UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Instituto de Biociências de Botucatu

Departamento de Física e Biofísica

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia

“Structural studies of PLA_2-like toxins and development of the structure solution method SEQUENCE SLIDER”

RAFAEL JUNQUEIRA BORGES

Botucatu – SP
January / 2017

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Instituto de Biociências de Botucatu

Departamento de Física e Biofísica

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Instituto de Biología Molecular de Barcelona

Departamento de Biología Estructural

PARC CIENTIFIC DE BARCELONA

Crystallographic Methods Laboratory

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA)

PROGRAMA DE DOCTORAT EM BIOTECNOLOGIA

“Structural studies of PLA_2-like toxins and development of the structure solution method SEQUENCE SLIDER”

Doctoral thesis presented by Rafael Junqueira Borges, Bachelor degree in Biomedical Sciences, Ms.C. in Genetics, for the degree of Doctor (Ph.D.) from the University of State of Sao Paulo and the University of Barcelona.

Thesis director: Dr. Marcos Roberto de Mattos Fontes, UNESP Prof.

Thesis co-director: Dr. Isabel Usón Finkenzeller, ICREA Prof.

Botucatu – SP
January / 2017

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Borges, Rafael Junqueira.
Structural studies of PLA_2-like toxins and development of the structure solution method SEQUENCE SLIDER / Rafael Junqueira Borges. - Botucatu, 2017

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Marcos Roberto de Mattos Fontes
Coorientador: Isabel Usón Finkenzeller
Capes: 20901003

1. Cobra venenosa - Veneno. 2. Fosfolipases. 3. Cristalografia. 4. Biologia molecular.

Palavras-chave: Crystallography; Methods development; PLA2-like proteins; Phasing; Snake venom.

To all those who fight to eradicate human ignorance and suffering.

Acknowledgements

I express my sincere acknowledgements to the institutions and people that contributed for the development of my thesis.

Dr. Marcos Fontes, for the incentive and teachings since my early undergraduate research and for the given background that allowed me to adventure over new scientific fields and laboratories

Dr. Isabel Usón, for the innumerous scientific discussions personally and via video-conference, for the patience, persistence and dedication to introduce me to phasing and programming, and for trusting my work.

Dr. Ney Lemke, for guiding me over mathematics, normal model analysis and scientific ideas with metaphoric examples.

Dr. Mario Oliveira, for the aid collecting SAXS experiments and interpreting results.

Dr. George Sheldrick, for the hospitality receiving me in his laboratory in University of Goettingen and for being always available to answer any complicated question from chemistry to applied crystallography with an astonishing simplicity.

Dr. Randy Read, who received me in his laboratory in University of Cambridge, for the impressive scientific clear thoughts and for the assertive suggestions over my ideas.

Henrique Barcellos Campanelli, for the help in the experiments and for the opportunity to allow me to teach what I have acquired over my scientific career.

Our collaborators: Dr. Bruno Lomonte, Dr. Maria Laura Fernadez, Dr. Andreimar Soares, Dr. Luis Ponce-Soto, Dr. Sergio Marangoni, Dr. Salomón Huancahuire-Veja, for the various supplied snake venom proteins and inhibitors, and Dr. Ehmke Pohl, Dr. Isabel dos Santos, Edson José Comparetti, for the supplied crystallographic datasets.

LNLS-MX1&2 scientific instrumentation team, Dr. Ana Carolina de Mattos Zeri, Me. Alexandre Lo Bianco Dos Santos and Luciano Braga Candido, for the help setting the X-ray beamline to better meet our needs.

LNBIO mass spectrometry technicians, Dra. Bianca Alves Pauletti and Romênia Domingues, for the effort to aid sample preparation and results interpretation.

FAPESP, for the financial support during a 4-year project (Process number 2013/00873-4 and 2014/11182-5).

My colleagues in the Structural Molecular Laboratory and the Crystallographic Methods Laboratory, in special to Claudia and Massimo, for the enriching discussion and for pushing me to my best.

My colleagues in the laboratories I visited, for the effort to fasten my adaptation and for extending my scientific visit into a cultural exchange and discovery.

All my friends over the cities I lived during this sandwich PhD, Botucatu, Barcelona, Goettingen and Cambridge, for the unique experiences, relaxing moments and amazing trips.

My parents, brother and family, for the endless love, care and support.

"To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science."

Albert Einstein

ABSTRACT

Phospholipases A_2 (PLA_2s) are one of the main component of *Bothrops* genus snakes and one of the main responsible for muscular necrosis in envenomation, a consequence not neutralized by administration of antiophidian serum. These proteins are myotoxic by disrupting membrane phospholipids by a catalytic mechanism dependent of calcium or by perturbing membrane integrity by an independent of calcium mechanism not fully known. Usually, snake toxins are purified directly from the natural source, extracted venom, and purity is a challenge due to the co-existence of various isoforms. The objectives of this thesis were to understand the non-catalytic membrane perturbation mechanism by studying the snake myotoxins and to propose a new crystallographic method to deal with impure samples, called SEQUENCE SLIDER. We performed structural X-ray crystallography studies, complemented with other biophysical techniques, such as small-angle X-ray scattering, with three bothropic myotoxins in *apo* state and complexed to natural products and inhibitors. We proposed local and global measurements to characterize and relate these states to toxin function. With the method SEQUENCE SLIDER, we aided elucidation of structures with partial sequence known by evaluating different side chain against real space correlation coefficient calculated from diffraction data. Furthermore, we developed SEQUENCE SLIDER to increase the scope of the crystallographic phasing program ARCIMBOLDO to lower resolution than the usual 2 Å. In such cases, different hypotheses of sequences of the partial traces of polyalanine are evaluated simultaneously and pushed through autotracing until the structure is phased. We improved the myotoxic mechanism comprehension and developed a structure solution method for challenging crystallographic structures.

Keywords: Structural molecular biology; Phospholipases A_2-like; Snake venom; phasing; ARCIMBOLDO.

RESUMO

As fosfolipases A_2 (PLA_2s) são um dos maiores constituintes protéicos do veneno botrópico e um dos responsáveis pela necrose muscular, consequência esta não eficazmente neutralizada pela administração do soro antiofídico. Estas proteínas são tóxicas através do rompimento ou perturbação da membrana celular em um mecanismo catalítico dependente de cálcio e outro independente, sendo este último não totalmente elucidado. Usualmente, estas toxinas são obtidas diretamente do veneno das serpentes, sendo sua purificação um desafio pela co-existência de diferentes isoformas. O objetivo desta tese foi compreender o mecanismo miotóxico independente de cálcio através de estudos estruturais e propor nova metodologia que trate de dados cristalográficos de toxinas provenientes de amostras impuras, chamada SEQUENCE SLIDER. Para tanto, cristalografia e outras técnicas biofísicas, como espalhamento de raios X a baixo ângulo, serão utilizados para estudar três miotoxinas ofídicas em estado nativo e complexado com produtos naturais e inibidores. Nós propusemos medidas locais e globais para caracterizar e relacionar a estrutura dessas toxinas a função. Com o SEQUENCE SLIDER, pudemos elucidar as estruturas de toxinas inéditas cuja sequência era parcialmente conhecida. Esta nova metodologia proposta consiste em avaliar diferentes cadeias laterais contra o coeficiente de correlação em espaço real calculado a partir dos dados cristalográficos. Em paralelo, desenvolvemos o SEQUENCE SLIDER no âmbito do método cristalográfico *ab initio* ARCIMBOLDO com objetivo de aumentar seu escopo a dados com resolução mais baixa que 2 Å. Nestes casos, diferentes hipóteses de sequências para as soluções parciais de polialanina são avaliadas mutualmente e enviadas a autotraçamento até que grande parte da estrutura seja faseada. Pudemos melhorar a compreensão do mecanismo miotóxico e desenvolver um programa que poderá auxiliar resolução de estruturas cristalográficas desafiadoras.

Palavras chave: Biologia molecular estrutural; Fosfolipases A_2 homólogas; Veneno de serpente; faseamento; ARCIMBOLDO.

RESUM

Les fosfolipases A_2 (PLA_2s) de les serps del gènere Bothrops són un dels principals responsables de la necrosi muscular en l'enverinament, una conseqüència no neutralitzada per l'administració de sèrum antiofídic. Aquestes proteïnes són miotòxiques pel seu efecte disruptiu sobre la membrana fosfolipídica, bé per un mecanisme catalític dependent de calci o bé per un mecanisme no depenent de calci poc descrit a la bibliografia. En general, les toxines de serp es purifiquen directament de la seva font natural, el verí extret, i la purificació és un repte a causa de la coexistència de diverses isoformes. Els objectius d'aquesta tesi van ser entendre el mecanisme no catalític de pertorbació de la membrana mitjançant l'estudi de miotoxines de serp i proposar un nou mètode cristal·logràfic per fer front a les mostres impures, anomenat SEQUENCE SLIDER. Es van realitzar estudis estructurals de cristal·lografia de raigs X, complementats amb altres tècniques biofísiques, com SAXS (dispersió de raigs X d'angle petit), amb tres miotoxines en estat apo i acomplexades amb productes naturals i inhibidors. Hem proposat mesuraments locals i globals per caracteritzar i relacionar aquests estats amb la funció de la toxina. Amb el mètode SEQUENCE SLIDER, hem facilitat l'elucidació de les estructures amb seqüència parcial coneguda mitjançant l'avaluació de diferents cadenes laterals contra el coeficient de correlació de l'espai real calculat a partir de dades de difracció. D'altra banda, hem desenvolupat SEQUENCE SLIDER per augmentar l'abast del programa ARCIMBOLDO amb resolució més baixa que l'habitual de 2 Å. En aquests casos s'avaluen simultàniament diferents hipòtesis de seqüències de les traces parcials de polialanina i es proven mitjançant l'autotraçat fins que l'estructura sigui resolta. Hem millorat la comprensió del mecanisme miotòxic i hem desenvolupat un mètode de solució estructural per a estructures cristal·logràfiques desafiants.

Paraules clau: Biologia molecular estructural; Fosfolipases A_2-like; Verí de serp; Phasing; ARCIMBOLDO.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ASU	Asymmetric unit
Atx	Ammodytoxin
B1MMV	Structure BthTX-I complexed to MMV
BPB	4-bromophenacyl bromide
BthTX-I	First protein identified from *Bothrops jararacussu* venom
BthTX-II	Second protein identified from *Bothrops jararacussu* venom
CC	Correlation Coefficient
CC7	7 Cα centroid coordinates
CC_i	initial CC
CC_f	final CC
CC_{mc}	RSCC of main chain atoms
CCSa	RSCC calculated by *EDSTATS*
CC_{sc}	RSCC of side chain atoms
CD	Circular Dichroism
dMDiS	distance between Cβ of MDiS residues 121, and 125
iFace	interface-binding surface
LNLS	Laboratório Nacional de Luz Síncrotron
MDiS	Membrane-Disruption Site comprehended by L121, and F125
MDoS	Membrane-Docking Site comprehended by K115, R118, and K20
MMV	12-methoxy-4-methyl-voachalotine
MR	Molecular Replacement
M_w	Molecular weight
NM	Normal Mode
NSLS-I	National Synchrotron Light Source I
PDB	Protein Databank
PEG	polyethilene glycol
PCC	RSCC calculated by *PHENIX.POLDER*
PLA_2s	Phospholipases A2
PrTX-III	Third protein identified from *Bothrops pirajai* venom
R_g	radius of gyration
RSCC	Real Space Correlation Coefficient
SAXS	Small Angle X-ray Scattering
SEQSLIDER	SEQUENCE SLIDER
R_h	Hydrodynamic radius
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
SS	Secondary Structure
$wMPE_i$	initial weighted mean phase error
$wMPE_f$	final weighted mean phase error

Amino acid	3-letter abbreviation	1-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 INTRODUCTION

THESIS OUTLOOK

This thesis focuses on the study of the mechanism of action of the phospholipases A_2 (PLA_2s) and PLA_2-like toxins from the snakes of Bothrops genus and on the development of the structure solution method called SEQUENCE SLIDER (SEQSLIDER).

The first chapter starts with a general introduction on the importance to study snake venom and its constituents, on ophidic accidents epidemiology and on the toxic effect of PLA_2 and PLA_2-like proteins that are still not neutralized but current antivenom. The structural information available is correlated with what is known on their toxic mechanism as a background for the discussion of our results. The methodologies that we proposed are shown as a result of this thesis and are not included in the Material and Methods.

We present results and discussions on PLA_2-like proteins in **section 2.8**. We start proposing three complementary methodologies that better characterize the more than 20 available structures. First, measuring the geometric orientation beween the two identical monomers of the dimeric PLA_2-like proteins structures using Euler Angles. Second, we measure the local variability of the monomers. And third, we calculate hydrophobic channel accessibility. We discuss the results of these three methodologies finding different toxin states related to how they damage membranes. The hypothesis raised in previous sections is tested with Normal Mode Analysis. Essential to the proposition of the previous analysis, we present the results of the complexes of BthTX-I, purified from *Bothrops jararacussu* venom, with the inhibitors zinc and MMV.

We present in **section 2.9** the experimental results of two basic PLA_2 that may resemble PLA_2-like proteins, BthTX-II and PrTX-III from *Bothrops jaracussu* and *Bothrops pirajai*, respectively. SAXS, DLS and crystallography are used to better characterize their structures. Their possible mechanism of action is discussed relating to the literature information.

Chapter 3 focuses on the second objective of this project, the proposition of a method that aids solution of crystallographic challenging datasets. A brief history of crystallography and its current challenges are described. The environment in which SEQSLIDER originated is outlined in the ARCIMBOLDO scope and in the complexity composition of venoms.

We describe SEQSLIDER focused in venoms in **section 3.7**. We describe the algorithm and how it evaluates different side chain using real space correlation coefficient. First, we test the algorithm on high resolution known structures from both natural and synthetic source. Second, we elucidate two previously unknown myotoxins with SEQSLIDER.

We describe SEQSLIDER focused in phasing within the ARCIMBOLDO scope in **section 3.8**. Again, the algorithm is first described, a known structure is elucidated, MltC, followed by study of an unknown case, FrmR E64H.

In the appendix, we attach the published articles related to this thesis. The proposition of a comprehensive myotoxic mechanism (FERNANDES et al., 2013), a review of the available structures of PLA_2-like proteins (FERNANDES et al., 2014) and the functional and structural study of BthTX-I with zinc (BORGES et al., 2017) are all related to **section 2.8**. At last, the application ARCIMBOLDO_LITE in a large set of structures (SAMMITO et al., 2015) is related to **section 3.3**.

This thesis present two symbiotic objectives, the understanding of toxins mechanism of action and the proposition of methodologies that aid solution of structures of medical importance. The structural description of PLA_2s and PLA_2-like proteins allowed the proposition of additional steps in these toxins mechanism of action. Three different bioinformatics tools are employed to evaluate both local and global structural features that may also be used to characterize structural movement from Molecular Dynamics models and other family of proteins. As a side chain evaluator, SEQSLIDER may aid structure solution of protein with unknown fold within ARCIMBOLDO scope or within sequence uncertainties.

2 STRUCTURAL STUDIES OF SNAKE VENOM PLA_2S AND PLA_2-LIKE PROTEINS

2.1 Study of venom

The study of venoms tracks from at least more than two millennia ago with the detailed synopsis “Historia Animalium” of Aristotle (384-322 BC). The idea of antidotes and the benefits of venom emerged on the ancient Rome and Greece, when venom-based drugs were used to treat smallpox, leprosy, fever and small wounds (UTKIN, 2015). It was in the 17 and 18 centuries that the study of venoms began to take its form. Francesco Redi, known as one the founders of toxinology, identified snake venom in reptile’s teeth and not in the bile as it was once believed and; a hundred years later, Felice Fontana identified the snake glands and started extracting venom (UTKIN, 2015). In the late 19th century, Albert Calmette discovered that after injection of small quantities of venom in animals, their blood serum could be used as an antidote (UTKIN, 2015), whose technique is still used to produce antivenom. At the beginning of next century, toxins started being isolated, but the numbers of scientific articles increased only after late 50’s (UTKIN, 2015). As an example, in 1949, a Brazilian group of scientists discovered a bradykinin-potentiating peptide purified from the venom of *Bothrops jararaca* (FERREIRA; ROCHA E SILVA, 1962; ROCHA E SILVA; BERALDO; ROSENFELD, 1949) that was later used to develop Captopril®, a medicine against hypertension. The most cited paper in the field came a few years later in 1963, the isolation of α-bungarotoxin by Chang and Lee (CHANG; LEE, 1963), as this toxin is still one of the most selective markers of nicotinic acetylcholine receptor. Later, an exponential growth on the number of venom studies came with the advent of the omics sciences (proteomics, transcriptomics, genomics and venomics) that allowed the collective characterization of complex samples.

2.2 Snake venom accidents in the World and Brazil

One of the focus of the study of venom is to aid the envenomings, particularly those caused by snakebites are an important global health issue. According to recent estimates, from 421,000 to 1,800,000 ophidian accidents occur each year, of which around 5% of the cases result in death, with the highest frequencies of accidents in Asia, Africa and Latin America (KASTURIRATNE et al., 2008). Based on this scenario, the World Health Organization classified snakebite envenomings as a neglected tropical disease (WORLD HEALTH ORGANIZATION, 2007).

In the Latin America region, the high estimates expect approximately 130,000 accidents *per* year, in which around 1.8% of cases result in death (KASTURIRATNE et al., 2008). Most of these accidents are caused by *Bothrops* genus snakes (American lanceheads from Viperidae family) (GUTIÉRREZ et al., 2009). In the last ten years in Brazil, it happened around 28,000 snake accidents per year (BRASIL, 2015a), in which only 0.46% resulted in death (BRASIL, 2015b). Similar to Latin America, most of the ophidic accidents in Brazil are bothropic (BRASIL., 2008) and their main complication is local myotoxicity that leads to permanent sequelae and disability if antivenom is not quickly administered (WARRELL, 2010). These consequences may incapacitate rural workers, as these accidents are concentrated in the countryside (GUTIÉRREZ; THEAKSTON; WARRELL, 2006). Instituto Butantan in Brazil is one of the biggest specialized research center of toxinology and antidotes and vaccine production in the world. The improvement of the current antivenom treatment against local myonecrotic effect in bothropic accidents will be highly beneficial.

2.3 Snake venom PLA_2 and its catalytic mechanism

The phospholipases A_2 (PLA_2) are one of the main responsible of local myonecrosis in bothropic accidents. PLA_2s rapid damage is not well reversed by antivenom therapy, thus it is one of the major difficulties in treating the accidents caused by these snakes (GUTIÉRREZ; THEAKSTON; WARRELL, 2006). It is one of the most abundant group of toxins in the venom of some snakes of the Viperidae family (CALVETE et al., 2010; CALVETE; JUÁREZ; SANZ, 2007).

PLA_2s are small (molecular weight around 14 kDa) and stable enzymes with seven disulfide bridges (**Figure 2.1 A_1**). It hydrolyses *sn*-2 ester bond of phospholipids in the bilipid interface of micelles, vesicles, and membranes, releasing lysophospholipids and fatty acid in a catalytic mechanism dependent of calcium ions (LOMONTE; GUTIÉRREZ, 2011; SCHALOSKE; DENNIS, 2006). Its enzymatic activity is related to two regions: i) the catalytic network comprehended by residues H48, Y52, Y73 and D99 (numbering suggested by Renetseder *et al.* (1985) and adopted in **chaper 2**) shown in **Figure 2.1 A_1**; and ii) the calcium binding loop comprehended by Y28, G30, G32 and D49 shown in **Figure 2.1 A_2**. In the catalysis, while the calcium ion holds substrate position by coordination of the phosphate head of the phospholipid, of the main chain oxygen of residues 28, 30 and 32 and the side chain oxygens of D49, the H48 extracts a proton of a water molecule that promotes hydrolysis

in the *sn*-2 ester bond (SCOTT et al., 1990a; VERHEIJ et al., 1980). The relevance of the other residues of the catalytic network, Y52, Y73 and D99, is to maintain the H48 position (SCOTT et al., 1990b).

Prior to catalysis, PLA_2 needs to be in an **active monomeric state** gained with its interaction with the membrane in a process known as interfacial activation (BAHNSON, 2005). The region of PLA_2 interface-binding surface (iFace) to the membrane is not the same as the catalytic residues (BAHNSON, 2005). In fact, it is a challenge to mimic the physiological environment of protein interacting to the membrane (BAHNSON, 2005) and PLA_2 has become the prototype of interfacial enzymology (BERG et al., 2001). The breakthrough to elucidate interfacial catalysis was the elucidation of the structure of a pancreatic PLA_2s in a dimer state with an inhibitor bound to the active site of both protein monomers and with five coplanar phosphate anions in between the dimer (BAHNSON, 2005). While the inhibitor mimics the tetrahedral intermediate in catalysis, the five anions mimic the phosphate head group of membrane phospholipids (**Figure 2.1 C**$_1$). Thus, PLA_2 iFace was experimentally determined, and this result is supported by spectroscopical experiments (LIN et al., 1998). Moreover, different states of porcine PLA_2 were identified in crystallographic structures: i) an **inactive state** before interaction with the membrane; ii) an **active state** triggered by the interaction with the membrane and complexed to a phospholipid, and iii) an **active state** complexed to the product of the catalysis. For a more detailed description on the catalytic mechanism and PLA_2 iFace, see the review by Bahnson (2005).

Besides the use of snake venom PLA_2 for the elucidation of its catalytic mechanism (SCOTT et al., 1990a), the study of these toxins attracted the attention of many researchers by its variety of pharmacological activity. These toxins are: neurotoxic (post e pre-synaptic), myotoxic (local and systemic) and cardiotoxic; able to initiate and inhibit platelet aggregation; hemolytic, able to induce anticoagulating, hypotensive and edematogenic effect; capable to cause convulsion and hemoglobinuria; able to damage different organs or tissues (i.e. liver, kidney, lungs, testis, pituitary damage) (KINI, 2003).

2.4 Snake venom PLA_2-like membrane perturbation mechanism

PLA_2s are subgrouped by their activity in catalytic (D49 and N49-PLA_2s) and non-catalytic (**Figure 2.1 A** and **B**). This last subgroup was given different names in time, here we adopted PLA_2-like proteins, but homolog PLA_2s or according to its 49th residue (K, S, and

R49-PLA_2s) are also common. The reason for these names was either the fact that they share a common ancestor or a similar tertiary structure to the catalytic snake venom PLA_2s. This natural substitution from D49K (from Asp to Lys in the position 49) is relevant to catalysis since the oxygens of the side chain are essential to calcium coordination (**Figure 2.1 A_2**), and the Nζ of K49 occupies the putative position of calcium (**Figure 2.1 B_2**).

The PLA_2-like proteins, similarly to PLA_2s, presents a variety of pharmacological activities. *In vivo*, it is observed local myotoxicity, edema, cytokine release, leukocyte recruitment, hyperalgesia and mechanical allodynia, analgesia and tumor growth inhibition (LOMONTE et al., 2009). In *in vitro* experiments, it is observed many toxic effects, such as cytotoxicity in different human cells, bactericidal, fungicidal and antiparasite action, liposome disruption, among others (LOMONTE et al., 2009). The pharmacological studies *in vivo* are important to understand the ophidic accident envenomation, while the *in vitro* studies guide comprehension of toxin mechanism of action and drug design, since the effects happen in a controlled environment (LOMONTE et al., 2009).

Since the identification of PLA_2-like toxins in the 1980s (MARAGANORE et al., 1984), these proteins have been investigated using many different approaches to elucidate their toxic mechanism (FERNANDES et al., 2014). The large charge distribution on the toxins' surface and their various pharmacological and toxic effects enhance the complexity and hinder our comprehension of their mechanisms of action. Functional studies highlighted the importance of their dimeric oligomerization (DE OLIVEIRA et al., 2001a; DE OLIVEIRA; FERREIRA; WARD, 2009), as well as the C-terminal region, which has a high distribution of basic and hydrophobic residues, for myotoxicity (CHIOATO et al., 2007; LOMONTE; ANGULO; CALDERÓN, 2003).

Various structural experiments led clues on the comprehension of PLA_2-like proteins action mechanism. At the elucidation of their crystallographic structures, two dimers, that coexist in the crystal packing, have been proposed as the biological unit for these toxins. One is called large dimer whose interface is small and composed of the end of the N-terminal helix and two antiparallel β-strands (left side of **Figure 2.2**), while the other is called compact dimer whose interface is larger and composed by the calcium binding loop surroundings (right side of **Figure 2.2**). It is still discussed which one is the biological dimer, while most evidence support the compact one (DOS SANTOS; SOARES; FONTES, 2009;

FERNANDES et al., 2010; MURAKAMI et al., 2005, 2007), few other authors suggest the large one (DE OLIVEIRA et al., 2001b; LOMONTE, 2012).

In 1998 when the compact dimer has not been seen, da Silva-Giotto *et al.* identified two different states of PLA_2-like structures in large dimer assembly, they called one open (*apo* or native in **Figure 2.2**) and the other closed (complex in **Figure 2.2**) based on the different orientation of their monomers. They correlate this quaternary flexibility to the capability to promote biological membrane damage as this movement would displace the membrane phospholipids. Later, the Lys122 in the C-terminal region, which is one of the few residues conserved among PLA_2-like proteins, was identified towards the solvent in the *apo* form and hyperpolarizing a calcium binding loop in the structure complexed to fatty acids (AMBROSIO et al., 2005). With calcium binding loop hyperpolarization, the residues surrounding Lys122 form a hydrophobic knuckle that was hypothesized to participate in membrane damage (AMBROSIO et al., 2005).

Based on the compact dimer and PLA_2-like dimeric structures, dos Santos *et al.* proposed the PLA_2-like iFace in the same strategy used by Bahnson (2005), sulfates ions forming a plane found in a complexed structure may mimic the phosphate head groups of membrane phospholipids. These sulfates are interacting to a cluster of basic residues, K115, R118 and K20 (in red, **Figure 2.1 B**$_2$) that are close together in the compact dimer (DOS SANTOS; SOARES; FONTES, 2009). They also proposed a different pattern observed in apo and complexed structures relating to the differences in the quaternary structure, i.e., the angles between monomers. This pattern is calculated by two angles, Aperture and Torsion. The former is the angle between the long α-helix (helix 3 in **Figure 2.1 A**$_1$) of each monomer. The latter is the angle between the helix 1 with the cross product vector between the long α-helix with α-helix 2 of the other monomer (**Figure 2.3**). Based on the different dimeric orientations observed in the available crystallographic structures, PLA_2-like proteins from the *Bothrops* genus were classified as being in i) an **inactive state**, whose toxins are native with void hydrophobic channels; and ii) an **active state**, whose toxins were complexed to either hydrophobic molecules or inhibitors.

