U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

New Antidotes for Viper Venom sPLA$_2$ Toxin

Mariana Passos Monteiro Mascarenhas Contreiras

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Mestrado em Bioinformática e Biologia Computacional

2021

MSc

2.º CICLO

FCUP
2021

U.PORTO

New Antidotes for Viper Venom sPLA$_2$ Toxin

Mariana Passos Monteiro Mascarenhas Contreiras

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

Agradecimentos

Quero em primeiro lugar agradecer ao Professor Doutor Pedro Alexandrino Fernandes, pela orientação que me ofereceu durante este ano. Todos os seus conselhos foram cruciais na elaboração desta dissertação, assim como a extraordinária dedicação que demonstrou.

À Professora Doutora Maria João Ramos por me ter aceitado no grupo de Química Teórica e Computacional, e pelos conhecimentos que partilhou comigo.

À Matilde, por todo o incansável apoio e indispensável ajuda. Tornaste melhor este ano atípico, melhor. Foi um prazer trabalhar contigo.

À Ana Filipa, por me entender com uma telepatia sobrenatural. A tua amizade é das melhores coisas que me acompanha de Coimbra.

À Denise e à Maria Eduarda, por estarem sempre perto mesmo a milhares de quilómetros de distância. Tenho muitas saudades vossas.

Ao Pedro, por ser a minha maior inspiração e mais longa gargalhada. Considero-me para sempre honrada de te chamar meu irmão.

E finalmente, aos meus pais, um simples “obrigada” nunca será suficiente, mas ainda assim: obrigada, devo-vos tudo o que sou e tenho.

Abstract

Snakebite envenoming is a clinical emergency. It is estimated to occur in 1.8–2.7 million people annually worldwide, resulting in over 400 thousand amputations and 81 to 138 thousand deaths per year. Despite its death toll, snakebite envenoming is often unreported. Yet, these gross estimations surpass the consequences of some of the world's recognised neglected tropical diseases (NTDs), including schistosomiasis and leishmaniasis.

Due to lack of infra-structures and protection, snakebite envenoming greatly impacts underdeveloped rural communities, while going unnoticed in higher-income countries of North America and Europe. As a consequence, snakebite envenoming has historically received little attention from funding bodies, public health authorities, the pharmaceutical industry and health advocacy groups, thereby impairing the development of effective interventions to reduce snakebite death toll.

Snake venoms are complex cocktails. Comprised of a unique mixture of enzymes and other proteins, snake venom secreted phospholipases A_2s stand out due to their presence in 95 % of the two most medically important snake families: elapids and vipers; and because they are not effectively neutralized by current therapies. *Bothrops Asper* (commonly known as Terciopelo) is one of the deadliest viper snake species in South America. Myotoxin-1 (MT-I) is a basic phospholipase A_2 (PLA_2s) secreted by *B. asper* that is responsible for binding to and disrupting the integrity of the plasma membrane of muscle fibers, and it's also the first to have been isolated, hence the name. Phospholipases A_2 catalyse the hydrolysis of the fatty acid ester in the 2-position of 3-sn-phospholipids and 4-Nitro-3-(octanoyloxy)benzoic acid (NOBA) is a synthetic substrate that is hydrolyzed by this enzyme at comparable rates.

In the present study, we built a model of MT-I using NOBA as a substrate analogue. MT-I is a basic homodimeric secreted PLA_2, with a calcium ion in each of its active site. Based on its UniProt sequence (UniProtKB ID: P20474), the corresponding 122 residues X-ray crystal structure (PDB ID: 5TFV, 2.54 Å resolution) was identified. However, it does not comprise the calcium ion needed for catalysis, affecting the conformation of its conserved calcium-coordinating loop. On a first stage, calcium ions were modelled into the crystal structure, and classical molecular dynamics (MD) were carried out. Analysis of the MD trajectory indicated that further modelling was required for the calcium coordinating loop to effectively accommodate the ion.

Through Swiss-Model homology modelling server, template structures based on the MT-I sequence were retrieved. As MT-I's are homodimers, homodimeric structures with conserved calcium coordinating loops were preferred. Several distinct models were generated and inspect through structure alignment. Analysis of i) calcium coordinating loop, ii) relative orientation of monomers in the homodimer, iii) conformation of active site residues and iv) presence of ligands at the active site indicated that the crystal structure PDB ID:1FXF is the best scoring template. The substrate NOBA was modelled in the resulting homology model and subjected to MD for further analysis.

These findings can be very useful for future studies on MT-I's catalytic mechanism and possible drug design, joining the efforts of creating low-cost, thermostable antivenoms to be delivered to those most in need, hopefully having an impact in these communities.

Keywords: Snake venom, Neglected Tropical Diseases (NTDs), *Bothrops Asper*, phospholipase A_2 (PLA_2), myotoxin 1 (MT-I), 4-Nitro-3-(octanoyloxy)benzoic acid (NOBA), homology modeling, molecular dynamics (MD).

Resumo

O envenenamento por mordedura de cobra é uma emergência clínica. Estima-se que ocorra em 1,8–2,7 milhões de pessoas anualmente em todo o mundo, resultando em mais de 400 mil amputações e 81 a 138 mil mortes por ano. Apesar da sua taxa de mortalidade, o envenenamento por mordedura de cobra muitas vezes não é reportado. Ainda assim, estas estimativas defetíveis superam as consequências de algumas das doenças tropicais negligenciadas (NTDs) mundialmente reconhecidas, como a esquistossomose e a leishmaniose.

Devido à falta de infraestruturas e auxílio, o envenenamento por mordedura de cobra tem um grande impacto nas comunidades rurais pouco desenvolvidas, passando despercebido em países mais privilegiados da América do Norte e Europa. Consequentemente, o envenenamento por mordedura de cobra sempre recebeu pouca atenção de entidades financiadoras, autoridades de saúde pública, indústria farmacêutica e grupos de apoio à saúde; prejudicando o desenvolvimento de intervenções eficazes que reduzam o número de mortes por mordedura de cobra de cobra.

Os venenos de cobra são misturas complexas. Compostos por combinações específicas de enzimas e outras proteínas, as fosfolipases A_2s (PLA_2s) secretadas no veneno de cobras e destacam-se por estarem presentes em 95 % das duas famílias de cobras mais importantes do ponto de vista médico: elapídeos e víboras; e porque não serem neutralizadas eficazmente por terapias atuais. *Bothrops Asper* (comumente conhecido como Terciopelo) é uma das espécies de cobra víbora mais mortais da América do Sul. A miotoxina-1 (MT-I) é uma fosfolipase A_2 básica secretada por *B. asper* que é responsável por se ligar e quebrar a integridade da membrana plasmática das fibras musculares, e também é a primeira a ter sido isolada, daí o nome. As fosfolipases A_2 catalisam a hidrólise do éster de ácido gordo na posição 2 de 3-sn-fosfolipídios e o ácido 4-nitro-3-(octanoiloxi) benzóico (NOBA) é um substrato sintético que é hidrolisado por esta enzima a velocidades semelhantes.

Neste estudo, construímos um modelo de MT-I usando NOBA como substrato análogo. MT-I é uma PLA_2 secretada, que é homodimérica e básica, com um ião cálcio em cada um de seus centros ativos. Com base na sua sequência UniProt (UniProtKB ID: P20474), identificou-se a correspondente estrutura cristalina de raios-X de 122 resíduos (PDB ID: 5TFV, resolução de 2,54 Å. No entanto, esta não possui o ião cálcio necessário

para a catálise, afetando a conformação do loop conservado, onde o cálcio está coordenado. Numa primeira etapa, os iões de cálcio foram modelados na estrutura do cristal e uma dinâmica molecular clássica (MD) foi realizada. A análise da trajetória da MD indicou que seria necessária modelação adicional para que o loop de coordenação do cálcio conseguisse acomodar o ião.

Com recurso ao servidor de modelação de homologia Swiss-Model, estruturas "template" baseadas na sequência de MT-I foram obtidas. Como as MT-Is são homodímeros, houve preferência por estruturas homodiméricas com loops de cálcio conservados. Vários modelos distintos foram gerados e inspecionados por meio do alinhamento das estruturas. A análise de i) loop de coordenação de cálcio, ii) orientação relativa de monómeros no homodímero, iii) conformação dos resíduos de centro ativo e iv) presença de ligantes no centro ativo demonstrou que a estrutura cristalina PDB ID: 1FXF é o melhor template mais adequado. O substrato NOBA foi inserido no modelo de homologia resultante e submetido a MD para análise adicional.

Estas descobertas podem ser muito produtivas para estudos futuros sobre o mecanismo catalítico de MT-I e possível drug design, juntando-se aos esforços para a criação de antivenenos termoestáveis de baixo custo para serem entregues aos que mais necessitam, com esperança de ter um impacto nessas comunidades.

Palavras-chave: Veneno de Cobra, Doenças Tropicais Negligenciadas (NTDs), *Bothrops Asper*, fosfolipase A_2 (PLA_2), miotoxina 1 (MT-I), ácido 4-nitro-3-(octanoiloxi) benzoico (NOBA), modelação por homologia, dinâmica molecular (MD).

Table of contents

1. Introduction 14

1.1. Why is This a Problem? 14

1.2. Current Solutions, Latest Developments and Challenges 16

1.3. Composition of Venom and Effects 17

1.4. *Bothrops Asper* 18

1.5. Phospholipases A_2 20

1.6. Myotoxin I 25

1.7. Catalytic Mechanism 25

1.8. Dissertation Outline 28

2. Methodology 29

2.1. Preparation of Protein Structure 29

2.2. Homology Modeling 31

2.3. Energy Minimization 32

2.4. Molecular Mechanics 32

2.4.1. Force Fields 33

2.4.1.1. AMBER Force Field 34

2.4.2. Molecular Dynamics 36

3. Results and Discussion 38

3.1. Initial X-Ray Structure 38

3.2. Homology Modeling 39

3.2.1. Choice of Template 39

3.2.2. Model Building 44

3.2.3. Protonation states and Waters 45

3.3. Molecular Dynamics 46

3.3.1. Energy Minimization/Relaxation 47

3.3.2. Equilibration 47

3.3.3. Production..... 48

4. Conclusions and future work..... 55

5. References 56

List of equations

Equation 1 - Total potential energy..... 32

Equation 2 - Time-independent Schrödinger equation 33

Equation 3 - Bonded term for potential energy in AMBER..... 34

Equation 4 - Nonbonded term for potential energy in AMBER 35

Equation 5 - Differential equation that describes the motion of a particle 35

List of figures

Fig. 1 - Geographical estimations of snakebite envenomings and deaths 14

Fig. 2 - Photograph of B. asper 19

Fig. 3 - Hydrolysis of the sn-2 acyl bond of glycerophospholipids..... 21

Fig. 4 - Comparison of the three-dimensional structure of a group I PLA$_2$ and a group II PLA$_2$ 22

Fig. 5 - A view of the structure of a Asp49 PLA$_2$ 23

Fig. 6 - Structures of the PLA$_2$ substrate NOBA 24

Fig. 7 - Scheme representing sPLA$_2$ interfacial catalytic turnover..... 26

Fig. 8 - Two proposed mechanisms of PLA$_2$ 27

Fig. 9 - Two views of 5TFV 30

Fig. 10 - Representation of intramolecular and intermolecular interactions 35

Fig. 11 - Structure of 1JIA, a basic PLA$_2$ 40

Fig. 12 - Ca^{2+}-coordinating loop in 1JIA 41

Fig. 13 - Structure of 5TF and 1JIA superimposed..... 41

Fig. 14 - Structure of 1FXF 42

Fig. 15 - Structure of the 1FXF inhibitor MJ33 42

Fig. 16 - Heptacoordination of Ca^{2+} in 1FXF..... 43

Fig. 17 - Structure of 5TF and 1FXF superimposed..... 43

Fig. 18 - Loop comparison between 5TF, 1JIA and 1FXF..... 45

Fig. 19 - RMSd of protein, throughout the initial 140 ns MD simulation..... 49

Fig. 20 - RMSd of NOBA, throughout the initial 140 ns MD simulation..... 49

Fig. 21 - RMSF plot of the initial 140 ns MD simulation..... 50

Fig. 22 - Initial MT-I conformation for MD..... 50

Fig. 23 - Heptacoordination of Ca^{2+} in minimized MT-I structure..... 51

Fig. 24 - Coordination of Ca^{2+} in minimized MT-I structure..... 52

Fig. 25 - Coordination of Ca^{2+} in MT-I structure after equilibration 52

Fig. 26 - Coordination of Ca^{2+} in MT-I structure after equilibration 53

Fig. 27 - RMSd of all protein, throughout the final 122 ns MD 53

Fig. 28 - RMSd of ligand NOBA, throughout the final 122 ns MD 54

Fig. 29 - Superimposition of visualizations belonging to initial model, minimization, equilibration and production..... 54

List of tables

Table 1 - Percentages of different protein families in the venoms of Caribbean and Pacific B. asper populations of Costa Rica 20

Table 2 - Most common force fields used in the simulations of different biomolecules..... 34

Table 3 - Model evaluation measurements for 1FXF template..... 44

Table 4 - Parameters for the MD simulations..... 47

Table 5 - Steps performed in minimization stage 48

List of abbreviations

3FTx - Three-finger Toxin

AMBER - Assisted Model Building with Energy Refinement

BLAST - Basic Local Alignment Search Tool

CHARMM - Chemistry at Harvard Macromolecular Mechanics

DNA - Deoxyribonucleic acid

FF - Force Fields

GAFF - General Amber Force Field

GMQE - Global Model Quality Estimate

i-face - Interface Binding Surface

LAAO - L-amino oxidase

MD - Molecular Dynamics

MDiS - Membrane-disrupting site

MDoS - Membrane-docking site

MM - Molecular Mechanics

MT-I - Myotoxin I

MT- II - Myotoxin II

NMR - Nuclear Magnetic Resonance

NOBA - 4-Nitro-3-(octanoyloxy)benzoic acid

NPT - Constant Temperature, Constant Pressure Ensemble

NTD - Neglected Tropical Disease

NVE - Constant Energy, Constant Volume Ensemble

NVT - Constant Temperature, Constant Volume Ensemble

PDB - Protein Data Bank

PLA$_2$ - Phospholipase A$_2$

QM - Quantum Mechanics

QSQE - Quaternary Structure Quality Estimate

RDF - Radial Distribution Function

RMSd - Root-Mean-Square deviation

RMSF - Root-Mean-Square Fluctuations

RNA - Ribonucleic acid

$sPLA_2$ - Secreted Phospholipase A_2

SVMP - Metalloprotease

SVSP - Serine Protease

VMD - Visual Molecular Dynamics software

1. Introduction

1.1. Why is this a problem?

Snakebite envenoming poses clinically challenging emergencies as a result of their potentially rapid lethality. It is considered a category A neglected tropical disease (NTD), a critical incentive towards recognition and allocation of resources, antivenom development and critical investor funding.[1,2]

Snakebite envenoming is estimated to occur in 1.8–2.7 million people annually worldwide (out of the approximate 5·4–5·5 million people that are bitten), resulting in 81 to 138 thousand deaths per year and over 400 thousand amputations and permanently disabled victims.[1,2,3] Half of the estimated deaths from snakebite, per year, derive from India, but since data from sub-Saharan Africa is scattered and restricted to the few medical reports provided by hospitals, these numbers might be severe underestimations of the reality.[4] Another factor contributing to the underreporting of deaths is that many of the victims who seek out traditional methods of treatment may not be included in the final reports.[5] Thus, even though snakebite envenoming is often unreported, these gross estimations surpass the yearly mortality of some of the world's most recognized NTDs, including schistosomiasis and leishmaniasis.[1,2]

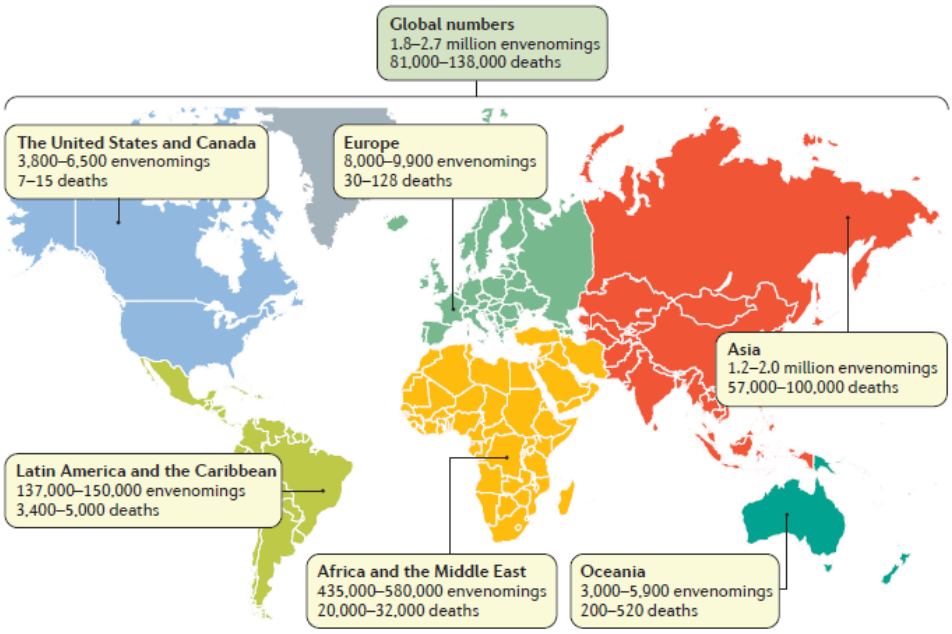


Fig. 1 - Geographical estimations of snakebite envenomings and deaths. From [1].

