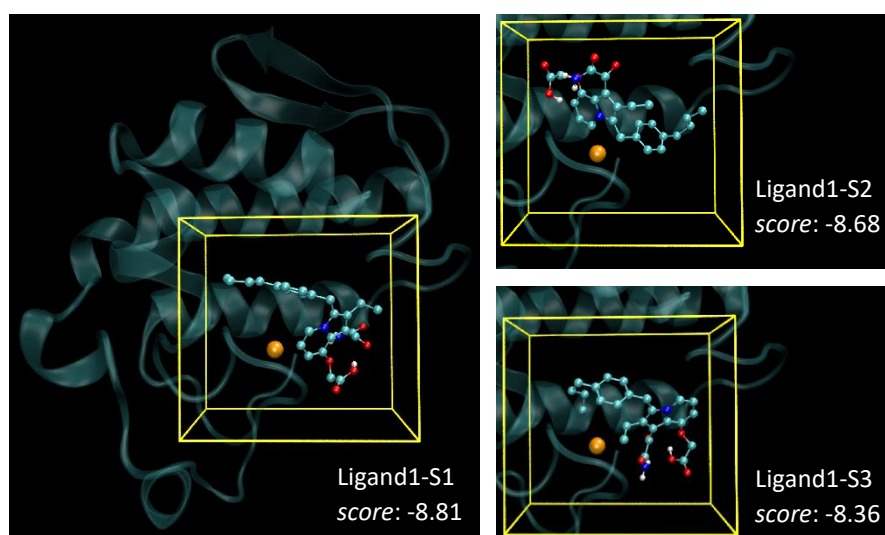
Um outro fator que deve ser tido em consideração é o facto de, aquando da ancoragem molecular não existirem águas no sistema. Ora, as zonas apolares do ligando tendem a procurar resíduos da mesma natureza, para que sejam rodeadas; enquanto zonas polares tendem a procurar átomos com os quais possam estabelecer pontes de hidrogénio. Por outras palavras, espera-se que, do ponto de vista bioquímico, a porção hidrofílica do inibidor esteja orientada para fora da cavidade da proteína (onde estariam moléculas de água) e a zona hidrofóbica para o seu interior. No entanto, na ausência de moléculas de solvente (retiradas para garantir que o espaço de ancoragem estaria livre), a procura deste tipo de orientação acaba por ser comprometida. A corroborar o supracitado, na fig.9 estão representadas três das geometrias do ligando com maior pontuação (-8.81, -8.68 e -8.36).

Ligand	Cluster	Score	Ki (M)
ligand1-S1	1	-7.48	3.29e-06
ligand1-S2	1	-6.85	9.51e-06
ligand1-S3	5	-6.23	2.71e-05
ligand1-S4	2	-5.34	1.22e-04
ligand1-S5	1	-5.24	1.44e-04

Ligand	Cluster	Score	Ki (M)
ligand1-S1	2	-8.81	3.48e-07
ligand1-S2	1	-8.68	4.33e-07
ligand1-S3	1	-8.36	7.44e-07
ligand1-S4	3	-8.22	9.42e-07
ligand1-S5	2	-8.15	1.06e-06
ligand1-S6	1	-7.49	3.23e-06