Recently, based on structural and functional studies, a myotoxic mechanism has been proposed that involves specific regions in the C-terminal region, the MDoS (Membrane-Docking Site comprehended by K115, R118, and K20) and the MDiS (Membrane-Disruption Site comprehended by L121 and F125) (FERNANDES et al., 2013). The entrance of a

hydrophobic molecule in the toxin hydrophobic channel would induce an allosteric modification of the protein, with dimer reorientation (DOS SANTOS; SOARES; FONTES, 2009) from the **inactive** to **active state,** enabling the MDoS to dock with the cell membrane and the MDiS to perturb cell membrane integrity (FERNANDES et al., 2013). After the toxin penetrates the lipid bilayer and disturbs the phospholipids and membrane integrity, ion homeostasis is disrupted (LOMONTE; GUTIÉRREZ, 2011). The PLA_2s also break ion homeostasis, although the catalytic mechanism plays an important role as the phospholipids are cleavaged instead of being perturbed. Consequently, in both group of toxins, the calcium influx increases membrane damage by inducing hypercontraction of myofilaments, mitochondrial damage, calcium-dependent proteases and cytosolic PLA_2s activity, leading to muscle necrosis (LOMONTE; GUTIÉRREZ, 2011; MONTECUCCO; GUTIÉRREZ; LOMONTE, 2008). All of these indirect effects of the PLA_2-like proteins or PLA_2s activity enhance cytotoxicity (MONTECUCCO; GUTIÉRREZ; LOMONTE, 2008).

Based on this myotoxic mechanism and on functional and structural experiments with PLA_2-like toxins and inhibitors, three different inhibition mechanisms were suggested for these proteins. First, ligands binding inside the hydrophobic channel or in its entrance, blocking the access of fatty acid molecules. Second, ligands interacting to specific residues of the MDoS or MDiS regions (FERNANDES et al., 2015), preventing the protein/membrane interaction. And third, ligands inducing toxin oligomerization (SALVADOR et al., 2015), which may result in a combination of the two previous processes. BPB and rosmarinic acid are examples of the first class, and they inhibit all or a substantial fraction of the myotoxic, paralyzing and edematogenic effects of different toxins (DOS SANTOS et al., 2011a; SOARES; GIGLIO, 2003; TICLI et al., 2005). Caffeic acid and aristolochic acid are examples of the second type of inhibitor, with partial myotoxic inhibition of PrTX-I from *Bothrops pirajai* venom (FERNANDES et al., 2015). The versatile suramin may bind in different regions inhibiting these protein in all of the three previous categories (MURAKAMI et al., 2005; SALVADOR et al., 2015), as it always possesses high inhibition against PLA_2-like proteins (DE OLIVEIRA et al., 2003; MURAKAMI et al., 2005; SALVADOR et al., 2015).

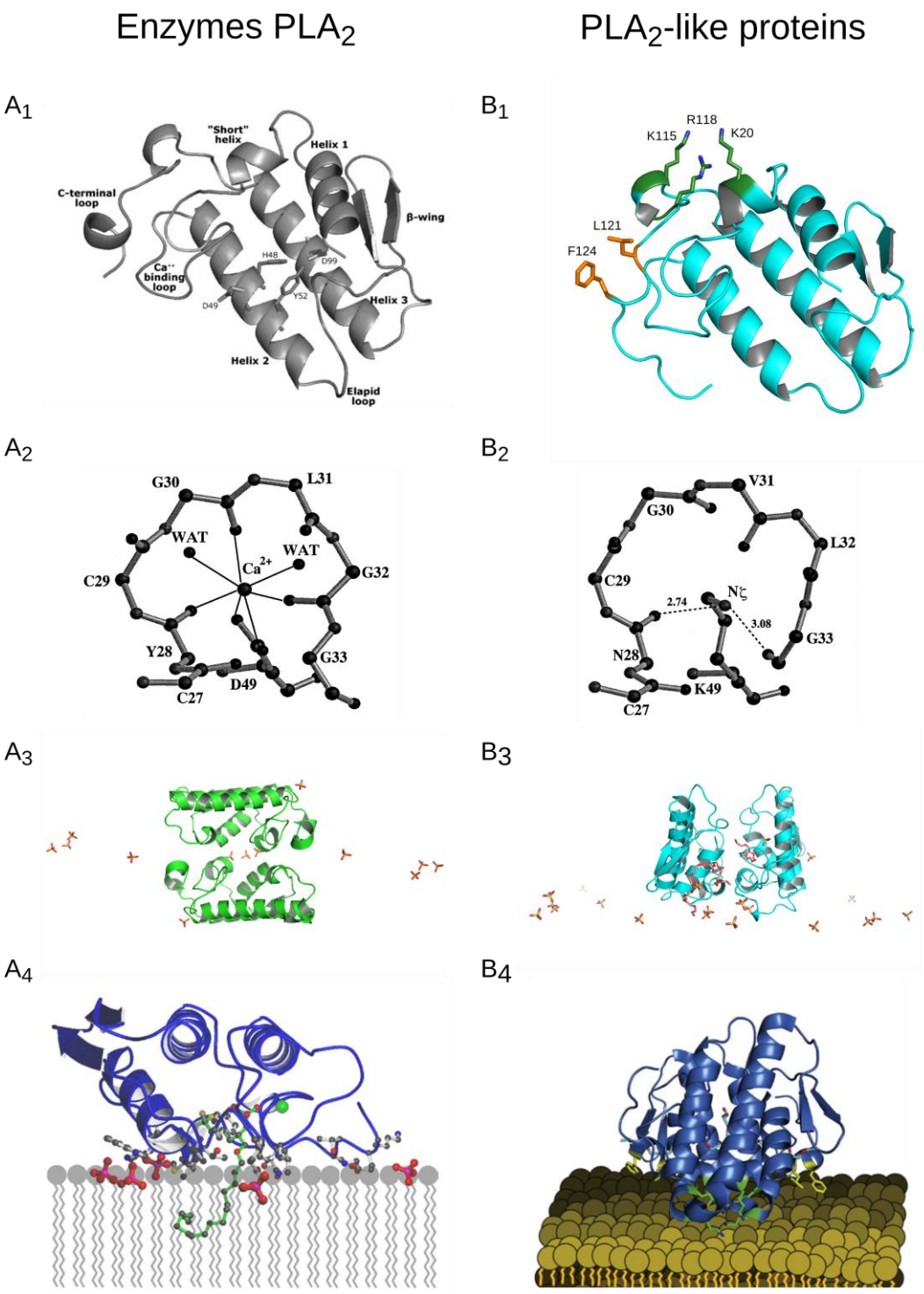


Figure 2.1 Structures and interaction site of snake venom PLA_2 (**A**) and PLA_2-like proteins (**B**).
In $\mathbf{A_1}$, cartoon representation of the tertiary structure of PLA_2 with residues of the catalytic network with side chains in sticks (H48, Y52, Y73, and D99) (extracted from Magro *et al.* (2009)). In $\mathbf{B_1}$, cartoon representation of the tertiary structure of the PLA_2-like protein in cyan, side chains in sticks of residues related to myotoxicity, MDoS in green and MDiS in orange. In $\mathbf{A_2}$, sticks representation of calcium binding loop with calcium coordinated and, in $\mathbf{B_2}$, distorted calcium binding loop with Nζ of K49 occupying calcium binding position (Extracted from Azevedo *et al.* (DE AZEVEDO et al., 1998)). In $\mathbf{A_3}$ and $\mathbf{B_3}$, the plane of sulfates formed in symmetry mates of crystallographic structures of PLA_2 and PLA_2-like protein mimic phosphate head of phospholipids present in membranes, represented in $\mathbf{A_4}$ (extracted from Bahnson (2005)) and $\mathbf{B_4}$.

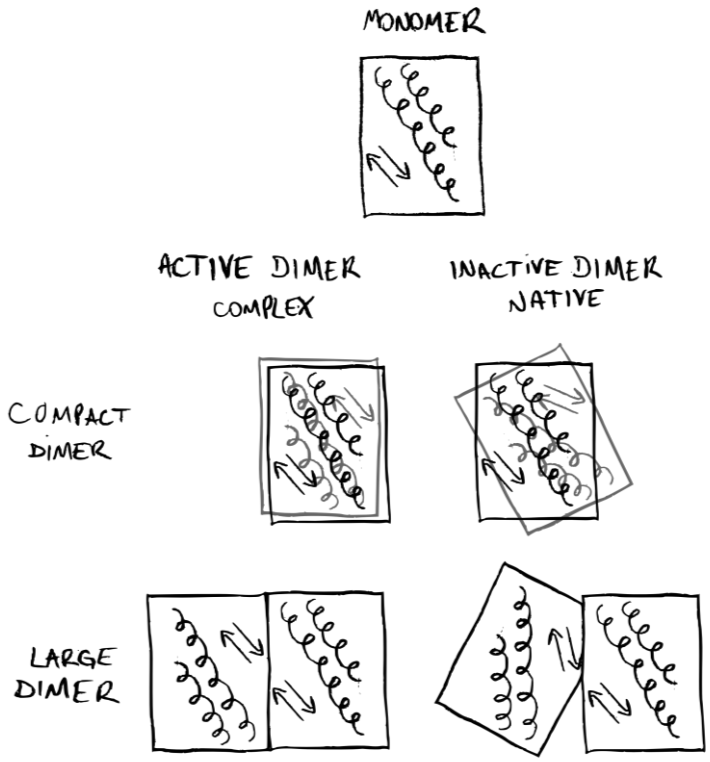


Figure 2.2 Drawing of the PLA_2-like monomer and dimers.
The rectangle with the two parallel α-helices in the diagonal with the antiparallel β-sheets represents the monomer. The large and the compact dimers are shown on the left-hand side and the right-hand side, respectively.

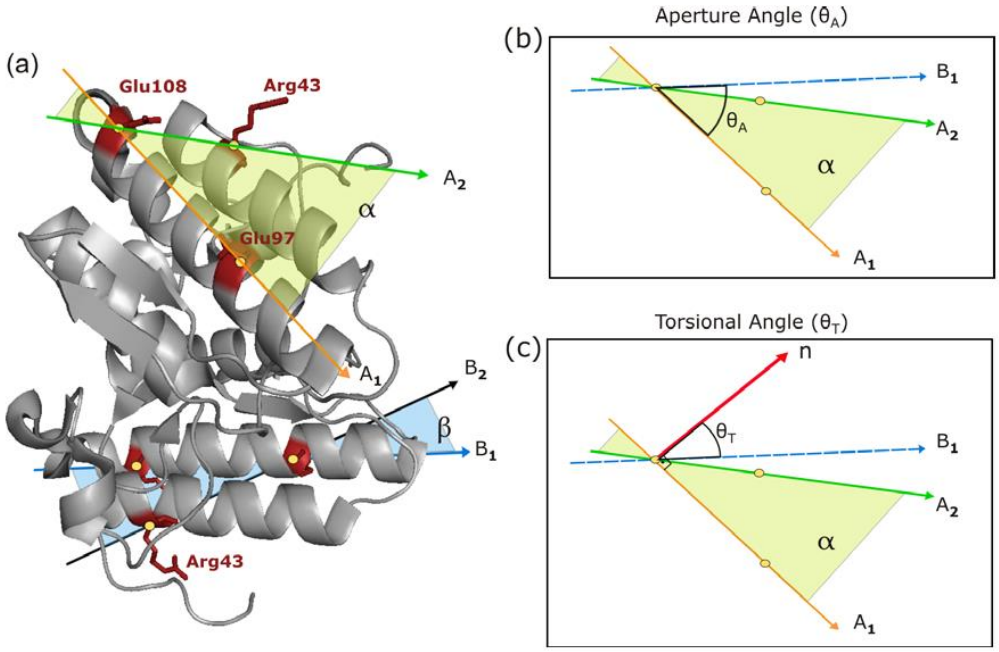


Figure 2.3 Compact dimer of PLA_2-like proteins and dos Santos, Soares and Fontes (2009) angle proposition.
In A, the compact dimer of PLA_2-like proteins with the respective vectors A_1, A_2, B_1 and B_2 proposed by dos Santos, Soares, and Fontes (2009). In B, the Aperture angle (θ_A). In C, the n vector, which is the cross product vector between the A_1 and A_2, and the Torsion angle (θ_T). Figure extracted from dos Santos, Soares, and Fontes (2009).

2.5 Myotoxicity of some bothropic PLA_2s and PLA_2-like proteins independent of calcium

Some Asp49-PLA_2s have a high distribution of basic residues, including within the C-terminal region, such as other PLA_2-like proteins do. In our laboratory, we have investigated BthTX-II recently through structural, functional and phylogenetic approach (DOS SANTOS et al., 2011b). We have observed that this toxin does not need calcium ions to develop its toxicity, which is different from other PLA_2s and is similar to PLA_2-like proteins. By the structural analysis, we have observed a distortion of the calcium binding loop that prevents calcium coordination (**Figure 2.4**). By phylogenetic studies, we have seen three basic PLA_2s (BthTX-II, PrTX-III and Mt-I) having a higher similarity with PLA_2-like proteins than other Asp49-PLA_2s. In the literature, these two other toxins are also identified as having toxicity even in calcium absence.

Despite these observation, it remains unknown how exactly are the activation of PLA_2-like proteins and PLA_2 independent of calcium and how they disturb the membranes. The action mechanism knowledge may aid antivenom treatment.

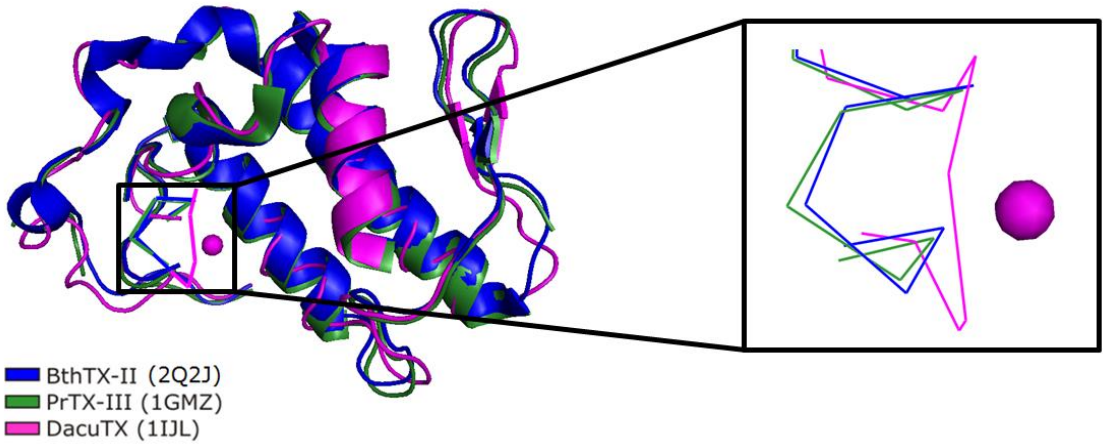


Figure 2.4 - Superposition in cartoon of two basic Asp49-PLA_2s, BthTX-II and PrTX-III, and an acid PLA_2, DacuTX.
In zoom and *sticks*, the basic toxins have a distortion in calcium binding loop when compared to the acid one which has this ion coordinated showed as a sphere.

2.6 Objective

The aim of this chapter is to propose a general methodology to describe the membrane perturbation mechanism independent of catalysis by reviewing all the models available in the Protein Databank including the structural results of two particular toxins of

Bothrops jararacussu in native and complexed state performed in this thesis. The first is BthTX-I, a PLA_2-like protein, complexed to the inhibitors zinc and MMV (12-methoxy-4-methylvoachalotine). And the second toxin is BthTX-II, an Asp49-PLA_2 in native and complexed to fatty acids.

2.7 Material and Methods

1. Compounds purification

We obtained the lyophilized venom of *Bothrops jararacussu* from *Centro de Extração de Toxinas Animais LTDA*. We purified the BthTX-I, a PLA_2-like toxin, and the BthTX-II, a basic Asp49-PLA_2 using gel filtration and reverse-phase liquid chromatography using AKTApurifier UPC-900 (GE) that was first reported in the thesis of a former student of our laboratory (CARDOSO, 2016). To ensure protein purity, we did another step of reverse phase and electrophoresis gel. Other students in our laboratory participated in protein purification, such as Henrique Barcellos Campanelli, Fábio Florença Cardoso and Eloah dos Santos de Biasi.

The ligand MMV was purified from the root bark of *Tabernaemontana catharinensis* by our collaborators Andreimar Martin Soares (Fundação Oswaldo Cruz, Rondônia, Brazil) e Paulo Sérgio Pereira (Universidade de Ribeirão Preto, Brazil) (PEREIRA et al., 2008).

2. Secondary structure evaluation

We performed Circular Dichroism (CD) experiments to evaluate if MMV was inducing any secondary structure modification in BthTX-I. We did it on a Jasco J-815 (JASCO Inc., Tokyo, Japan) spectropolarimeter, equipped with a Peltier temperature control PTC-423-S. We prepared BthTX-I samples with a concentration of 350 $\mu g.ml^{-1}$ and the MMV of 88 $\mu g.ml^{-1}$, which is a molar ratio of 1:8 protein ligand. After pre-incubation of 30 min at 20 ºC, we took all CD measurements in a spectral range of 190–260 nm at 20 ºC, using an optical path length 0.05 mm and scanning speed of 100 nm/min, the bandwidth of 2 nm and response time of 1 s. To obtain the final spectrum, we used the resultant of 20 spectra that were accumulated, averaged and corrected from the baseline for buffer solution contribution and normalized to residual molar ellipticity [θ]. We corrected the complex spectrum using the baseline for the ligand solution in the same concentration. We processed and analyzed the acquired data using Spectra ManagerTM II software.

3. Quaternary structure evaluation

We used SDS gel in non-reducing conditions, Dynamic Light Scattering with DynaProTitan (Wyatt Technology Corporation) and Small Angle X-ray Scattering (SAXS) to evaluate quaternary structure and polydispersity.

a. SDS gel in non-reducing conditions

We performed SDS-PAGE gel electrophoresis without the denature conditions of heating sample and mixing it with β-mercaptoethanol. We prepared the samples containing a mass of approximately 15 μg of the toxin. In the case of the complex of toxin and MMV, 3,5 μg of this ligand was used, which is in a molar ratio of approximately 1:8 protein ligand. The complex was pre-incubated for 30 min at 20 ºC.

b. Small-angle X-ray scattering

We performed the SAXS experiments in SAXS2 beamline in LNLS. We analyze the scattering with package *ATSAS* (KONAREV et al., 2006). Firstly, we used *GNOM* program to calculate the distance distribution function *p(r)* and the radius of gyration (R_g) of the studied toxin. Using program *CRYSOL*, we simulated the theoretical scattering of crystallographic models. With the comparison of the experimental curve *I(q)* with the theoretical curve of the crystallographic structures, we may estimate precisely the oligomeric state of the toxin in solution. We were aided by Dr. Mário de Oliveira Neto (UNESP Botucatu, Brazil) to perform experiments and data analysis.

4. Crystallographic determination

We performed the crystallization trials manually or automatically by the robot Oryx4 (Douglas Instrument) in different methods: vapor diffusion (hanging and sitting drop) or Microbatch. We used the sparse matrix method or we optimizated the crystallization conditions already published in the literature. We flash cooled the crystals in liquid nitrogen and collected in one of the following crystallography beamlines: MX-I and II at the Laboratório Nacional de Luz Síncrotron (LNLS) in Campinas, São Paulo, Brazil or X6a and X25 in National Synchrotron Light Source I (NSLS-I).

We processed the data with either *XDS* (KABSCH, 2010), *HKL2000* (OTWINOWSKI; MINOR, 1997), *iMosflm* (BATTYE et al., 2011) or *autoPROC* (VONRHEIN et al., 2011) depending on how each one manages better the diffraction pattern measurement and indexing. The diffraction pattern was measured to a maximum resolution in which low divergence of equivalent measurements was seen (as judged by $CC_{1/2}$ (KARPLUS;

DIEDERICHS, 2012)). For *XDS* data processing, we manually performed merging with *aimless* (EVANS, 2011). For *iMosflm* data processing, we scaled and merged reflections with *SCALA*. We evaluate the crystallographic data with *Xtriage* (ZWART; GROSSE-KUNSTLEVE; ADAMS, 2005), *AIMLESS* (EVANS, 2011) and *XPREP* (written by Sheldrick GM and proprietary of Bruker AXS). We obtained the phases by molecular replacement with *Phaser* (MCCOY et al., 2007) using the native models of the studied toxin. We performed the refinement with *Refmac* (WINN et al., 2011), *Phenix.refine* (ADAMS et al., 2010) or *Buster* (BRICOGNE et al., 2011), the modeling with *Coot* (EMSLEY et al., 2010); model validation with *MolProbity* (DAVIS et al., 2007); and structure analysis, comparison and images with *Coot*, *Pymol* (http://www.pymol.org/) and *Chimera* (PETTERSEN et al., 2004).

5. Bioinformatics tools

We used *PISA* (KRISSINEL; HENRICK, 2007), an algorithm that evaluates the interfaces in macromolecular crystallographic structures using thermodynamics, to foresee the toxin biological unit. We used the *MOLE* program (SEHNAL et al., 2013) in default options to locate and characterize the chemical properties of molecular tunnels and pores in the hydrophobic channel of model BthTX-I/Zn. We used CAVER 3.0 (CHOVANCOVA et al., 2012) to characterizes the tunnels accessing the hydrophobic channels of all available PLA_2-like crystallographic models. To improve convergence of evaluated different tunnels, we increased the CAVER option probe radius to 1.8 and shell radius to 4. For the structural comparisons in **section 2.8**, we used the following structures, with the Protein DataBank identification code in parentheses (PDB id): *apo* BthTX-I (3HZD and 3I3H), BthTX-I/PEG4k (3IQ3), dimeric BthTX-I/BPB (*p*-bromophenacyl bromide) (3HZW) (FERNANDES et al., 2010), BthTX-I/PEG400 (2H8I) (MURAKAMI et al., 2007), BthTX-I/*α*-tocopherol (3CXI), PrTX-I/*α*-tocopherol (3CYL), *apo* PrTX-I (2Q2J) (DOS SANTOS; SOARES; FONTES, 2009), PrTX-I/BPB (2OK9) (MARCHI-SALVADOR et al., 2009), PrTX-I/rosmarinic acid+PEG330 (3QNL) (DOS SANTOS et al., 2011a), PrTX-I/caffeic acid+PEG4k (4YU7), PrTX-I/aristolochic acid+PEG4k (4YZ7) (FERNANDES et al., 2015), PrTX-II/*n*-tridecanoic acid (1QLL) (LEE et al., 2001), *apo* BnSP7 (1PA0) (MAGRO et al., 2003), BbTX-II (4K09), MTX-II/PEG4k (4K06) (FERNANDES et al., 2013), BaspTX-II/suramin (1Y4L) (MURAKAMI et al., 2005), BnIV/myristic acid (3MLM) (DELATORRE et al., 2011), MjTX-II/stearic acid (1XXS) (WATANABE et al., 2005), MjTX-II/PEG4k (4KF3)

(SALVADOR et al., 2013b), and MjTX-II/suramin+PEG4k (4YV5) (SALVADOR et al., 2015). We did not include the PDB ids 3I03, 3I3I, 1PC9, 1CLP, 4DCF in our analysis, the first two because a monomer composes their asymmetric unit (ASU) content and the others because the resolution was worse or equal to 2.5 Å. For the structural comparisons in **section 2.9**, we used the two available bothropic basic PLA_2 models, BthTX-II (2OQD) (CORRÊA et al., 2008) and PrTX-III (1GMZ) (RIGDEN et al., 2003). We used *PyMOL* (The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.) to create cartoon and sticks images, *Coot* to create electron density figures, and *MOLE* local version 2.13.9.6 to create tunnels images.

6. Normal Mode Analysis

We chose 2H8I (BthTX-I/1PEG) and 3IQ3 (BthTX-I/3PEG) as representatives of the **inactive** and **active state**, respectively, to analyze the normal mode using CHARMM v.36b1 program (BROOKS et al., 2009). We generated the topology and parameter files with the CHARMM-GUI server (www.charmm-gui.org) employing an additional energy minimization. We applied the conjugated gradient algorithm with harmonic constraints that were progressively decreased from 250 to 5 kcal/mol·$Å^2$, with 100 steps of minimization at each decrease. We removed the constraints to carry out an additional of 10,000 steps of conjugated gradient that follow by 300,000 steps of the basis Newton-Raphson algorithm. We used the final minimized structure to calculate the 7-15 lowest frequency NMs, using the VIBRAN module of CHARMM for both of the chosen structures.

We generated structures along each NM based on a short MD simulation at a low temperature (30 K) using VMOD facility of CHARMM and a minimization afterward. We establish an initial maximum displacement range of 3 Å for each direction of the NM with a 0.1 Å projection step based on the values of mass-weighted root mean square. If necessary, we increase the maximum displacement range to 6 Å in the desired NM. For each of these steps, we applied a harmonic force constant over the Cα atoms (increasing from 1,000 until 10,000 kcal/mol·$Å^2$), and we carried out a short MD simulation for 1 ps for each constant value, totaling 10 ps of simulation. We obtained the final structure employing 1,000 steps of CG energy minimization keeping the restraints. We were aided by Dr. Ney Lemke (UNESP Botucatu, Brazil).

2.8 Comparison of available snake venom PLA_2-like proteins structures

One of the main themes in current toxinology is the development of alternatives to the conventional serum therapy that does not efficiently neutralize snake venom injuries. In the bothropic envenomation, which is the majority of the ophidic accidents in Latin America, the myonecrosis caused by the myotoxins PLA_2, PLA_2-like proteins and metalloproteinases is still an issue. Despite the identification of the snake venom PLA_2-like proteins in 1984 (MARAGANORE et al., 1984) and a large number of crystallographic structures elucidated (FERNANDES et al., 2014), there is still a lack of knowledge on how to inhibit those toxins and how they molecularly disrupt the membranes. The objective of this section is to compare the 23 available snake venom PLA_2-like protein structures including the structure BthTX-I/MMV elucidated during this thesis to fill this gap.