The effects of snakebite envenoming are extensive and range from local tissue damage and bleeding to respiratory paralysis, acute kidney injury, rhabdomyolysis, cardiotoxicity and thrombosis.[1] The social consequences are also devastating: in Africa, a farmer might suffer such terrible injuries that he's rendered fit only to beg, or the stigma associated with scars, and other disfigurements, may wreck a girl's marriageable status.

Snakes are cold-blooded animals and abundant in warmer climates. This means the most affected regions by snakebite are tropical countries of the developing world. On the other hand, people of wealthier countries have far fewer encounters with venomous snakes and are therefore uninformed of the significance of this problem and the degree it endangers other communities. As a consequence, snakebite envenoming has hardly received the financial backing and consideration it deserves from funding bodies and public health authorities, thereby impairing the development of effective interventions to reduce the astronomical morbidity and mortality rates.[1]

A key component of neglected tropical diseases is their incessant link to poverty, as is the case of snakebite. In fact, no other disease is so directly linked to poverty-stricken and agro-based communities who suffer from it as a daily occupational hazard. The survival of many of these people is reliant on their inexpensive, non-mechanised farming techniques and to add insult to injury it is these methods that leave them vulnerable to being bitten in their feet, legs and hands.[4]

Snakebite presents a higher incidence and mortality in the 10-40 years old age group, which is composed of the most productive members of rural communities. Children are particularly at risk, by either helping in farming chores or playing in the fields. To make matters worse, contact between snakes and humans happens frequently during the wet season, where agricultural activity coincides with the snakes' breeding season. Snakebite changes not only the lives of those directly affected but also the families and communities they are a part of. Good physical health is key to those devoted to agricultural activities, making the consequent disability or even death of victims detrimental to the already precarious economic state of entire households. In addition, it's in places most economically unstable that medical assistance is inordinately priced, making it even harder to break the cycle of poverty.[1]

1.2. Current solutions, latest developments and challenges

Snake bite is preventable and treatable through access to antivenoms and by training healthcare providers, and by community education to reduce occupational and environmental risk, however, lack of funding hinders progress. As snakebite envenoming is not infectious and it cannot be eradicated by vaccination, it's not an attractive target for funders as vector-borne diseases.[2,3]

Currently, the only effective treatment for snakebite envenoming is antibody-based therapies. Large, domesticated animals, mostly horses, are inoculated with venom for a period of months or even years. Horse plasma rich in specific antibodies is then purified. As immunoglobulins are incredibly specific, they neutralize only the toxins in the venoms used in the immunization process - this means identification of snake species is critical for choosing the correct antivenom. To tackle this issue, the manufacture of polyvalent antivenoms, that contain antibodies effective against venom from multiple species, is standard. Nonetheless, these therapies are expensive and require refrigeration, and may lead to severe side effects: the immune system can generate an exacerbated response and the patient might go into anaphylactic shock.

Other difficulties such as the uncertainty about the species identity and age, which influences the composition and quantity of venom injected, make already difficult decisions much harder. The fear caused by trauma and followed rushed assistance can also cause other misleading symptoms like vomiting and tachycardia.[1]

However, current protocols are not directed to specific venom proteins, and do not consider variations in the molecular mass and immunogenicity of different venom components. Instead, antivenoms include redundant antibodies to non-toxic molecules, effectively 'diluting' potent neutralizing immunoglobulins in the mix, which leads to the need to administer high volumes of antivenom (20-200 ml). As an expected consequence, the victim's immune system responds extensively, leading to anaphylaxis and other early and late adverse effects.[5,6]

Another element to consider is the price and availability of antivenom. Antivenom production is government-subsidised in many Latin American countries, and its supply is free to the patient leading to a low mortality rate due to snakebite in Latin America (2,300 deaths) despite the high incidence (129,000).[4] However, not all countries whose population is affected by snakebite envenoming have the same resources, as is the case of India, Nigeria and Nepal.[5]

1.3. Composition of venom and effects

Of the more than 100,000 existing venomous species, roughly 725 are snakes and they make up for 80 % of deaths by envenomation world-wide. Snake venoms are complex cocktails of typically 30 to over 100 toxins (mostly proteins) that disrupt the activity of vital enzymes, receptors, and ion channels of the prey, destroying the central and peripheral nervous systems, the cardiovascular and circulatory systems and also the neuromuscular system. Besides prey immobilization, venoms also carry out digestive functions. Some of these proteins exhibit enzymatic activities, whereas several others are non-catalytically active proteins and polypeptides. Snake diets are very diverse and include snails, fishes, frogs, toads, lizards, chickens, mice, rats and even other snakes.[7]

It is hypothesized that because of the unpredictability of the kind of prey snakes might face, they've evolved to have a variety of toxins at their disposal. In spite of the extent of different components, venom proteins belong to only a few major families. These proteins need to retain activity across a wide range of temperatures (between the snake and the prey), and are in general more heat stable and more resistant to proteolysis due to the presence of additional disulphide bridges that stabilize the molecular scaffold.[8] Venom toxins often work synergistically, amplifying each other's effects and dissemination in the body. Due to high target specificity, these toxins are also key elements in the development of new research tools and drugs of potential clinical use. Captopril and Prialt used in the treatment of renovascular hypertension and chronic untreatable pain, respectively, are examples of drugs designed from venom toxins. Venom composition is contingent on many factors, including age, gender, and the geographical region where the snake is found on, that are responsible for contrasting clinical symptoms in victims of envenoming by the same snake species. The most abundant constituents are the primary targets for immunotherapy. [6]

Some of the most common local effects of snake bite are: radiating pain, swelling, blistering and bruising in the bitten site; and overall extensive local tissue damage. Systemic effects include early syncope and collapse with transient loss of vision and consciousness; hypotension and shock; arrhythmia; spontaneous systemic bleeding from the nose and gums, the respiratory, gastrointestinal and genitourinary tracts; stroke and (in pregnant women) haemorrhages leading to abortion and foetal death; acute kidney injury and rhabdomyolysis. The symptoms of myotoxicity include muscle pain, tenderness, weakness, elevated serum creatine kinase levels, and (if systemic

myotoxicity is prompt) myoglobinuria, which is often associated with renal alterations. Haemostatic disturbances are caused by the hinderance of the coagulation cascade and platelet functions. Venom toxins are both pro-coagulant, causing intravascular clotting (blocking blood flow to vital organs), and anti-coagulant, turning blood incoagulable.[9]

Additionally, depending on the snake family responsible for the snakebite, consequences might vary. Elapids mostly induce neurotoxic manifestations, such as neuromuscular paralysis. On the other hand, viperids cause myotoxicity, local and systemic effects like bleeding, coagulopathies and hypovolaemic shock. Viperid venoms distinctively cause local haemorrhaging due to the depletion of blood vessel protein components, which become fragile and break down. The resulting hypoxic state furthers myotoxicity culminating in the destruction of both sarcoplasmic reticulum and muscle fibrils of skeletal muscles.[1]

In a study where 132 species of snakes were analysed, which included 65 of the known 239 pit viper species (27 % of all pit vipers), 59 families of proteins were classified based on their compositional abundance and ubiquity and, of those, ten families of dominant and secondary proteins were found. Four of the most prevalent include: phospholipase A_2s (PLA_2), metalloproteases (SVMP), serine proteases (SVSP) and three-finger toxins (3FTx). Venoms of vipers include mostly metalloproteases, PLA_2s and serine proteases. Elapid venoms were found to be more variable in the number of different protein families compared to viper venoms. Despite this, PLA_2s were the most widespread protein family present in both elapids and vipers, making it possible to diagnose snakebite through a simple phospholipase A_2 assay in human body fluids which supports efforts to develop universal antivenoms.[10,11]

1.4. *Bothrops asper*

The species provoking the most bites and deaths are *Echis* species in Northern Africa, *Bothrops asper* in Central and South America, and *Naja* and *Bungarus* species in Asia.[5] In Costa Rica, *Bothrops asper* (belonging to the Viperidae family) or terciopelo, as is commonly known, is responsible for almost half of all snakebite accidents and therefore, from a public health point, the most important.[12] With proper medical care, the fatality rate drops from 7 % to 0.5-3 %. Even so, after *B. asper* snakebite many victims face life threatening sequelae due to the tissue damaging effects of the venom.[9] *B. asper* inject a relative large amount of venom (mean of 187 mg of venom per bite) and are extremely

aggressive.[13,14] The symptomatology of *B. asper* snakebite includes prominent local tissue damage, consisting of swelling, blistering, prominent oedema, haemorrhage, dermo and myonecrosis with clinical manifestations such as bleeding, effects on platelet aggregation, coagulopathy, hypotension, hemodynamic alterations, pulmonary oedema, and acute renal failure. Other less common effects include intravascular haemolysis, acute myocardial damage multiple organ failure, and death.[12,14]



Fig. 2 - Photograph of *B. asper*, one of the deadliest snakes in Central America. From https://www.biodiversity4all.org/taxa/30798-Bothrops-asper.

Reported differences in the constitution of the venoms of specimens of *B. asper* from the Caribbean and Pacific regions of Costa Rica were determined through proteomic analysis. The similarity between venom protein profiles of the two neighbouring populations is estimated to be around 52 %. A fundamental difference is the ratio of PLA$_2$ in the two populations. These proteins can reach up to 45 % of total venom proteins (Pacific), and no less 29 % (Caribbean), in adult individuals. Intraspecific variation is a general feature of snake venoms, that has evident clinical and therapeutic implications, as victims bitten by the same species often experience very different symptoms, calling for more specific antivenoms.[9]

Table 1 - Percentages of different protein families in the venoms of Caribbean and Pacific *B. asper* populations of Costa Rica. PLA$_2$ -: phospholipase A$_2$; SVSP - serine protease; SVMP - metalloprotease; LAAO - L-amino oxidase. Adapted from [9].

	% of total venom proteins			
	Caribbean		Pacific	
Protein Family	Adult	Neonate	Adult	Neonate
PLA$_2$	**28.8**	**23.7**	**45.1**	**27.7**
SVSP	18.2	6.7	4.4	2.6
SVMP	41	63	44	65.5
LAAO	9.2	2.5	4.6	3.4
Other components	2.8	4.1	1.9	0.8

Victims of *B. asper* bites develop local tissue damage characterized by dermonecrosis, blistering, edema, local haemorrhage and myonecrosis, and, in severe cases, defibrinogenation, thrombocytopenia, platelet hypoaggregation, bleeding distant from the bite site, disseminated intravascular coagulation, cardiovascular shock and acute renal failure.[9,15]

1.5. Phospholipases A2

The Phospholipase A$_2$ (PLA$_2$s) superfamily describes enzymes that specifically catalyse the hydrolysis of the ester bond at the sn-2 position of glycerophospholipids producing lysophospholipids and fatty acids.[7]

Snake venom phospholipases A$_2$ (PLA$_2$s) are multi-toxic enzymes that can cause both local and systemic effects. PLA$_2$s from snake venom exerting their action on skeletal muscles are referred to as myotoxic PLA$_2$s. The majority of viperid myotoxic PLA$_2$s act locally. In cases where PLA$_2$s carry out systemic myotoxicity, renal complications occur due to myoglobin accumulation in the kidneys.[16,17] While some myotoxic PLA$_2$s may also induce systemic myotoxicity, beyond local sequelae, snakes of the Bothrops genus cause mostly prominent local tissue damage, rarely systemic myotoxicity, in spite of the abundance of myotoxic PLA$_2$s in the venoms of these snakes.[15]The deleterious action of myotoxic PLA$_2$s is enhanced by other toxins, further promoting muscle damage and impairing its regeneration.[18]

Representatives of this superfamily of enzymes are widely distributed, in viruses, mammals all the way to snakes.[19] PLA$_2$s have been subdivided in different groups on the basis of their structural and biochemical properties, like size, sequence, localization, function, and divalent cation requirements.[16] Other classifications can be made of PLA$_2$s according to criteria like their isoeletric point.[20] Although toxicity in PLA$_2$s is linked to basic PLA$_2$s rather than acidic, with few exceptions, acidic PLA$_2$s are in general stronger enzymatically and have seemingly a digestive role.[16] Basic PLA$_2$s molecules are very strong myotoxins, provoking acute muscle tissue necrosis, and are more abundant than acidic ones. [9,13,21]

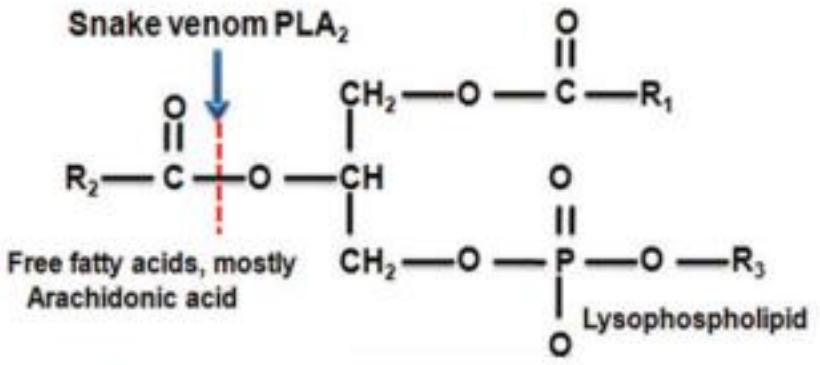


Fig. 3 - Hydrolysis of the sn-2 acyl bond of glycerophospholipids, catalysed by PLA$_2$s. From [22].

Venom PLA$_2$s carry out a variety of toxic and pharmacological effects, which include neurotoxic, myotoxic and hemolytic actions.[18] Snake phospholipases are secreted PLA$_2$s (sPLA$_2$s), and they are multi-toxic enzymes that cause both local and systemic effects. They belong to either group I or II, depending on the snake belonging to the Elapidae or Viperidae family, respectively.[16,21] Nonetheless, the two groups share a conserved structure and identical catalytic machinery, albeit having some differing regions, as well as functions and toxic mechanisms.[16,23] Group I proteins display a characteristic “elapid loop”, while group II enzymes present a special C-terminal extension of 5–7 amino acids.[15] Group II proteins can present as monomers, or as homodimers, heterodimeric complexes, trimers, or larger oligomers. [16]

Despite being first discovered in snakes, sPLA$_2$s are present in bee, scorpion or wasp venoms, and all mammalian tissues.[7,17]

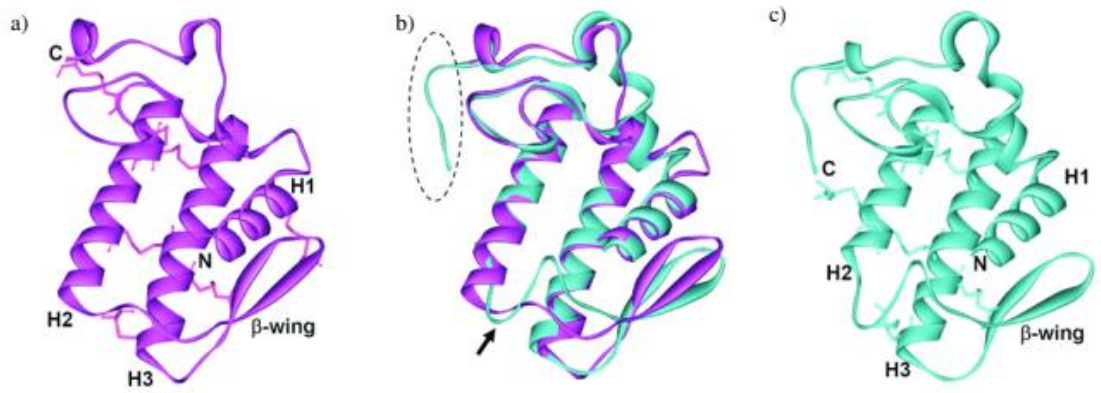


Fig. 4 - Comparison of the three-dimensional structure of a group I PLA$_2$ (a) (PDB ID: 1AE7) and a group II PLA$_2$ (c) (PDB ID: 1VAP). In (b) we observe the superimposition of the two structures and there overall high conservation, despite the distinctive extended C-terminal in (c) and the "elapid loop" in (a), N and C terminal are represented, as well as the helices (H1, H2, H3) and β-wing region. From [16].