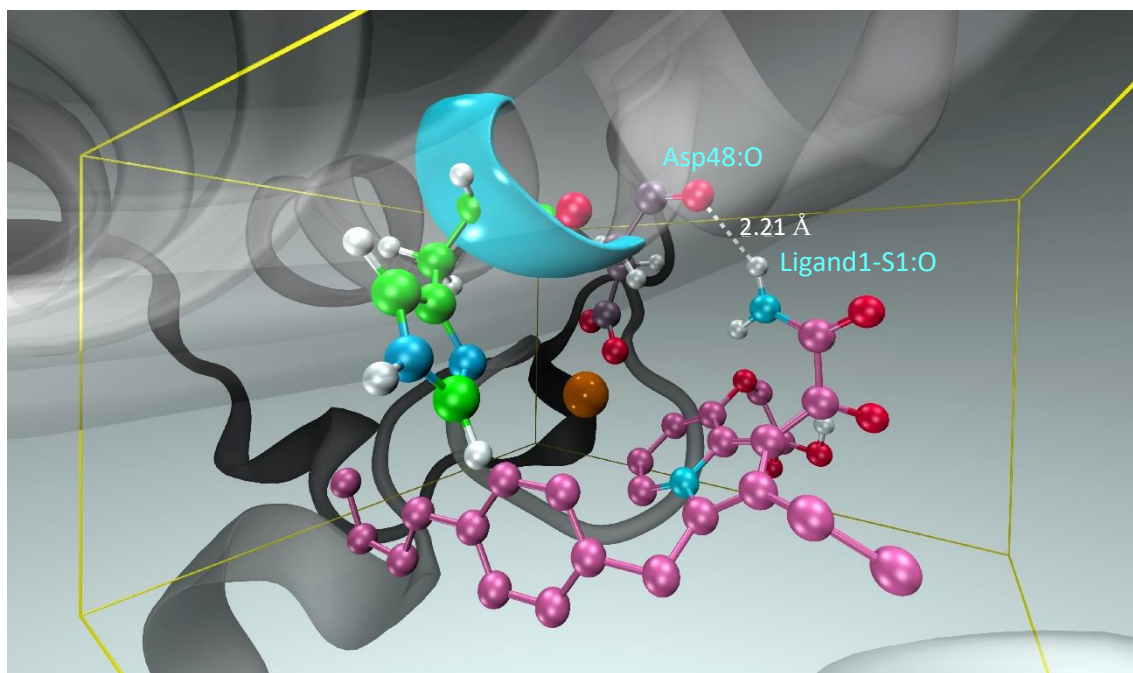
**Fig.8** – Tabelas de soluções de *Docking*. Geometrias ordenadas e agrupadas consoante *score*. Relevo para as poses melhor pontuadas, resultantes do primeiro (fig.7) (a) e do segundo (b) procedimento de ancoragem, após deslocar a caixa de pesquisa (fig.9 e fig.10).



**Fig.9** – Representação gráfica das geometrias de ligando com melhor pontuação, evidenciando a orientação relativa dos grupos polares e apolares da molécula. Ligando *[1-Aminooxalyl-3-benzyl-2-ethyl-indolizin-8-yloxy]-acetic acid* representado em CDK (consoante tipo de átomo). Estrutura da proteína sPLA2 representada em NewCartoon (a ciano). Íão  $\text{Ca}^{2+}$  a laranja. Área de *docking* a amarelo. *Software* de visualização molecular: VMD.

Apesar do ligand1-S2 aparecer à frente do ligand1-S3 na seriação, a orientação do último acaba por ser mais correta (cauda apolar no interior do *pocket* proteico, porção