2.8.1 Quaternary structure possibilities

Essential to the comprehension of a toxin mechanism of action is the knowledge of its exact protein tertiary and quaternary structure. Bothropic PLA_2-like proteins have a similar tertiary structure as PLA_2, but instead of being monomers in solution like the latter, most of these non-catalytic proteins are dimers. Such quaternary structure was seen in SAXS (MURAKAMI et al., 2007), DLS (FERNANDES et al., 2010, 2013; SALVADOR et al., 2013b), and non-reducing SDS-Page (COSTA et al., 2008; FRANCIS et al., 1991; LOMONTE; GUTIÉRREZ, 1989; RODRIGUES et al., 1998; TOYAMA et al., 1995) (**Table 2.1**).

All the authors depositing bothropic PLA_2-like crystallographic structures agree that these toxins are dimers (Biological Unit (BU) in **Supplementary Table 2.11**), with the exception of Delatorre *et al.* (2011). In the crystal packing of the crystallographic structures of PLA_2-like toxins, two dimers coexist and are hypothesized to be the biological unit by different authors (**Figure 2.2**). To establish evidence on which dimer is biologically relevant, we calculated the energy necessary to dissemble (ΔG^{diss}) the large and the compact dimer (**Figure 2.2**) of each evaluated crystallographic models using PISA (**Table 2.2**). It is evident how the compact dimer has an energetically more favorable interaction than the large dimer, as all sampling of the latter has negative and lower values. Negative values indicate that additional energy is needed to keep the oligomeric assembly together. Since all the ΔG^{diss}s of the large dimer models are less than -6 kcal/mol, it is unlikely that this is the oligomeric

assembly in solution and thus biologically relevant. On the contrary, all compact dimer models have positive ΔG^{diss}, with the exception of the MjTX-II models (4KF3, 4YV5, 1XXS, and 1Y4L) whose ΔG^{diss} varies from approximately -1 to 0 kcal/mol. Such range is described as the gray zone, where stability in solution is uncertain. 4YZ7 model (BthTX-I) was another exception with a ΔG^{diss} of -1.3 kcal/mol, although after improving the model by modeling the 43 omitted side chains, this value improved to 2.5 kcal/mol, a value in agreement with the others. Omitting few atoms in the model is a standard approach in crystallography due to structural flexibility or radiation damage. But when computation techniques, such as Molecular Dynamics, use those models, introduction of missing atoms is required. The procedure was not necessary for the other structures, as only a few side chain atoms of their models were removed.

Table 2.1 - Dimeric evidence of the PLA_2-like toxins whose crystallographic structures were elucidated.

Toxin	Snake specie	Evidence of dimer		
		Non-reducing SDS-PAGE	DLS	SAXS
BaspTX-II	*Bothrops asper*	(FRANCIS et al., 1991; LOMONTE; GUTIÉRREZ, 1989)	(FERNANDES et al., 2013)	(MURAKAMI et al., 2007)
BbTX-II	*Bothrops brazili*	(COSTA et al., 2008)		
BnIV Bnsp7	*Bothrops pauloensis*	(RODRIGUES et al., 1998)		
BthTX-I	*Bothrops jararacussu*	In **section 2.8.9**	(FERNANDES et al., 2010) (SALVADOR et al., 2013b)	
MjTX-II	*Bothrops moojeni*			
PrTX-I	*Bothrops pirajai*	(TOYAMA et al., 1995)		

To obtain more clues on these quaternary structures, we evaluated their crystallization condition, unit cell and space group (**Supplementary Table 2.11**). All the crystals were grown in an environment of high concentration of PEG, with the exception of MjTX-II/stearic acid (1XXS) that was grown in sodium citrate. All those environments had a high ionic force and a wide range of pH, from 4.6 to 8.5. We did not find any correlation of the unit cell/space group to crystallization conditions. Although, structures whose datasets are of primitive and orthorhombic crystal systems are correlated with abundance of negative and hydrophobic molecules in the crystal structure (**Supplementary Table 2.12**), whether these ligands are rare for trigonal. Moreover, the mutual presence of the compact dimer in different

crystal forms supports it as the biological unit. This approach was systematically implemented in the Protein Common Interface Database (ProtCID) (XU; DUNBRACK, 2011), as these authors have shown that an interface present in various crystal forms, especially in homologous non-identical proteins, are likely to be biologically relevant (XU et al., 2008). ProtCID evaluation identifies the compact dimer interface in the three crystal forms available in our evaluated structures (orthorhombic, monoclinic and trigonal) and an additional four crystal forms of PLA_2-like toxins from other *genera*. These are from *Viridovipera stejnegeri* (PDB id: 4H0S), *Gloydius halys* (PDB ids: 1A2A and 1BJJ) and *Agkistrodon contortrix laticinctus* (PDB id: 1S8G). Thus, the compact dimer is likely to occur in solution, and it is supported by SAXS experiments (MURAKAMI et al., 2007).

Table 2.2 - Summary of PLA_2-like crystallographic models and their dissociation calculations.

PDBs	ΔG^{diss} (kcal/mol)	
	LD	CD
1PA0	-7.2	0.1
2Q2J	-7.6	3.2
3HZD	-6	4.9
3I3H	-6	4.3
4K09	-7.5	1.5
4WTB	-6.9	7.2
2H8I	-7.1	5
3HZW	-8.3	0.9
4KF3	-8.8	-0.6
4YV5	-8	-0.2
1XXS	-8	-1.1
1Y4L	-8.2	-1.2
3QNL	-8.9	4
4YZ7	-6.2	2.5[#]
3CYL	-8.1	1.2
3CXI	-8.6	1.2
4YU7	-8.4	0.1
2OK9	-9	0.7
4K06	-8.7	1.7
3IQ3	-8.5	1.1
3MLM	-9.7	2.3
B1MMV	-9.7	2.9
1QLL	-9.6	6.6

ΔG^{diss} stands for the energy necessary to dissociate the dimer. [#] refers to the ΔG^{diss} of the model 4YZ7 after completing side chain atoms of 44 incomplete residues in the model.

Table 2.3 - Summary of local and global measurements of PLA_2-like crystallographic models

PDBs	Ligands	Angles			dMDiS (Å)				Tunnel			toxin	Res	SG
					CanMon		N-CanMon		CanMon		N-CanMon			
		ψ	θ	φ	A	B	A	B	A	B	A			
1PA0*	-	141	26	54	-	4.8	10.8	-	-	331	0	Bnsp-VII	2.20	$P3_121$
2Q2J	1 sulfate / 2 TRIS	141	26	54	-	4.9	10.9	-	-	267	0	PrTX-I	1.65	$P3_121$
3HZD	-	141	26	54	-	4.8	11.2	-	-	270#	0	BthTX-I	1.91	$P3_121$
3I3H	-	140	27	53	-	4.6	9.3	-	-	251	0	BthTX-I	2.17	$P3_121$
4K09	-	142	26	53	-	4.5	10.8	-	-	233	0	BbTX-II	2.11	$P3_121$
4WTB*	3 zinc / 2 sulfates	141	27	55	-	4.7	10.3	-	-	265	0	BthTX-I	2.16	$P3_121$
2H8I	1 PEG	141	27	55	-	4.9	8.5	-	-	242#	0	BthTX-I	1.90	$P3_121$
3HZW	2 BPB / 1 isopropanol	170	24	46	-	6.2	9.7	-	-	385	405	BthTX-I	2.28	$P2_12_12_1$
4KF3	4 PEGs / 6 isopropanol	169	23	49	4.4	4.4	-	-	408	406	-	MjTX-II	1.92	$P2_12_12_1$
4YV5*	2 suramin / 3 PEGs / 7 sulfates	170	24	48	4.9	5	-	-	405	446	-	MjTX-II	1.90	$P2_12_12_1$
1XXS*	4 stearic acids / 5 sulfate	170	24	47	4.7	4.4	-	-	483	471	-	MjTX-II	1.80	$P2_12_12_1$
1Y4L*	1 suramin / 2 PEGs / 5 isopropanol	167	23	43	4.8	4.6	-	-	459	484	-	BaspTX-II	1.70	$P2_12_12_1$
3QNL	1 rosmarinic acid / 1 PEG / 8 isopropanol	173	24	34	4.8	5.2	-	-	364	417	-	BthTX-I	1.77	$P2_12_12_1$
4YZ7	1 aristolochic acid / 1 PEG / 5 sulfates	181	28	27	4.7	4.8	-	-	532	406	-	PrTX-I	1.95	$P2_12_12$
3CYL	2 vitamin Es / 1 PEG / 5 sulfates	177	28	27	4.5	5.3	-	-	415	432	-	PrTX-I	1.87	$P2_1$
3CXI	2 vitamin Es / 1 PEG / 4 sulfates	178	28	26	4.7	5.1	-	-	509	433	-	BthTX-I	1.83	$P2_1$
4YU7	4 caffeic acids / 3 PEGs / 1 sulfate	177	27	28	5	4.7	-	-	458	387	-	PrTX-I	1.65	$P2_1$
2OK9	2 BPBs / 2 isopropanol	180	29	25	4.5	5.3	-	-	458	429	-	PrTX-I	2.34	$P2_1$
4K06	3 PEGs / 5 sulfates	178	28	28	5.1	4.9	-	-	473	391	-	BbTX-II	2.08	$P2_1$
3IQ3	3 PEGs / 2 sulfates	177	27	27	4.7	5.2	-	-	312	460	-	BthTX-I	1.55	$P2_1$
3MLM	2 myristic acids / 4 sulfates	178	28	28	4.6	5.4	-	-	399	441	-	BnIV	2.20	$P2_1$
B1MMV/BB'	2 MMV	171	16.6	55.1	-		9.3	9.3	-	-	-	BthTX-I	2.10	$P3_221$
1QLL	2 N-tridecanoic acids	163	19.8	80.3	5.2	5.2	-	-	-	-	-	PrTX-II	2.04	$P2_1$

The abbreviations are related to the distance between the two Cβ MDiS residues (dMDiS), canonical monomer (CanMon), non-canonical monomer (N-CanMon), resolution (Res), and Space Group (SG). The toxins (TX) Bnsp, Pr, Bth, Bb, Mj, Basp, and Bn were purified from the venom of *Bothrops pauloensis*, *Bothrops pirajai*, *Bothrops jararacussu*, *Bothrops brazili*, *Bothrops moojeni*, *Bothrops asper* and *Bothrops neuwiedi*, respectively. In PDBs, angles Tunnel and dMDiS columns, font colored in magenta, red and dark green are a representation of asymmetric, exception and symmetric PLA_2-like properties, respectively. In Ligands column, PLA_2-like protein inhibitors, and natural compounds are colored in orange, light blue and dark green, respectively. * Chain letter A and B were inverted. # Side chain of H120 had to be remodeled for tunnel calculation.

2.8.2 Global measurement: orientation and translations between almost identical objects

The compact dimer is present in all the bothropic PLA_2-like structures of **Table 2.2**, although the orientation between their monomers may be different. Dos Santos *et al.* (2009) using two angles calculated from the vector of the two parallel helices of each monomer (**Figure 2.3**) differentiate those structures in two different states; one called **active** and another **inactive** (**Figure 2.2**). Herein we update this angle description in a more intuitive, general and informative methodology.

Any 3D object in a 3D Euclidian space may be perfectly placed into another identical object by one 3D rotation and one 3D translation; this is common and known in crystallography as it is the base concept for Molecular Replacement. The 3D translation may be easily separated in 3 translations, one in each axis (usually called X, Y, and Z). The 3D rotation is not so simple as they may be described in different ways depending on the application, such as rotation matrixes used in crystallography, quaternions in computer graphic games, Euler angles in robotics and aeronautics. Euler angles are a sequence of three rotations around the objects' local coordinate axes (JANOTA et al., 2015). There are different conventions of Euler angles as the order and choice of rotating axis changes the resulting rotation (DIEBEL, 2006). We believe that the most intuitive 3D rotation description is using once each of the 3 axes, like the Tait-Bryan angles used in aeronautics. These angles are related to the movement made by airplanes and are described as row, pitch and yaw (JANOTA et al., 2015) illustrated as X, Y, and Z axes, respectively, in **Figure 2.5 A, B, C** and **D**. For our description, the nomenclature of the Greek letters φ, θ and ψ will be used.

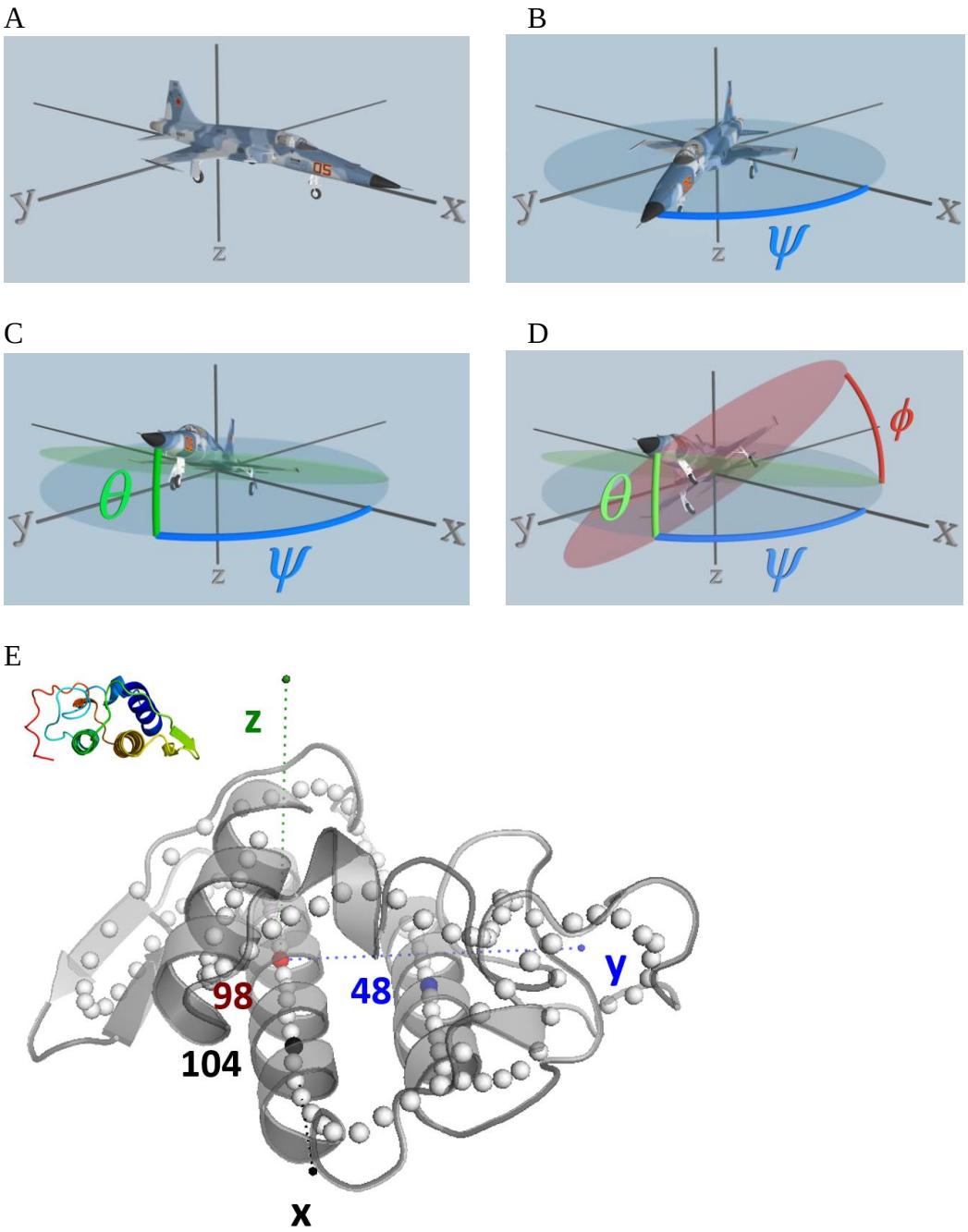


Figure 2.5 - Representation of the aeronautical (Tait-Bryan) angles of movement in an airplane and its similarity to our proposed PLA_2-like protein angles description.
In **A**, the airplane in an original position of with X, Y and Z axes. In **B** it is first rotated ψ degrees in Z axis, then it is rotated θ degrees in Y shown in **C** and at last rotates φ degrees in X axis shown in **D**. In **E**, PLA_2-like protein is represented by grey cartoon with its 7 Cα centroid coordinates (CC7) represented by white spheres. The axes are established by CC7/98 as origin colored red, CC7/104 as X axis colored black, CC7/48 as Y axis colored blue and the normal vector to the XY plane colored in green.

Thus, the first step is to establish immutable and standard orthogonal axes for the monomer-monomer description. Particular and immutable coordinates have to be chosen to ensure that same axes will be considered in each evaluated dimer. The most common way to

compare overall structural similarity is by calculation of the root-mean-square deviation (RMSD) of Cαs after superposition of corresponding residues. Since we want to look for the variation locally, we chose to use distances of corresponding Cαs after an optimum superposition. One strategy is calculating the root-mean-square fluctuation (RMSF), which is the standard deviation of the distances of points to a reference. RMSF is commonly used in Molecular Dynamics as the coordinate positions are varied over time. Since insertion and deletion are common in toxins, as evolutionary forces are stronger in this class of proteins, two different numbering of residues are used, one continuous (last residue number 121) and another containing gaps (following Renetseder *et al.* numbering (1985) with last residue number 133). As this choice depends on the depositor, the different models contain both numbering. Thus, prior to superposition, we renumbered all polypeptide chains to exclude gaps and we calculated correspondence of residues to monomer A of PDB id 3IQ3, that we chose as reference. We superposed all monomers available in **Table 2.2** to the reference and extracted the distance between the correspondent Cα. We calculated the standard deviation of the distances of each residue of all superposition (RMSF) (purple line in **Figure 2.6**).

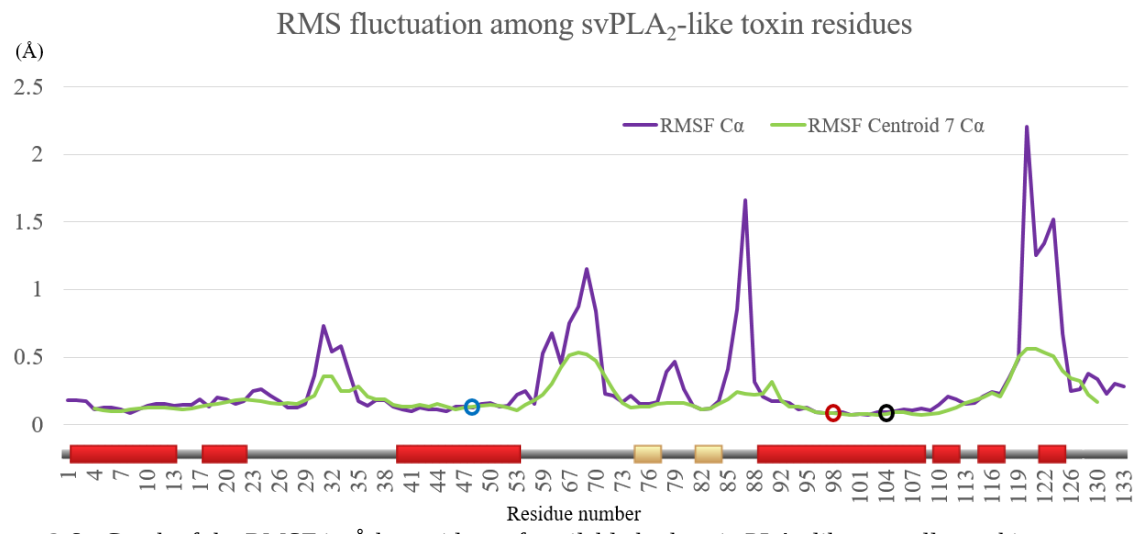


Figure 2.6 - Graph of the RMSF in Å by residues of available bothropic PLA_2-like crystallographic structures. The RMSF of Cα and centroid Cα (7 consecutive Cα, CC7) are colored in purple and green, respectively. Circles highlight the chosen CC7 as coordinate system for angle calculation, CC7/98 as the origin in red, CC7/104 as X axis in black and CC7/48 as Y axis in blue. Secondary structure is shown above residue numbers with helix and sheet colored in red and yellow, respectively.

As expected, the α-helices are the most rigid structure in the available models, being two of them almost parallel (helices 2 and 3 in **Figure 2.1 A_1**). These two helices are easily seen in the simple drawing representation in **Figure 2.2**. If one tries to see the arrangement

between the monomers (represented by rectangles) in the **active state** dimer, there is a twofold symmetry whose axis is almost inside the longest helix. Simplifying, one rectangle of the **active** dimer may be placed in the other after a rotation of 180^{o} of the diagonal (from the top right to the left) and after a translation. If the large dimer is chosen, a translation to the right is required and, if the compact dimer is chosen, to below the page. Taking into account a more realistic representation of the dimer and the membrane (or the sulfate plane) in **Figure 2.1 B_2** and **B_3**, the four α-helices are almost parallel to each other and perpendicular to the plane of sulfates (membrane). These α-helices look like four drills perpendicular to the soil. In this metaphor, the longest α-helix has the MDoS on its edge and the MDiS close by in approximately a parallel plane to the plane of sulfates (membrane). Thus, we have chosen these two helices as a reference for the geometric description, the X axis is the center of the longest α-helix, the Y axis is formed by a vector that connects X axis to the other parallel α-helix, and the last Z axis is composed by the normal vector of the plane XY. Such description may be compared to the airplane axes of movements (**Figure 2.5 D** and **E**), being the chosen X axis (long α-helix) the body, the Y axis the wings and the Z axis the vertical stabilizer.

To accomplish these, we generated a centroid coordinate obtained by the mean coordinate of 7 consecutive Cα atoms (named herein as CC7). We chose it because 7 Cα atoms compose approximately two α-helix complete turns (7.2 residues) and when their centroid coordinate will be close to the α-helix center (gray spheres in **Figure 2.5 E**). As an example, the first coordinate of a model composed by CC7 would be the CC7/4, which is the centroid position of Cα of residues 1 to 7. To best choose the coordinates, we calculated the same RMSF using CC7 coordinates of all evaluated models (**Figure 2.6** in green line). As it was already expected, this centroid coordinate takes into account more atoms and thus has a lower fluctuation. Afterward, we searched for three local minima: two farthest apart in the longest α-helix, to reduce sensibility to local variation, to compose one axis; and one local minimum of the other parallel α-helix to constitute another axis. We chose the axes by taking the CC7/98 as origin, CC7/98 to CC7/104 as X axis, CC7/98 to CC7/48 as Y-axis (**Figure 2.5 E**). The Z axis is automatically generated by the normal vector of the plane XY.

It is no coincidence that the coordinates composing the axes are similar to Dos Santos *et al.* (2009) choice (**Figure 2.3**), i.e. Cα of residues 97 and 108 of the longest α-helix and residue 43 of the other parallel helix. Our new angle description is an update of previous proposed methodology, and it is related to our current knowledge of the monomers

rearrangement gathered in hours and hours spent analyzing the different states of these toxins that was didactically represented in **Figure 2.2**.

Our second step was to place the template, 3IQ3/A, into the desired axes, i.e., its CC7/98 in the origin of the coordinate system in the PDB and so on. This step was done manually moving molecule as a rigid group in *COOT*. Then, we prepared a *python script* that standardizes PDB models numbering and inverts monomers chain letters, generates CC7 coordinates, applies superpositions and summarizes results into a table. To run this *script*, 3 arguments are requested in the following order: i) path of a folder containing protein coordinate files (in format and extension of PDB); ii) template (already positioned at the XYZ axes desired position); and iii) chosen name for tables to be saved. Each of the steps of the *script* is described below:

1. Correction of PDBs renumbering by removal of gaps;
2. Generation of CC7 coordinates;
3. Calculation of distance between each Cα of the evaluated PDBs and the template correspondent Cα
4. Calculation of distance between each CC7 of the evaluated PDBs and the template correspondent CC7
5. Creation of a PDB where chain A is called B and vice-versa
6. Superposition of monomers A of given dimer models (including the ones that monomer B is called A) into the template
7. Superposition of monomer A into B of each dimer model
8. Calculation of translation of monomer A gravity of mass to monomer B gravity of mass
9. Extraction rotation matrix from superposition log file done in the 7th step and conversion into Euler angles in ZYX convention (same order as shown in **Figure 2.5 A**, **B**, **C** and **D**)
10. Generate distance plots and angle of each structure

The steps 4 and 5 are done using *lsqkab* (KABSCH, 1976) asking to output deltas (distance calculation), and the steps 7 and 9, requesting to output *xyz* (coordinates) and *log* (contain rotation matrix). The step 8 is using the gravity of mass function available in *PYMOL*. The conversion from matrix rotation to ZYX convention was implemented using equation 449 available **in section 6** of topic 8.11 Euler Angle Sequence (3,2,1) of Diebel's publication (DIEBEL, 2006). Functions of each step were generated, and *script* flow may be easily altered. All the outputs are saved in separate folders:

- step 1 is saved in ./0_renumb folder;
- step 2 in ./1_Cc7
- step 3 in ./0_1_allCA

- step 4 in ./1_1_Cc7
- step 5 in ./1_BA
- step 6 in ./2_sup
- step 7 in ./3_itself
- step 8, 9 and 10 are saved into given path and given name when calling the *script*

Assuming that our structures are homodimers, our idea is to describe the rotation and translation necessary to superimpose, or to place, one monomer in the same position as the other. We chose *lsqkab* available in CCP4 suite, but any superimposing algorithm should suffice this task as the monomers are almost identical.

Table 2.3 includes these new proposed angles and other evaluations that will be described in next sections. The θ is the angle with lower divergence with a value close to 25 in all the sampling, with the exception of 1QLL (PrTX-II/fatty acid) and B1MMV (BthTX-I/MMV) that were colored in red. These two exceptions may be a new different state that has not been described yet, and they will be discussed in concluding remarks. For the structures with θ~25, two groups may be found. One (colored in magenta) composed of 1PA0, 2Q2J, 3HZD, 3I3H, 4K09, 4WTB and 2H8I that has ψ~140 and φ~55. And another (colored in dark green) consisting of 4KF3, 4YV5, 1XXS, 1Y4L, 3QNL, 4YZ7, 3CYL, 3CXI, 4YU7, 2OK9, 4K06, 3IQ3 and 3MLM that has ψ~180 and φ from 25 to 50.

As suggested by Dos Santos *et al.* (2009), there is a correlation between the monomer-monomer orientation and ligand binding. The first group is almost entirely composed of *apo* structures (**Supplementary Table 2.12**), with the exception of 4WTB and 2Q2J that have charged molecules or ions bound and 2H8I that has a hydrophobic molecule inside its protein cavity. As 2H8I is the only structure to contain a hydrophobic molecule bound in group one, we colored it red.