Structurally, sPLA$_2$s from snake venom are very conserved and they comprise three main alpha helices, a beta sheet and seven disulphide bonds[17] and similar molecular weight (12–15 kDa), are composed of 119–134 amino acids, although the location of their seven disulphide bonds is different. Their activity is dependent on the availability of a calcium ion for the hydrolysis of phospholipids. There are various quaternary structures these proteins can adopt, mainly monomers or non-covalent homodimers, but some form complexes through covalent and/or non-covalent interactions with other PLA$_2$s or with other proteins. [7,16,21]

PLA$_2$s with the amino acid Aspartate in position 49 - Asp49 - at the active site, are catalytically active and they catalyse hydrolysis of membrane phospholipids at the sn-2 position of phospholipids (of sn-3 glycerophospholipids to release lysophospholipids and fatty acids) in the plasma membrane of various cells.[9,22] sPLA$_2$s with a Lysine substitution in position 49, or other amino acids (less frequently Ser, Arg, Gln, or Asn) - Lys49 - are devoid of enzymatic activity. These proteins are not true PLA$_2$s *sensu stricto*. Yet, they are known as sPLA$_2$s homologues because of their structural similarity with Asp49 sPLA$_2$s, presenting the same disulphide brigdes and C-terminal extension. So far, all studied Lys49 are basic proteins.[23]

A synergism between both kinds of sPLA$_2$s has been found in vitro and in vivo models. This synergism occurs unidirectionally: Lys49 myotoxicity is amplified by the Asp49 enzyme, but not the other way around. This effect isn't observable for the acidic Asp49

sPLA$_2$s. The enhancement phenomenon can be conducive to an easier and effective digestion of mammalian prey, granting a selective advantage considering the high energetic costs of venom protein production.[24]

Experimental data indicates that the first site of action of myotoxic sPLA$_2$s is the sarcolemma (plasma cell membrane of skeletal muscle fibers), followed by a slew of degenerative events occurring in skeletal muscle fibers, further amplifying the myotoxic effect of sPLA$_2$s.[15] This results a prominent calcium influx, which in turn prompts hypercontraction of myofilaments, mitochondrial and overall irreversible cell damage, and cell death.[12]

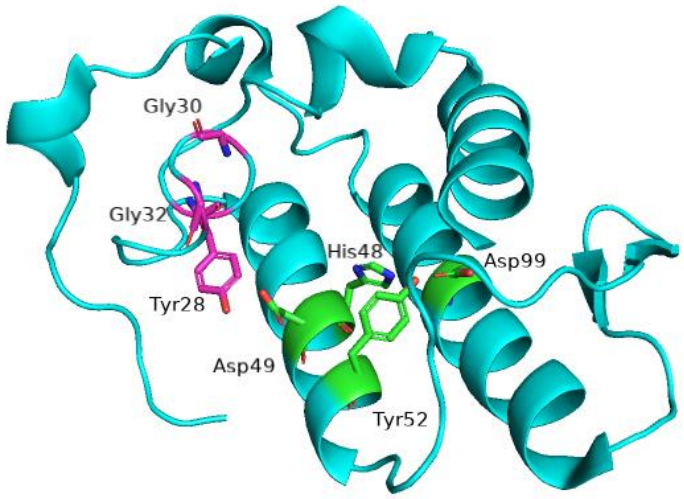


Fig. 5 - A view of the structure of a Asp49 PLA$_2$ (PDB ID: 5TFV)[25] showing the active site residues in green (His48, Asp49, Tyr52, Asp99) and the residues in the Ca^{2+} loop in pink (Tyr28, Gly30, Gly 32). Image obtained using PyMOL.[26]

Both protein subgroups induce skeletal muscle necrosis, although through different means. Asp49 sPLA$_2$s damage plasma membrane by phospholipid hydrolysis while Lys49 homologues disrupt the integrity of membranes through the interaction of hydrophobic and cationic residues, in C-terminal region of the molecule, with membrane phospholipids.[16,18]

In whichever way the sarcolemma is affected, alteration of membrane structure occurs, causing the exchange of ions from the extracellular medium, mostly Ca^{2+}, as well as other deleterious effects, that leads to a stereotyped series of events. These include: a) rapid depolarization; b) myofilaments hyper contract, inducing further mechanical damage; c) mitochondria swells, causing severe impairment of their function; d) activation of Ca^{2+}-dependent proteinases, which contribute to the degradation of cytoskeletal components; e) activation of cytosolic PLA$_2$s, which could contribute to

phospholipid hydrolysis in the plasma and other organelles' membranes; f) over activation of calpains that degrade cytoskeletal components. [15,16, ,22]

sPLA$_2$s also disrupt the membrane of sarcoplasmic reticulum and T tubules.[15] Histological evidence shows that cells present lesions consistent with disruption of membrane.[12] Additionally, there's a loss of ATP and K^+ causes pain and promotes more damage to cells. ATP acts as a "danger signal" and causes the influx of Ca^{2+}, damaging myotubes.[16] There is also an efflux of molecules such as myoglobin, creatine kinase and lactate dehydrogenase.[15]

sPLA$_2$s are also implicated in phagocytosis which represents the first step in triggering host defence and inflammation.[20] The inflammatory events evoked by sPLA$_2$s are primarily associated with enzymatic activity (Asp49) that release arachidonic acid - a precursor for the biosynthesis of several lipid mediators of inflammation - and lysophospholipid - which enters the inflammatory pathway through a platelet activating factor. Inflammatory mediators are formed enzymatically from arachidonic acid, progressing inflammation from the envenomed region to surrounding tissues. [22] The free fatty acids and lysophospholipids released from the phospholipid hydrolysis catalysed by Asp49 sPLA$_2$s can, by themselves, damage the sarcolemma through their detergent activity.[12,15]

Prior studies have determined that carboxylate-based compounds are true substrates for PLA$_2$s and that, for 4-nitro-3-(octanoyloxy)benzoic acid (also called NOBA), the specificity constants, k_{cat}/K_m with three representative PLA$_2$s range from 0.34 to 7.01 x 10^{-3} M^{-1} s^{-1}.[27] These compounds make the experimental analysis of enzymatic activity easier, as they are more easily controllable than the micelles or monolayers these enzymes prefer. [28]

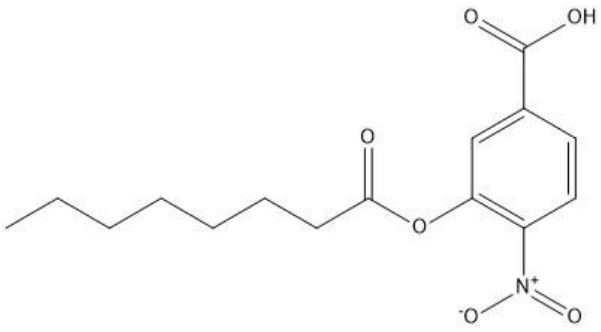


Fig. 6 - Structure of the PLA$_2$ substrate NOBA. Figure obtained using ChemDraw 20.1

1.6. Myotoxin I

Myotoxin I,II and III are secreted phospholipases A_2 present in the venom of *Bothrops asper*. Myotoxin I (MT-I) is a basic Asp49 PLA_2 and it was first reported in the venom of *Bothrops Asper* by Gutiérrez et al.[12], whereas myotoxin II is a Lys49 PLA_2 homologue and myotoxin III is also a basic PLA_2[29,30]. Researchers categorized this toxin, its biochemical properties and physiologic effects on skeletal muscle of mice. Isolated MT-I induces myonecrosis in mice, not unlike the action of the crude venom, causing similar muscle lesions observed by histological pictures. All of them MT-I and MT-II induce sarcolemma alterations, leading to the influx of Ca^{2+} in muscle cells, compromising mitochondrial activity, myofibrillar hypercontraction and additional degenerative events. Follow-up studies also show evidence that the mode of action of basic Asp49 enzymes occurs through direct catalytic action towards the membrane phospholipids of muscle. [23]

Studies done on cultures of mice myotubes have shown that only the catalytically active myotoxins induce significant phospholipid hydrolysis. Nonetheless, these in vitro studies also show that the acute damage done by Lys49 is morphologically indistinguishable from that of Asp49 PLA_2-induced myonecrosis.[13] This proves that, albeit having similar results, Lys49 PLA_2s induce membrane disorganization and damage through a distinct molecular mechanism, yet to be fully understood. As these homologues retain all the structural characteristics of PLA_2s, including their interfacial binding surface, recognition by particular membrane phospholipids or receptors could be a possible explanation of their toxic activity.[17]

There is also the possibility that these PLA_2s may be internalized, therefore having an intracellular site of action, generating other cellular degenerative events. [18,16] However there is no reports on lipidic or glucidic acceptors, or protein receptor that could explain the specific myonecrotic activity has been identified so far.[16,21]

1.7. Catalytic Mechanism

Compared with many other soluble enzymes, PLA_2s have the particular feature of acting preferably on aggregated substrates, such as micelles, monolayers or vesicles which requires what is known as ‘interfacial activation’.[18] Interfacial catalysis occurs when an enzyme needs to associate with the interface of a substrate to gradually access the substrate for the catalytic cycle.[31] This means that its activity toward aggregated

substrates is considerably higher than that toward monomeric substrate molecules. The interface binding surface, or 'i-face', of these enzymes promotes the accessibility of substrate and mediates the allosteric effects of the interface on the steps of the interfacial turnover cycle. This is structurally and functionally distinct from the active site, where the catalytic cycle is carried out.[32]

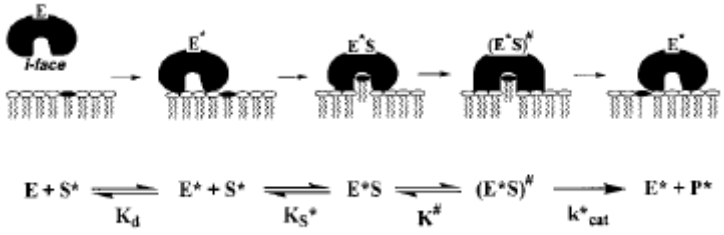


Fig. 7 - Scheme representing sPLA$_2$ interfacial catalytic turnover. The functionally active monomer E* will access the substrate S* and make specific interaction along its i-face to become fully activated, resulting in the E*S* complex that undergoes further change E+S*# to catalyse the lipid hydrolysis. From [31].

PLA$_2$s can present a variety of states of oligomerization, from monomers to homodimers, heterodimeric complexes or even larger oligomers.[16] MT-I is a homodimer, a quaternary structure that provides larger binding surfaces, increasing the affinity of interactions with target receptors.[25] Regarding its tertiary structure, the conserved structural fold, common to both groups, presents a short N-terminal α-helix, followed by a loop implicated in Ca^{2+} binding, a second long α-helix, a β-wing region, and a third long α-helix ending with a C-terminal loop. A sevenfold pentagonal coordination stabilizes the Ca^{2+} ion in the structure: two carboxylate oxygen atoms of Asp49, three main chain oxygen atoms of Tyr28, Gly30 and Gly32, and two oxygen atoms of two structurally conserved water molecules. The Ca^{2+} ion is fundamental for the catalytic activity of PLA$_2$s, binding the substrate and polarizing its coordinated carbonyl group in order to facilitate hydrolysis. [33] The catalytic reaction mechanism follows a general base-mediated attack on the sessile bond through the involvement of a conserved water molecule which serves as a nucleophile.

Scotte and associates have proposed that His48, stabilized by Asp99, removes a proton from a water molecule, and the formation of a tetrahedral intermediate may occur. The substrate interacts with the enzyme on an internal surface that is shielded from bulk solvent, through a hydrophobic channel whose opening is in the i-face. Sn-1 and sn-2

alkyl substituents were parallel in a hydrophobic channel extending from the catalytic site to the surface of the molecule close to the N-terminus.[18]

Currently, two different catalytic mechanisms are proposed: 1) a “catalytic triad”, or single-water mechanism, and 2) a “calcium-coordinated oxyanion”, or water-assisted mechanism. The presumed “catalytic triad” is composed of Asp99, His48, and a nucleophilic water present in the active site, whereas in the “calcium-coordinated oxyanion” mechanism two different water molecules are involved.

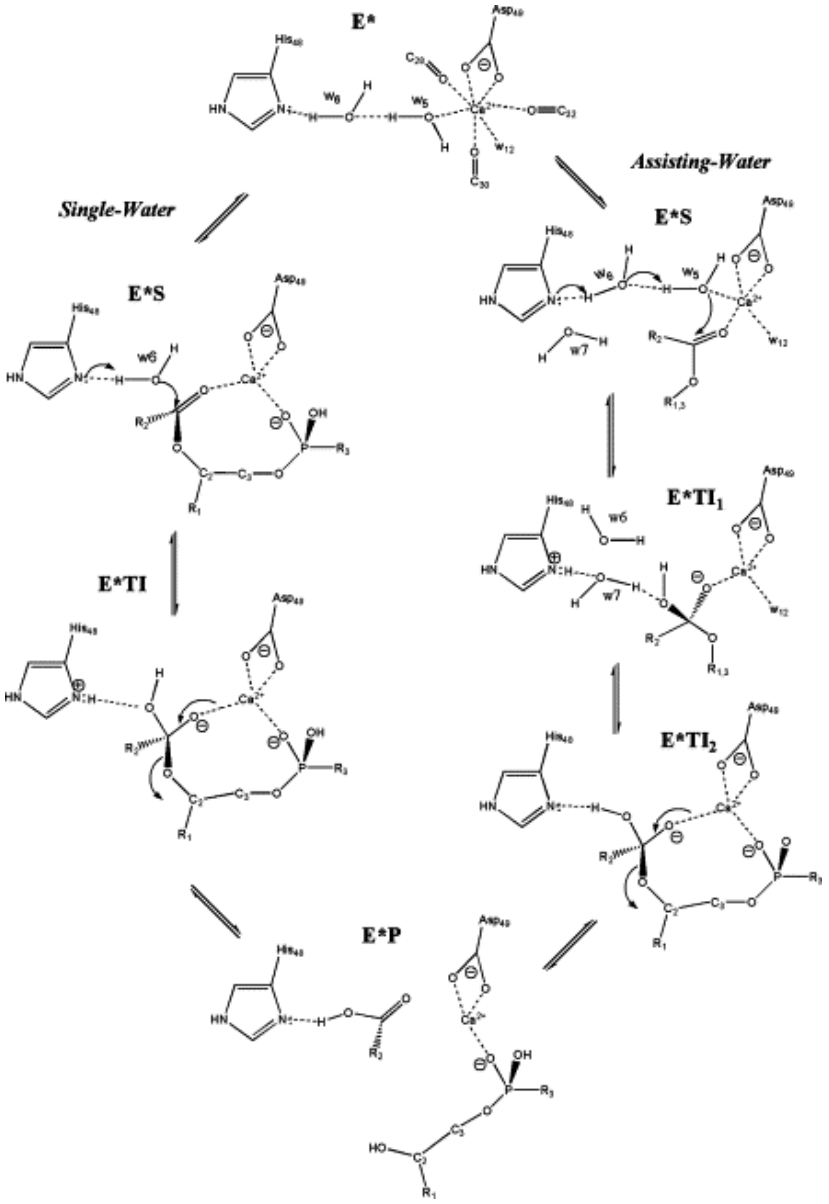


Fig. 8 - Two proposed mechanisms of PLA$_2$, the single-water (left) or assisting-water (right). From [34].