polar voltada para as águas a adicionar) – o que evidencia, novamente, a falibilidade das funções de pontuação, na simulação da realidade biológica.

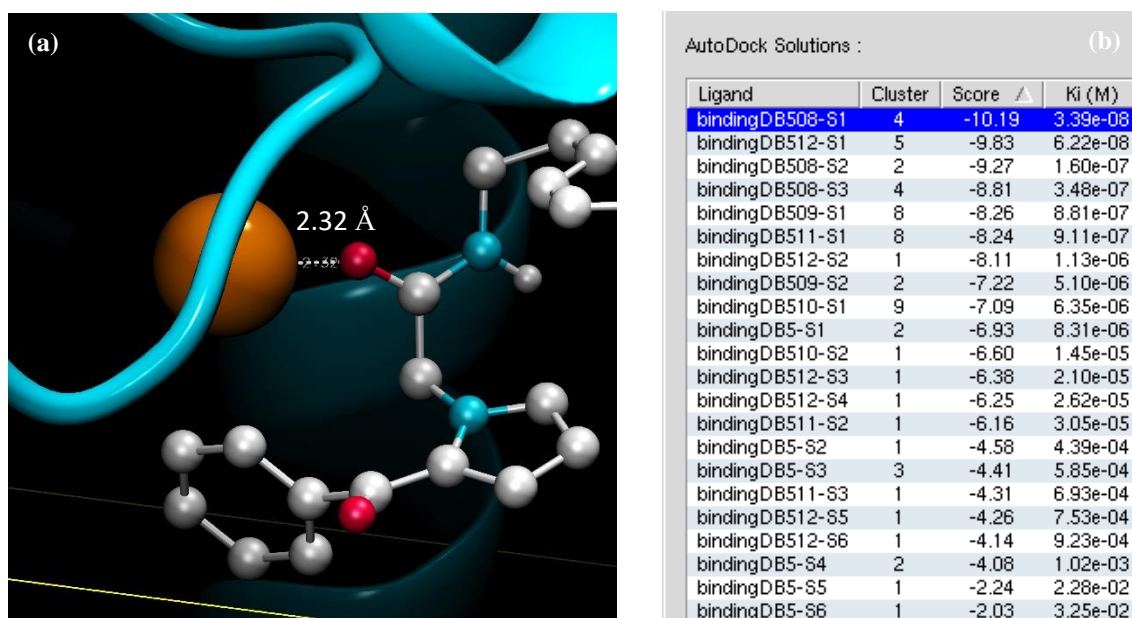
Na fig.10, é possível ver um *close-up* da conformação do inibidor com *score* mais baixo (-8.81) e a única ponte de hidrogénio estabelecida com a proteína. “À primeira ideia”, talvez fosse de esperar que a estrutura expectavelmente mais estável estabelecesse mais interações de hidrogénio. No entanto, e como referido anteriormente, a sua contribuição para o cálculo da energia de associação é reduzida e, na realidade, como são forças muito intensas, 1 ou 2 pontes são suficientes para uma boa inibição. Para além do mais, a constante de dissociação associada ( $K_i = 3.48 \times 10^{-7} \text{ M}$ ) é inferior ao limiar de grandeza considerado para um ligando ser promissor ( $10^{-6}$ ), i.e., ter uma forte afinidade.