All the structures in the second group (dark green) are in a complexed form, with both of their hydrophobic channel occupied by an inhibitor molecule or natural compound(s) (**Supplementary Table 2.12**). Most of these structures have the abundant presence of negative molecules surrounding the protein, being R34 and PLA_2-like iFace (considered herein as the side chains of residues 16, 17, 20, 115 and 118 (MDoS)) the most common region. In Dos Santos *et al.* description, the first group would be considered **inactive** and the second **active state**. We will support such classification in further sections, but the angles of the **active state** near to 0 or to 180^o are closer to a symmetrical orientation, which may be

optimal to the expose same residues in both monomers in protein iFace. Thus, in such a symmetrical orientation, hydrophobic and basic residues (such as MDiS and MDoS) may be exposed together in a single plane.

2.8.3 Tertiary structure variability: Local measurement using Cα and Cβ distances

PLA_2-like proteins have a tertiary structure almost identical to enzymes PLA_2s sharing few small deviations between structures. Locally, two specific regions have a higher variation with RMSF above 1.5 Å (purple line in **Figure 2.6**). The peak of 86 and 87 (with 2.2 and 1.6 Å variation, respectively) were already reported as a region with two different conformations but, from our knowledge, no correlation to functional properties has been found so far. The most prominent variated residues are in toxin's C-terminal region. Among the residues in this region, the hydrophobic ones were described by Ambrosio *et al.* as able to form a hydrophobic knuckle, conserved in PLA_2-like proteins, and related to toxicity (AMBROSIO et al., 2005). Subsequently, two residues of this region (L121 and F125) were suggested to be responsible for toxin-induced membrane disruption and called the MDiS (FERNANDES et al., 2013). The MDiS are two hydrophobic and exposed residues stabilized by one hydrogen bond between the main chain oxygen of L121 and main chain nitrogen of F125 and by the rigid P123 that, together with K122, compose a short 3_{10} helix. These two residues, in special, are within the five most varied residues. Flexible regions are usually related to function as dynamic structural alterations are related to substrate binding, enzymatic action or interaction (FUGLEBAKK; ECHAVE; REUTER, 2012). Therefore, the variance in this local region highlights its importance.

To characterize the MDiS region by a simple measurement, we evaluated the distances of all the atoms of these two residues in each available structure and chose empirically the distance Cβs of residues 121 and 125. We support this choice with the facts that Cβ is the most conserved atom in different hydrophobic side chain atoms and it contributes to the residue's hydrophobic property. By this simple atomic distance measurement, we could differentiate the monomers into two states that we chose to call **canonical** and **non-canonical monomers**. The first has MDiS (orange in **Figure 2.7**), while the second has a stretched loop conformation where 121 and 125 are far from each other (brown in **Figure 2.7**) (BORGES et al., 2017).

The distances between the Cβ atoms from L121 and F125 from all **canonical monomers** are less than 5.5 Å, whereas the distances for the residues from all **non-canonical monomers** are usually larger than 10 Å (**Table 2.3**). The structures in the **inactive state** have its monomers asymmetrical, as one is in **canonical** conformation, while the other is in **non-canonical**. On the contrary, the structures in the **active state** have both monomers symmetrical and **canonical**. As the MDiS is essential for toxin membrane disruption (CHIOATO et al., 2007; FERNANDES et al., 2013), the presence of hydrophobic residues far from each other in the **non-canonical monomer** and its asymmetrical dimer orientation, that interferes in MDiS exposition in PLA_2-like iFace, supports classification of *apo* structures and 4WTB and 2H8I as an **inactive state**.

The 2H8I/A (BthTX-I/PEG400/A) is the only **non-canonical monomer** having its distance below 9 Å and the only structure in the **inactive state** having the hydrophobic channel occupied by molecules. The reduced distance between the MDiS Cβ atoms is probably a consequence of the presence of a small PEG molecule in the **canonical monomer** because that is the main difference compared to *apo* BthTX-I (**Supplementary Table 2.12**) and we observed that the electron density for the other PEG400 molecule in the **non-canonical monomer** is ambiguous. Therefore, 2H8I (BthTX-I/PEG400) structure may represent an intermediate state (colored in red in **Table 2.3**) between the **inactive** and **active states**.

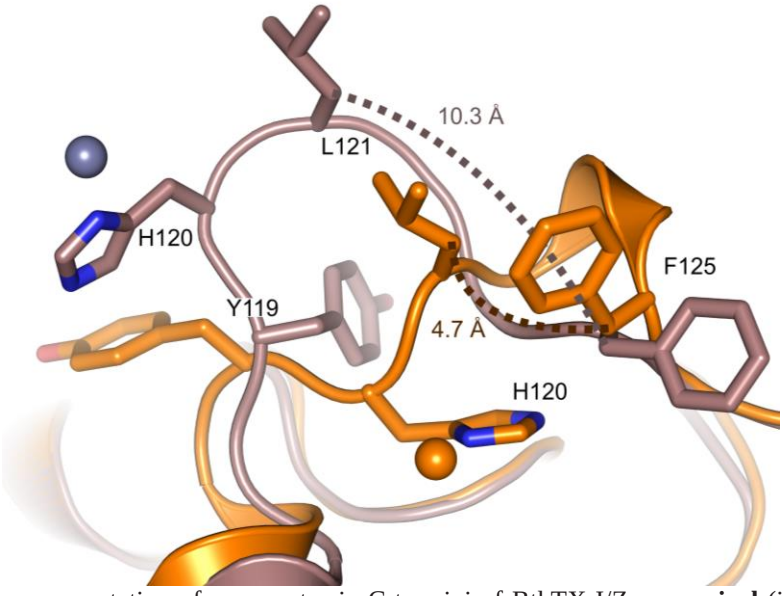


Figure 2.7 - Cartoon representation of asymmetry in C-termini of BthTX-I/Zn **canonical** (in orange) and **non-canonical monomer** (in brown).
Distances between MDiS Cβ residues (121 and 125) are showed on brown dots. Extracted from (BORGES et al., 2017).

2.8.4 PLA_2-like toxins hydrophobic channel characterization and accessibility

Bothropic PLA_2-like proteins have a cavity surrounded by positive and hydrophobic residues that are related to the entrance of hydrophobic or negative molecules in the hydrophobic channel (FERNANDES et al., 2014). The residues in this cavity are thought to be fundamental for the myotoxic activity of PLA_2-like proteins (*i.e.*, MDoS, MDiS, and H48) (**Supplementary Table 2.13**), supporting the importance of the hydrophobic channel, N and C-terminal regions for toxin activity. The entry of hydrophobic molecules into this channel is fundamental to induce dimer reorientation from the **inactive** to **active state**, a key step of the myotoxic mechanism of action (DOS SANTOS; SOARES; FONTES, 2009; FERNANDES et al., 2013). In this context, *apo* PLA_2-like toxins (empty hydrophobic channel) are in the **inactive state** (DOS SANTOS; SOARES; FONTES, 2009), and their dimers exhibit an asymmetrical conformation (one monomer is **canonical,** and the other is **non-canonical**). In contrast, the crystallographic structures of the PLA_2-like proteins with ligands bound in the hydrophobic channel are in the **active state** (DOS SANTOS; SOARES; FONTES, 2009) and their monomers in a symmetric conformation (**canonical**).

To better characterize these toxins activation by the entrance of hydrophobic molecules, we evaluated the accessibility of the hydrophobic channel through protein iFace with tunnel calculation using two algorithms, *MOLE2* and *CAVER3.0*. The characterization relevant tunnel of the models in the **inactive state** is more complicated than the **active state**, as there is no molecule inside the hydrophobic channel guiding the tunnel's path. For such a purpose, we used *MOLE*, that calculates the chemical properties, to characterize one representative of the **inactive state**. We chose the BthTX-I/Zn structure as this is one of the topic in this thesis (**section 2.8.8.1**).

The BthTX-I/Zn dimer has a large cavity of 2757 $Å^3$ surrounding the interior and the interface of both monomers, according to the calculations from the *MOLE2* algorithm. This cavity is composed mostly by N-terminal, calcium binding loop and C-terminal residues (**Supplementary Table 2.13**). Inside it, we assessed hydrophobic channel accessibility through tunnel computations using each H48 residue as the starting point. This amino acid is the deepest residue that interacts with different ligands found in the hydrophobic channel, such as zinc ions, *p*-bromophenacyl bromide (BPB), fatty acids and polyethylene glycol (PEG) molecules (**Figure 2.8 A**). We identified three different tunnels (**Supplementary Table 2.14**) in this region: **Tunnel 1** (**Figure 2.9 A**, green surface and **Figure 2.9 B**) and

Tunnel 2 (**Figure 2.9 A**, dark red surface and **Figure 2.9 B**) in the hydrophobic channel of the **canonical monomer** and **tunnel 3** (**Figure 2.9 A**, purple surface) in the **non-canonical monomer**. Both **tunnels 1** and **2** are long, 19.1 and 22.8 Å long, respectively, and are accessible to the solvent through the N-terminal region of the **canonical monomer** and C-terminal region of the **non-canonical monomer** (**Supplementary Table 2.14**). **Tunnel 3** in the **non-canonical monomer** is rather short, 6.0 Å long, and accesses exterior through residues that are essential only for catalytic PLA_2s (calcium binding loop, K49, and Y52) (**Supplementary Table 2.14**) and that are not related to PLA_2-like protein iFace (**Figure 2.1 B**$_3$ and **B**$_4$). The hydropathy values of **tunnels 1**, **2** and **3** are 0.29, -0.07 and -0.98, respectively; thus, only **Tunnel 1** is hydrophobic (**Supplementary Table 2.14**). With a superimposition between the monomers of the BnIV/myristic acid and the BthTX-I/Zn structures, we observe the myristic acid exactly inside **Tunnel 1** of the BthTX-I/Zn **canonical monomer** (zoomed square in **Figure 2.9 B**$_1$). On the contrary, the fatty acid clashes with the **non-canonical monomer** residues. Thus, the **inactive state** of PLA_2-like proteins, represented here as BthTX-I/Zn, has only one accessible hydrophobic channel in the **canonical monomer** (**Tunnel 1**) through PLA_2-like protein iFace.

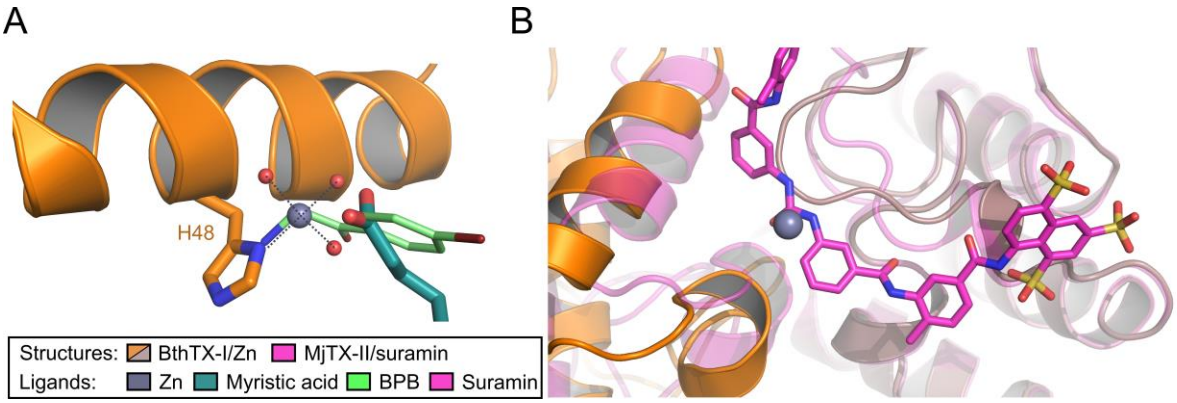


Figure 2.8 - Comparison of ligands interacting to PLA_2-like protein H48 (in **A**) and to MDiS (in **B**).
BthTX-I/Zn is shown in orange cartoon and zinc ions are represented by purple spheres. In **A**, Zn, myristic acid in dark green sticks and BPB are shown interacting to H48. In **B**, zinc shares similar position with urea group of suramin in purple sticks (from MjTX-II/suramin shown in purple cartoon). Adapted from (BORGES et al., 2017).

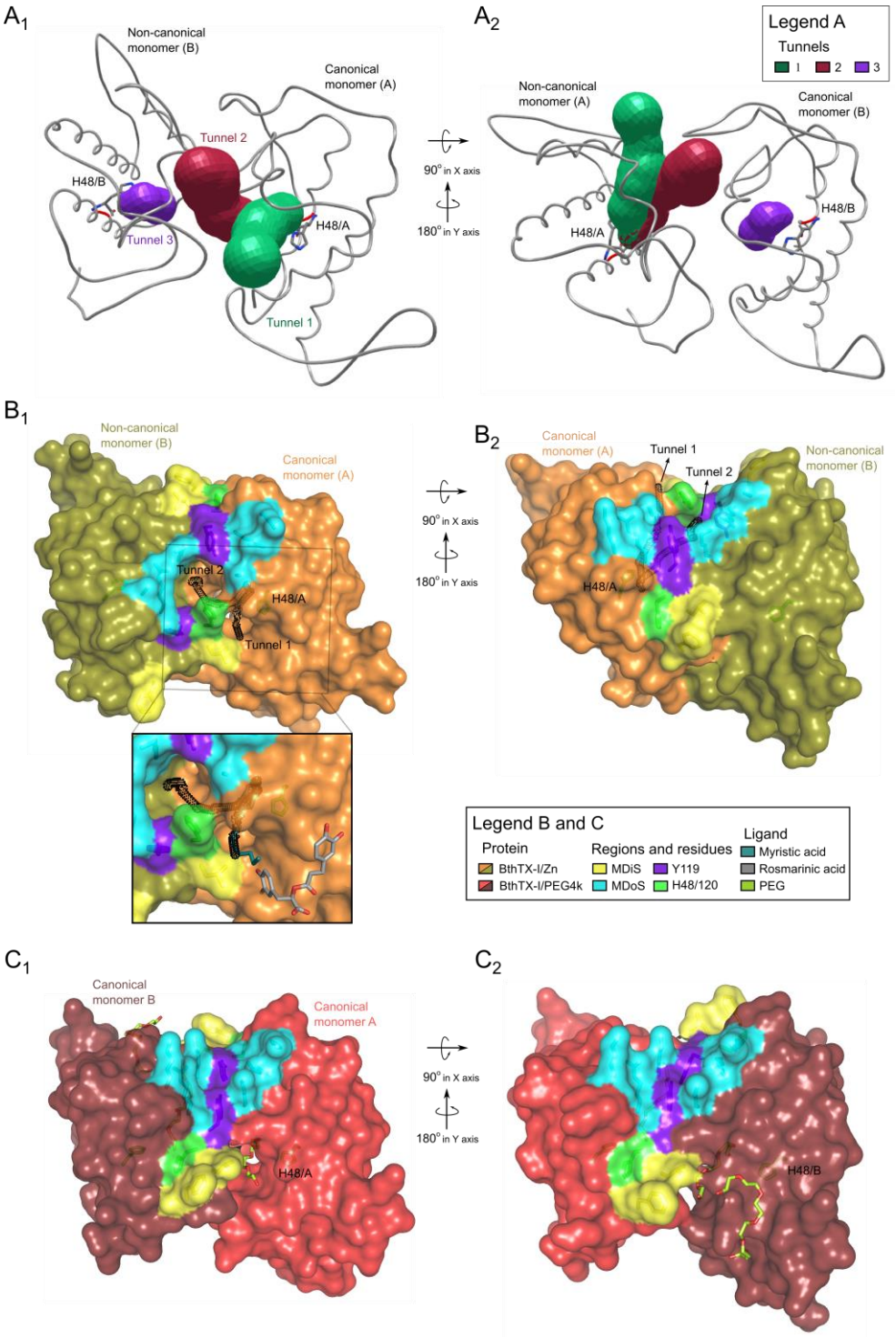


Figure 2.9 - Tunnels analysis for bothropic PLA_2-like crystal structures in the **inactive** (**A** and **B**) and **active states** (**C**).
Tunnel 1, which is the only tunnel with a hydrophobic feature, exists in **canonical monomers** (**A_1**, **B_1**, **C_1**, and **C_2**) and is accessible to the solvent through the N-terminal region of the **canonical monomer** and C-terminal region of the **non-canonical monomer** (**A** and **B_1**). (Zoom region in B_1) After superposition, myristic acid (PDB id: 3MLM) occupies **tunnel 1,** and rosmarinic acid (PDB id: 3QNL) is in **Tunnel 1** entrance region. In **B_2**, the hydrophobic channel of the **non-canonical monomer** is inaccessible. In **B** and **C**, MDiS is in yellow, MDoS is in cyan, Y119 residue is in purple, H48 and H120 residues are in green. Myristic acid and rosmarinic acid ligands are represented in dark green and gray sticks, respectively. Extracted from (BORGES et al., 2017).

We extended the tunnel's accessibility calculation for each H48 through *CAVER (command line version)*. We manually selected the calculated tunnels in *PYMOL*, and we calculated their volume using the tunnel' diameter and length generated by this program. When PEG or fatty acid were present in the model, we chose the tunnel whose path fills this molecule. The ligands PEG or fatty acid may be seen as a category of molecules capable of activating the protein as they are both found on the same site in toxin hydrophobic channel, their structures are similar, their effect in toxin structure are the same and their discrimination by electron density and B-factor analysis alone is difficult. Thus, the choice to model a fatty acid, as this ligand may have come from venom snake purification, or a PEG, if present in crystallization condition, may be a matter of the crystallographer choice, as none of these models contain atomic resolution diffraction data to distinguish. We chose *CAVER* for the systematic characterization because it produced a better consistency between the calculated tunnels than with the *MOLE*. With this latter algorithm, we could not find few of the tunnels and some of the calculated cavities had different volumes which interfere in the tunnels' depths.

All the **canonical monomers** of bothropic PLA_2-like proteins contain **tunnel 1** with a volume close to 400 $Å^3$ (**Table 2.3**), which structurally connects the N-terminal region and H48 with the C-terminal region of the other monomer. As a matter of fact, the N-terminal with its hydrophobic residues have been suggested to be important for toxin activity (DÍAZ et al., 1994; SOARES et al., 2000), and it may be related to the membrane interaction, similar to the MDiS. PLA_2-like structures in the **active state** display symmetrical and **canonical monomers**, and **tunnel 1** occupied by ligands, such as fatty acids or PEG molecules (**Supplementary Table 2.12**). Interestingly, a rosmarinic acid inhibitor is at the entrance of one of the PrTX-I/RA **tunnel 1** regions interacting with the N-terminal residues of one monomer and the hydrophobic residues of the C-terminal region of the other monomer (gray sticks in the zoom of **Figure 2.9 B1** and **Supplementary Table 2.12**), preventing PEG4k entrance.

PLA_2-like structures in the **inactive state** only exhibit **tunnel 1** for one monomer (**canonical monomer**). Curiously, the hydrophobic tunnels of BthTX-I/PEG400 and *apo* BthTX-I (3HZD) are closed by the position of the H120 side chain, whereas they are open in the available models of the other bothropic PLA_2-like proteins. The comparison between **tunnels 1** of the structures in the **active** and **inactive states** shows that their cores are

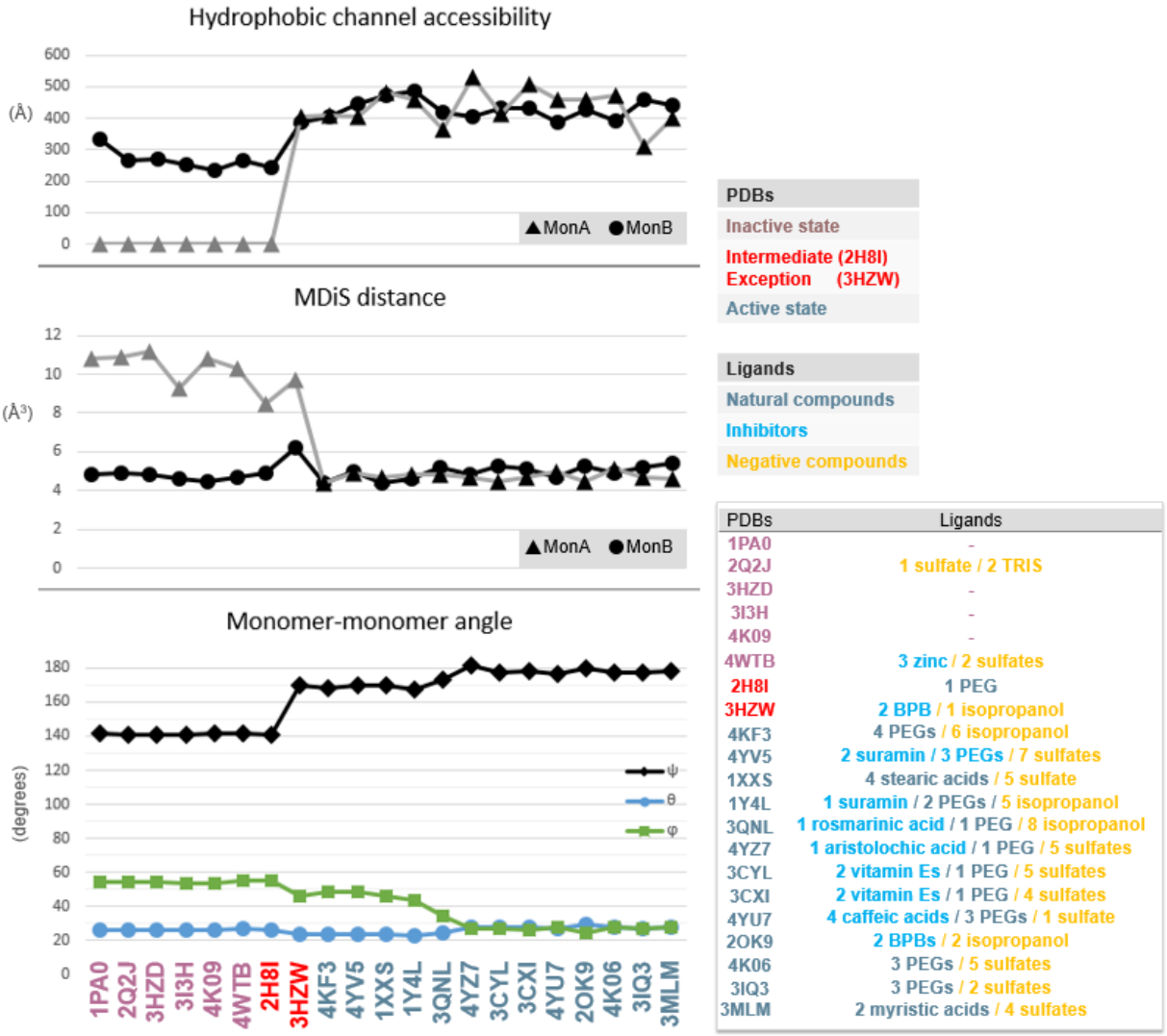
identical; however, their entrances are different due to the different dimeric orientations and differences in the C-termini. MjTX-II has an insertion in residue 120 in relationship to other bothropic toxins (BaspTX-II, BbTX-II, BnIV, Bnsp-VII, BthTX-I, PrTX-I and PrTX-II), thus instead of tunnel exits through the middle of the N-terminal helix, it is shifted to the beginning of N-terminal helix (SALVADOR et al., 2013b). This difference could be related to the higher φ on MjTX-II dimer than what is seen for other toxins (**Table 2.3**).

2.8.5 Results and discussion of global and local measurement

Herein with these global and local analysis, we can better describe the different states of bothropic PLA_2-like protein structures (**Figure 2.10**). The **inactive state** of bothropic PLA_2-like protein has the monomers in asymmetrical conformation. Their monomers A are in the **non-canonical** with dMDiS greater than 9 Å, and with its hydrophobic channel inaccessible (seen by triangles of magenta structures and 2H8I in the dMDiS and hydrophobic channel accessibility graphs in **Figure 2.10**). The geometric relationship of their monomers is far from a symmetrical orientation, as ψ and φ are distant to either 0^o or 180^o, respectively (seen on angles of magenta structures and 2H8I in monomer-monomer angle graph in **Figure 2.10**). Both monomers of these structures have no hydrophobic molecule inside its hydrophobic channel, with the exception of 2H8I.

The opposite is seen for **active state** PLA_2-like proteins. Both of their monomers are symmetrical with **canonical monomers** with dMDiS smaller than 6 Å, and both hydrophobic channels accessible (seen on the dark green structures in MDiS distance and hydrophobic channel accessibility graphs in **Figure 2.10**) and occupied by hydrophobic molecules. Their geometric relationship to one another is close to a symmetrical orientation, as ψ and φ are close to 0^o and 180^o, respectively (monomer-monomer angle graph in **Figure 2.10**).

A few exceptions of these observations may be seen. 2H8I is an intermediate state between **inactive** and **active**, as it has a stable hydrophobic molecule in the accessible hydrophobic channel (**canonical monomer**) and the dMDiS of the **non-canonical monomer** (8.5 Å) is an intermediate value between 9-10 Å and 5-6 Å. 3HZW is a model of the H48 chemically modified by the inhibitor BPB and one of its monomers has a **non-canonical monomer** (dMDiS greater than 9 Å), although with a symmetrical geometric relationship between monomers and with both hydrophobic channels accessible.