In the mechanism labelled single-water mechanism (left scheme in Fig. 7) the attacking water is activated by His48 and polarization of the substrate ester by coordination with

the active site calcium, and only one intermediate is formed (E*TI). Whereas in the assisted-water mechanism (right scheme in Fig. 7), the calcium coordinated catalytic water (w5) is bridged through a second water sphere water (w6) to His48, then the assisting water (w7) bridges His48 and the oxygen of the tetrahedral intermediate (E*TI$_1$), with a second intermediate (E*TI$_2$) formed before the product conversion. Experimental results endorse the water-assisted mechanism as the most favourable conformation.[33] Note should be made of the fact these findings derive from calculations on a monomeric crystallographic model of a venom PLA$_2$ (ID PDB: 1POB).[31]

These two disparate mechanisms could involve specific membrane targets in cells, like glycerophospholipids or glycolipids, with variable affinity. However, no targets for these PLA$_2$s have been found yet, awaiting further studies. [15,16] Studies have observed that, upon injection in the muscle, MT-I remains at the injection site and, when injected intravenously it doesn't spread to induce systemic muscle damage, further supporting the hypothesis that there are unknown receptors involved. [15]

1.8. Dissertation outline

Phospholipases A$_2$ are challenging research targets. The binding affinities of all known ligands of PLA$_2$ are in the range from 10^{-4} to 10^{-8} M, making them poor to moderate candidates as drugs.[7] In addition, ligands are able to take up only specific subsites within the entire substrate-binding space, leading to the few and far between numbers of interactions with the protein. All the unique characteristics of this kind of enzyme have made their catalytic mechanism difficult to solve.

There are, as of September 2021, 322 structures that contain PLA$_2$s currently available on PDB, but a model of MT-I of *B. asper* with known synthetic substrate NOBA is still to be determined. With this work, we attempt to build a robust model of a *B. asper* MT-I, using homology modelling and molecular dynamics simulations. Having laid part of the groundwork for future research, this is yet another step towards the full understanding of these complex proteins.

2. Methodology

In order to create a model of MT-I, several computational techniques have been used. In this chapter details of the implementations of these methodologies are provided.

2.1. Preparation of Protein Structure

Firstly, the sequence of a basic MT-I belonging to a *B. Aspe*r (UniProt ID: P20474) was found on UniProtKB, an online protein database containing information such as their amino acid sequence.[35] As previously stated, this is a basic sPLA$_2$, presenting as a homodimer, and containing a calcium ion in its active site.

The corresponding crystallographic structure of this sequence had already been established and published on Protein Data Bank (PDB), that archives structural data of proteins.[36] The total number of structures in PDB is 181969 (as of September 2021), of which 159633 are X-ray structures (88 %)[37]. X-Ray crystallography is one of a few experimental techniques used for identifying the atomic and molecular structure of a protein, the second one most widely used being nuclear magnetic resonance (NMR) spectroscopy, although it has its drawbacks. Crystallizing proteins is difficult due to their variety of molecular weights, shapes, aggregate states, and surface features that change with pH and temperature.[38] Still, as NMR isn't able to handle large proteins, X-ray crystallography is the most commonly used method for gaining protein structural information.[39]

In X-Ray crystallography, the atoms that form the crystal of the studied molecule scatter a beam of X-rays into many specific directions, forming a diffraction pattern, from which the structure can be inferred. Through the measurement of the angles and intensities of the diffracted beams, a 3D picture can be built. Since the scattering is determined by the electron density, small atoms, with low electron density, may not appear in the diffraction patter. For that reason, X-Ray structures in the PDB database do not have hydrogen atoms. Consequently, hydrogens atoms have to be added in the structure at a certain point.

However, as a result of the fact that the diffraction experiment must be performed with the protein in the solid state (given that it is protein crystals and not individual molecules or a protein solution that give rise to an X-ray diffraction pattern), the analysis of a specific

structure may be compromised due to the loss of information concerning the true dynamics of the molecule. [39]

The X-ray structure (PDB ID: 5TFV) encompasses 122 amino acid residues, with a 2.54 Å resolution, a TRS ligand employed in crystallography methods, but doesn't include the Ca^{2+} ion needed for catalysis.[27] The absence of Ca^{2+} ions in the X-ray structure, and the fact that there were parts of the structure that had only been partly modelled, meant that homology modelling was required. Swiss-Model is a web-server for structural bioinformatics [40], with an integrated database and interface, allowing the search of "templates", that is, proteins whose structure can be used to build as a template for a "target".

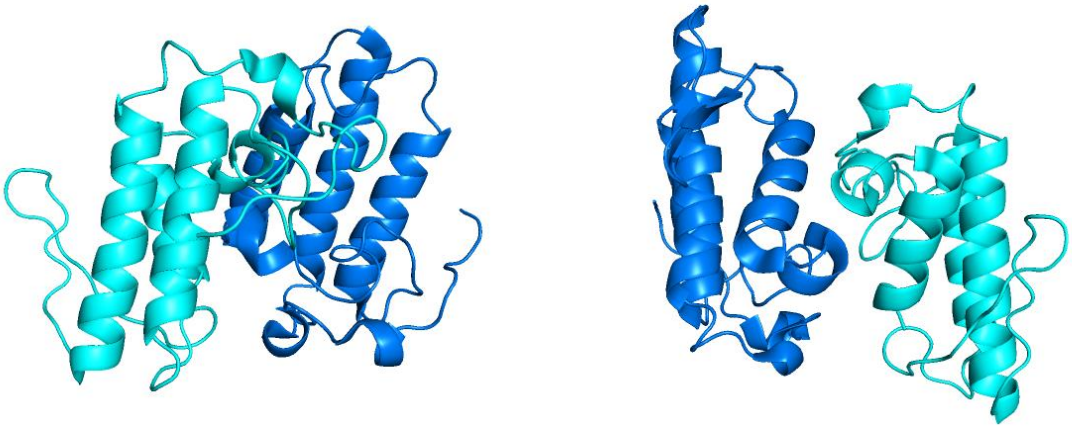


Fig. 9 - Two views of 5TFV, with each monomer coloured a different shade of blue. Image obtained using PyMOL.[26]

The sequence of MT-I serves as query to search for homologous protein structures against the Swiss-Model template library, through two database search methods: BLAST and HHblits. BLAST (Basic Local Alignment Search Tool) is fast and very accurate for closely related templates, while HHblits is reliable for distantly related sequences.

After the search is complete, templates are ranked by measures of expected quality of subsequent models, like Global Model Quality Estimate (GMQE) and Quaternary Structure Quality Estimate (QSQE), depending on a template's structure and target-template alignment. Review of sequence similarity, template alignment and three-dimensional structures is necessary to proceed with the modelling step. After selecting the template, an atomistic protein model is generated by assigning conserved atom coordinates as defined by the target-template alignment. Insertions and deletions, as defined by the sequence alignment, are resolved by first searching for viable candidate fragments in a structural database. To quantify modelling errors, this modelling engine

relies on the QMEAN scoring function, that generates global and per residue quality estimates. To facilitate interpretation of the obtained model quality estimates, the QMEAN global score is transformed to a Z-score, indicating whether the model scores as it would be expected from experimentally determined structures of similar size.[40]

Upon search for templates using the MT-I sequence (UniProt ID:20474) as a target, we retrieved several templates. Through visual inspection and through criteria such as the presence of ions, ligands, number of residues, identity percentage, we gathered the template of choice was PDB ID: 1FXF [31], the structure of the crystallographic dimer of porcine pancreatic PLA$_2$. Using the X-ray crystal structure PDB ID: 1FXF as a template, an homology model was generated.

2.2. Homology Modeling

Knowledge of the three-dimensional structure of proteins is of essential importance for many fields, including pharmacology and medicine. Since a protein's function is determined by its structure (which is in turn determined by its aminoacidic sequence) and experimental methods have the drawbacks already discussed, several computational methods can be employed. Conveniently, although protein sequences can present numerous possible combinations, structural diversity is reduced, given there are only so many folds proteins can adopt because of stereochemistry. Stereochemistry dictates the relative spatial arrangement of atoms that form the structure of molecules (in this case, proteins) are limited. In addition to this, homologous proteins are more disposed to conserving their structure than their sequence, and this effect is seen even if they've diverged from their common ancestral a long time ago.[41] These observations gave rise to homology modeling.

Homology modeling is a technique employed in the making of a proteins 3D structure from its amino acid sequence and 3D structures belonging to homologous proteins. This approach can assume several workflows, but in this work, it consists of: i) search of proteins homologues used as templates; ii) alignment of the sequence of the protein homologues with our target protein (in this case, MT-I), iii) construction of a model for the structure of the target protein from the chosen template; and finally iv) refinement and analysis of the quality of the modeled structure.

Despite having a homology modelling model obtained from Swiss-Model, additional modelling was necessary in certain regions of the MT-I model.

2.3. Energy Minimization

Before conducting a Molecular Dynamics (MD) simulation it is necessary to carry out an energy minimization protocol of the generated model, using the Sander module of Amber, a suite of biomolecular simulation software.[42]

This method relaxes the system of the tensions between misdirected atoms, which aren't energetically favourable. If not removed, the energy in that region will be unrealistically high and that can either crash the simulation or cause the trajectory to proceed in a misleading direction. This is an inherent issue of the homology modelling process. Thus, the minimization process is crucial to finding the minimum potential energy in relation to all atomic coordinates of the protein. The resulting minimized structures are more likely found in nature, ensuring the quality of the model we chose to work with.

In the present work, the energy minimization protocol was performed in four consecutive steps, in which restraints (used for fixing the cartesian coordinates of selected atoms) exerted on the system are gradually removed: i) water molecules; ii) hydrogen atoms (added in by xleap); iii) side-chain and iv) all free. The relaxation of the system is done in steps, where the entire structure is fixed except for the selected atoms.

2.4. Molecular Mechanics

Molecular Mechanics (MM) is an empirical method that use classical mechanics to describe molecular systems: the atom is the smallest particle, and all molecules are represented by particles representing atoms or groups of atoms bonded by springs.[43] Systems are described by combining a force field with a set of parameters. The Born-Oppenheimer approximation is accepted, which neglects the motion of the atomic nuclei when describing the electrons in a molecule and the inclusion of the electrons and all quantic elements is done by the parametrization of the force field. [44]

MM methods rely on a set of parameters for all bonded and non-bonded interactions in a system to produce the total potential energy of the system:

$$U(R) = \sum U_{bonded} + \sum U_{nonbonded} \quad (1)$$

This method contrasts with Quantum Mechanics (QM), where electrons are explicitly represented in the model and interaction energy is calculated by solving the electronic structure of the molecules in the system. MM simulations are much faster than QM, and therefore vital for the study of very large and complex systems, like most proteins and other biomolecular structures.[45] Nonetheless, this simplification compromises computational speed for atomic precision, this not appropriate for bond-breaking reactions. To form or break bonds a more detailed description is required, and the estimation of the Schrödinger equation [46]:

$$H\psi = E\psi \tag{2}$$

With H being the Hamiltonian operator and describing the total energy of the system. ψ represents the wavefunction and E the energy values of ψ.

2.4.1 Force Fields

Force Fields (FF) calculate the forces that determine a system's dynamics, as a function of the atomic coordinates, and describe the bond lengths, bond angles, dihedral angles, improper planes, and electrostatic and van der Waals forces. FFs are defined by a set of parameters (also called topologies) and equations.[44,47]

The parameters of molecular FFs derive from both experimental and computational chemistry methods and each FF has their own distinct ones. The method in which the different energy terms of the potential energy function are treated also differs from one FF to the next.

There are many FFs available to choose from and this selection depends on the type of system in study. As mentioned before, MM applies FFs that estimate forces on atoms and bonds (represented by rigid spheres connected with springs) and protons and electrons are not described explicitly but rather implicitly by the defined atomic parameters. In the present work we used the AMBER (Assisted Model Building with Energy Refinement) Force Field.[48] CHARMM (Chemistry at HARvard Macromolecular Mechanics)[49] and GROMOS (GROningen MOlecular Simulation)[50] are examples of other widely used force fields in biomolecular studies.

2.4.1.1 AMBER Force Field

Amber is the collective name for a suite of programs that perform molecular dynamics simulations, with a focus on biomolecules.[51] These programs operate together to setup, perform and analyse MDs. Amber is also the name of the classical molecular mechanics force fields implemented by this software.

Table 2 - Most common force fields used in the simulations of different biomolecules. [42]

Composition	Protein	DNA	RNA	Carbohydrate	Lipid	Water model[a]	General[b]
Force field	ff14SB	OL15	OL3	GLYCAM_06j	lipid17	tip3p	gaff2

[a] for a water model with associated atomic ions. Not needed if you are using an implicit solvent model

[b] for organic molecules like ligands

For molecules not present in current FFs, Amber also has tools to parametrize them and multiple FFs can be carefully combined if the system contains components of different natures. Amber 12 SANDER was used

AMBER ff14SB, TIP3P and GAFF were used to perform the Molecular Mechanics simulations.

For AMBER, the general mathematic equation to calculate these properties is:

$$E_{bonded} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)] \quad (3)$$

and

$$E_{nonbonded} = \sum_{\substack{nonbonded \\ pairs\ ij}} \left(\varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^2 - 2 * \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right) \quad (4)$$

The bonded or intramolecular term (Equation 3) defines the interactions of the stretching of bonds, the bending of valence angles, and the rotation of dihedrals. In the first term, that describes the stretching of the bonds, the parameters K_b represent the stifness of the bond, b the bond length and b_0 the equilibrium length of the bond. In the second term, that involves triplets of atoms (b) and describes the bending of angles, the θ is the angle formed by the two vectors, K_θ and θ_0 characterize the stiffness and equilibrium geometry of the angle. The third term is sum over quadruplets of atoms (c) that defines the energetics associated with rotation of the dihedral angle defined by those four atoms, χ is the value of the dihedral, K_χ the energetic parameter that determines barrier heights, n the multiplicity and σ is the phase.

The non-bonded or intermolecular term (Equation 4) represents the Van der Waals and Coulomb interactions that occur between atoms in separate molecules or atoms separated by three or more bonds in the same molecule (d).

The first part of this equation is known as the Lennard-Jones (LJ) equation (commonly the van der Walls term), which models attractive dispersion and repulsive Pauli exclusion interactions. The second part of the equation, the Coulomb's law, is used to model the electrostatic interaction between nonbonded pairs of atoms. In this term, the ε_{ij} prefactor is parameter based on the types of the two interacting atoms, $R_{min,ij}$ is the distance at which the LJ energy is a minimum, r_{ij} is the interatomic distance between nonbonded pairs of atoms, q_i and q_j are the parameters that describe the effective charges on atoms i and j.

For a given compound, all parameters have to be known and consistently defined, for all atoms that are part of the simulation.

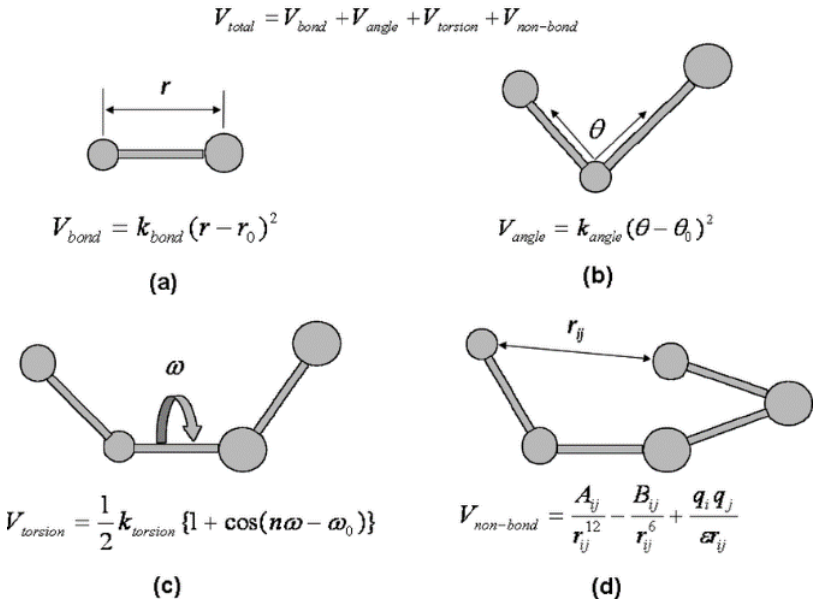


Fig. 10 - Representation of intramolecular ((a), (b) and (c)) and intermolecular (d) interactions. From [52].

2.4.2. Molecular Dynamics

Molecular dynamics (MD) strive to reproduce the dynamics of an atomic system by iteratively solving a discretized version of Newton's equations of motion.

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{5}$$

This describes the motion of a particle with mass m_i, along x_i coordinate, with F_{x_i} force (Equation 5), for each system's particle.[53]

In this approach, we select a model system, comprising of *N* particles and solve Newton's equations of motion for this system until the properties of the system no longer change with time (we equilibrate the system). After the equilibration stage, we perform the actual measurement.[43]

To measure an observable quantity in an MD simulation, we must first be able to express this observable as a function of the positions and momenta of the particles of the system: i) assign initial positions and velocities to all particles in the system; ii) compute the forces acting on all particles (using periodic boundary conditions) - having computed all forces between particles we can; iii) integrate Newton's equation of motion - repeated until time evolution of the system has been completed for the desired length of time; iv) average measurements can now be computed.