**Fig.10** – Representação gráfica do centro de reação do sistema sPLA2–([1-Aminooxalyl-3-benzyl-2-ethyl-indolizin-8-yloxy]-acetic acid), com ênfase para a interação por ponte de hidrogénio Asp48:O e ligand1-S1:H. Estrutura da proteína representada em NewCartoon (a cinza), com destaque para os resíduos catalíticos His47 (Carbonos a verde) e Asp48 (Carbonos a cinza) – a coordenar o ião  $\text{Ca}^{2+}$  (a laranja). Geometria do ligando com maior pontuação representada em CDK (Carbonos a rosa). Átomos de Hidrogénio e Nitrogénio a vermelho e ciano, respetivamente. Área de *docking* a amarelo. *Software* de visualização molecular: VMD.

## Triagem Virtual

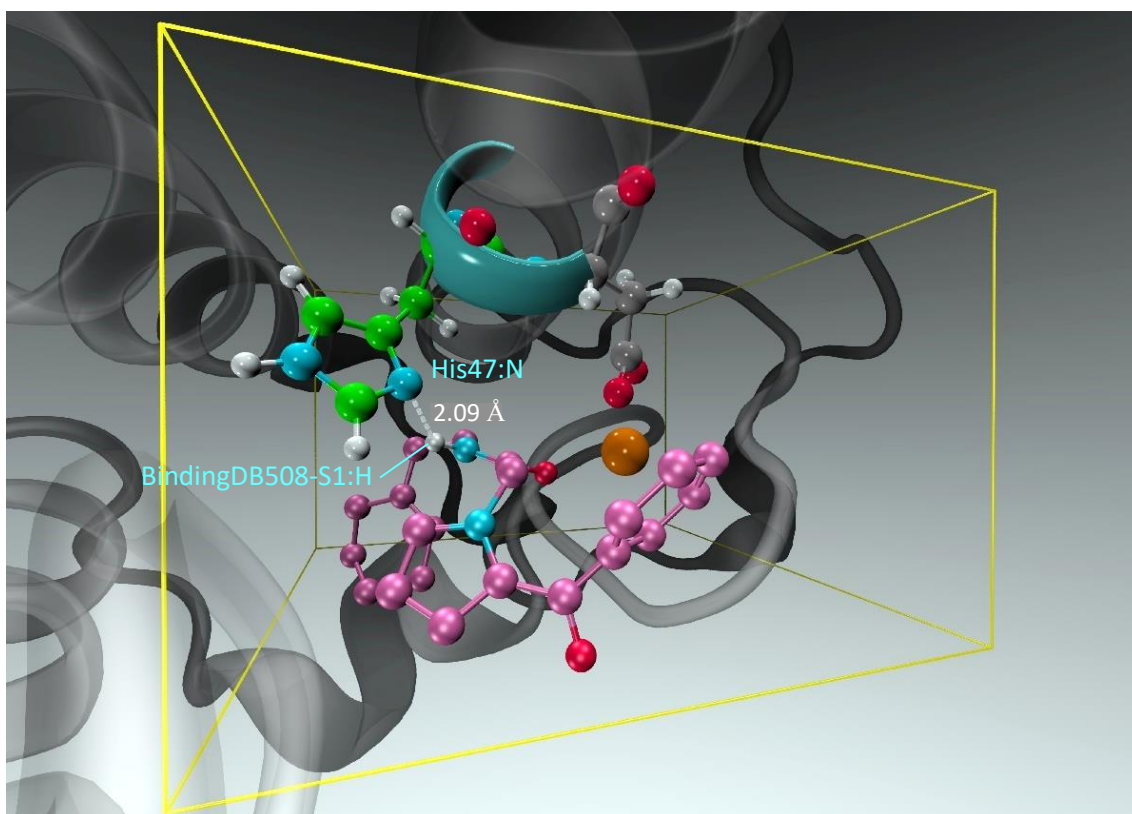
Para dar início ao protocolo de Triagem Virtual, acedeu-se à BindingDB – uma base de dados de afinidades de ligação (experimentalmente determinadas), relativas a uma ampla série de complexos proteína-ligando. As estruturas candidatas a *hit* para o recetor sPLA2 (com estrutura previamente refinada) foram ordenadas segundo  $IC_{50}$  e descarregadas, em formato \*.sdf. O ficheiro foi convertido para \*.mol2, tornando-se compatível com o *software* de cálculo, e filtrado, segundo o objetivo do trabalho: comparar os 5 compostos com menor capacidade de inibição aparente, com a molécula ancorada no protocolo anterior – o melhor inibidor para a hidrolase em estudo. A seleção foi feita com base na concentração de cada composto quando se atinge metade da resposta inibitória – o tal  $IC_{50}$ . Quanto maior, menor a afinidade, menor a inibição e, portanto, maior a probabilidade das moléculas encontradas não serem, na verdade, ligandos. Por último, procedeu-se a ancoragem de cada composto na cavidade da proteína. Na fig.9, encontram-se tabeladas as poses obtidas e representada uma das melhores geometrias de ligando, com ênfase para a interação eletroestática estabelecida entre o  $Ca^{2+}$  e o átomo de O, *perfeitamente* orientado para o ião.



**Fig.11** – (a) Geometria de um dos ligandos com melhor pontuação (-10.19), com ênfase para a ponte eletroestática, estabelecida entre o Oxigénio (a vermelho) e o ião Cálcio (a laranja). (b) Setação das poses de *Virtual Screening* encontradas.

Notavelmente, num protocolo de *Virtual Screening* ideal, começar-se-ia por ir ao PDB, retirar as estruturas proteína-ligando, removê-lo e executar o *Docking*, de forma a verificar se as dimensões escolhidas para a área de pesquisa estão de acordo com o esperado (adicionando o ligando que foi retirado). Assim, seria possível saber, *à priori*, se a função de *score* está correta e podia prosseguir-se com a segurança de que a caixa de ancoragem não era muito pequena, nem muito grande. No entanto, no âmbito deste protocolo, não se fez isso e os resultados obtidos revelaram que a caixa escolhida foi muito pequena. Nesta situação, o composto é forçado a dobrar-se para satisfazer as dimensões ou, quando não possível, a ficar fora da caixa. Quando isso acontece, é atribuída uma penalização. Precisamente, por isto – pelas dimensões e localização da caixa serem altamente influentes na geometria e distribuição espacial do possível ligando –, nem sempre o menor *score* corresponde à molécula com maior potencial de inibição (*menor*\* IC<sub>50</sub>). Desta feita, entende-se o porquê do BindingDB5-S1 não apresentar a melhor pontuação (-6.93 em comparação a -10.19). Outro fator que condiciona esta análise é o facto de \*nem todos os IC<sub>50</sub> serem determinados nas mesmas condições experimentais e, como tal, acabam por não ser equiparáveis.