PDBs	Ligands
1PA0	-
2Q2J	1 sulfate / 2 TRIS
3HZD	-
3I3H	-
4K09	-
4WTB	3 zinc / 2 sulfates
2H8I	1 PEG
3HZW	2 BPB / 1 isopropanol
4KF3	4 PEGs / 6 isopropanol
4YV5	2 suramin / 3 PEGs / 7 sulfates
1XXS	4 stearic acids / 5 sulfate
1Y4L	1 suramin / 2 PEGs / 5 isopropanol
3QNL	1 rosmarinic acid / 1 PEG / 8 isopropanol
4YZ7	1 aristolochic acid / 1 PEG / 5 sulfates
3CYL	2 vitamin Es / 1 PEG / 5 sulfates
3CXI	2 vitamin Es / 1 PEG / 4 sulfates
4YU7	4 caffeic acids / 3 PEGs / 1 sulfate
2OK9	2 BPBs / 2 isopropanol
4K06	3 PEGs / 5 sulfates
3IQ3	3 PEGs / 2 sulfates
3MLM	2 myristic acids / 4 sulfates

Figure 2.10 - The hydrophobic channel accessibility with tunnel volume calculation, the distance of MDiS residues and monomer-monomer angle for all bothropic available PLA_2-like toxins.
The points in the abscissa are referred to the structure shown in the abscissa of the third graph, which is in the same order of the table in the legend on the right. In the first two graphs, monomer A and B are represented in triangle and circle points, respectively. In the monomer-monomer angle chart, the ψ, θ, and φ angles are colored in black, blue and green, respectively. The structures are colored according to their state, being **inactive** magenta, intermediate/exception in red and **active** dark green. Toxin in **inactive state** possess the asymmetrical monomers, as one is **non-canonical** with MDiS distance greater than 8 Å, its hydrophobic channel closed (tunnel with volume 0), and its orientation with the other monomer do not tend to the symmetrical relationship, as numbers are far from 0 and 180º. Toxins in the **active state** possess both **canonical monomers** with MDiS distance close to 5 Å, both of the hydrophobic channels opened, and the monomer-monomer orientation tends to a symmetrical relationship as the angles tend to either 0 or 180º.

Another exception, the monomers of the structures 1QLL and the biological unit of B1MMV (BthTX-I/MMV) have a geometric relationship far from what is seen for other structures since θ is not close to 24 (**Table 2.3**). Consequently, the calculated tunnel, the one that goes through calcium binding loop and N-terminal helix of one monomer and MDiS of the other, does not exist in these structures. By this reason, we did not include their results in

the graphs of **Figure 2.10**. Moreover, the crystallographic models of one particular toxin elucidated by Ph.D. student Lino F. G. de Lima, from our group, have a similar geometric orientation to 1QLL. This different geometric orientation far from a symmetrical relationship (φ~80^o) unable MDoS and MDiS residues to be exposed on a single plane. This asymmetry may be related to the smaller toxicity for this toxin in neuromuscular preparation in comparison to other bothropic PLA_2-like proteins, that Lino Lima together with our collaborator Dra. Marcia Gallaci have seen. This unusual asymmetrical geometric relationship could be related to a non-ideal exposition of MDoS and MDiS that reduces its activity.

2.8.6 Evaluation of flexibility of PLA_2-like protein in active and inactive states

With the evaluation of the available bothropic PLA_2-like proteins in the previous sections, we identified i) an **inactive state** in an asymmetrical conformation (ψ~140, θ~25 and φ~55) with a **non-canonical monomer** and both hydrophobic channels empties; ii) an intermediate in an asymmetrical conformation but one hydrophobic channel occupied; and iii) an **active state** symmetrical (ψ~170-180, θ~25 and φ~25-50) with both hydrophobic channels filled. By such analysis and observations, we could hypothesize that the entrance of a hydrophobic molecule in the accessible hydrophobic channel was the important step for protein activation and the main responsible for the geometric change in monomers that opens the other inaccessible hydrophobic channel.

To address this hypothesis, we used Normal Mode Analysis. Low-frequency Normal Mode (NM) can describe real-world protein motions and relate to fundamental biological properties (ALEXANDROV et al., 2005; THOMAS et al., 1999). We evaluated two structures. The only representative of the intermediate state (BthTX-I/1PEG, PDB id: 2H8I, and herein we will refer it as *Inac*), whose angles are ψ~140, θ~25 and φ~55, And the **active state** BthTX-I/3PEGs (PDB id: 3IQ3, and herein we will refer it as *Act*), whose angles are ψ~175, θ~25 and φ~25. For each case, we evaluated the NM with (complexed) and without ligands (*apo*). We chose the myristic acids as ligands that we extracted from 3MLM and introduced into the **canonical monomer(s)** of *Inac* and *Act* after monomer-monomer Cα superposition. *Inac* has only one **canonical monomer**, therefore its complexed state had only one fatty acid.

After minimizing each structure, we generated and evaluated the NM 7-14 manually and by the angles proposed in **section 2.8.2**. By our analysis, the presence of a hydrophobic

molecule inside toxin cavity is necessary to stabilize the dimer. Thus we chose to focus on results of the complexes. The graphs in **Supplementary Figure 2.26** show the angles between monomers in a degree of movement between -3 to 3 Å along Normal Modes 7 to 10 of *Inac* and *Act*. The NMs of these two structures are different. For BthTX-I *Inac*, we desire to observe the toxin activation, which is measured by an increase of ψ, stability of θ and decrease of φ (**Supplementary Figure 2.26 C**). The NM09 from the initial state (0 highlighted by vertical gray line) to -3 Å is the NM that best describes such movement (**Figure 2.26 A**). For inactivation of *Act* (**Supplementary Figure 2.26 C**), we looked for the opposite, the NM08 from the initial state (0) to -3 Å best describes such movement (**Supplementary Figure 2.26 B**). For the sake of comparison, the reference graph of morphing video needs to be looked backward since the chosen movement goes from 0 to -3 Å (**Supplementary Figure 2.26 C**).

To obtain a complete transition between the **active state** (*Act*) and the **inactive/intermediate state** (*Inac*) or vice-versa, we had to increase the deviation for generating the models. From structures of -6 until 1.5 Å, we obtained satisfactory values. To evaluate that *Inac* NM09 is the contrary movement of *Act* NM08, we plotted them together (**Figure 2.11**), inverting the NM direction of *Act* structures from negative to positive to opposite (1.5 to -6 Å which is in upper X axis in **Figure 2.11**). Thus, as seen by the similar movement pattern in both *Act* and *Inac* NMs, these states may reach one another. Our proposed geometrical description with the Tait-Bryan angles have been proven useful for a detailed description of the dimeric rearrangement between the two identified PLA_2-like toxin states and for the characterization of movement in Molecular Dynamics.

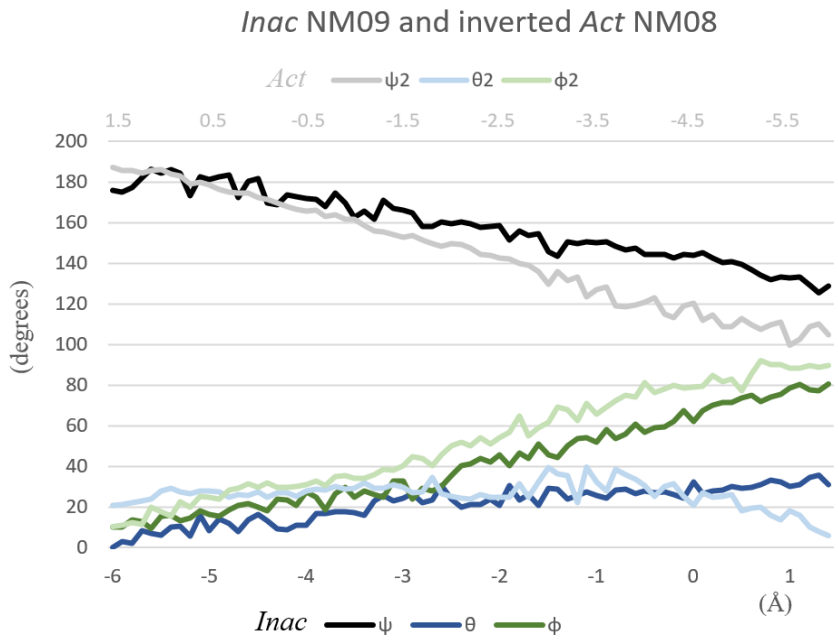


Figure 2.11 - Angles between monomers along the range of -6 to 1.5 Å of NM09 of *Inac* and along the range of 1.5 and -6 Å of NM08 of *Act*.
Inac ψ, θ, and φ are colored black, dark blue and dark green, respectively. *Act* ψ, θ, and φ are colored gray, light blue and light green, respectively. The Same pattern of the three angles is seen in both cases.

2.8.7 New steps in the myotoxic mechanism of PLA_2-like toxins

We observed new structural features of PLA_2s-like toxins with the observation of the MDiS distance, hydrophobic channel accessibility and geometric orientation between monomer of available structure. Thus, we suggest the following myotoxic mechanism for PLA_2-like proteins, in which item ii was added and additional details were added to items i, iii and iv of the previously proposed myotoxic mechanism (FERNANDES et al., 2013):

i) The activation starts with the entrance of **one** hydrophobic molecule/fatty acid through **tunnel 1** of the **canonical monomer**.

ii) Approximation of L121 to F125 residues from the **non-canonical monomer** (**Figure 2.12 A**).

iii) The dimer is reoriented with an increase of ψ and a decrease of φ, and the **non-canonical monomer** adopts its canonical conformation by inverting the Y119 and H120 positions and further approximating the L121 and F125 positions (**Figure 2.12 B**).

iv) The inaccessible and vacant hydrophobic channel is opened by the more symmetrical dimer orientation and a hydrophobic molecule/fatty acid is bound, stabilizing the toxin **active state**.

v) Protein-membrane docking through the MDoS.

vi) Membrane destabilization by its interaction with the MDiS.

vii) Uncontrolled influx of ions and, consequently, cell death.

Steps i to iv of the proposed myotoxic mechanism are represented by two morphing videos based on the different identified states of the BthTX-I crystallographic structures (Video 1 and 2 available in https://goo.gl/nNEQMY and https://goo.gl/iyhplp, respectively) (BORGES et al., 2017).

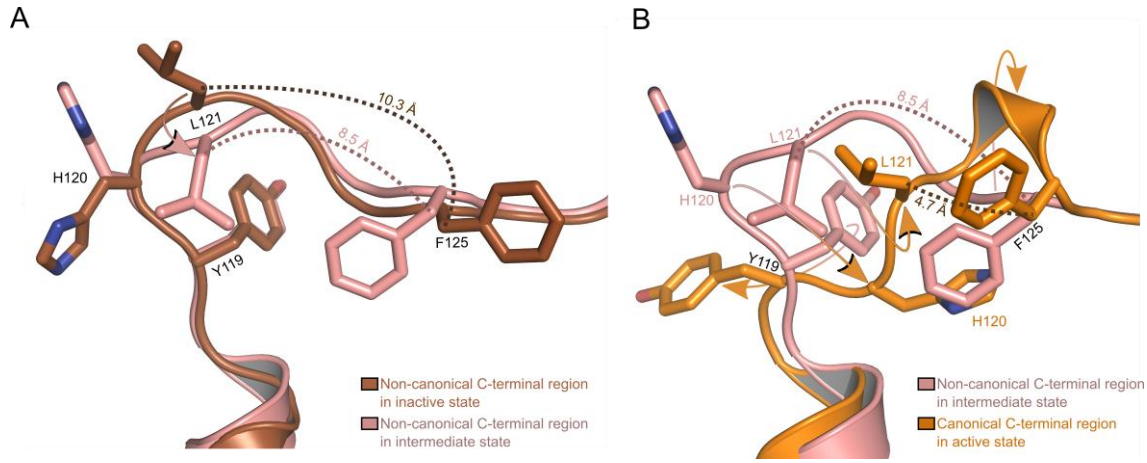


Figure 2.12 - Structural transitions of the C-termini from the BthTX-I that occurs during hydrophobic molecule entrance.
The first state is represented by BthTX-I/Zn **non-canonical monomer** showed as a brown cartoon (**A**). An intermediate second state is found only in the **non-canonical monomer** of the BthTX-I/PEG400 (PDB id: 2H8I) showed as a pink cartoon (**A** and **B**). Finally, a third state is represented by the **canonical monomer** of the BthTX-I/PEG4k (PDB id: 2H8I) showed as orange cartoon (**B**). In **A**, the transition from the first to the second state is characterized by a reduction of 10.3 Å to 8.5 Å from L121-F125 Cβ distance resulted from the structural change of Leu121 caused by hydrophobic molecule entrance in the **canonical monomer** channel. In **B**, the transition from the second to the third state is characterized by the change of **non-canonical monomer** to **canonical** because Y119 and H120 residues inverts their direction (180º rotation), leading to the approximation of the L121 and F125 residues (4.7 Å) and the formation of a short 3_{10} helix. Side chains of Y119, H120, L121 and F125 residues are shown in sticks. Meaningful conformation change pointed in arrows.

2.8.8 Structural studies of BthTX-I complexed with zinc

One of the main themes in current toxinology is the development of alternatives to the conventional serum therapy that does not efficiently neutralize snake venom injuries. Currently, the majority of studies with this objective focus on compounds isolated from plants used in folk medicine in developing countries (SÁNCHEZ; RODRÍGUEZ-ACOSTA, 2008; SOARES et al., 2005). On the other hand, only a few studies were performed to evaluate the inhibitory effects of divalent cations against snake venoms and, more particularly, PLA_2-like

toxins (HELUANY et al., 1992; RODRIGUES-SIMIONI et al., 1995; SOARES et al., 2002), despite their role in physiological functions and their use in the treatment of diseases (PRASAD, 2012; PRASAD et al., 2009). Zinc ions are relatively harmless to humans and have been employed in different therapies. It is an effective antioxidant and anti-inflammatory agent, and it may have beneficial effects on myocardial pathologies and atherosclerosis (LITTLE et al., 2010). Zinc has been administered to prevent specific types of cancer (DHAWAN; CHADHA, 2010; GRATTAN; FREAKE, 2012; PRASAD et al., 2009) and blindness in patients with age-related macular degeneration, to complement the treatment of diarrhea, to assist the immune system, and to treat the common cold, Wilson's disease, and sickle cell disease (PRASAD, 2012).

Recently, we have published an article reporting the complex between bothropstoxin I (BthTX-I) and zinc ions by functional, calorimetric and structural methods to obtain new insights into the mechanism by which the PLA_2-like proteins are inhibited (BORGES et al., 2017). We have shown the zinc is able to protect the mouse phrenic nerve-diaphragm preparation against the toxic effect of BthTX-I by the functional essays, twitch-tension, and membrane resting potential. Exploring the zinc anomalous peak by crystallography, we identified zinc ions interacting with three different sites in the dimeric BthTX-I, the H48/A, H120/A and H120/B (**Figure 2.13**). We confirmed this stoichiometry by calorimetric experiments. We are not including the description of the crystallographic data in this thesis, as it was already described in a Master thesis (BORGES, 2012), but the structural analysis was, and it follows below. The entire article is attached at the end of this thesis.

2.8.8.1 Zinc interaction site and its relationship to inhibition

Based on previous functional and structural experiments with PLA_2-like toxins and inhibitors, three different inhibition mechanisms were suggested for PLA_2-like proteins. First, ligands, such as BPB, rosmarinic acid, and suramin, binding inside the hydrophobic channel or in its entrance, blocking the access of fatty acid molecules. Second, ligands, such as caffeic acid, aristolochic acid, and suramin, interacting to specific residues of the MDoS or MDiS regions (FERNANDES et al., 2015), preventing the protein/membrane interaction. And third, suramin inducing toxin oligomerization (SALVADOR et al., 2015). Our study identified that the zinc ions interact with the BthTX-I structure at two different sites (**Figure 2.13**): i) H48 in the hydrophobic channel from the **canonical monomer** (Zn1) and ii) H120 (near the MDiS)

from both the **canonical** (Zn2) and **non-canonical monomers** (Zn3). These interacting regions would be related to the inhibition sites for the first and second inhibition processes, which will be discussed below.

The conservation of the catalytic network in PLA_2-like toxins has been attributed to the maintenance of the hydrophobic channel (AMBROSIO et al., 2005; FERNANDES et al., 2014) for entrance of a fatty acid that would lead to an oligomeric change for PLA_2-like proteins, a fundamental step in exerting myotoxicity (FERNANDES et al., 2013, 2014). Supporting this idea, different ligands have been shown to interact with H48, such as the BPB inhibitor, fatty acids, and PEG molecules (**Figure 2.8 A**). Although the dimer orientation of the BthTX-I/Zn and dimeric BthTX-I/BPB structures are different (monomer-monomer angle graph in **Figure 2.10**), which may lead to the binding of Zn ions only to the **canonical monomer** H48, the zinc and C atoms of BPB share almost same position interacting to toxin's H48 Nπ atom (**Figure 2.8 A**); thus, their inhibitory mechanisms may be related.

Zn2 and Zn3 interact with the H120 Nπ atom from both monomers (**Figure 2.13 B** and **C** and **Figure 2.8 B**) in the vicinity of the MDiS region. Similarly, the PrTX-I/aristolochic acid+PEG4k and MjTX-II/suramin+PEG4k crystallographic structures possess inhibitors bound to the MDiS region, whose interactions were related to their inhibitory abilities (FERNANDES et al., 2015; SALVADOR et al., 2015). With the superposition of these structures with the BthTX-I/Zn monomers, we reveal that zinc is located close to aristolochic acid and this ion shares the same position as the suramin urea group (**Figure 2.8 B**). Therefore, the binding of Zn ions to H120 may prevent the interaction between the MDiS residues (particularly L121) and the membrane.

Zinc is the only inhibitors which interacted to a PLA_2-like protein in the **inactive state**. The structures in this state are characterized by asymmetrical monomers, whether only the hydrophobic channel of the **canonical monomer** is accessible through the **Tunnel 1** (for more details, see **section 2.8.4**). The zinc ions are interacting to both beginning and ending part of **Tunnel 1**, as it interacts to H48 of **canonical monomer** and the H120 (in the vicinity of MDiS residues) of the **non-canonical monomer**. As the entrance of a hydrophobic molecule through the **Tunnel 1** of the only accessible hydrophobic channel is require for toxin activation, the presence of zinc in this tunnel may block this process, inhibiting toxicity.

In addition, the interaction of zinc ions with the H120 in the **non-canonical monomer** may interfere with the oligomeric change of the toxin to its **active state**, as a

position inversion between Y119 and H120 is necessary (**Figure 2.12**). Although H120 is conserved for many bothropic PLA_2-like proteins, a natural variant of tyrosine is found in a few bothropic snake venoms (ARNI et al., 1995; LIZANO; LAMBEAU; LAZDUNSKI, 2001; NÚÑEZ et al., 2004; PONCE-SOTO et al., 2007; SALVADOR et al., 2013a; WATANABE et al., 2005; YUNES QUARTINO; BARRA; FIDELIO, 2012), with which zinc would probably not interact, reducing its inhibitory capability.

Finally, it is interesting to highlight that H48 is a strictly conserved residue for all PLA_2-like proteins, and H120 is conserved for many bothropic PLA_2-like proteins. Thus, the inhibitory mechanism proposed here would be similar for several bothropic PLA_2-like toxins, and zinc could be a potential inhibitory agent for anti-bothropic snake venom therapy.

Based on these observations, we can hypothesize that the Zn ions simultaneously inhibit BthTX-I through two different mechanisms: i) preventing fatty acid binding to H48 and thus avoiding the state transition from the **inactive** to **active state**, and ii) by binding directly to the MDiS region, preventing the membrane disruption process.

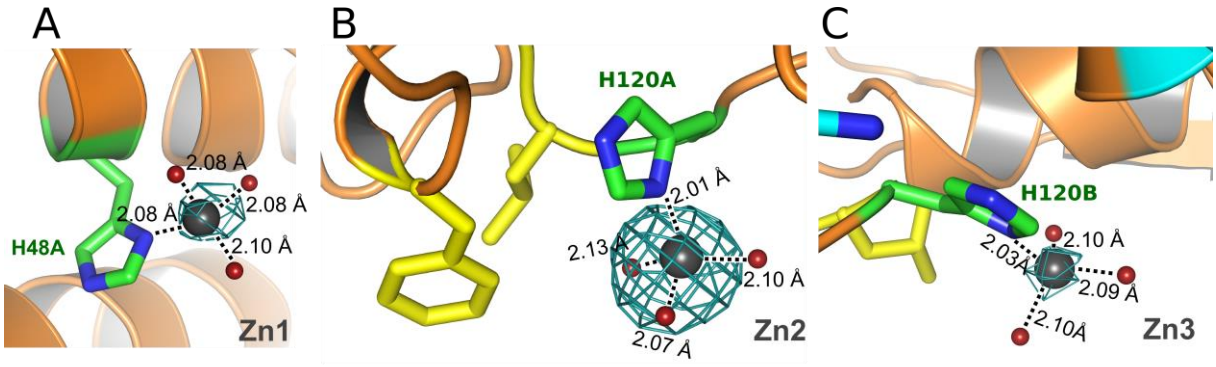


Figure 2.13 - BthTX-I and zinc ions interacting regions and coordination distances (black dashes).
The anomalous difference map is contoured at 5 σ and shown in cyan. Zinc spheres are represented in gray, water spheres in red, side chain of interacting histidines in green sticks, MDoS and MDiS side chains in cyan and yellow, respectively. In **A**, Zn1 and H48/A (monomer A); In **B**, Zn2, C-terminal H120/A and the MDiS/A; and in **C**, Zn3 and H120/B. Extracted from (BORGES et al., 2017).

2.8.9 Structural studies of BthTX-I complexed with MMV

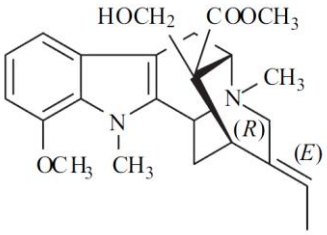
BthTX-I, the most abundant toxin in the venom of *Bothrops jararacussu*, induces myotoxicity and participate in the myonecrosis in bothropic accidents. Despite these consequences not being well neutralized by antivenom treatment, many plants have been used in folk medicine to treat ophidic accidents and may fill this gap (SOARES et al., 2005). One of them is *Tabernaemontana catharinensis* that possess inhibitory effect against snake venom of *Bothrops jararacussu* and its main toxins, BthTX-I e BthTX-II (VERONESE et al., 2005).

Among the active principles of this plant and the responsible for this antibothropic activity, MMV (**Figure 2.14**, M_w=431.5 g/mol) is an outstanding candidate (BATINA et al., 2000). MMV consists in a different inhibitor since it possesses a basic charge with a quaternary nitrogen (PEREIRA et al., 2008) and most of the inhibitors have either positive charge or are neutral (SOARES et al., 2005). Recently, our collaborators from the pharmacology laboratory of Márcia Gallaci (UNESP Botucatu, Brazil) evaluated the inhibitory effects of MMV against BthTX-I in one of the most sensitive assays of PLA_2 and PLA_2-like protein myotoxicity, mouse phrenic nerve-diaphragm preparations. Unfortunately, only a partial inhibition is observed (CARVALHO, 2011), MMV may not be the best candidate for an antivenom complement, but together with structural analysis, they may aid the comprehension of the PLA_2-like proteins myotoxic mechanism of action.

MMV did not induce any considerable quaternary structure alteration in BthTX-I, as we observe similar bands of SDS-PAGE gel electrophoresis on both the toxin itself and on the complex (**Figure 2.15 A**). These bands refer to BthTX-I dimeric structure, which is 27.5 kDa calculated from its theoretical molecular weight, and it is in agreement with other PLA_2-like protein studies (FERNANDES et al., 2014).

MMV also is not able to change BthTX-I secondary structure, as we did not observe significant differences in the CD spectra of BthTX-I when the inhibitor is introduced into toxin solution (**Figure 2.15 B**). The well-defined peaks at 208 and 222 nm in both spectra indicate that snake venom PLA_2-like structure kept its typical helical content.

We obtained crystals from cocrystallization experiments using lyophilized samples of BthTX-I and MMV grown in the hanging drop vapor diffusion method in 18ºC (MCPHERSON, 2009). The drops setup was 0.9 μL of protein (21.3 mg/mL), 0.45 μL of MMV (10.5 mg/mL), 0.9 μL of reservoir solution. The molar ratio of the drop was approximately 1:8 protein ligand. The reservoir solution was 0.1 M sodium citrate, pH 5.6, 20% PEG4k, 22% isopropanol similar of previous crystallization conditions (FERNANDES et al., 2010). We cryoprotected the crystals in reservoir solution containing 10% glycerol, collected it in nylon loops and flash-cooled in liquid nitrogen. We collected X-ray diffraction of this crystal at LNLS-MX1, on a MAR CCD imaging plate detector to a dataset of resolution of 2.2Å (**Table 2.4**).



MMV: 12-methoxy-4-methyl-voachalotine

Figure 2.14 - Molecular structure of MMV (12-methoxy-4-methyl-voachalotine)

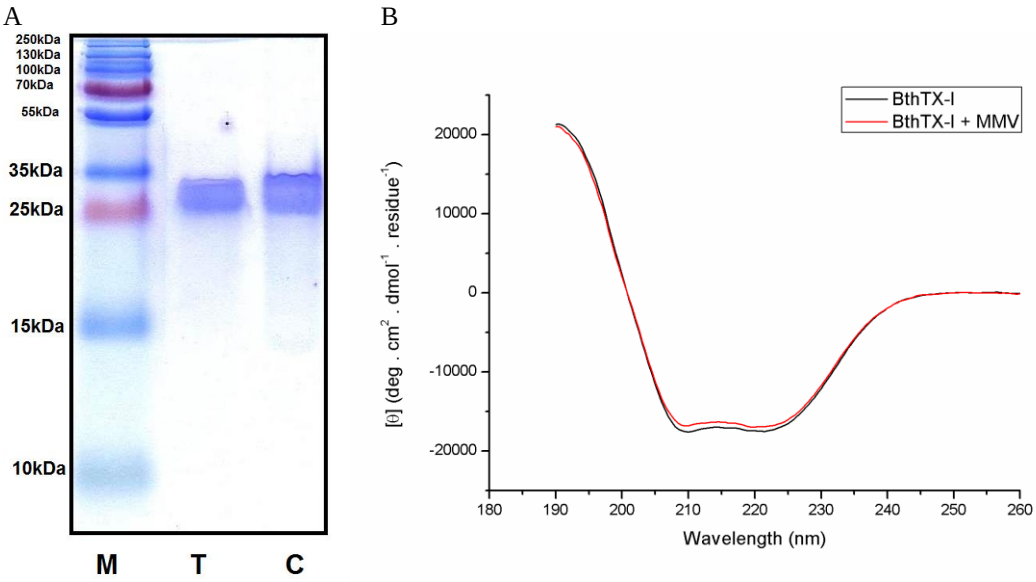


Figure 2.15 - Effect of MMV in BthTX-I quaternary and secondary structure.
In **A**, SDS-PAGE electrophoresis gel in non-denaturing conditions of BthTX-I (Lane T) and its complex with MMV (Lane C). The marker is on the left-hand side (Lane M) with its molecular weights as a reference (Thermo Scientific PageRuler™ Plus Prestained Protein Ladder). A dimer is observed in both toxin and complex. In **B**, Far-UV circular dichroism spectra of BthTX-I (black line) and its complex with MMV (red line). No significant differences are observed when the ligand is present.