Both thermodynamic properties and structural properties do not depend on the time evolution of the system: these are static equilibrium averages that can be obtained by MD simulations.

In such simulations, the total energy E is a constant. The choice of an adequate time step in MD simulations is essential and a trade-off between computational resources, and the time length necessary to analyse the chemical phenomenon in study, has to be made. In this work, a timestep of 2 fs was used.

A microstate is a microscopic configuration that a thermodynamic system may occupy. A macrostate of a system refers to its macroscopic properties or state functions, such as volume, temperature, pressure, and density. An ensemble estimates the probability distribution for the state of the system, taking a large number of replicas of the system, each microstate, referring to a possible state the real system may exhibit.

The most used types of ensembles are: the canonical ensemble (number of particles (N), volume (V) and temperature (T) constants); isothermal-isobaric (N, pressure (P) and

T constants); microcanonical (N, V and total energy (E) constants); and the grand canonical (V, T and chemical potential (μ) constants).

For the specific scope of this thesis, MD simulations have been employed to analyse protein and substrate dynamics. These interactions are described in atomistic details (including hydrogens), neglecting any QM influence effect.

3. Results and Discussion

3.1. Initial X-Ray Structure

Over the years, many three-dimensional structures of snake venom sPLA$_2$s have been determined and made available through the Protein Data Bank (www.rcsb.org).[16]

PDB ID: 5TFV is the crystallographic structure of MT-I, a basic sPLA$_2$ from the venom of B. Asper.[25] The study where the 5TFV structure was published focused on two other regions of the protein[25], which they named MDoS (for "membrane docking site") and MDiS (for "membrane-disrupting site"), that could be responsible for non-catalytic myotoxic activity, similarly to Lys49 homologues. Along with this hypothesis, the researchers add that the inexistence of Ca^{2+} in the crystal, and consequent distortion of its loop, was proof of this secondary mechanism non-Ca^{2+}-dependent. Evidence for this "alternative" activity is scarce, however there is unequivocal experimental data[23,24] that confirms the catalytic nature of MT-I's myotoxicity is dependent on the presence of Ca^{2+} in the medium.
Although it doesn't encompass the calcium ion needed for catalysis (altering the conformation of both the catalytic loop and the coordinating residues), we chose this X-Ray structure as it appertains to the species in analysis thus being an adequate starting point. To circumvent structural alteration induced by the absence of Ca^{2+} in the *B. Asper* sPLA$_2$ X-Ray structure (PDB ID:5TFV), there are two methods:

1. Plainly correcting the coordinates of the catalytic loop in the PDB ID: 5TFV file. However, the method wasn't feasible because, in the absence of the cation, the residues that coordinate to it aren't oriented correctly, as well as other catalytically important surrounding important residues outside of the loop (His48, Asp49, Tyr52, Asp99).

2. The second method is through homology modeling.

3.2. Homology Modeling

Templates that possess similar ligands to NOBA are ideal. Since the structure of this substrate is simpler and less challenging than the micelles, monolayers or water-soluble phospholipids this enzyme binds, we opt for including it in our model.

3.2.1. Choice of Template

The majority of snake venom PLA$_2$s form oligomers, particularly homodimers, as dissociation into monomers often impairs toxicity.[54]

Studies on Lys49-PLA$_2$s have found two different conformations, which they referred to as “conventional dimer” and “alternative dimer”. Although some X-ray scattering experiments points in the direction of an “alternative dimer” occurring in solution, no conclusion has been drawn yet. These proteins adopt many conformations due to the flexibility between the monomers.[55]

The catalytically active, Ca^{2+}-dependent, PLA$_2$ from the venom of *Crotalus atrox* forms a dimer in which the catalytic sites of the two monomers face an internal cavity, effectively shielding the access of the lipid substrate to the active site.[56] Although robust kinetic evidence indicates that these PLA$_2$s must maintain dimeric state to function, the dimer could potentially separate or rearrange, during the catalytic process.[57] An activated quaternary structure transition has been suggested which would expose the active sites, thereby permitting substrate binding.

5TFV adopts the “alternative dimer” like conformation[25], as seen by other Asp49 PLA$_2$s. This dimer is a flat, oblate ellipsoid, with the two subunits tightly.[57] Remarkably, despite the need to retain a dimeric structure for function, the formation of a dimer, where the catalytic sites of the two monomers face an internal cavity and are shielded from bulk solvent is not conducive to its catalytic activity. This is surprising, especially when considering that this type of enzymes prefer aggregates, such as micelles, rather than freely single phospholipids. Most probably these enzymes exist as dimers in solution but dissociate into monomers upon binding the cell membrane, as it was shown to happen with the homolog human sPLA$_2$ enzymes.

Initially the template PDB ID: 1JIA[58] was considered. This structure was presented on the Swiss-Model workspace after inputting the UniProtKB accession code of 5TFV (UniProt ID: P20474). While not on top of the template result ranking, 1JIA seamed promising because it is a homodimer in solution, so building a homo-dimer model was

possible, and a better positioning of NOBA in a dimer conformation is expected in a dimer conformation. Additionally, this structure has the two crucial Ca^{2+} ions, in their correct position. Higher scoring templates lacked structural relevance for various reasons, but the most preponderant motif is the absence of the Ca^{2+} ion, which impacts the conformation of the Ca^{2+} coordinating loops. Considering Ca^{2+} binds loosely with the enzyme, it can be difficult to crystallize and consequently, several structures are deficient of this cofactor.

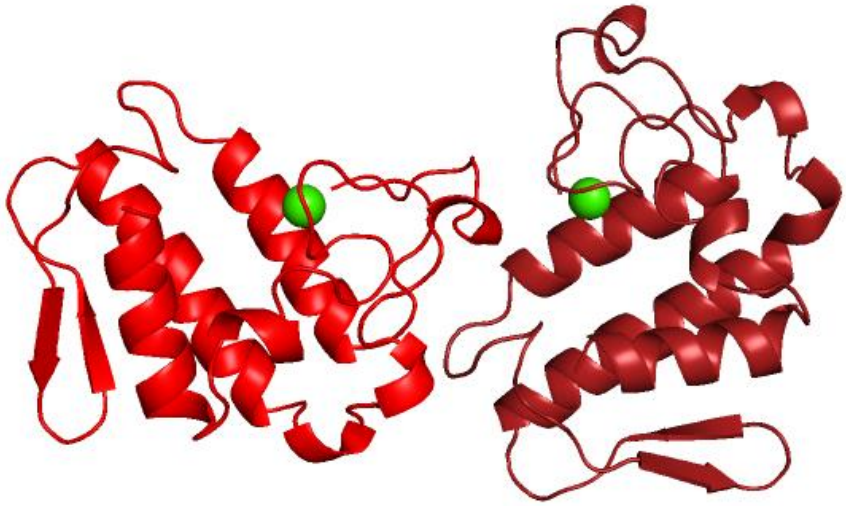


Fig. 11 - Structure of 1JIA[58], a basic PLA$_2$. Each monomer coloured a different shade of red, and the two Ca^{2+} (green) in CPK colouring. Image obtained using PyMOL.[26]

Furthermore, because this template is crystallographic structure with 2.1 Å resolution, with sequence similarity of 0.53, GMQE value of 0.74, a 63.93-64.46 % sequence identity with HHblits or BLAST, respectively, a model was built with Swiss-Models automated mode. After making some corrections on the model and introducing NOBA substrate, we proceeded to run a standard minimization and molecular dynamics. The system crashed, both substrate and calcium were released from the protein and dimerization was lost.

The relative orientation of the monomers was found to be the cause for this outcome: 1JIA's dimer is positioned in a twofold axis, with both monomers perpendicular to each other, as observable by Fig.12. This made labile coordination of the Ca^{2+} ion in this structure, impossible.

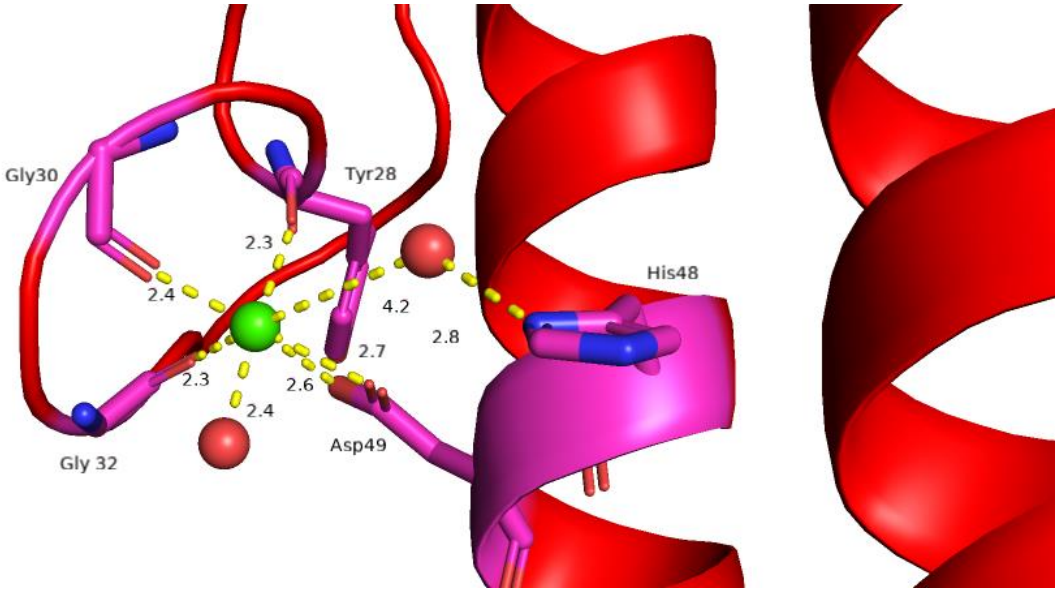


Fig. 12 - Ca^{2+}-coordinating loop in 1JIA. Bonding residues (pink) coloured by element, Ca^{2+} (green) and catalytic water and an assisting water (red) in CPK colouring. Bond distances in Ångström. Image obtained using PyMOL.[26]

Further inspection of both 1JIA and 5TFV was conducted in PyMol. We obtained a RMSd (root-mean-square deviation) of 0.915 Å, when aligning chain A (646 atoms), 0.582 Å when aligning chain B (614 atoms) and 15.723 Å (1589 atoms) when aligning the entirety of both chains, showing that the folding of each chain was preserved but the quaternary structure was lost. From this primary template and model, we learned the importance of the correct oligomerization and also verify that the Ca^{2+} indeed coordinates in the loop defined by the residues Tyr27, Gly29 and Gly31, although not for a considerable time. We suspect that the conformation of the loop seen in the 5TFV structure is the result of the lack of Ca^{2+}, not the reason for its absence.

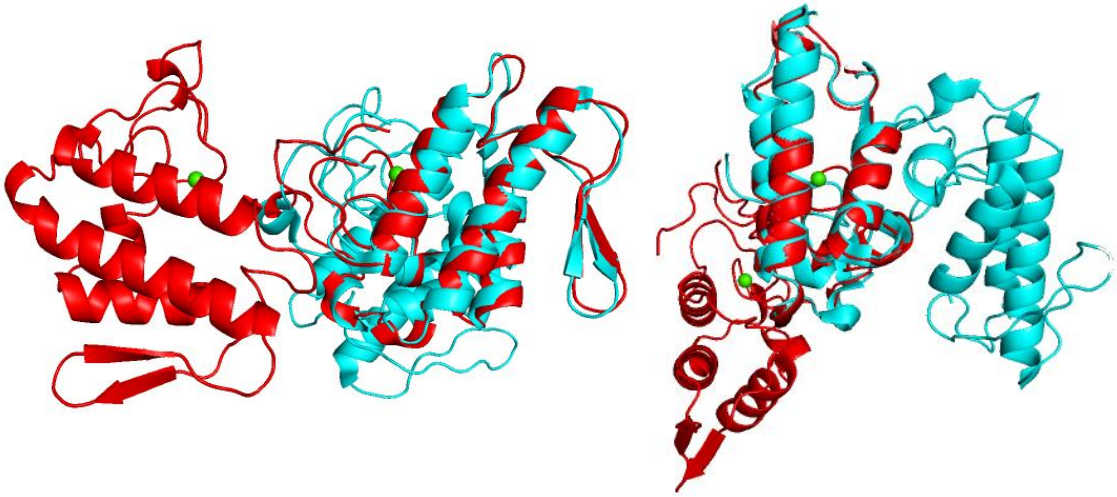


Fig. 13 - Structure of 5TF (blue) and 1JIA (red) superimposed. We can observe the different orientation of the two chains, in each structure. Image obtained using PyMOL.[26]

Following the review of other research papers[31], PDB ID: 1FXF was explored. This structure wasn't on the initial Swiss-Model top ranking, appearing listed on the footnotes. The absence of residues, in both the C- and N- terminals, in this structure its sequence identity with P20474 lower than the top 50 template options listed.

1FXF is the X-Ray structure of a crystallographic dimer of porcine pancreatic sPLA$_2$ and, like most mammalian sPLA$_2$s, it shares a very high degree of similarity with snake PLA$_2$s.[21] Additionally, the overall three-dimensional structure of pancreatic PLA$_2$s is very similar to that of another Asp49 PLA$_2$.[57]

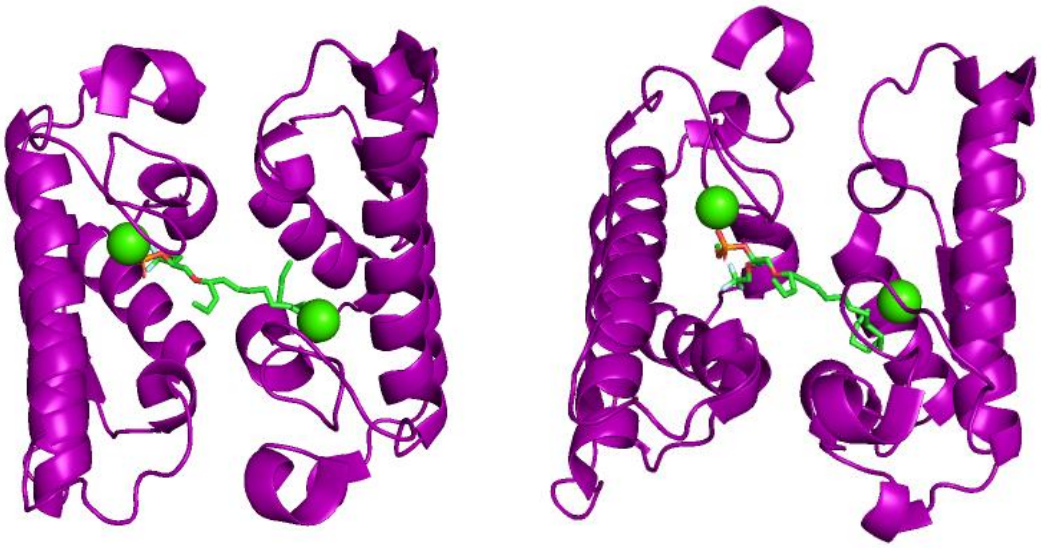


Fig. 14 - Structure of 1FXF (purple). MJ33 is shared between the two monomers. Backbone (purple), Ca^{2+} (green) in CPK colouring and MJ33(green) coloured by element. Image obtained using PyMOL.[26]

This structure holds five phosphate anions bound and one molecule of a tetrahedral mimic MJ33 (1-hexadecyl-3-(trifluoroethyl)-sn-glycero-2-phosphomethanol) harbored between the two subunits of the dimer. The sn-2 phosphate of MJ33 is coordinated in the active site of one subunit, to the calcium, and the alkyl chain extends into the active site slot of the second subunit across the subunit-subunit interface. [31]

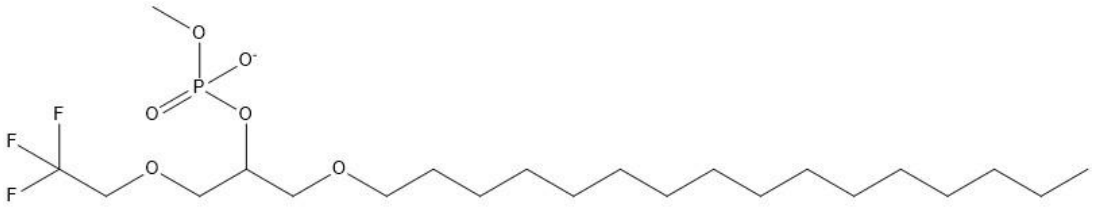


Fig. 15 - Structure of the 1FXF inhibitor MJ33. Figure obtained using ChemDraw 20.1

PLA$_2$s are structurally very well conserved. Despite having significant lower sequence identity than other templates, 1FXF retains key structural elements other templates lacked, so a compromise was made. Comparison of RMSd between models created with BLAST and HHblits was done using PyMOL. 1FXF and 5TFV have a RMSd of 1.130 Å (591 atoms), when aligning chain A, 1.049 Å (501 atoms) when aligning chain B and 4.478 Å (1362 atoms) when aligning the entirety of both proteins.