Por outro lado, comparativamente aos resultados obtidos no protocolo anterior, pode verificar-se que aos primeiros 9 compostos no *ranking* (agrupados em *clusters* de igual pontuação) correspondem K<sub>i</sub> na ordem dos  $10^{-8} \text{ M} < 10^{-7}$ , obtido para o que se sabe ser o melhor inibidor, e  $< 10^{-6}$ , equivalente ao limite superior para a identificação de um *hit*. Desta feita, e invocando os constrangimentos anteriores, pode concluir-se que moléculas que se esperava apresentarem uma menor afinidade com o recetor, na verdade, acabaram por se revelar os inibidores mais fortes. Isto não quer, necessariamente, dizer que a capacidade de seriação do *Virtual Screening* executado não foi a ideal, mas sim, que o procedimento não é exato. Por exemplo, dada a dimensão da biblioteca de compostos disponível, é comum que potenciais *hits* tenham sido descarregados com estados de protonação diferentes daqueles que apresentam quando em associação. O mesmo pode acontecer com resíduos ionizáveis da proteína. Ora, grupos químicos diferentes podem originar diferentes formas de ligação e, conseqüentemente, diferentes geometrias. Na fig. 12, encontra-se representada a conformação do composto com menor *score* (-10.19), enfatizando – à semelhança do *docking* anterior – a única ponte de hidrogénio estabelecida – precisamente, com um resíduo ionizável (His47).



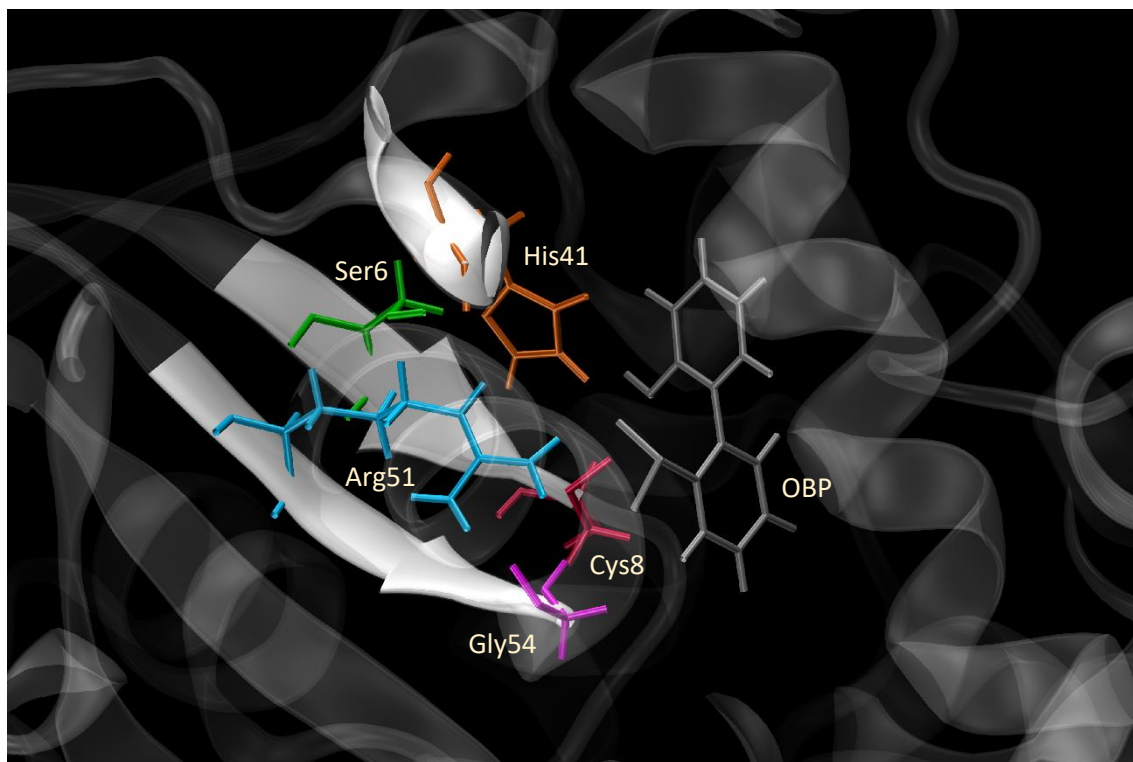
**Fig.12** – Representação gráfica do centro de reação do sistema sPLA2 com ênfase para a interação por ponte de hidrogénio His47:N -- H da estrutura BindingDB508-S1. Estrutura da proteína representada em NewCartoon (a cinza), com destaque para os resíduos catalíticos His47 (Carbonos a verde) e Asp48 (Carbonos a cinza) – a coordenar o ião  $\text{Ca}^{2+}$  (a laranja). Geometria do ligando com maior pontuação representada em CDK (Carbonos a rosa). Átomos de Hidrogénio e Nitrogénio a vermelho e ciano, respetivamente. Área de *docking* a amarelo. *Software* de visualização molecular: VMD.