We indexed and integrated the dataset with *iMosflm* and scaled through *SCALA*. The highest symmetry in the dataset was P622, but the Molecular Replacement failure finding a solution. As the intensity statistics looked unusual, an indication of twinning as evaluated with *XTRIAGE*, and no twin law is possible in this space group, we lowered the space group symmetry to P321. The intensity statistics indicated the merohedral twinning (Wilson ratios ($\langle I^2\rangle/\langle I\rangle^2$ = 1.615 (perfect twin: 1.5)) and moment of E ($\langle|E^2-1|\rangle$=0.574 (perfect twin: 0.375) which explains the apparent higher symmetry. We successfully obtained a Molecular Replacement solution in space group $P3_221$ using a native BthTX-I structure and the use of the correct twin law (-h, -k, l) was necessary for reducing R factors and solving the structure. We obtained a dimer in the ASU with *PHASER*, we refined the structure with *Refmac5* and

manually built with *Coot* using Fo-Fc and 2Fo-Fc maps. In final stages of refinement, we decreased the weighting term to 0.1 to obtain reasonable angle and bonds rms and a R-R_{free} gap of 0.5%.

We found a large electron density area in omit map close to monomer A, and it matched the MMV structure perfectly (Figure 2.16 A). Its overall B factor is in agreement with protein monomer A and B factor (Table 2.4). We substituted three neighboring water molecules to an isopropanol molecule in monomer B. The electron density of the protein MDiS region was not clear in both of the monomers. Thus we reduced the occupancy to 0.0 of the side chains of the residue 120 until the 125. Moreover, we did the same procedure for the side chains of K16/A, V31/B, K36/A, K53/A, K70/A, K93/B, K115/A, K116/B and K127/A due to the lack of electron density. All the other atoms were refined with full occupancy (1.0) including ligands. The N67/A is the only residue out of allowed Ramachandran Plot region, and this is justified by their its interaction to the inhibitor MMV. The final model reached R/R_{free} 16.0/21.1% and molprobity score of 2.0, which is in the 80th best percentile of all structures elucidated from datasets of similar resolution (2.1 ± 0.25 Å). In the Ramachandran plot, 93.7% and 5.9% are in the favored and allowed regions, respectively.

MMV partially inhibits the myotoxicity of BthTX-I as shown in mouse phrenic nerve-diaphragm preparations (CARVALHO, 2011). We analyzed its complex and native quaternary, tertiary and secondary structure through non-reducing SDS-Page, crystallography, and CD, respectively. Curiously, MMV did not induce any alteration in secondary and quaternary BthTX-I structure, although it caused drastic tertiary structural changes in loop regions of one monomer.

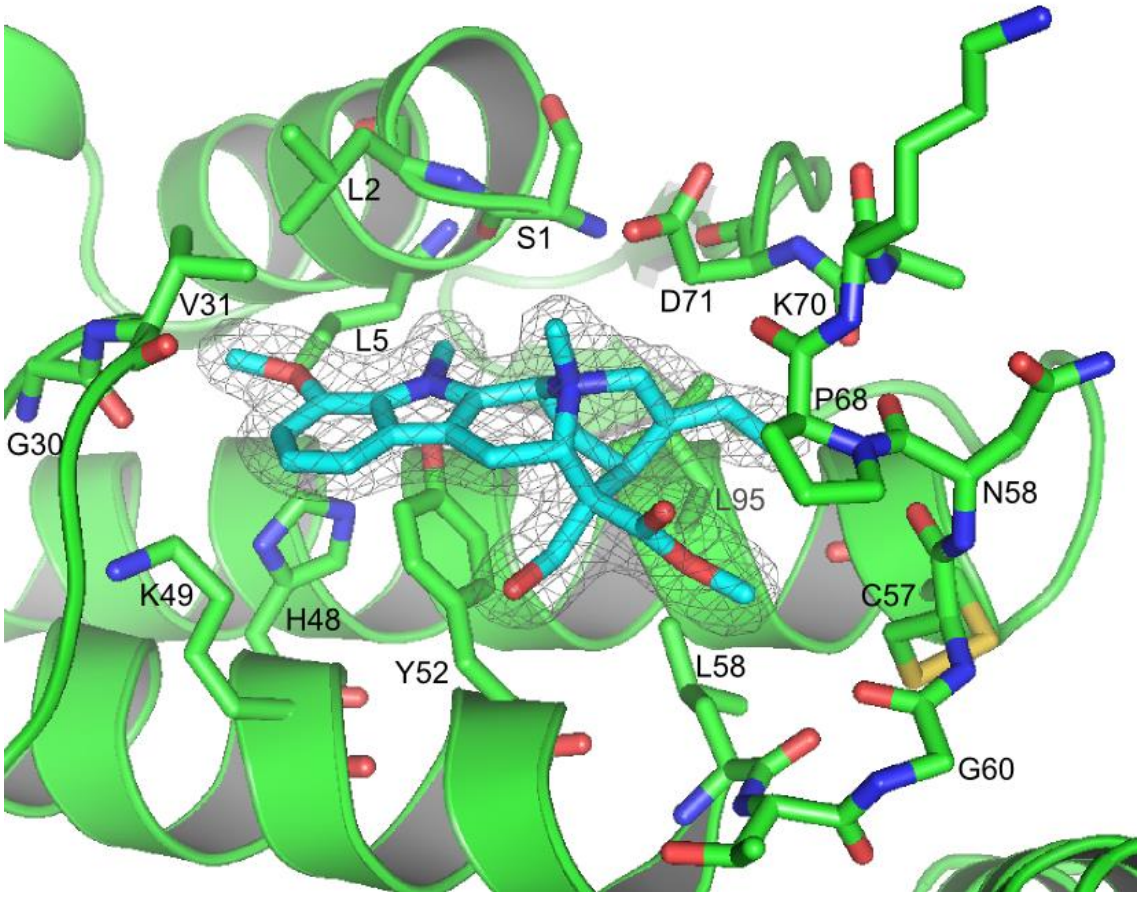
The MMV interacts to only one monomer of the toxin N-terminal (leucine residue in position 2, L2), calcium binding loop (G30, V31), α-helix 2 (H48, K49 and Y52), loop that connects α-helix 2 and antiparallel β-sheets (L58, G60, C61, N67, P68, K70, D71) and α-helix 3 (L85) (**Figure 2.16** and **Supplementary Table 2.12**). Possibly this unique BthTX-I interaction site is related to the MMV basic and hydrophobic character, as the ligand quaternary nitrogen interacts electrostatically to the protein D71 while the other interactions are with the toxin's hydrophobic channel residues. By these interactions, MMV completely shifts the position of the highlighted regions inside the red circles in **Figure 2.17 A**.

Table 2.4 - Data collection statistics of the X-ray diffraction of the crystals of BthTX-I/MMV

Beamline in LNLS	LNLS-MX1
Wavelength (Å)	1.43
Space group	$P3_221$
Unit cell	
abc (Å)	a=b=104.9, c=64.4
αβγ (º)	α= β=90, γ=120
Resolution (Å)	26.2 - 2.10 (2.21 - 2.10) [a]
Total reflections	224240
Unique reflections	24142 (1959) [a]
Multiplicity	9.3
R_{meas} [b] (%)	12.1 (35.6) [a]
CC (1/2)	99.6 (29.0) [a]
Completeness (%)	99.8 (99.5) [a]
I/σ (I)	17.1 (7.8) [a]
Twin law (fraction)	-h, -k, l (0.51)
R [c] (%)	16.0
R_{free} [c] (%)	21.1
RMS deviations from ideal values	
Bond lengths (Å)	0.008
Bond angles (º)	1.3
Ramachandran plot (%)	
Residues in most favorable regions	93.7
Residues in accepted regions	5.9
Residues in disallowed regions	0.4
Clashes	12.9
Molprobity score	2.0
Number of molecules/atoms and their averaged B factor (Å^2)	
Protein chain A	1 (26.4)
Protein chain B	1 (34.4)
MMV	1 (19.7)
Isopropanol	1 (36.8)
H_2O	261 (36.8)

[a] Number in parenthesis are for highest resolution shell. $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$, $\hat{I}_h = \sum_h \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$ redundancy independent R-factor. Calculated for I > -3σ (I). [c] $R_{cryst} = R_{hkl} \frac{||Fobs_{hkl}| - |Fcalc_{hkl}||}{|Fobs_{hkl}|}$. [d] R_{free} is equivalent to R_{crist}, but calculated with reflections (5%) omitted from the refinement process. RMS *root-mean-square*. Anomalous correction: percentage of correlation between random half-sets of anomalous intensity differences. Anomalous signal: mean anomalous difference in units of its estimated standard deviation $R_{hkl} \frac{|F(+) - F(-)R_{cryst}|}{\sigma}$. F(+), F(-) are structure factor estimates obtained from the merged intensity observations in each parity class.

A



B

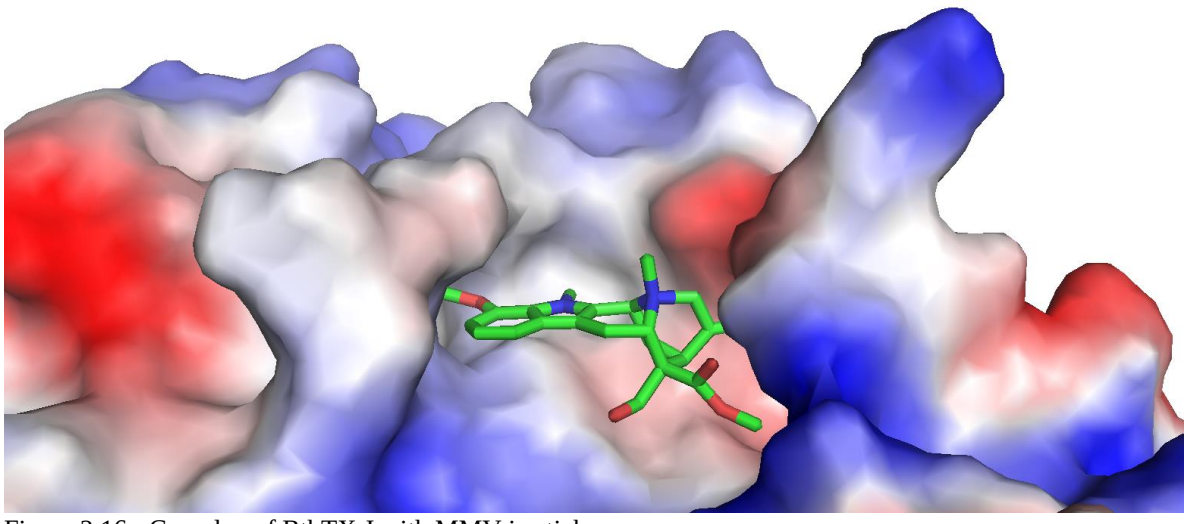


Figure 2.16 - Complex of BthTX-I with MMV in sticks.
In **A**, cartoon with interacting residues in sticks and omit map of the MMV in gray (4 σ), and in **B**, electrostatic surface.

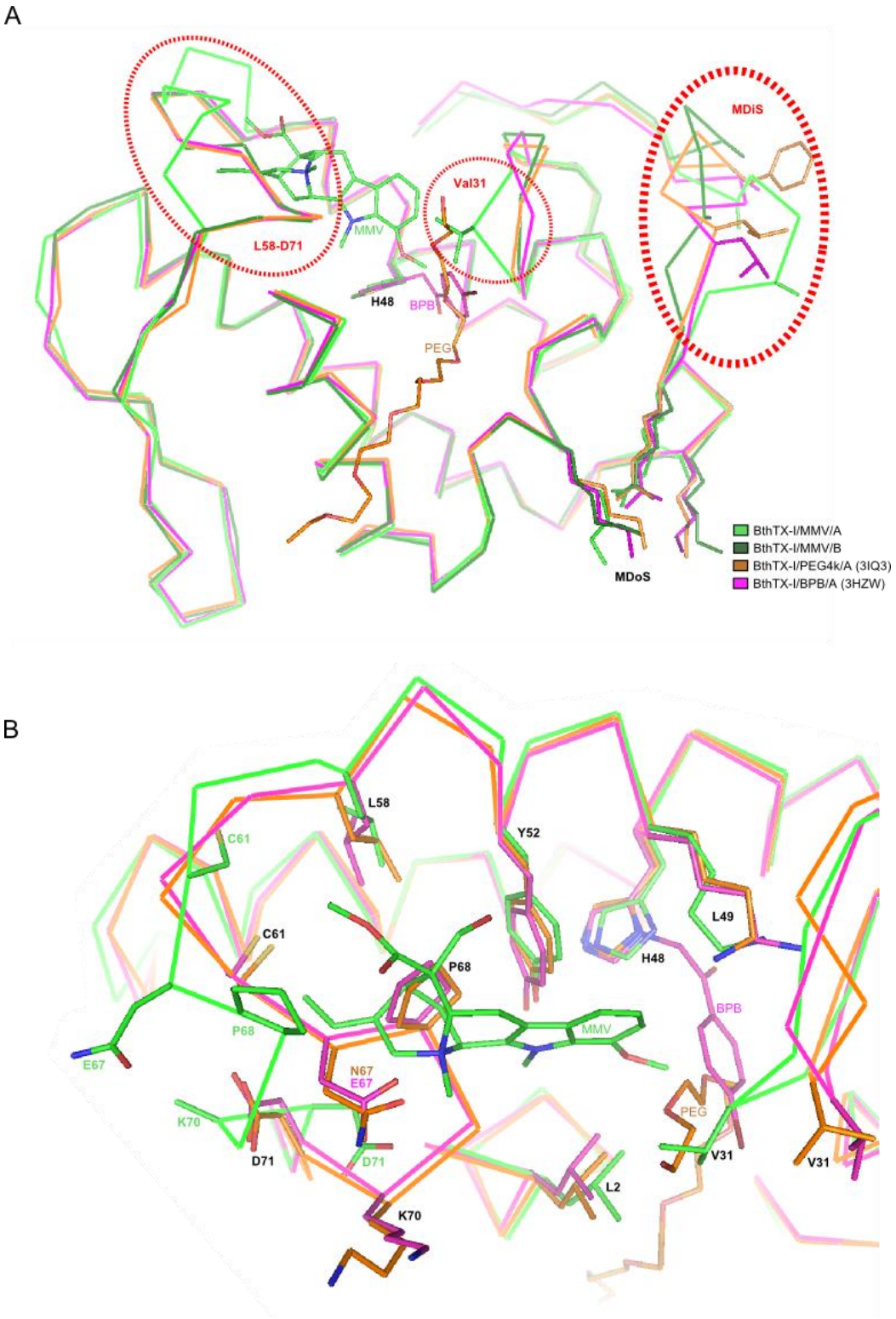


Figure 2.17 - Structural comparison between BthTX-I/MMV with other PLA_2-like protein structures.
In **A**, superposition of Cα in ribbons of BthTX-I/MMV monomer A (light green) and monomer B (dark green), BthTX-I/PEG (PDB id: 3IQ3) in orange and BthTX-I/BPB/A (PDB id: 3HZW) in magenta. The ligands are represented in sticks in the same colors of the protein from their complexes. Three main differences are observed in dashed red circles in the region of L58-D71, calcium binding loop (V31) and MDiS. In **B**, zoom and different angle of the first two distinct areas, side chain of MMV interacting residues in sticks, MMV pulls toxin V31 closer and occupies the ordinarily P68 position.

The BthTX-I monomer interacting to MMV is the one that most deviates from other bothropic PLA_2-like structures, as the lowest RMSD from superpositions (using all Cα atoms) to available bothropic PLA_2-like toxin is 1.5 Å (**Supplementary Table 2.15**). In the other hand, the monomer B with no MMV molecule is in a canonical conformation (MDiS distance of 3.5 Å) as seen by small RMSD values with **canonical monomers** (**Supplementary Table 2.15**). BthTX-I/MMV monomers are asymmetrical with a RMSD of 1.8 Å similarly to **inactive state** PLA_2-like protein models. MDiS side chains lack electron density in both monomers, although it is in a stretched conformation (L121/A/Cβ-F125/A/Cβ distance of 9.3 Å) in the **non-canonical monomer** containing MMV (right red circle in **Figure 2.17 A**). Thus, by monomer structural comparison with other PLA_2-like structures available in the PDB, we confirm that these three different regions highlighted in the **Figure 2.17 A** are unique.

Analyzing ligand site interaction with the complexed models available in PDB, we find MMV region interaction close to suramin and α-tocopherol with some atoms overlapping each other. Although, differently to α-tocopherol, MMV does not block the fatty acid site and it does not interact mutually to both monomers, which is also the case for suramin. Thus, MMV is the first inhibitor to be found close to N-terminal, calcium binding loop, and L58-D71 region compared with other inhibitor sites determined by crystallography (**Supplementary Table 2.12**).

The dimeric assembly of BthTX-I/MMV in the ASU is different from the compact and the large dimer (**Figure 2.2** and **2.18**). Such a different dimer may be a consequence of the different crystal packing, since the space group and unit cell of this structure ($P3_221$) (**Figure 2.18 A**) is different than BthTX-I **active** and complexed form (monoclinic or orthorhombic) (**Figure 2.18 B**) and different than BthTX-I in *apo* and **inactive** form ($P3_121$ in PDB id: 3HZD) (**Figure 2.18 C**). As PISA algorithm aids determination of the biological unit, i.e. the quaternary structure present in solution, by evaluating interaction within existing symmetry mates, we evaluated BthTX-I/MMV structure. Three main complexes were found and they are described in **Table 2.5**. The one in the ASU composed of monomer A with ligand MMV and by monomer B and ligand isopropanol (represented as $A_1MMV_1B_1POL_1$). Second, the monomer B and its symmetry (B' is generated by the symmetry operation -x, -x+y, -z+2/3) composed by B_2POL_2. And third, the monomer A and its symmetry (A' is generated by the symmetry operation x-y, -y, -z+1/3) composed by A_2MMV_2. Of these, the only two stable in

solution are the A/A' and B/B' that are thermodynamically favorable (positive values for ΔG^{diss} available in **Table 2.5**). The interface between A and B are not significant for the complex formation as shown by low CSS value (Complex Significance Score), another indication that the dimer in the ASU is not the biological unit. In the other hand, the interfaces connecting A/A' and B/B' are essential to the formation of the complex (CSS=1.0). Moreover, they are similar to each other and to the compact dimer (**Figure 2.18 B, C,** and **D**). Thus, both A/A' and B/B' dimers are possible to exist in solution, although A/A' is more stable as seen by higher ΔG^{diss}, lower energy, P-value of Δ^iG, and lower ΔG^{int}. Thus, we believe that BthTX-I in the presence of MMV adopts the dimeric structure of A/A' in solution, where MMV interacts identically to both monomers.

The structure of BthTX-I/MMV A/A' is in the compact dimer assembly (**Figure 2.18 E**), although their exact geometric orientations are different (**Table 2.3**). The ψ of 171° is similar to the **active form**, the φ of 55 is similar to the **inactive state,** and the θ is a bit smaller than both, with a value of 17°. Moreover, the hydrogen pattern of the dimeric interface is more similar to the **active state**, as Y119 is pointed to each other forming a hydrogen bond between their Oηs. Thus, BthTX-I/MMV resembles both BthTX-I **inactive** and **active form**.

Therefore, MMV has the capability to inhibit partially BthTX-I myotoxicity as observed in the functional experiments (CARVALHO, 2011) and, by our structural analysis, we conclude that these compounds form a dimeric complex in which one ligand molecule interacts to each monomer of the toxin. The MMV induces a significant and unique structural change on different loop regions, such as V31, L58 until D71 and MDiS residues, that also changes the general geometric orientation between the monomers. The MDiS residues have been shown to be important to the BthTX-I myotoxicity (CHIOATO et al., 2007; FERNANDES et al., 2013). Based on these observations, we suggest that the protein loses its capability to induce myotoxicity by the presence of MMV ligand in calcium binding loop and N-terminal region and by its effect in toxin's MDiS which is in **non-canonical** conformation.

Table 2.5 - Composition of complexes within crystal contacts of BthTX-I/MMV structure

Composition	Stable	Assembly				Interface of monomers				
		ΔG^{diss}	Surface area	Buried area	ΔG^{int}	Sym op	Interface surface	Δ^iG Energy	Δ^iG P-value	CSS
A MMV B POL	No	-1.9	13000	2540	-4.0	x,y,z	785.0	-6.9	0.443	0.136
B_2/POL_2	Yes	2.0	12930	1850	-1.5	-x,-x+y,-z+2/3	695.8	-12.6	0.091	1.000
A_2/MMV_2	Yes	2.8	13690	2620	-19.2	x-y,-y,-z+1/3	568.7	-13.9	0.069	1.000

The unit of ΔG and surface are in kcal/mol and $Å^2$, respectively. ΔG^{diss} is the energy necessary to dissociate the complex. Δ^iG and ΔG^{int} are the solvation free energy gained upon formation of interface and the complex, respectively.

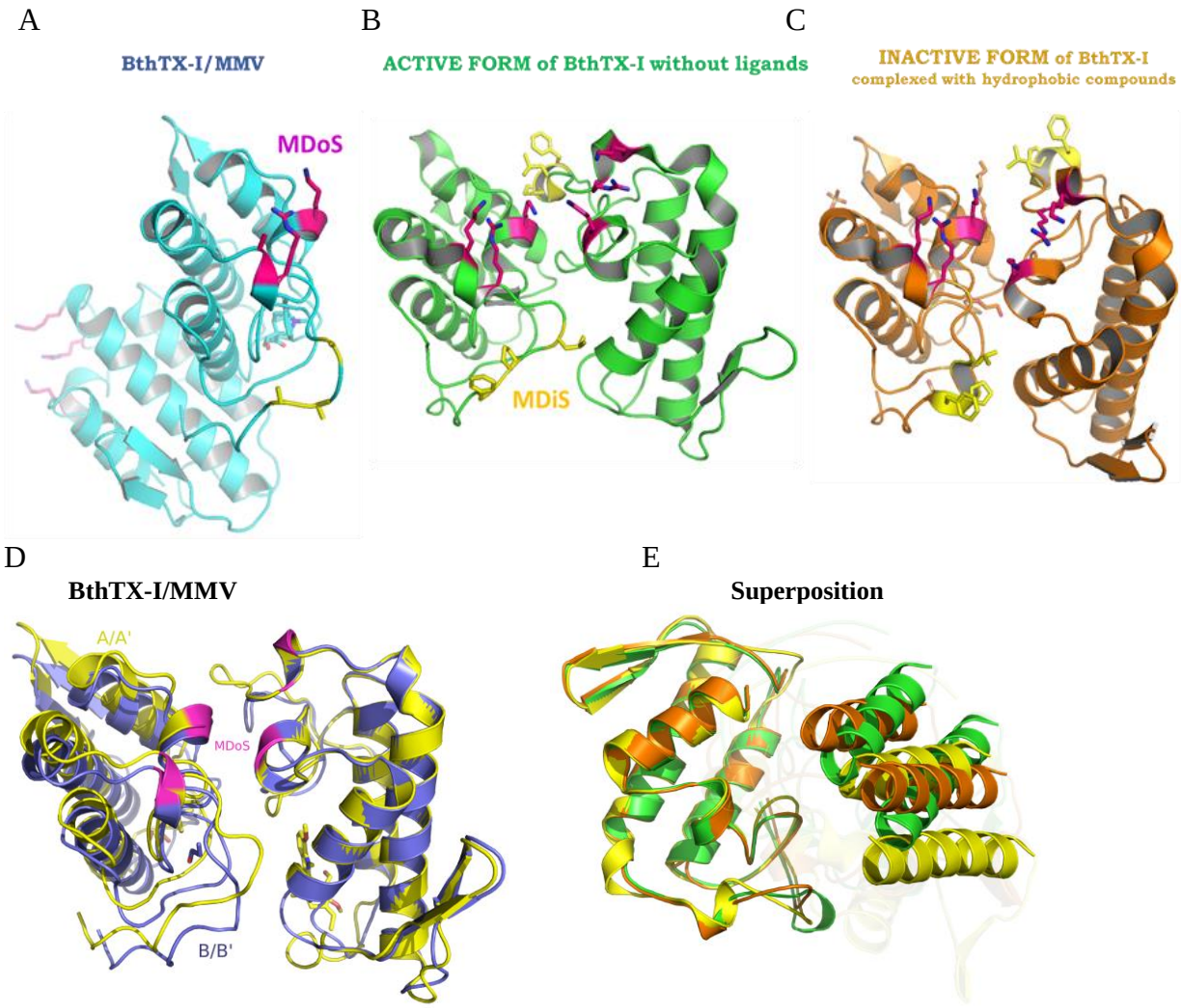


Figure 2.18 - BthTX-I different dimers in cartoon representation.
In **A**, the dimer of BthTX-I/MMV present in ASU colored in cyan; in **B**, the **active state** of BthTX-I colored in green (PDB id: 3IQ3); in **C**, the **inactive state** of BthTX-I (PDB id: 4WTB) colored in orange. In **A**, **B,** and **C**, sidechain of MDiS and MDoS are represented in sticks yellow and magenta, respectively. In **D**, two different compact dimers of BthTX-I/MMV present in the crystal mates, one formed by A and its symmetry (A/A') with MMV colored in yellow and the other B with its symmetry (B/B') colored in dark blue. In **A**, **B**, **C** and **D**, one of the monomers is in the same orientation. In **E**, comparison of the different orientation of the monomers of the **active**, **inactive state** of BthTX-I and BthTX-I/MMV (A/A'), where monomers in the left are superposed and the ones in the right are transparent with the exception of two long and parallel α-helices. In **E** and looking at the right monomers, it is possible to see that the **active state** of BthTX-I and BthTX-I/MMV are in same rotation, but in a different translation, while the opposite happens with **inactive** and **active state** BthTX-I.

MMV is a BthTX-I inhibitor that interacts with a unique toxin region, and it induces the highest structural deviation in available bothropic PLA_2-like crystallographic structures. Unfortunately, MMV partial protection against BthTX-I toxicity in mouse phrenic nerve-diaphragm preparation keep out its use to complement the antivenom serum. Although, it could be used to evaluate other candidates with similar molecular structure but inert to biological membranes.