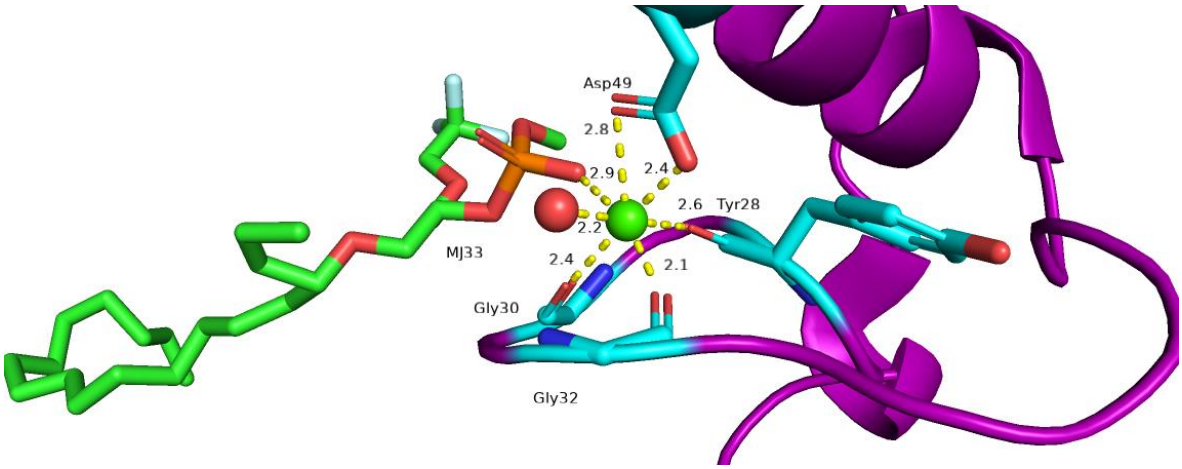


Fig. 16 - Heptacoordination of Ca^{2+} in 1FXF. The sn-2 phosphate of MJ33 is coordinated to the calcium of subunit A, the hexadecyl tail extends into subunit B (not represented). Backbone (purple), Ca^{2+} (green) in CPK colouring, coordinating residues (blue) and MJ33 (green) coloured by element. Bond distances in Ångström Image obtained using PyMOL.[26]

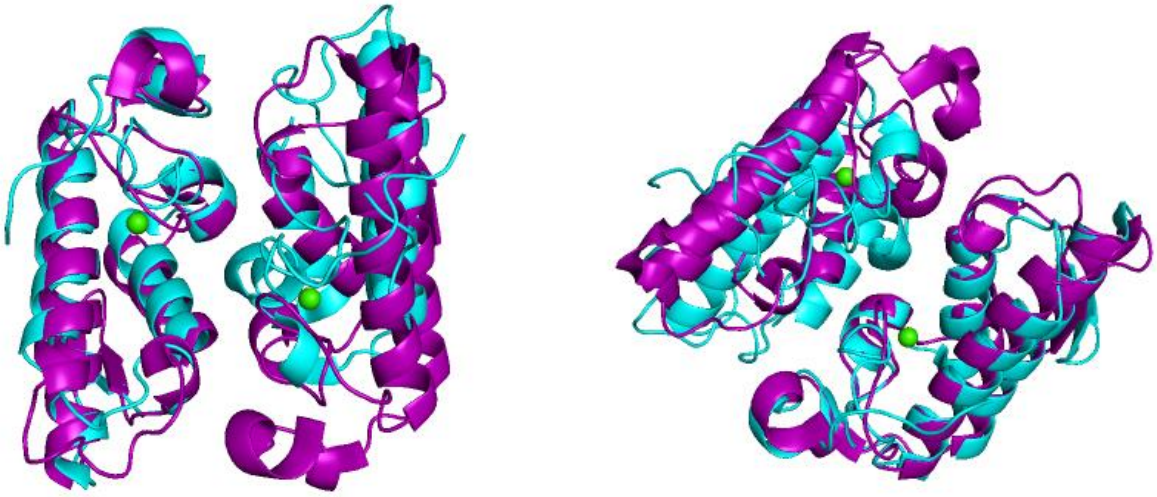


Fig. 17 - Structure of 5TF (blue) and 1FXF (purple) superimposed. The conformation of the two dimers is analogous. Image obtained using PyMOL.[26]

Four models were created in total, one for each chain of the template used in the modelling process. Despite having a higher sequence identity, Model 2 and Model 4 were monomeric structures and, therefore, not of interest. Model 1 and Model 2 had the same performance in all measurements, and upon inspection they were identical. We downloaded the PDB file of Model 1.

Table 3 - Model evaluation measurements for 1FXF template.

1FXF	**Model 1**	Model 2	**Model 3**	Model 4
Template	1fxf chain A	1fxf chain B	1fxf chain A	1fxf chain B
Found By	HHblits	HHblits	BLAST	BLAST
GMQE	0.61	0.29	0.61	0.29
Sequence Identity	39.82 %	49.21 %	39.82 %	49.21 %
RMSd	0.957 (649 atoms) all protein 1.402 (727 atoms) chain A 0.985 (656 atoms) chain B	0.988 (327 atoms) only chain A	0.957 (649 atoms) all protein 1.402 (727 atoms) chain A 0.985 (656 atoms) chain B	0.988 (327 atoms) only chain A

3.2.2. Model Building

The model obtained by Swiss-Model require substantial processing before running a minimization and dynamic.
The NOBA cofactor is not identified by the AMBER force-field by default, so parametrization was necessary, identifying all constituting atoms, using XLEaP.

NOBA was positioned in accordance with the orientation of MJ33 in 1FXF, which we supposed is the correct positioning as this is the location of a hydrophobic pocket. The PyMol tool pair_fit, which fits pairs of selected atoms between two models, was used and

these coordinates were saved in the models PDB file. NOBA is not a good candidate for docking, due to its very flexible aliphatic chain so the docking of NOBA was not the preferred approach. Since previous research doesn't explore this substrate and docking is particularly difficult for this molecule.

Ca^{2+} ions were added, by aligning both the model with 1FXF, and transferring those coordinates to the PDB file. In Fig. 18, the distortion of the Ca^{2+}-coordinating loop of the initial structure is made apparent when contrasted to both 1JIA and 1FXF.

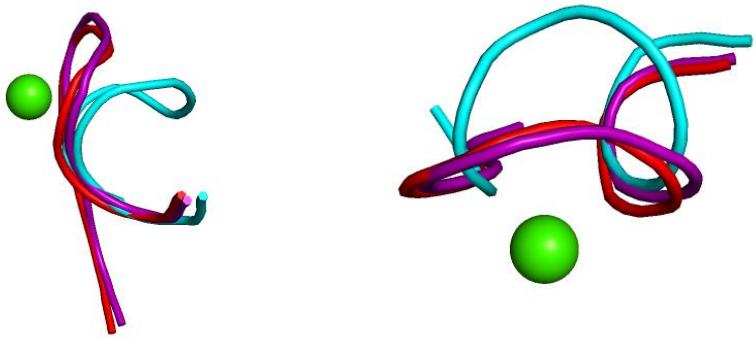


Fig. 18 - Loop comparison between 5TF (blue), 1JIA (red) and 1FXF (purple). The absence of this cation alters drastically this structure. Ca^{2+} (green) in CPK colouring. Image obtained using PyMOL.[26]

The PDB file was altered to correct the exchange of a leucine (in UniProt ID: P20474), in the position 113, for a valine (in PDB ID: 5TFV), due to the necessary modeling of the side chains performed at the time of the crystallization, because of the poor resolution of data (most likely with another toxin that had valine in that position).

And finally, a region, composed of nine residues on the C-terminal of 5TFV, was not present in 1FXF, and therefore these chains were added in the model, with the same technique

3.2.3. Protonation states and waters

Through xleap a waterbox was created around the protein, to encompass the structure and the addition of counterions necessary to neutralize the system, and to reproduce the salt concentration used in experiments

The protonation state of residues must be checked due to the lack of hydrogens in initial crystallographic structures. The server ProteinPrepare[59] was used, a web application

that allows the estimation of charge states and optimization of the hydrogen bonding network of protonated structures. For some residues that were flagged in this site, therefore inspection and analysis of the surrounding environment of a few residues were necessary.

We check the protonation state of our molecule at a pH of 7.4 (physiological pH). Lys76 and His49 are marked and need manual revision, which requires the attention of the surrounding residues and previous knowledge of the protein. Histidines, due to their charge being much more delocalized, require evaluation, to know where the hydrogen should be added in (two potential sites, 3 protonation states, HID, HIE and HIP). His49 should be in the HIE state of protonation (neutral, ε-nitrogen protonated) so it can bind to Asp99 with that hydrogen. We also modeled Lys76 protonated.

Editing the name of cysteine residues involved in the 7 disulphide bridges (from CYS to CYX) also ensures that LEaP chooses the correct residue template for cysteines linked via disulphide bonds. These cysteines were identified in the UniProt entry of P20474.

XLEaP, after adding missing hydrogens, solvent water molecules, counterions and defining disulphide bonds, creates the prmtop and inpcrd files (topology and parameters) of the prepared system. The prmtop file contains information about parameters and topology, such as the FFs used, while inpcrd carries the coordinates of all atoms system, as well as velocities in case of molecular dynamics.

3.3. Molecular Dynamics

The protein coordinates described in the correct PDB file of the model are used as starting configurations for the MD simulations.
Simulations were conducted using Sander, an Amber module which carries out energy minimization, molecular dynamics, and NMR refinements. All simulations were performed with a timestep of 2 fs
Simulations were first initialized by minimizing the system's energy, followed by molecular dynamics. Visual inspection was executed throughout these processes utilizing the VMD (Visual Molecular Dynamics software) and PyMol programs.

Table 4 - Parameters for the MD simulations.

Parameter	Value
Forcefield	Amber ff14SB
Water model	TIP3P
Temperature	310 K
Temperature control	Langevin thermostat
Pressure	1 bar
Box dimension	Free to ft the desired pressure
Periodic boundary conditions	Isotropic position scaling
Integration step	2 fs
Harmonic force constant	50 kcal mol^{-1}Å^2

3.3.1. Energy Minimization

After homology modelling, positioning of the substrate, completion of the C-terminal and addition of waters, it is important to relax the system from existing tensions that result from the modelling process.

Energy minimization is done successionally: in the first step (10,000 steps), we minimize the solvent, by restraining everything except waters; secondly, the hydrogens (10,000 steps); followed by minimization of the side-chains of the protein (30,000 steps); and in the final stage (80,000 steps) the entire system is minimized, no constraints are applied.

3.3.2. Equilibration

Because we seek to run a simulation in a particular thermodynamic ensemble at a specific state point (in this case, physiological conditions) and collect data for analysis, it is necessary to bring the system to the said state point in the most relevant configuration in the equilibrium ensemble of interest.

This procedure is also computed in steps. Equilibrating first at constant volume, bringing the system to the desired target temperature, in this case 310 K, before equilibrating to the target pressure, 1 bar. A thermostat will add or remove heat from the system (Langevin thermostat).

In this study, equilibration lasted 86080 ps (86.080 ns) where an NVT (constant temperature, constant volume ensemble) ensemble ran for 80 ps, allowing the heating

of the system (a); several NPT (constant temperature, constant pressure ensemble) equilibrations were performed, for each type of selection we introduced in the model: (b) the solvent restrictions are relaxed, as pressure remains constant, (c) the Ca^{2+} coordinating residues are relaxed, (d) NOBA and Ca^{2+} ions are relaxed and (e) the disulphide bridges are relaxed.

Table 5 - Steps performed in minimization stage.

Minimization step	Simulation length
NVT	80 ps
NPT - not water	6 ns
NPT- not loops	20 ns
NPT - not active site	20 ns
NPT - not ligands	20 ns
NPT - not disulphide bridge	20 ns

3.3.3. Production

Once equilibration is complete, we enter a production phase, which ran in total for 40 ns (10 steps of 4 ns) and where both temperature and pressure are kept constant, but volume varying.

Analysis of results includes the calculation and comparison of RMSd and RMSF (root-mean-square fluctuation), as well as visual inspection of the system.

An initial MD was carried out, with the parameters defined previously, with the exception of the length of the production phase. In this simulation, we implemented 56 ns, extending it to give enough time for the molecule to assume a favorable conformation. As a result, dimerization was preserved, as well as Ca^{2+} coordination. NOBA, however, in the production step of the dynamic started to break away from the initial position (distance to Ca^{2+} of 6.09 Å) to the point where ester cleavage was no longer feasible.

Cpptraj is an Amber program for processing coordinate trajectories and data files, for the analysis of simulations. RMSd of both the protein and the substrate were computed, as well as RMSF. RMSF measures the average deviation of a particle (e.g. a protein residue) over time from a reference position (typically the time-averaged position of the particle). Thus, RMSF analyzes the portions of structure that are fluctuating from their mean structure the most (or least).

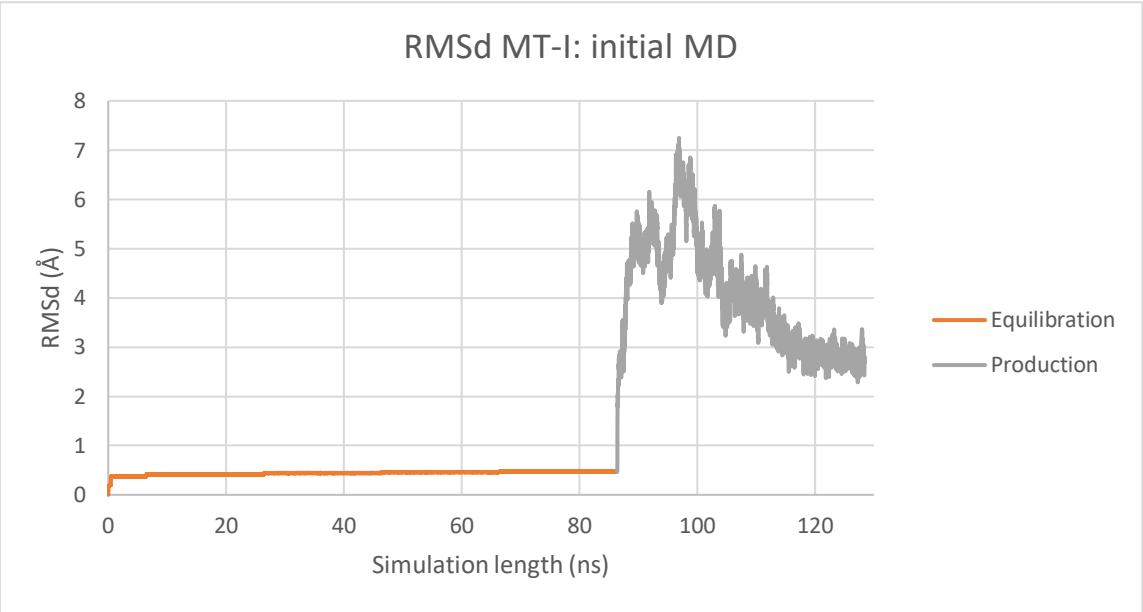


Fig. 19 - RMSd of protein, throughout the initial 140 ns MD simulation.

Some frames of the simulation were removed, due to anomalies caused by not defining periodic boundary conditions. We see that after equilibration ends (86 ns) and production begins there is a big increase in RMSd values, this is due to the release of constraints. The RMS deviations for the substrate indicate NOBA is stable during the simulation.