## Métodos QM/MM

Antes de iniciar o protocolo de QM/MM, acedeu-se ao VMD e carregou-se o ficheiro da estrutura reactante (Reactant.pdb). O substrato (OBP) e os resíduos do centro ativo (Ser6, Cys8, His41, Arg51 e Gly54) foram identificados (fig.13).

### Definição das regiões QM e MM

De seguida, com recurso à extensão MolUp, definiu-se que átomos seriam descritos por QM e MM. Para tal, carregou-se o ficheiro QMMM\_prep.com – um arquivo previamente fornecido, com informação sobre o sistema molecular (como coordenadas cartesianas, nome e numeração dos resíduos) e devidas instruções para os cálculos de energia. Os átomos envolvidos na reação química foram destacados para constituir a *High Layer* (OBP + Ser6, Cys8, His41, Arg51 e Gly54), mantendo o resto do sistema fixo (porção de proteína sobranante + solvente). A *Low Layer* foi “descongelada” e as instruções necessárias para a minimização energética do sistema introduzidas no programa.

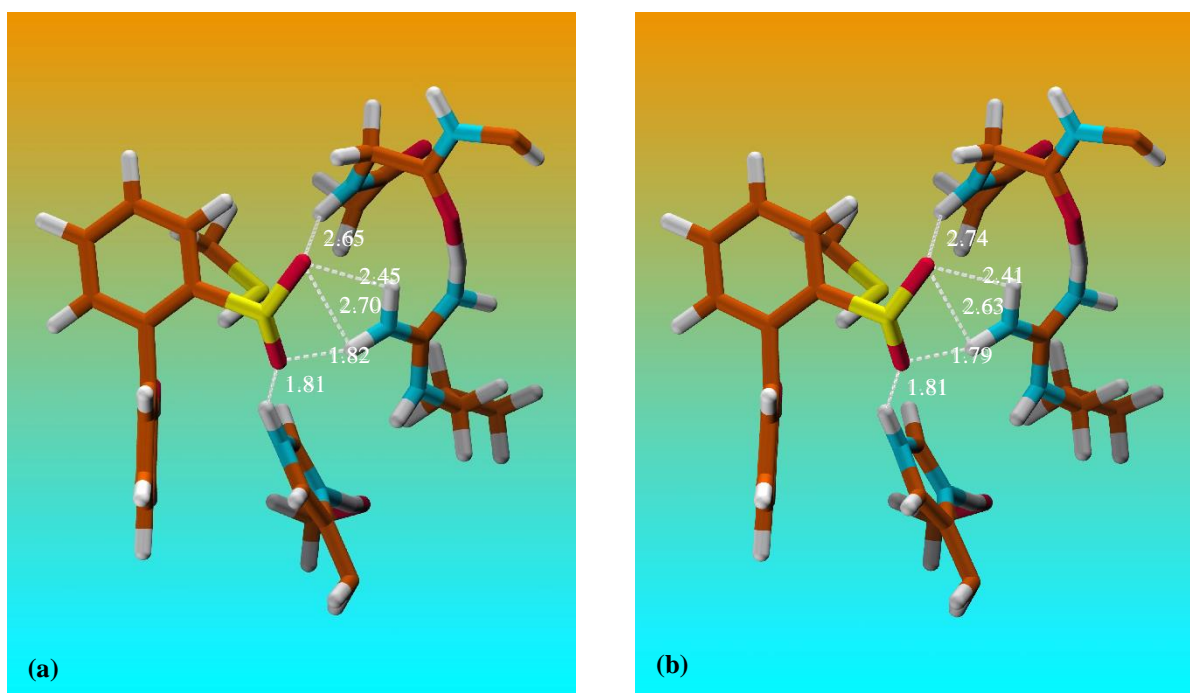


**Fig.13** – Representação gráfica do cerne reacional do sistema DszB – OBP. Estrutura da proteína representada em NewCartoon (a cinza), com destaque para os resíduos do centro ativo Ser6 (verde), Cys8 (rosa), His41 (laranja), Arg51 (azul) e Gly54 (magenta). Substrato representado em *lines* (a cinza). PDB ID: 2DE3. *Software* de visualização molecular: VMD.