2.8.10 Concluding remarks

With the results presented here in **Section 2.8**, we identified exciting new structural features related to the oligomeric changes between the **inactive** and **active states** of PLA_2-like proteins. From the detailed structural analysis of the MDiS region, we identify entirely different MDiS conformations in the **inactive** and **active states**, which may be measured by the distance between the L121 and F125 residues. Furthermore, from the comprehensive structural bioinformatics study of the protein tunnels, we could identify the relevant tunnels for the PLA_2-like proteins and their relationships to the transition between both states. The entrance of hydrophobic molecules in **Tunnel 1** of **canonical monomer** leads to a dimer reorientation and a structural change from asymmetrical to symmetrical monomers. The dimer reorientation may be measured by Bryan-Tait angles, and it is in agreement with NM analysis in both states of the toxin.

The symmetrical relationship in the **active state** is related to an ideal exposition of MDiS and MDoS residues in the PLA_2-like iFace. Comparing those **active** structures, the higher variance in the rotation of the φ angle may be related to the toxin interaction to and disruption of the membrane. This particular movement is described only by the lowest frequency NM of *Act* (7). We will submit this angle measurement methodology to the models generated by Molecular Dynamics simulations including Normal Mode Analysis done by Angelo Magro (FCA, UNESP, Botucatu) and David Perahia (ENS Cachan, France) to address this hypothesis as this is also their objective.

The local and global measurements of these toxins aided the comprehension of the inhibition mechanisms of zinc ions and MMV. Zinc interacts to Tunnel 1 residues in the only **canonical monomer** of BthTX-I (residue H48 and MDiS region), while MMV is bound in an unique region. With the highest structural deviation seen in all bothropic PLA_2-like proteins, MMV induces a different arrangement between monomers. We hypothesized that these observations are related to their inhibition mechanism. The same strategy with Normal Mode analysis used to correlate **active** and **inactive state** of BthTX-I may be applied to study the different available inhibitors. Our proposed methodology to evaluate local and global features of these proteins are easily extended to new structures are solved in our laboratory or the ones generated by Molecular Dynamics.

2.9 Structural studies of snake venom basic D49-PLA$_2$s independent of calcium

Some Asp49-PLA$_2$s have a high distribution of basic residues, including within the C-terminal region, similar to other PLA$_2$-like proteins. For BthTX-II and PrTX-III, whose crystallographic structure has been elucidated, two dimers coexist in their unit cell, the large that has been described by Rigden *et al* (2003) and the compact (DOS SANTOS et al., 2011b). In the latter, the calcium binding loop distorted participates in the dimeric interface, this observation together with phylogetic analysis has correlated these two toxins to PLA$_2$-like proteins (FERNANDES et al., 2014).

2.9.1 Small-angle X-ray scattering studies of BthTX-II

Essential to BthTX-II toxic mechanism description is the knowledge of its quaternary structure. SAXS is a low-resolution technique that helps quaternary structure assignment by the generation of the envelope of a monodisperse sample by its X-ray scattering in solution. A better resolution is obtained by generating the theoretical scattering of a crystallography models and fitting them at the experimental scattering. As different dimers were seen, SAXS experiments may answer this uncertainty.

We measure SAXS experiments with BthTX-II in 1 and 5 mg/mL concentration in 50 mM buffer of HEPES pH 7.4 and in 50 mM sodium citrate pH 5.6. In HEPES buffer, we combined the scattering of both concentrations, as they had same shape and the high concentration curve had a small effect of concentration. The curves were scaled before merging and we used initial of the low concentration curve, the average of intermediate region and the high-resolution curve for the values close to q 3.0, as lower dispersion is seen (**Figure 2.19 A**). In sodium citrate, the shape of the curves is different, and, thus, previous merging procedure is not advisable as samples are in different structural states (**Figure 2.20 A** and **2.21 A**).

We estimated the R_g with the Guinier approximation with the software *PRIMUS* (KONAREV et al., 2003) and we used it to define if initial scattering points should be discarded (Small and top right-hand side graph in **Figure 2.19 A**, **2.20 A** and **2.21 A**). Thus, we excluded the first 3, 9 and 9 points of the scattering curves of toxin in two concentrations in sodium citrate (1 and 5 mg/mL) and of the toxin in HEPES (merged), respectively, and we obtained R_g 18.4, 21.2 and 16.1 nm, respectively (Experimental scattering in **Table 2.6**). The different R_gs in the two different toxin concentration in sodium citrate confirms the

impossibility to merge curves. We generated a graph of distance distribution (ρ(r)) with the maximum diameter (D_{max}) of 48, 98 e 63 Å for the scattering curves of the toxin in 1 and 5 mg/mL of sodium citrate and of the toxin in HEPES (**Figure 2.19 B**, **2.20 B** and **2.21 B**), respectively, using the *GNOM* (SVERGUN, 1992). The shape of the distance distribution is typical of globular proteins (FEIGIN; SVERGUN, 2013), whether toxin in citrate condition is less globular.

Using *CRYSOL* software (SVERGUN; BARBERATO; KOCH, 1995), we generated the theoretical scattering curves of the monomeric and possible dimeric crystallographic models using PDB models elucidated in the next section (**Figure 2.22**) to measure their discrepancy to experimental measurement. In HEPES, the protein is in a monomeric state as shown by small discrepancy (χ^2) of 1.35, whether for the dimers χ^2 are greater than 5. In sodium citrate, none of the calculated discrepancies were satisfactory. Thus, we compared the experimental R_g to those generated using the theoretical scattering of BthTX-II monomer and dimers with *CRYSOL* (Theoretical scattering in **Table 2.6**). Sodium citrate condition probably induces a mixture of oligomers.

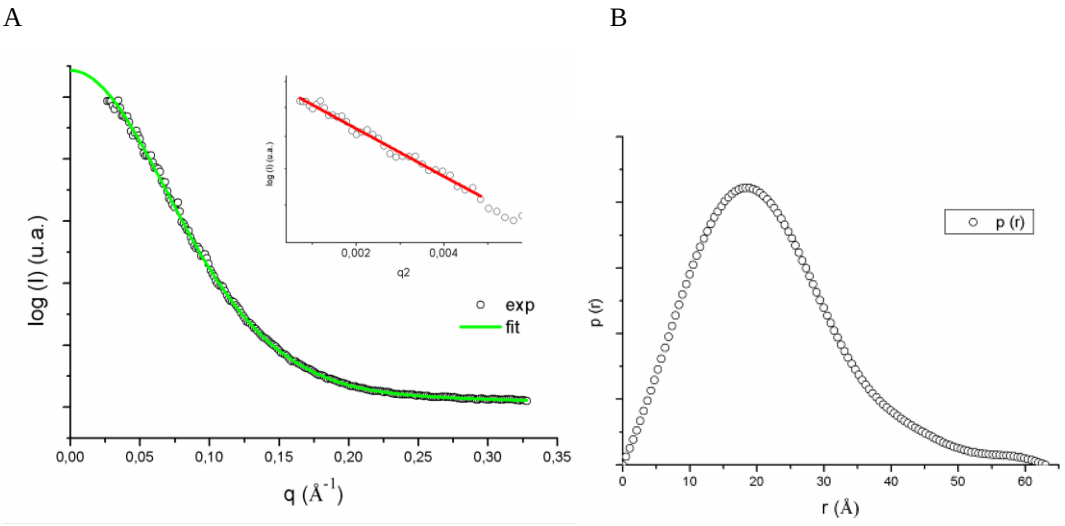


Figure 2.19 - Experimental SAXS curves of BthTX-II in 50 mM HEPES pH 7.4.
In **A**, the open circles are the fusion of the scattering curves of BthTX-II in low and high concentration (1 and 5 mg/mL) and the green line is the adjustment done to generate the distance distribution (p(r)) in **B**. The insert in A is Guinier analysis.

To confirm it, we evaluate the hydrodynamic radius (R_h) of BthTX-II in the same buffer conditions by DLS measurements. We evaluate them only in the concentration of 1.5 and 1.8 mg/mL, in the temperatures 4, 10, 18, 25 e 37 ºC, with 100 acquisitions (**Table 2.7**).

A monomeric state is confirmed for the toxin in HEPES buffer, but in citrate the higher R_h, M_w, and polydispersity are compatible with a mixture of oligomers (**Table 2.6**), which is increased by temperature.

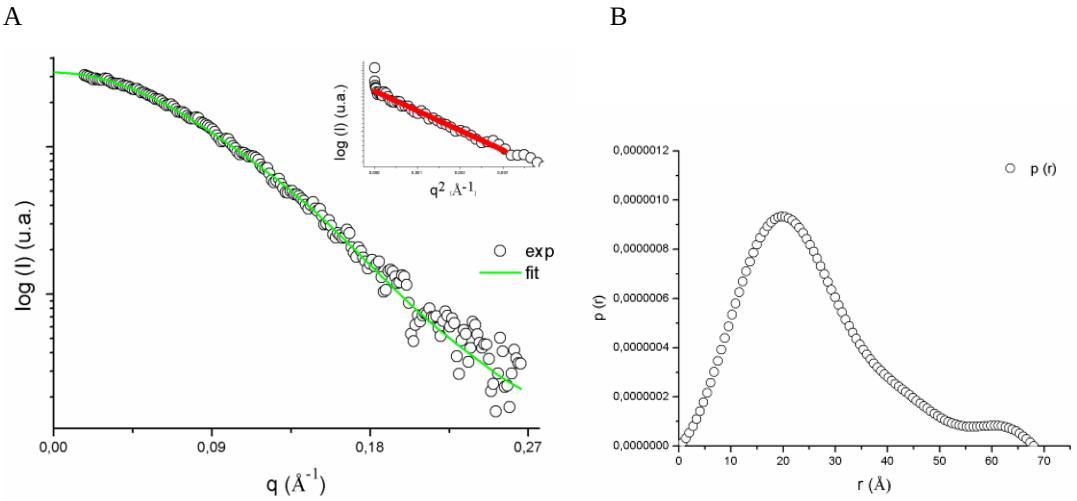


Figure 2.20 - Experimental SAXS curves of BthTX-II in low concentration (1 mg/mL) and in 50 mM sodium citrate pH 5.2.
In **A**, the open circles are experimental scattering and the green line is the adjustment to generate the distance distribution (p(r)) in **B**. The insert in A is Guinier analysis.

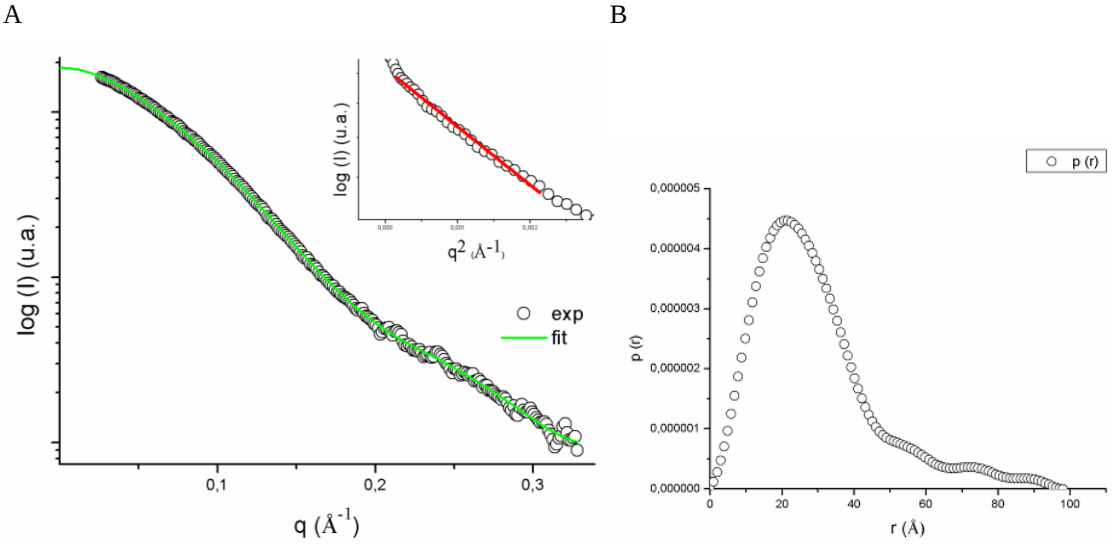


Figure 2.21 - Experimental SAXS curves of BthTX-II in high concentration (5 mg/mL) and in 50 mM sodium citrate pH 5.2.
In **A**, the open circles are the experimental scattering and the green line is the adjustment to generate the distance distribution (p(r)) in **B**. The insert in A is Guinier analysis.

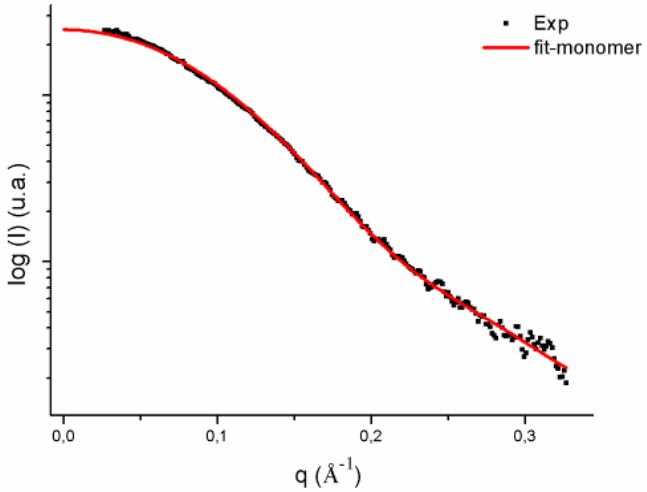


Figure 2.22 - Experimental SAXS curves of BthTX-II in 50 mM HEPES pH 7.4 and theoretical curve of monomeric BthTX-II structure.
The red line is the theoretical BthTX-II monomeric curve fitted into the experimental curve represented as black dots. The discrepancy of the curves is low (χ^2 of 1,35).

Table 2.6 - SAXS measurement of BthTX-II in citrate and HEPES buffer and theoretical scattering of different BthTX-II quaternary assemblies.

Protein	C (mg/mL)	R_g (nm)
Theoretical scattering		
Monomer		14.4
Large Dimer		20.7
Compact dimer		18.3
Experimental scattering		
BthTX-II Citrate	1	18.4
BthTX-II Citrate	5	21.2
BthTX-II HEPES	Merged	16.1
BthTX-II HEPES	1	16.0
BthTX-II HEPES	5	17.1

Rg stands for radius of gyration. Theoretical scattering of crystallographic model obtained in next section generated with *CRYSOL*.

Table 2.7 - Experimental data of dynamic light scattering (DLS) of BthTX-II in buffer of sodium citrate and HEPES.

Buffer/Temperature	R_h (nm)	%PD	Mw (kDa)	% Mass
HEPES pH 7.4				
4	17	7.6	12	100
10	18	30.8	14	99.9
18	17	15.5	12	99.4
25	18	22.8	14	98.8
Sodium citrate pH 5.2				
4	25	22.8	29	99.7
10	28	38.4	37	99.7
18	27	41.6	35	99.7
25	29	45.9	41	99.3

The abbreviations stand for hydrodynamic radius (R_h), percentage of polydispersity (%PD) and Molecular weight (Mw).

Thus, we conclude that the BthTX-II in 50 mM HEPES pH 7.4 is in a monomeric state based on SAXS and DLS measurement (**Table 2.6** and **2.7**). In 50 mM sodium citrate pH 5.6, the toxin is in a mixture state, as seen by high polydispersity in DLS experiments. In general, the toxin seems to be in a monomeric state in neutral pH, whereas in citrate and acid environment, it exists in a mixture of monomers and dimers. The increase in protein concentration in citrate buffer increases R_g, therefore, the presence of higher oligomeric assemblies is concentration dependent. In a polydisperse sample, to infer its quaternary structure is not precise, therefore we did not obtain a definitive answer on which is the dimer present in solution. More experiments are needed to definitively define the oligomeric state of BthTX-II or to better characterize the dynamic between monomer and dimer equilibrium.

2.9.2 Crystallographic studies of basic PLA_2s

2.9.2.1 Monomeric BthTX-II complexed to fatty acids

We obtained successfully crystals of BthTX-II complexed with stearic acids using the matrix sparse method and Microbatch technique. The cocrystallization drop was composed of 1.0 µL of protein (7 mg/mL), 1.0 µL precipitant and 0.2 µL stearic acid (4 mg/mL diluted in octane) (**Figure 2.23 A**). The precipitant solution was composed of 0.1 M HEPES pH 7.5, 22% polyacrylic acid 5100, and 20 mM $MgCl_2$ (**Table 2.8**). The best dataset was collected in X25 beamline in NSLSI and processed using XDS. Crystals diffracted to 2.1 Å, in $R3_2$ space group with a monomer inside its ASU (**Table 2.9**). We obtained its phases using the same toxin structure available in (PDB id: 2OQD).

Table 2.8 - Crystallization setup of BthTX-II and PrTX-III crystals

Protein/complex	Crystallization drop	Precipitant solution	Crystallization Technique
BthTX-II/STE	1.0 µL prot (7 mg/mL) + 1.0 µL prec + 0.2 µL STE (4 mg/mL in octane)	0.1 M HEPES pH 7.5 22% polyacrylic acid 5100 20 mM $MgCl_2$	Microbatch 70% paraffin and 30% silicon oil
Apo BthTX-II	1.0 µL prot (7 mg/mL) + 1.0 µL prec	0.1 M sodium citrate pH 5.2 25% PEG4000 0.21 M ammonium sulfate	Vapor diffusion – Hanging drop
Apo PrTX-III	0.5 µL prot (12 mg/mL) + 0.5 µL prec	0.3 M tri-potassium citrate 20% PEG3350	Vapor diffusion – Hanging drop

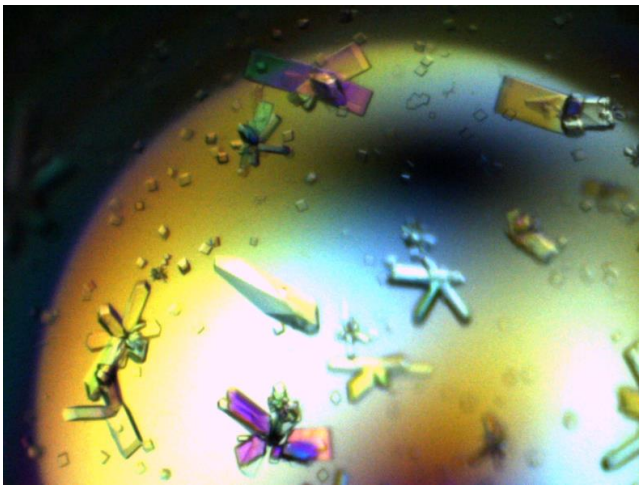
When solvent used in dilution was not described, water was used. Abbreviations stand for stearic acid (STE), protein (prot) and precipitant solution (prep)

Table 2.9 - X-ray data collection and refinement of BthTX-II/STE and *apo* BthTX-II and PrTX-III

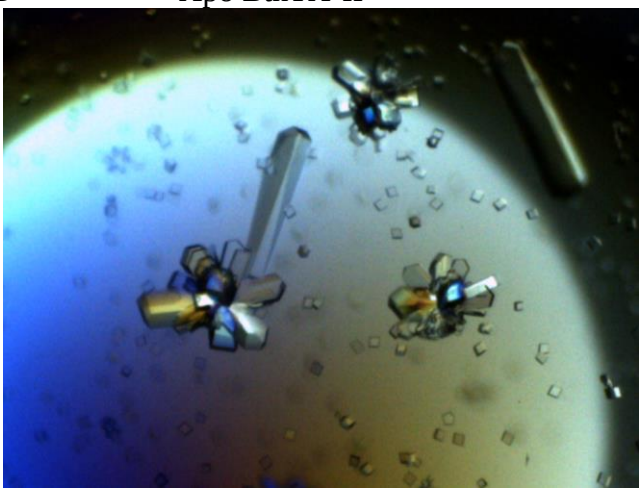
Datasets	BthTX-II/STE	*Apo* BthTX-II	*Apo* PrTX-III
Number of monomer per ASU	1	2	2
Beamlines	NSLSI-X25	NSLSI-X6A	LNLS-MX1
Data processing software	XDS	HKL2000	XDS
Space group	$R3_2$	$P2_12_12$	$P22_12_1$
Unit cell abc (Å) αβγ (º)	a=b=82.9 c=120.6 α=β=90 γ=120	a=60.3 b=107.3 c=44.3 α=β=γ=90	a=45.4 b=59.9 c=97.1 α=β=γ=90
Resolution (Å)	40-2.10 (2.22-2.10) [a]	26.2-1.79 (1.85-1.79) [a]	40.0-2.26 (2.39-2.26) [a]
Total reflections	86378	196445	106369
Unique reflections	9325 (1296) [a]	27379 (2499) [a]	22439 (2632) [a]
Average multiplicity	9.3	7.2	4.7
R_{meas} [b] (%)	11.4 (137.4) [a]	9.2 (89.5) [a]	10.7 (161.6)
CC (1/2)	99.6 (33.4)	99 (69.4) [a]	99.7 (45.9)
Completeness (%)	97.2 (83.8)	98.1 (91.8) [a]	93.4 (67.9)
Average I/σ (I)	16.1 (1.0)	19.9 (2.3) [a]	9.5 (0.9)
R [c] (%) R_{free} [c] (%)	20.4 23.7	17.7 21.5	21.1 25.4
RMS deviations from ideal values Bond lengths (Å) Bond angles (º)	 0.010 1.04	 0.010 0.95	 0.009 0.97
Ramachandran plot (%) Residues in most favorable regions Residues in accepted regions	 96.7 3.3	 95.8 4.2	 96.2 3.8
Molprobity score	1.66	0.79	1.11
Number of molecules/atoms, their averaged B factor ($Å^2$) Protein Ligand H_2O Na^+	 1 (39.9) 1 STE (39.8) 96 (47.5) 1 (34.9)	 2 (27.3) 2 BEZ (34.1) 337 (40.9)	 2 (63.2) 2 ACY (60.5) 88 (57.4)

[a] Number in parenthesis are for highest resolution shell. [b] Number in parenthesis and in brackets are B factor and occupation, respectively. $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$, $\hat{I}_h = \sum_h \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$ redundancy independent R-factor. Calculated for I > -3σ (I). [c] $R_{cryst} = R_{hkl} \frac{||Fobs_{hkl}| - |Fcalc_{hkl}||}{|Fobs_{hkl}|}$. [d] R_{free} is equivalent to R_{crist}, but calculated with reflections (5%) omitted from the refinement process. RMS *root-mean-square*. Anomalous correction: percentage of correlation between random half-sets of anomalous intensity differences. Anomalous signal: mean anomalous difference in units of its estimated standard deviation $R_{hkl} \frac{|F(+) - F(-)R_{cryst}|}{\sigma}$. F(+), F(-) are structure factor estimates obtained from the merged intensity observations in each parity class. The abbreviation of SHV, BEZ and ACY are related to heptanoic acid, benzoic acid, and ethanoic acid, respectively.

A BthTX-II/STE



B Apo BthTX-II



C Apo PrTX-III

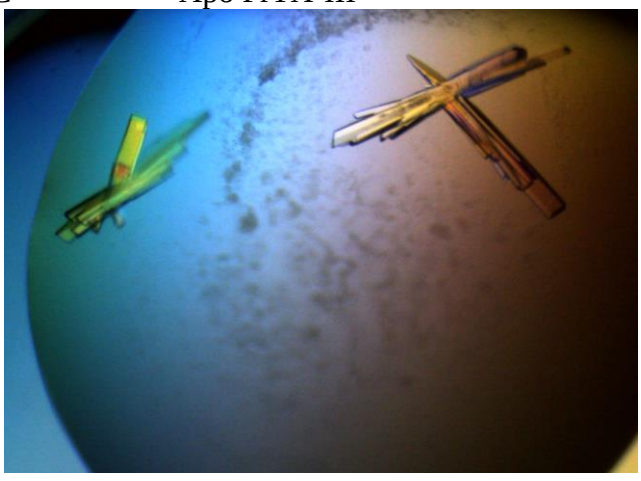


Figure 2.23 - Crystal of basic and bothropic PLA_2s.
In **A**, BthTX-II cocrystallized with stearic acid, in **B**, *apo* BthTX-II crystals, and in **C**, *apo* PrTX-III crystals.

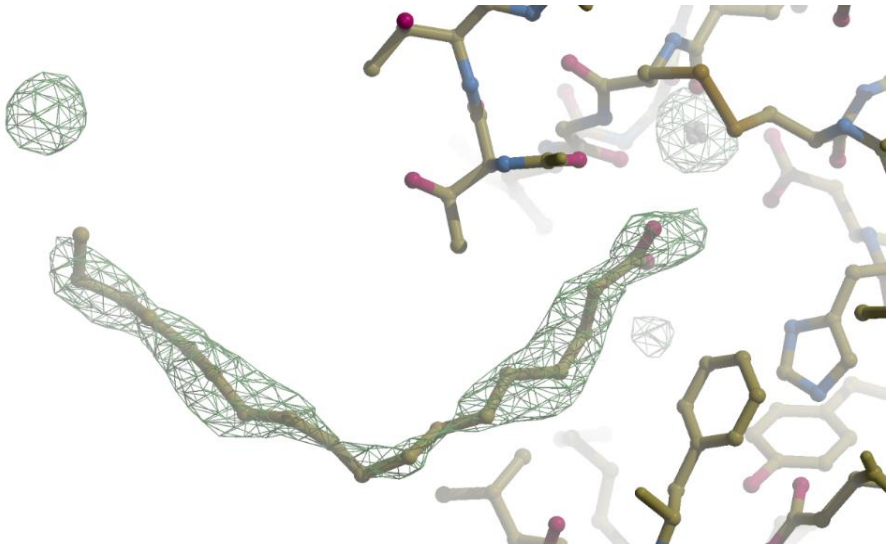
We found large electron densities in Fo-Fc map in regions where no atoms were included. The electron density of BthTX-II cocrystallized with stearic acid (BthTX-II/STE) matched perfectly this ligand. Its overall B factor is in agreement with protein monomer A and B factor (**Table 2.9**). As the center of the stearic acid is exactly in one of the symmetry axis, its occupancy was reduced to 0.5, thus one stearic acid interacts with two molecules of BthTX-II (**Figure 2.24 A**). Moreover, we have found a sodium ion being coordinated by toxin calcium binding loop (**Figure 2.24 B**). Due to poor electron density, we removed side chain atoms of 35Q, 56N, 105K, 106R, 116K, and 121K. We refined the model with *BUSTER* until R/R_{free} reached 20.4/23.7% and molprobity score of 0.5, which is in the 100th best percentile of all structures elucidated from datasets of similar resolution (2.1 ± 0.25 Å). In the Ramachandran plot, 98.3% and 2.7% are in the favored and allowed regions, respectively.