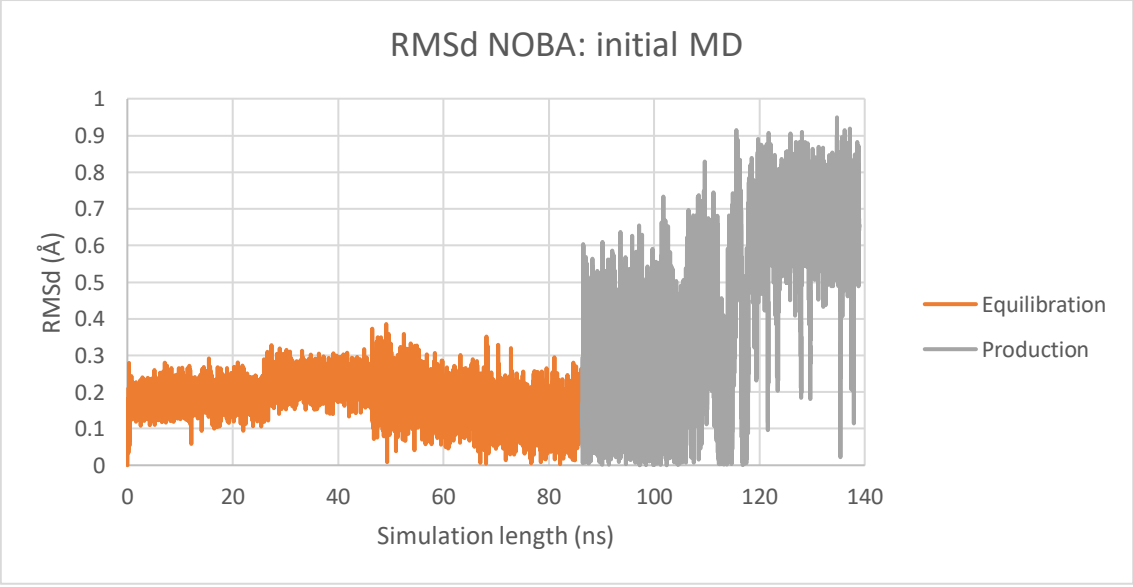


Fig. 20 - RMSd of NOBA, throughout the initial 140 ns MD simulation.

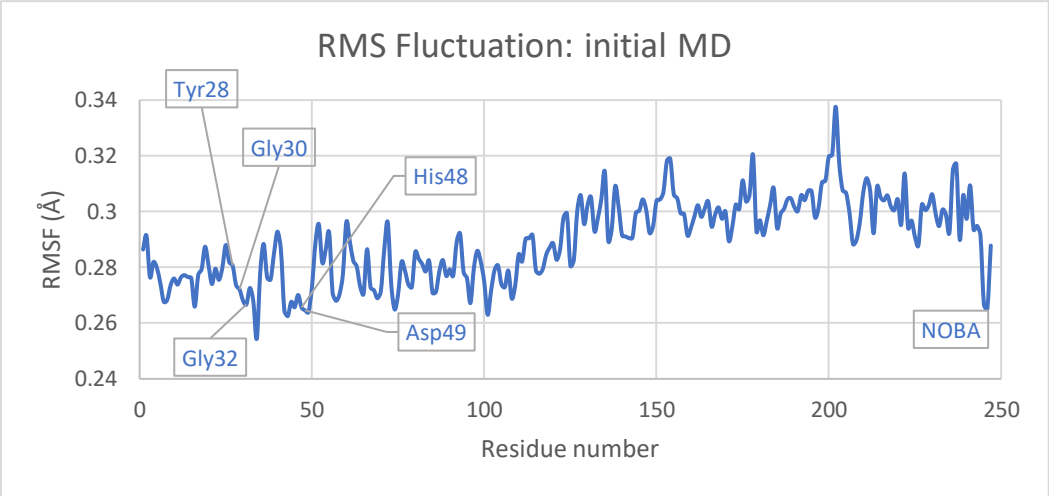


Fig. 21 - RMSF plot of the initial 140 ns MD simulation.

In RMSF, high values correspond often to loop regions, due to having more conformational flexibility. NOBA had a stabilizing effect in the Ca^{2+}-coordinating loop, which is made apparent by lower fluctuation values in the binding residues (Fig. 21), when compared to the other chain, lacking NOBA. This effect is characteristic of introducing a ligand at a protein binding pocket and doesn't translate conformational relevance: ligand bonding stiffens and stabilizes the protein regardless. In fact, visual inspection through the MD revealed the substrate acyl ester bond to be too distant from calcium to be hydrolyzed.

We run a second MD, restraining NOBA to an initial 2.6 Å distance from Ca^{2+}.

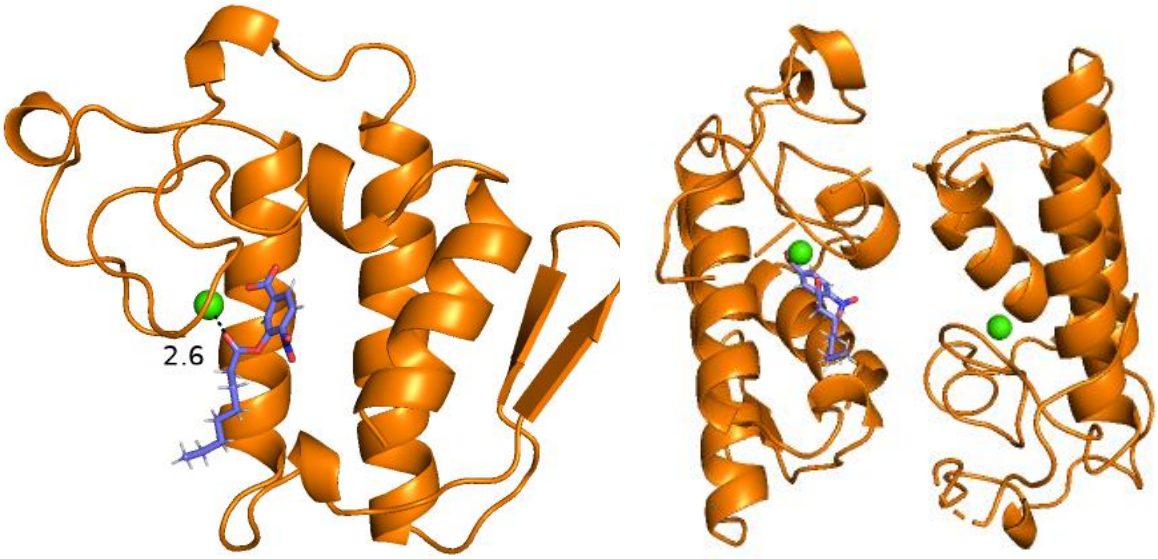


Fig. 22 - Initial MT-I conformation for MD. NOBA (deep blue) coloured by element, Ca^{2+} (green) in CPK colouring, and the beginning structure of MT-I in orange. Bond distances in Ångström. Image obtained using PyMOL.[26]

In the absence of substrate, the Ca^{2+} ion is held within its loop in a heptacoordinate fashion by two water molecules, three mainchain carbonyls and the side-chain oxygens of Asp49. In the presence of substrate, the water molecules are replaced by the non-bridging oxygen of the ester together with the oxyanion as described above.

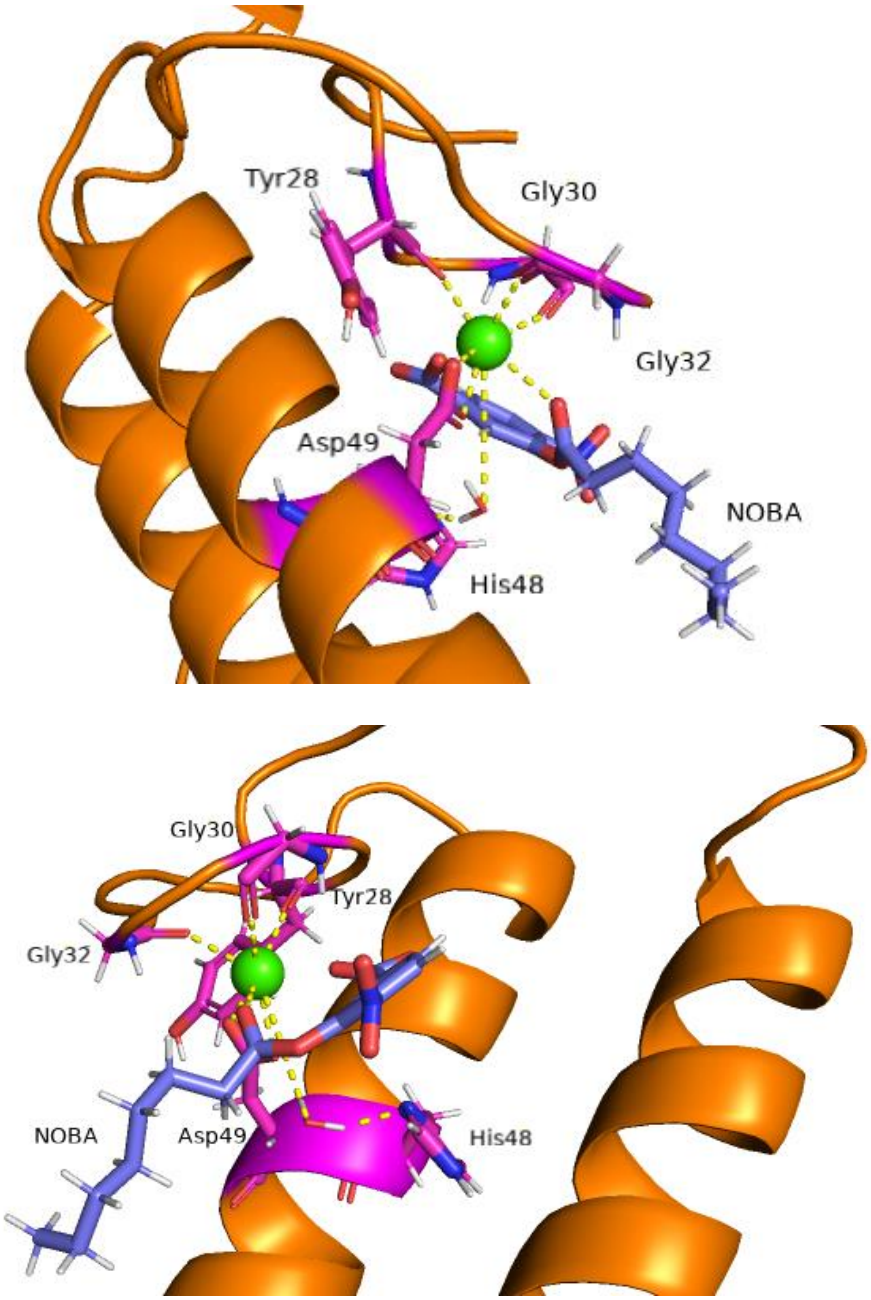


Fig. 23 - Heptacoordination of Ca^{2+} in minimized MT-I structure. NOBA (deep blue) colored by element, the beginning structure of MT-I in orange, Ca^{2+} (green) and water in CPK colouring. Image obtained using PyMOL.[26]

After minimization, Ca^{2+} heptacoordination remains: (a) by the three mainchain carbonyls residues, bond distances range from 2.4 to 2.5 Å; (b) the two side-chain oxygens of Asp49, bond distance of 2.4 Å; (c) NOBA's ester, at 2.9 Å; and finally, an 'assisting' water, at 4.8 Å, that bridges with His48, at 1.9 Å.

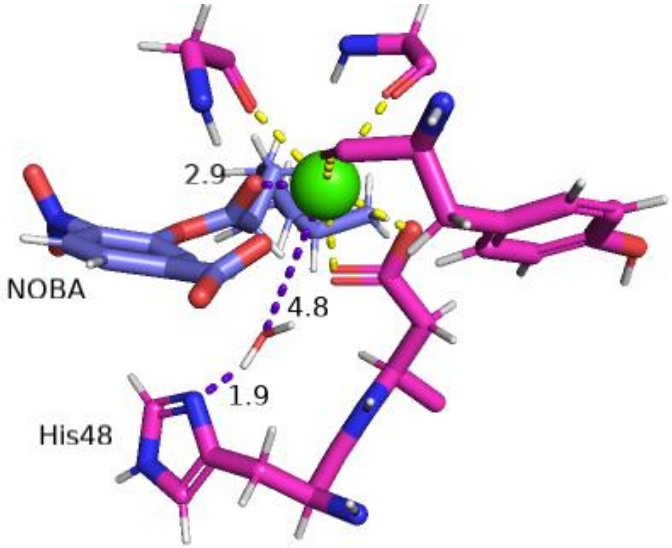


Fig. 24 - Coordination of Ca^{2+} in minimized MT-I structure. NOBA (deep blue) and protein residues (pink) colored by element, Ca^{2+} (green) and water in CPK colouring. Bond distances in Ångström. Image obtained using PyMOL.[26]

After equilibration, the Ca^{2+}-coordinating loop maintains previous bonds, however, NOBA's acyl group is now at 3.0 Å and the bridging water at 5.0 Å, which is in turn at a distance of 2.2 Å from His48.

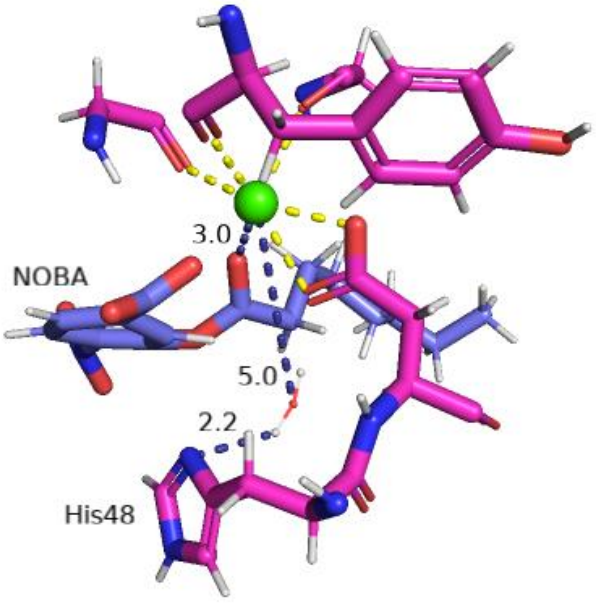


Fig. 25 - Coordination of Ca^{2+} in MT-I structure after equilibration. NOBA (deep blue) and protein residues (pink) colored by element, Ca^{2+}(green) and water in CPK colouring. Bond distances in Ångström. Image obtained using PyMOL.[26]

In the last step of production, we observe that NOBA's acyl group is now at 2.4 Å and the bridging water at 2.5 Å, which is in turn at a distance of 3.8 Å from His48. Residues Tyr28, Gly30, Gly32 and Asp49 hold previous bonds.

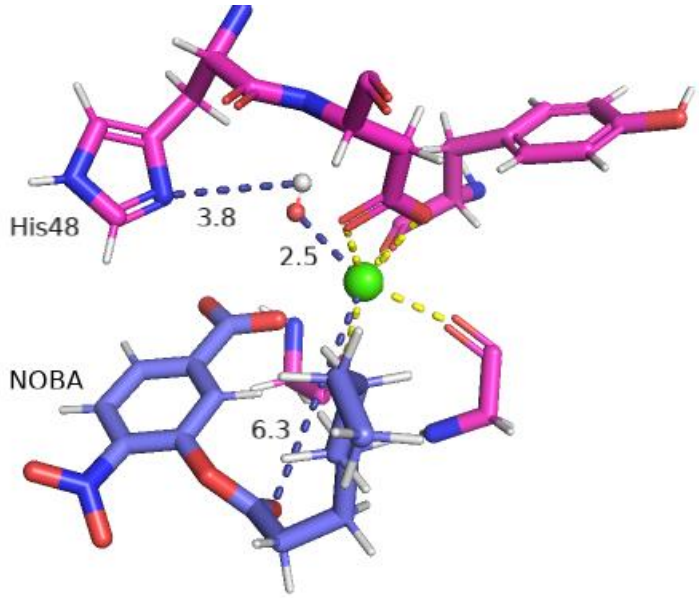


Fig. 26 - Coordination of Ca^{2+} in MT-I structure after equilibration. NOBA (deep blue) and protein residues (pink) colored by element, Ca^{2+} (green) and water in CPK colouring. Bond distances in Ångström. Image obtained using PyMOL.[26]

Similarly to before, in the plots RMSd of both the protein and NOBA, increase drastically as the simulation enters the production phase, however we observe overall lower values of deviation in MT-I.