## Otimização da geometria do estado reactante

Por ser um passo moroso, o *output* dos cálculos QM/MM foi antecipadamente disponibilizado (*initial\_opt.log*). As pontes de hidrogénio foram preservadas ao longo de toda a minimização. Estas estabelecem-se entre pares de átomos eletronegativos (dador e aceitador de eletrões) e são necessárias para a ancoragem estável do substrato, no local de reação. 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 protão e o seu aceitador. Na fig.14 encontram-se representadas potenciais pontes de hidrogénio, estabelecidas na *High layer* do sistema não minimizado e num estado intermédio da otimização. *Spot the differences!* São praticamente inexistentes.



**Fig.16** – Comparação visual de possíveis pontes de hidrogénio estabelecidas entre o sistema inicial não minimizado (a) e a sexta conformação otimizada (b) – onde, apesar de reduzidas, as alterações nas distâncias de interação (em Å) foram as mais significativas no decorrer dos cálculos QM/MM. *High Layers* representadas em *licorice* – átomos de C, a laranja; S, a amarelo; N, a ciano; O, a vermelho; e H, a branco. *Software de visualização molecular*: VMD.

### Cálculo do perfil de energia potencial e identificação da estrutura correspondente ao estado de transição do primeiro passo do mecanismo de reação

Para identificação da estrutura correspondente ao estado de transição, procedeu-se ao cálculo do PEP (em inglês, *Potential Energy Profile*) – uma função que retorna o valor de energia total de um sistema, a partir de uma única coordenada de reação. No contexto do mecanismo, corresponde à distância entre o H eletrofílico e o C-SO<sub>2</sub> (gradualmente, aproximados, em incrementos de 0.1 Å). Os centros de catálise correspondentes aos estados de reação mais relevantes (reagente, transição e produto) encontram-se representados na fig.16, deduzidos a partir da informação das fig.15 e 17.

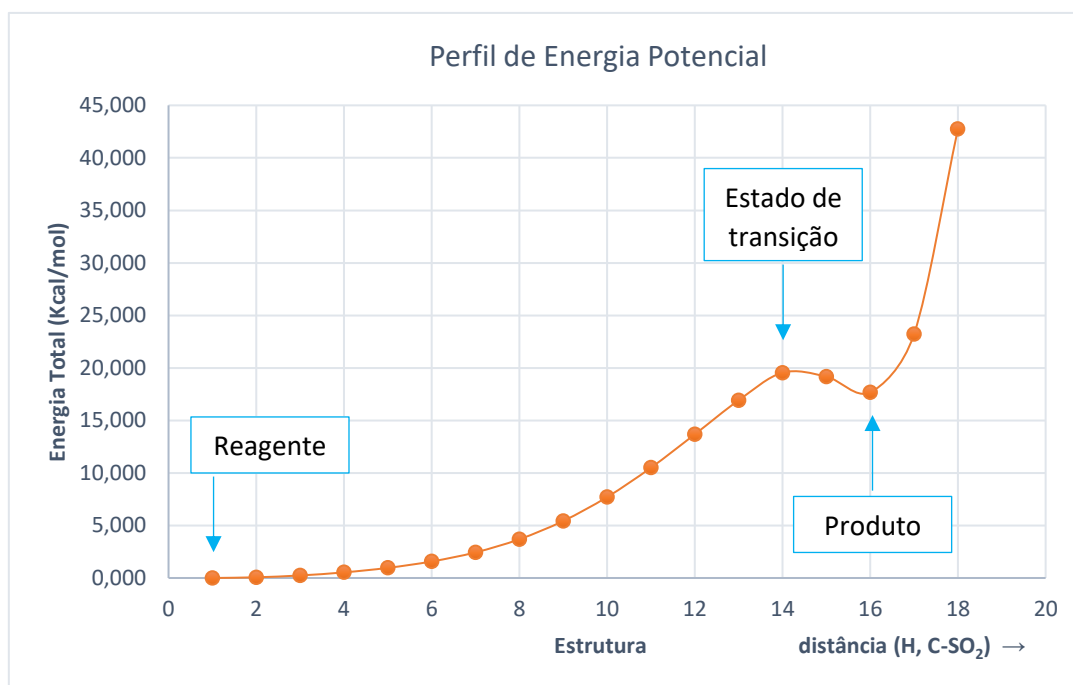
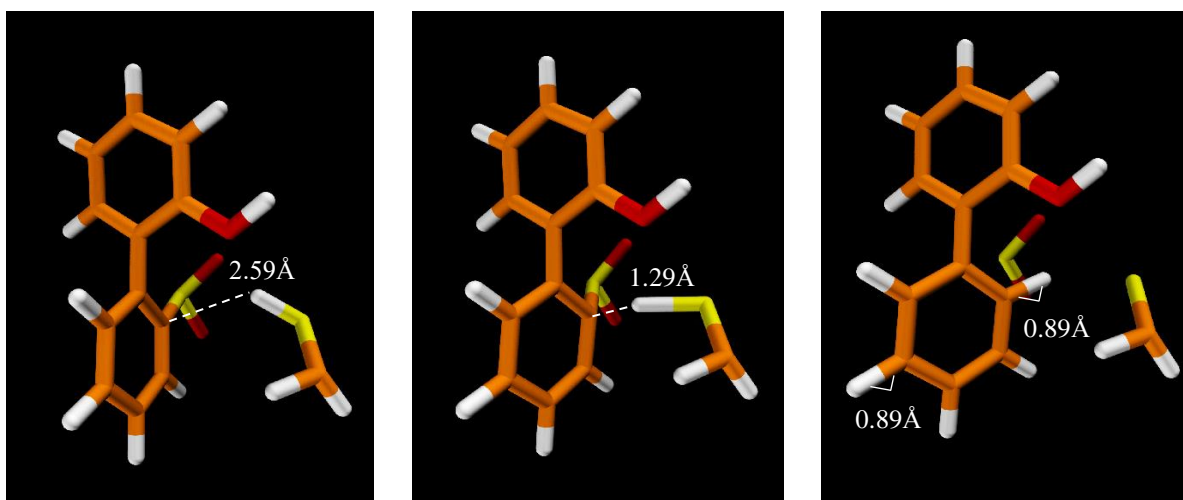


Fig.15 – Representação gráfica do Perfil de Energia Potencial.

Dos dados apresentados, concluiu-se que a Energia Livre de Ativação calculada ( $\approx 19.5$  kcal/mol, correspondente ao estado intermediário) é muito próxima do valor experimental (19.0 kcal/mol) – o que faz sentido, visto que a reação é dominada pela variação da entalpia.





**Fig.16** – Representação gráfica dos sistemas 1, 14 e 16, correspondentes aos estados mais relevantes da reação química: reagente, transição e produto, respetivamente. Nota: as ligações S–H e C–SO<sub>2</sub> representadas no estado transitório, correspondem, na verdade, a ligações intermediárias (em quebra e em formação). Distâncias entre o H eletrofílico e C do anel assinaladas a branco.

Structure	Total Energy (Hartree)	Total Energy (Kcal/mol)
1	-2741,855	0,000
2	-2741,854	0,067
3	-2741,854	0,243
4	-2741,854	0,534
5	-2741,853	0,963
6	-2741,852	1,574
7	-2741,851	2,440
8	-2741,849	3,691
9	-2741,846	5,429
10	-2741,842	7,704
11	-2741,838	10,501
12	-2741,833	13,678
13	-2741,828	16,920
14	-2741,823	19,546
15	-2741,824	19,171
16	-2741,826	17,679
17	-2741,818	23,224
18	-2741,786	42,754

$E(\text{Hartree}) * 627.5094 = E(\text{kcal/mol})$	$E(\text{Structure} = 1) \text{ em kcal/mol}$
627,5094	-1720539,484

**Fig.17** – Dados de energia tabelados e respetivas constantes de conversão. Energia de ativação salientada a ciano.

## **Bibliografia**

- (1) KINI, R. M. Excitement ahead: structure, function and mechanism of snake venom phospholipase A(2) enzymes. *Toxicon*, v. 42, n. 8, p. 827- 840, 2003.
- (2) MURAKAMI, M. T.; ARNI, R. K. A structure based model for liposome disruption and the role of catalytic activity in myotoxic phospholipase A(2)S. *Toxicon*, v. 42, n. 8, p. 903-913, 2003.