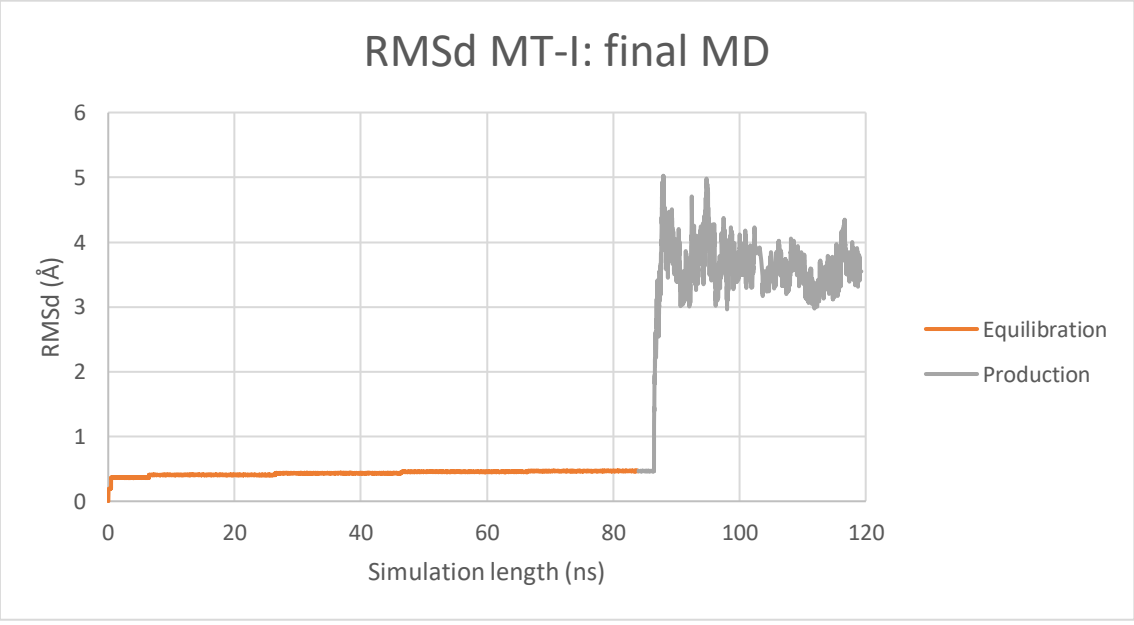


Fig. 27 - RMSd of all protein, throughout the final 122 ns MD.

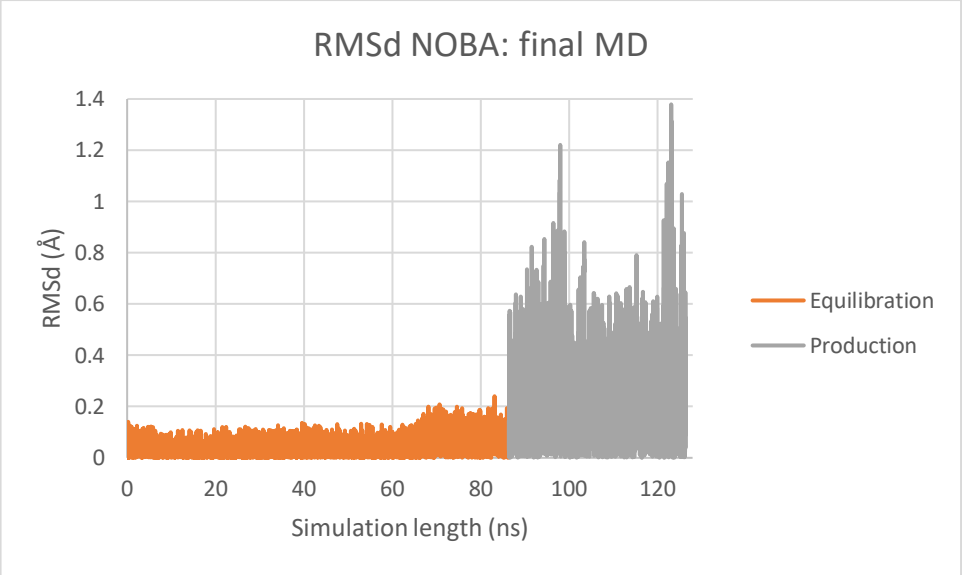


Fig. 28 - RMSd of ligand NOBA, throughout the final 122 ns MD.

The final MD structure, albeit not ideal, seems to be a fair representation of MT-I and an adequate start for more sophisticated simulations.



Fig.29 - Superimposition of visualizations belonging to initial model (green), minimization (blue), equilibration (pink) and production (yellow). Image obtained using PyMOL.[26]

4. Conclusions and future work

Basic sPLA$_2$s are powerful and prevalent enzymes in snake venoms Their ineffective neutralization by current antivenoms make them appealing research targets. The goal of this dissertation was to create a model for MT-I, a Ca^{2+} dependent, catalytic active Asp49 sPLA$_2$.

These results in this work elucidate structural and functional details of MT-I. Most likely, the binding pose occurs so rarely that it's very difficult to detect it, as the protein spends most of the time in its inactive form. The structural conformation that permits the reaction is rare. As has been evidenced by experimental data, low catalytic efficiency is unavoidable as there are only a few binding modes that are conducive to catalysis.[27] However, the mobility of the enzyme-substrate complex should allow the substrate with suboptimal structure to react with enzyme. The fact that MT-I bind to such "nonspecific substrates", like NOBA, is also evidence of the flexibility of the molecule, which is essential for catalysis to ensue, preventing the substrate from being locked into a nonproductive binding mode.

Enzyme conformational rearrangements, such as the probable dissociation into monomers of sPLA$_2$s upon contact with phospholipid membranes, might occur in a larger time scale, in the milliseconds (μs), than those covered during standard MD simulations, in the femto-seconds (fs) to microseconds (ms). This is also the case for HIV-I protease.[61] In these situations, running longer MDs isn't realistic, given the computational cost. Repeated observations ensures more statistically significant conclusions regarding a particular phenomenon.[60] Molecules have many degrees of freedom, making it very hard to find the global minimum energy conformation and simulations can get trapped in local minima, so sampling is one way to strengthen results.

With a reasonable model of MT-I, we can, in the future, apply QM/MM methods to determine its catalytic mechanism and virtual screening to search for possible inhibitors. The creation of synthetic inhibitors for basic PLA$_2$s, a major family of proteins in snake venoms, will further efforts of designing therapies more efficacious than the current available.

5. References

1. Gutiérrez, J. M. *et al.* Snakebite envenoming. *Nat. Rev. Dis. Prim.* **3**, 17063 (2017).

2. The Lancet. Snake-bite envenoming: a priority neglected tropical disease. *Lancet* **390**, 2 (2017).

3. Williams, D. *et al.* The Global Snake Bite Initiative: an antidote for snake bite. *Lancet* **375**, 89–91 (2010).

4. Harrison, R. A., Hargreaves, A., Wagstaff, S. C., Faragher, B. & Lalloo, D. G. Snake envenoming: A disease of poverty. *PLoS Negl. Trop. Dis.* **3**, (2009).

5. Gutiérrez, J. M., Theakston, R. D. G. & Warrell, D. A. Confronting the neglected problem of snake bite envenoming: The need for a global partnership. *PLoS Med.* **3**, 0727–0731 (2006).

6. Calvete, J. J., Sanz, L., Angulo, Y., Lomonte, B. & Gutiérrez, J. M. Venoms, venomics, antivenomics. *FEBS Lett.* **583**, 1736–1743 (2009).

7. Kang, T. S. *et al.* Enzymatic toxins from snake venom: Structural characterization and mechanism of catalysis. *FEBS J.* **278**, 4544–4576 (2011).

8. Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A. & Fry, B. G. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* **28**, 219–229 (2013).

9. Alape-Girón, A. *et al.* Snake venomics of the lancehead pitviper bothrops asper. Geographic, individual, and ontogenetic variations. *J. Proteome Res.* **7**, 3556–3571 (2008).

10. Tasoulis, T. & Isbister, G. K. A review and database of snake venom proteomes. *Toxins (Basel).* **9**, (2017).

11. Maduwage, K., O'Leary, M. A. & Isbister, G. K. Diagnosis of snake envenomation using a simple phospholipase A2 assay. *Sci. Rep.* **4**, 1–4 (2014).

12. Gutiérrez, J., Ownby, C. L. & Odell, G. V. Isolation of a myotoxin from Bothrops asper venom: Partial characterization and action on skeletal muscle. *Toxicon* **22**, 115–128 (1984).

13. Guti, M., Montecucco, C. & Postle, A. D. Muscle phospholipid hydrolysis by Bothrops asper Asp49 and Lys49 phospholipase A 2 myotoxins – distinct mechanisms of action. **280**, 3878–3886 (2013).

14. Rueda, A. Q. *et al.* Biochemical Characterization , Action on Macrophages , and Superoxide Anion Production of Four Basic Phospholipases A 2 from Panamanian Bothrops asper Snake Venom. **2013**, 1–9 (2013).

15. Gutiérrez, J. M. & Ownby, C. L. Skeletal muscle degeneration induced by venom phospholipases A 2: Insights into the mechanisms of local and systemic myotoxicity. *Toxicon* **42**, 915–931 (2003).

16. Lomonte, B. & Gutiérrez, J. M. Phospholipases A2 from viperidae snake venoms: How do they induce skeletal muscle damage? *Acta Chim. Slov.* **58**,

647–659 (2011).

17. Massimino, M. L. *et al.* Cell surface nucleolin interacts with and internalizes Bothrops asper Lys49 phospholipase A 2 and mediates its toxic activity. 1–14 (2018) doi:10.1038/s41598-018-28846-4.

18. Gutiérrez, J. M. & Lomonte, B. Phospholipases A2: Unveiling the secrets of a functionally versatile group of snake venom toxins. *Toxicon* **62**, 27–39 (2013).

19. Šribar, J. & Križaj, I. Secreted phospholipases A2 - Not just Enzymes. *Acta Chim. Slov.* **58**, 678–688 (2011).

20. Zuliani, J. P., Gutiérrez, J. M. & Teixeira, C. Signaling pathways involved in zymosan phagocytosis induced by two secreted phospholipases A2 isolated from Bothrops asper snake venom in macrophages. *Int. J. Biol. Macromol.* **113**, 575–582 (2018).

21. Tonello, F. & Rigoni, M. Cellular Mechanisms of Action of Snake Phospholipase A2 Toxins. *Snake Venoms* 1–14 (2015) doi:10.1007/978-94-007-6648-8_26-1.

22. Urs, N. A. N. *et al.* Implications of phytochemicals in snakebite management: Present status and future prospective. *Toxin Rev.* **33**, 60–83 (2014).

23. Mora-Obando, D., Díaz, C., Angulo, Y., Gutiérrez, J. M. & Lomonte, B. Role of enzymatic activity inmuscle damage and cytotoxicity induced by Bothrops asper Asp49 phospholipase A2 myotoxins: Are there additional effector mechanisms involved? *PeerJ* **2014**, (2014).

24. Mora-Obando, D., Fernández, J., Montecucco, C., Gutiérrez, J. M. & Lomonte, B. Synergism between basic Asp49 and Lys49 phospholipase A2 myotoxins of viperid snake venom in vitro and in vivo. *PLoS One* **9**, (2014).

25. Salvador, G. H. M., dos Santos, J. I., Lomonte, B. & Fontes, M. R. M. Crystal structure of a phospholipase A2 from Bothrops asper venom: Insights into a new putative “myotoxic cluster”. *Biochimie* **133**, 95–102 (2017).

26. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

27. Cho, W., Markowitz, M. A. & Kézdy, F. J. A New Class of Phospholipase A2 Substrates: Kinetics of the Phospholipase A2 Catalyzed Hydrolysis of 3-(Acyloxy)-4-nitrobenzoic Acids. *J. Am. Chem. Soc.* **110**, 5166–5171 (1988).

28. Cho, W. & Kézdy, F. J. Chromogenic substrates and assay of phospholipases A2. *Methods Enzymol.* **197**, 75–79 (1991).

29. Xiao, H., Pan, H., Liao, K., Yang, M. & Huang, C. Snake Venom PLA2, a Promising Target for Broad-Spectrum Antivenom Drug Development. *Biomed Res. Int.* **2017**, (2017).

30. Bultrón, E., Gutiérrez, J., & Thelestam, M. Effects of Bothrops asper (terciopelo) myotoxin III, a basic phospholipase A2, on liposomes and mouse gastrocnemius muscle. *Toxicon* **31**, 217–222 (1993).

31. Pan, Y. H., Epstein, T. M., Jain, M. K. & Bahnson, B. J. Five coplanar anion binding sites on one face of phospholipase A2: Relationship to interface binding. *Biochemistry* **40**, 609–617 (2001).

32. Tsai, Y. C., Yu, B. Z., Wang, Y. Z., Chen, J. & Jain, M. K. Desolvation map of the i-face of phospholipase A2. *Biochim. Biophys. Acta - Biomembr.* **1758**, 653–665 (2006).

33. Schürer, G., Lanig, H. & Clark, T. The mode of action of phospholipase A2: Semiempirical MO calculations including the protein environment. *J. Phys. Chem. B* **104**, 1349–1361 (2000).

34. Bahnson, B. J. Structure, function and interfacial allosterism in phospholipase A2: Insight from the anion-assisted dimer. *Arch. Biochem. Biophys.* **433**, 96–106 (2005).

35. Consortium, T. U. UniProt : the universal protein knowledgebase in 2021. 1–10 (2021) doi:10.1093/nar/gkaa1100.

36. Berman, H. M., Westbrook, J. & Feng, Z. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

37. PDB Data Distribution by Experimental Method and Molecular Type. https://www.rcsb.org/stats/summary.

38. Li, X. X., Xu, X. D., Dan, Y. Y. & Zhang, M. L. The factors during protein crystallization: A review. *Crystallogr. Reports* **53**, 1261–1266 (2008).

39. Souza, D. H. F., Selistre-De-Araujo, H. S. & Garratt, R. C. *Determination of the three-dimensional structure of toxins by protein crystallography*. *Toxicon* vol. 38 (2000).

40. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

41. Wang, Z. How Fold. **191**, 186–191 (1996).

42. Amber, C. Amber 2021 Reference Manual. (2021).

43. Frenkel, D. & Smit, B. Understanding molecular simulation _ from algorithms to applications (2002, Academic Press).pdf. (2002).

44. Harrison, J. A. *et al.* Review of force fields and intermolecular potentials used in atomistic computational materials research. *Appl. Phys. Rev.* **5**, (2018).

45. Braun, E. *et al.* Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, 1–28 (2019).

46. Unknown - Unknown - A.R.Leach---Molecular Modelling Principles an.pdf.pdf.

47. Guvench, O. & MacKerell, A. D. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **443**, 63–88 (2008).

48. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Ka, and P. A. K. (2021). Amber 2021, University of California, San Francisco.

49. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., … Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30(10)**, 1545–1614 (2009).

50. Scott, W. R. P. *et al.* The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).

51. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**,

198–210 (2013).

52. Ngan, S. C., Hung, L. H., Liu, T. & Samudrala, R. Scoring functions for de novo protein structure prediction revisited. *Methods Mol. Biol.* **413**, 243–281 (2007).

53. Fernandes, H. S. & Bioquímica, M. Computational studies addressed to Histidine decarboxylase Computational studies addressed to Histidine decarboxylase. (2016).

54. Fernandes, C. A. H., Borges, R. J., Lomonte, B. & Fontes, M. R. M. A structure-based proposal for a comprehensive myotoxic mechanism of phospholipase A2-like proteins from viperid snake venoms. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 2265–2276 (2014).

55. Borges, R. J. *et al.* Functional and structural studies of a Phospholipase A2-like protein complexed to zinc ions: Insights on its myotoxicity and inhibition mechanism. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 3199–3209 (2017).

56. Da Silva Giotto, M. T. *et al.* Crystallographic and spectroscopic characterization of a molecular hinge: Conformational changes in Bothropstoxin I, a dimeric Lys49- phospholipase A2 homologue. *Proteins Struct. Funct. Genet.* **30**, 442–454 (1998).

57. Brunie, S., Bolin, J., Gewirth, D. & Sigler, P. B. The refined crystal structure of dimeric phospholipase A2 at 2.5 Å. *J. Biol. Chem.* **260**, 9742–9749 (1985).

58. Zhao, K., Song, S., Lin, Z. & Znou, Y. Structure of a basic phospholipase A2 from agkistrodon halys pallas at 2.13 å resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 510–521 (1998).

59. Martínez-Rosell, G., Giorgino, T. & De Fabritiis, G. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **57**, 1511–1516 (2017).

60. Grossfield, A. *et al.* Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, 1–24 (2019).

61. Calixto, A. R., Ramos, M. J. & Fernandes, P. A. Conformational diversity induces nanosecond-timescale chemical disorder in the HIV-1 protease reaction pathway. *Chem. Sci.* **10**, 7212–7221 (2019).