2.9.2.1 *Apo* BthTX-II

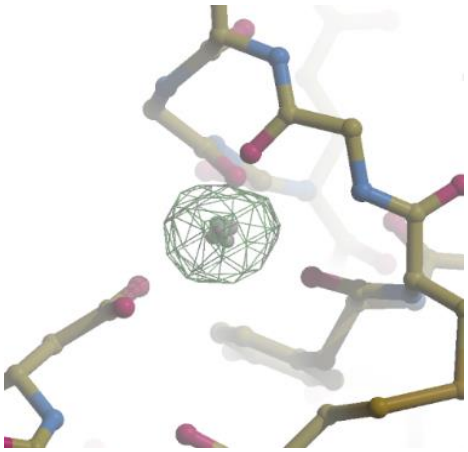
We obtained successfully crystals of *apo* BthTX-II in native state using the matrix sparse method and vapor diffusion technique. The crystallization drop was composed of 1.0 μL of protein (7 mg/mL) and 1.0 μL of precipitant solution (**Figure 2.23 B**). The latter was composed by 0.1 M sodium citrate pH 5.2, 25% PEG4000, 0.21 M ammonium sulfate (**Table 2.8**). The best dataset was collected in X6A beamline in NSLSI and processed using HKL2000. Crystals diffracted to 1.79 Å, in $P2_12_12$ with a dimer inside its ASU (**Table 2.9**). We obtain its phases using the same toxin structure available in (PDB id: 2OQD).

We found large electron densities in Fo-Fc map in regions where no atoms were included. It did not perfectly match any of the compounds available in the crystallization condition, thus we chose to model a benzoic acid that matched perfectly electron density shape (**Figure 2.24 C** and **D**). Its overall B factor is in agreement with protein monomer A and B factor (**Table 2.9**). Due to poor electron density, we removed side chain atoms of 56N/A&B, 105K/B, 116K/B, and 119E. We refined the model with *BUSTER* until it reached R/R_{free} of 17.7/21.5% and molprobity score of 0.79, which is in the 100th best percentile of all structures elucidated from datasets of similar resolution (1.79 ± 0.25 Å). In the Ramachandran plot, 97.1% and 1.7% are in the favored and allowed regions, respectively.

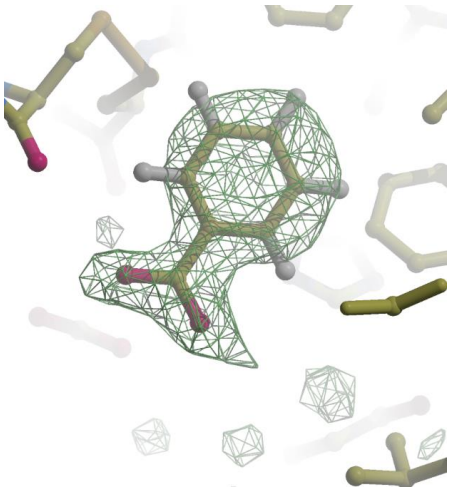
A



B



C



D

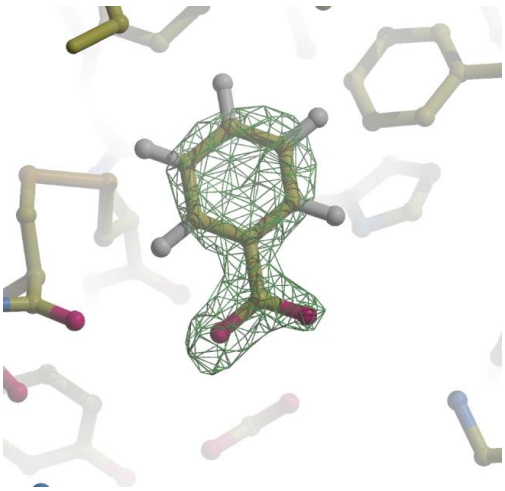


Figure 2.24 - Ligands found in the BthTX-II crystallographic structures.
In **A**, BthTX-II/STE is represented in sticks with stearic acid and sodium ion modeled in its omit map. One of the symmetry axes is in the center of fatty acid, thus its occupancy was reduced to 0.5. In **B**, only the sodium is shown being coordinated by oxygen atoms of the calcium binding loop of BthTX-II/STE. In **C** and **D**, apo BthTX-II is represented in sticks with benzoic acids modeled in omit map. The map contour used is 4 σ.

2.9.2.3 *Apo* PrTX-III

We obtained successfully crystal of native PrTX-III using the matrix sparse method and vapor diffusion technique. The crystallization drop was composed of 0.5 μL of protein (12 mg/mL) and 0.5 μL of precipitant solution (**Figure 2.23 C**). The latter was composed by 0.3 M tri-potassium citrate, and 20% PEG3350 (**Table 2.8**). We collected the best dataset in MX1 of LNLS and processed it using XDS. Crystals diffracted to 2.26 Å, in $P22_12_1$ space group with a dimer inside its ASU (**Table 2.9**). We obtained its phases using the same toxin structure available in (PDB id: 1GMZ).

Due to poor electron density, we removed side chain atoms of 11K/A&B, 30W/A, 33Q/A, 37K/A, 53K/A, 55T/A&B, 56S/A&B, 68K/A, 69D/A&B, 71T/A, 73V/A, 76E/A, 77E/B, 82K/A, 89K/A, 101D/B, 103N/A, 104K/A, 105K/A&B, 108S/A, 109S/B, 111K/A, 112S/A&B, 113L/A, 116K/A&B, 117A/A&B, 118P/A&B, 119D/B, 120D/A&B. We refined the model with *BUSTER* until R/R_{free} reached 21.0/25.4% and molprobity score of 1.11, which is in the 100th best percentile of all structures elucidated from datasets of similar resolution (2.1 ± 0.25 Å). In the Ramachandran plot, 94.9% and 4.7% are in the favored and allowed regions, respectively.

2.9.3 Discussion

The BthTX-II and PrTX-III are two basic PLA_2s from *Bothrops jararacussu* and *Bothrops pirajai* venom, respectively. BthTX-II quaternary structure has been shown to be dimeric in non-reducing SDS-Page, DLS (DOS SANTOS et al., 2011b) and SAXS (MURAKAMI et al., 2008). For PrTX-III, a dimer has been seen in non-reducing SDS-Page and analytical size-exclusion chromatography (XIMENES et al., 2012). Two dimers coexist in the unit cell, one large and one compact, the bioinformatics PISA tool and the SAXS *ab initio* reconstruction suggested the latter (DOS SANTOS et al., 2011b; MURAKAMI et al., 2008).

In our experiment in neutral pH, we observe a different result. The scattering curve (SAXS) of BthTX-II in 50 mM HEPES pH 7.4 in 1 and 5 mg/mL has a R_g of 16.0 Å and the shape identical to the monomeric structure of BthTX-II. The theoretical scattering of both compact and large dimers of BthTX-II do not fit in this curve. Contrarily, Murakami *et al.* (2008) concluded that BthTX-II has a dimeric assembly in a similar buffer condition, 20 mM

TRIS HCl pH 7.0 with the protein in 4-10 mg/mL. The apparent conclusion of a dimer to the *ab initio* reconstruction of SAXS curve whose Rg is 15.5 Å, a value correspondent of monomer, could be due to their use of a homology model, instead of the real BthTX-II crystallographic structure that was not available in 2008. The crystallographic model of BthTX-II/STE from crystals grown in HEPES pH 7.4 buffer, also supports this monomeric assembly. Therefore, our SAXS experimental analysis with the most accurate model, the DLS and crystallography results suggest a monomer in neutral pH.

We obtain a different result in the acid pH of sodium citrate buffer. The protein has an increased polydispersity with a probable mixture of monomers and dimers, being the quaternary structure increased with the increase of protein concentration or the temperature. The *apo* BthTX-II and PrTX-III structures elucidated from crystals grown in citrate buffer support the dimer, as this quaternary structure is found in the ASU and it is suggested by PISA analysis (DOS SANTOS et al., 2011b). Therefore, dimers may be present in different conditions, such as citrate and the one present in non-reducing SDS-PAGE gel.

Our elucidated *apo* structures by crystallography are dimers in their ASU with both of their monomers similar (RMSDs<0.6 Å accounting all Cα in **Table 2.10**) (**Figure 2.25 A**). These monomers are different than BthTX-II/STE (RMSDs>1.3 Å accounting all Cα in **Table 2.10**). Their main difference is the calcium binding loop regions (**Figure 2.25 B**), while BthTX-II/STE has a sodium coordinating this region (**Figure 2.24 B**), *apo* monomers have this loop in a distorted conformation that unable ion coordination. *Apo* monomers are also similar to deposited structures (PDB ids 2OQD and 1GMZ), as their RMSD are between 0.7 and 1.0 Å with no significant difference from our analysis. Rigden *et al.* hypothesized the crystallized state of *apo* PrTX-III is an **inactive state** that co-exists with an **active state**, which would have a canonical calcium binding loop. BthTX-II/STE model is a representative of the **active state** of catalysis, which was not seen previously. Curiously, BthTX-II and PrTX-III have been showing to display lower catalytic activity compared to other snake venom PLA_2 (ANDRIÃO-ESCARSO et al., 2000; SOARES et al., 2001; TOYAMA et al., 1999).

The classification of *apo* dimeric BthTX-II and PrTX-III in **inactive state** is only related to its catalytic activity, as we observed the capability of BthTX-II to exert myotoxicity independent of calcium (DOS SANTOS et al., 2011b). This could also be the case for other basic PLA_2s toxins, like the PrTX-III and Mt-III that, together with BthTX-II, have a closer

similarity to PLA_2-like protein than to other catalytic and monomeric snake venom PLA_2s (DOS SANTOS et al., 2011b). As the most likely dimer to occur in solution is the compact dimer as calculated by PISA tool (DOS SANTOS et al., 2011b), the distortion of the calcium binding loop may be related to this dimeric assembly, as W31 from both monomers are interacting (DOS SANTOS et al., 2011b). Moreover, C-termini of the dimer of *apo* BthTX-II and PrTX-III are conveniently side-by-side in the same plane in protein surface and present charged (MDoS) and hydrophobic residues (MDiS) (FERNANDES et al., 2014; SALVADOR et al., 2017). Therefore, BthTX-II and PrTX-III in dimeric state may have a similar iFace and myotoxic mechanism in comparison to PLA_2-like proteins (FERNANDES et al., 2014; SALVADOR et al., 2017).

Table 2.10 - BthTX-II/STE and apo BthTX-II and PrTX-III comparison in RMSD (Cα).

Structure		BthTX-II/STE	apo BthTX-II		apo PrTX-III	
	monomer	A	A	B	A	B
BthTX-II/STE	A	-	1.4	1.3	1.4	1.3
apo BthTX-II	A	1.4	-	0.5	0.2	0.6
	B	1.3	0.5	-	0.6	0.2
apo PrTX-III	A	1.4	0.2	0.6	-	0.6
	B	1.3	0.6	0.2	0.6	-
2OQD	A	1.4	0.7	1.0	0.7	1.0
	B	1.3	0.8	1.1	0.7	1.0
1GMZ	A	1.3	0.7	0.9	0.7	0.9
	B	1.3	0.7	1.0	0.7	1.0

High, medium and lower RMSD are colored in orange, yellow and green, respectively.

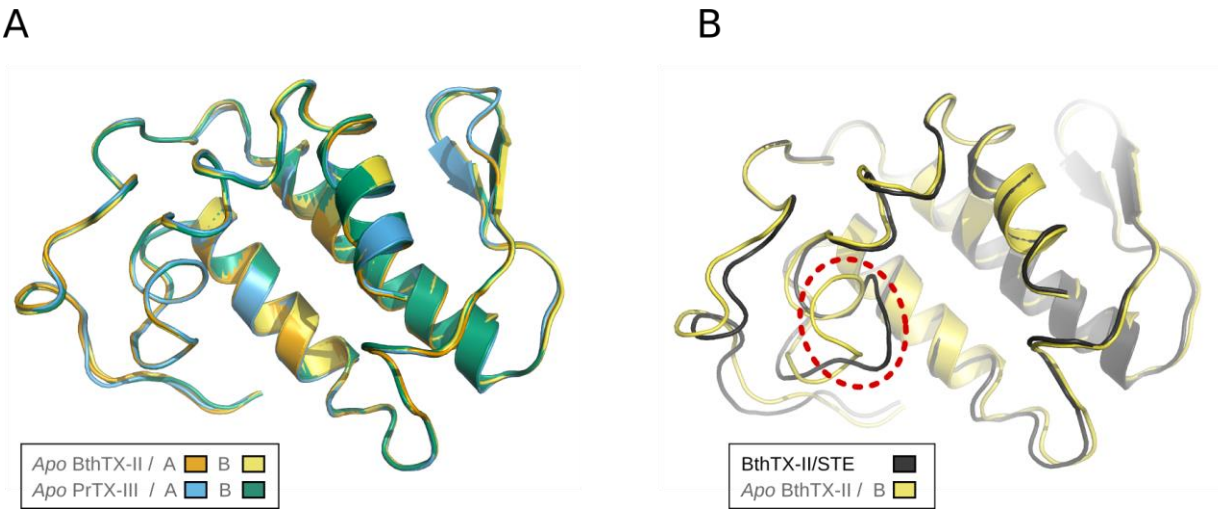


Figure 2.25 - Structural comparison of cartoon representation of monomers of BthTX-II/STE with canonical calcium binding loop and *apo* BthTX-II and PrTX-III with distorted calcium binding loop.

Despite all these observations, Mora-Obando *et al.* raised the hypothesis that the responsible for the calcium-independent myotoxic activity seen in the toxins BthTX-II, PrTX-III and Mt-III would be a contamination of PLA_2-like proteins in these samples (MORA-OBANDO et al., 2014). To purify single isoform from the complex snake venom that includes different isoforms of PLA_2s and PLA_2-like proteins with similar biophysical properties is a challenge. These authors have proven this hypothesis to be true for the study of Mt-III of *Bothrops asper* (MORA-OBANDO et al., 2014). These authors showed purity improvement by adding a reverse-phase chromatography step into the previous purification protocol, single ion exchange, which separated PLA_2s isoforms from PLA_2-like proteins. Despite this new PLA_2 peak was not described as one single toxin 100% pure, they showed that this is not capable of inducing myotoxicity independently of calcium through CK experiments after toxin is chemically modified by BPB. This observation for Mt-III could be similar to BthTX-II, although it would contradict our experimental results published in 2011 (DOS SANTOS et al., 2011b). As our new purification protocol includes a reverse-phase chromatography that ensures BthTX-II higher purity, we will address such hypothesis in the future.

A good experimental model for the study of the monomeric and dimeric BthTX-II is the liposome preparations, as they are stable in different buffers and pH. It has been shown that in liposome preparation in which phospholipid composition are easily changed, basic PLA_2s and PLA_2-like proteins increase its capability to disrupt this artificial membrane when they are enriched with anionic phospholipid (BULTRÓN; GUTIÉRREZ; THELESTAM, 1993; DÍAZ et al., 1991; PEDERSEN et al., 1995). In these last published articles, the site of membrane damage has been separated from its catalytic site. Similar observation for these toxins was seen evaluating human erythrocytes toxicity when changing their phospholipid composition (DÍAZ et al., 2001).

The snake venom contain a high concentration of metal-ion chelator, ~120 mM of citrate, which is enough to chelate the ~3 mM endogeonous calcium ions (FRANCIS; SEEBART; KAISER, 1992). Citrate, by complexing to divalent ions, inhibit enzymatic activity of arthropod venom PLA_2s and may represent a mechanism of snakes to protect themselves from their own toxins (FENTON et al., 1995; FRANCIS; SEEBART; KAISER, 1992). The study of PLA_2s in citrate buffer may aid understanding of how these toxins may

be inhibited. In contrast, HEPES is one of the buffer used to study the human physiological condition, which in the case of those toxins may be used to study envenonamation.

Our structural results suggests that the dimer structure, whose contact is a distorted calcium binding loop, is likely to occur in high concentration of citrate and low pH. As the catalytic activity has been dissociated from liposome disruption, citrate environment may represent a environment to study other basic PLA_2 damaging sites and other toxic activities independent of catalysis. The chemical modification of H48 with BPB is a already known method to prevent catalysis and may used as a control experiment (ANDRIÃO-ESCARSO et al., 2000; SOARES; GIGLIO, 2003).

Similarly to PLA_2-like proteins, in the compact dimer, it can be observed that C-termini of the toxins conveniently side-by-side in the same plane in protein surface with its positive (MDoS) and hydrophobic residues (MDiS) exposed (FERNANDES et al., 2014; SALVADOR et al., 2017). Therefore, dimeric BthTX-II and PrTX-III may have an independent catalysis toxic mechanism similarly to PLA_2-like proteins. Whether, in neutral pH, these proteins may adopt a monomeric assembly with calcium binding loop ordered and with catalytic activity.

2.9.3 Conclusion

The monomeric assembly of BthTX-II complexed to stearic acid and sodium has not been described in the literature or in the PDB, while two similar dimeric BthTX-II has been elucidated (CORRÊA et al., 2008; DOS SANTOS et al., 2011b). These two quaternary structures correlate to the DLS and SAXS measurement that we described in the previous subsection, as a dimeric assembly was seen in citrate low pH buffer and a monomer was seen in HEPES sodium neutral pH. We did not obtain clear experimental result on the most likely BthTX-II dimer that occur in solution by SAXS, we plan to study other conditions that the protein is monodisperse and dimeric, such as changing buffer and varying pH.

By the ours findings, we raise the hypothesis that the myotoxic mechanism of these proteins (BthTX-II and PrTX-III) when in dimeric state would be analogous to PLA_2-like proteins, which implies that perturbation membrane would be related to C-terminal region and to oligomeric conformation changes in monomer-monomer geometry (BORGES et al., 2017; DOS SANTOS et al., 2011b). In monomeric state, its structure is similar to acid PLA_2s and catalytic activity may play important role cleaveging membranes phospholipids and exerting

its toxicity. These hypotheses will be addressed in the future with the toxin in these two buffer conditions and chemically modification by BPB in a catalytic essay *in vitro*. Moreover, we will extend the methodology used to characterize PLA_2-like protein to obtain new insight into a similar toxic mechanism.

3 SEQUENCE SLIDER

3.1 Crystallography

Crystallography plays an important role in the comprehension of physiological mechanisms and drug design as it is the current technique that provides conclusive structural information down to the atomic resolution. The observation of different states of proteins, complexes, and chemical reactions has provided the base for major breakthroughs in the last six decades.

X-ray crystallography was born with the discovery of X-rays in 1895 by Wilhelm Conrad Röntgen and X-ray diffraction by Max von Laue 17 years later. The interpretation of the X-ray diffraction came only with William Lawrence Bragg, who proposed the still valid Bragg's Law, constructed the first X-ray spectrometer and solved the crystal structure of diamond. For long years, only structures of crystals of small compounds were solved, even though proteins crystals were already available. Earthworm hemoglobin crystals, per example, were described as early as 1840 (HÜNEFELD, 1840 *apud* JASKOLSKI; DAUTER; WLODAWER, 2014). Larger molecules started being solved in the middle of 20th century, as Dorothy Crowfoot Hodgkin solved the structure of penicillin in 1949 (CROWFOOT et al., 1949) and of vitamin B-12 in 1957 (HODGKIN et al., 1957). On 28 February 1953, Francis Crick announced in the Eagle, a small pub close to the old Cavendish Laboratory in Cambridge where scientist usually had lunch, that he and James Watson had finally "discovered the secret of life". This anecdote is told in Watson's book "The Double Helix" published in 1962 and they could not have discovered the structure of DNA (WATSON; CRICK, 1953) if it was not with the work of Dorothy Franklin with diffraction of DNA fibers.

The first protein structure came in 1958, as the structure of myoglobin was solved by Kendrew *et al* (KENDREW et al., 1958). In 1962, Max Perutz together with John Kendrew subsequently solved the structure of hemoglobin (PERUTZ et al., 1960). A protein databank (PDB) was created to gather the elucidated structures in 1971, although it started with only 7 structures, most of them enzymes extracted and purified directly from animals or plants (BERNSTEIN et al., 1977). Until the end of 80's, a crystallographic determinations could take years and there were few groups of crystallography and thus a low number of proteins were elucidated, 365 crystallographic models to be more precise (RCSB PDB, 2016). Available samples was one of the issues.

The first proteins studied and whose structure solved by crystallography were those easily available for the researchers, as the idea for recombinant protein production only came

in the 70's (RUSSO, 2003). As venom, at least for large animals such as snakes, is a natural and abundant source of active protein and peptides, enzymes from venom played an important role in the elucidation of physiological mechanisms. This was particularly the case for the superfamily of PLA_2s, as its catalytic mechanism of action was first described in 1990 based on three PLA_2 crystallographic models of venom proteins, two from a snake and one from a bee (SCOTT et al., 1990a).

Nowadays, more sensitive, lower noise detectors, the use of synchrotron radiation, cryocrystallography, anomalous signal, better algorithms, and faster computers have elevated the number of PDB entries to more than 120 thousand macromolecular structures, in which almost 90% were solved by X-ray crystallography (JASKOLSKI; DAUTER; WLODAWER, 2014). Thus, X-ray crystallography importance is unquestionable and no less than 42 scientist won the Nobel Prizes for their contribution to this technique or to structure description (JASKOLSKI; DAUTER; WLODAWER, 2014).

3.2 The phase problem and structure elucidation

X-ray crystallography is the current technique that provides better resolution and details of molecules. It is an alternative to the absence of lens capable of observing directly an object of the size of molecules. First, a compound has to be first crystallized, i.e. adopt a tridimensionally organized state in which the position of their molecules shares a common pattern in a repeated conformation, then it is possible to reconstruct its structure from the crystal measured diffraction. The light diffraction is a physical phenomenon of interference of waves occurring when the light's wavelength is of the same order of the size of the measurable object, that is why X-rays (order of Å) are used to measure atoms. Thus, the intensities of the diffracted waves are measured in a photon-sensitive detector. The intensities are converted to amplitudes, which correspond to the square root of the former. The ultimate result of crystallography is the electron density map that is calculated using the mathematical Fourier transform operation using the structure factor amplitudes (F_{hkl}) and the phases (α_{hkl}) of the diffracted pattern of the crystal structure according to Equations 1 and 2. Although, in the diffraction experiment, the phases are lost originating what is known as "phase problem". This is the most difficult problem in crystallography after crystallization, and many different techniques and algorithms have been proposed to solve it (HENDRICKSON, 2013) .

$$\rho(x,y,z) = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l} F_{hkl} \cdot e^{[-2\pi i(hx+ky+lz-\alpha_{hkl})]} \quad (1)$$

$\rho(x,y,z)$ is the electron density in position x, y and z in real space. F_{hkl} is the structure factor amplitude. h, k and l are the Miller Indices that defines a reciprocal space point. α_{hkl} is the phases.

$$F_{hkl} = \sum_{j=1}^{atoms} f_j \cdot e^{[2\pi i(hx_j+ky_j+lz_j)]} \quad (2)$$

x_j, y_j and z_j: coordinate position of j^{th} atom. f_j scattering factor of j^{th} atom.

Beyond deduction used by Bragg in the first determinations for structures of one or two atoms, the first general solution for the "phase problem" was obtained by the Patterson Function, a derivation of the electron density function that does not require phases (PATTERSON, 1934). It corresponds to a map of the vectors between each pair of atoms with their height being proportional to the product of the height of these atoms. If few atoms are within the unit cell, such as for small molecules, it is possible to deconvolute the Patterson Function and get their position. Afterward, the proposition of the probabilistic phase relations and the atomicity as restraints allowed to derive the reflection phases directly from intensities (EALICK, 1997). This is developed in direct methods and it is the most common technique to solve small compound structures.

Phasing by direct methods do not work in crystal containing high number of independent atoms, such as macromolecules. One of the ways to solve the "phase problem" is to reduce its complexity to the location of specific atoms. A substructure may be found using the differences in the intensity of two Isomorphous data sets (same crystal structure), one native and another derivative crystal that contains heavy atoms usually introduced by soaking. By small molecule methods, the substructure from these heavy atoms may be obtained using the Patterson Function of the difference of the intensities in the two datasets. Such strategy is known as Isomorphous Replacement. Another alternative is using the wavelength or energy necessary to excite the electrons in the inner shells of a specific atom type, these atoms will scatter x-ray differently (anomalous scattering) and their position in the crystal may be retrieved directly by the shift they produced in the measured intensities of Friedel pairs

(TAYLOR, 2003). Protein phases may be obtained from the deductions that the known position of the heavy atoms allow.

The most common approach to address the phasing “problem” is not experimental phasing, but approximating the phases based on a homologous protein with its structure already solved. This method is known as Molecular Replacement (MR) (ROSSMANN; BLOW, 1962). Highly automated, fast, and efficient MR became the more popular choice , as there is an abundance of different folds available from the PDB (EVANS; MCCOY, 2008). As proteins with similar sequences have the tendency to have similar structures, a homologous structure may be used as a search model. It is placed into the ASU and its phases account for the experimental data recorded. The model may be located by a rotation search followed by a translation search, whereas different solutions are scored based on a likelihood function (EVANS; MCCOY, 2008). Accounting the quality of the measured x-ray diffraction and the error and completeness of the search model in respect to the unknown structure, sophisticated likelihood functions distinguish wrong and correct rotation and translations (MCCOY et al., 2005; STORONI; MCCOY; READ, 2004). This is implemented in the PHASER and it has proven to be highly efficient (MCCOY et al., 2007).

After approximate phases are obtained from Isomorphous Replacement or Anomalous Scattering, the electron density map is calculated (ρ in **Equation 1**). As all atoms contribute to each structure factor, the map calculated from the initial phases of a partial model and the experimentally determined amplitudes may reveal missing atoms. If the initial phases are not enough to reveal missing features, they and the calculated map may be improved by density modification algorithms. This is obtained by enforcing protein features. Per example, the large and disordered solvent content in the macromolecule crystal is accounted for by solvent flattening (WANG, 1985). At high resolution, the sphere of influence algorithm is implemented in the phasing software SHELXE, as it differentiates protein electron density from water if peaks of density are within a minimum distance of 2.42 Å, which is the common 1-3 interatomic distance in macromolecules (THORN; SHELDRICK, 2013). The density coming from protein regions is sharpened while the one from the solvent is flattened.

If electron density is interpretable, a new or improved peptidic chain can be built in the map, which is known as building or autotracing. In SHELXE for example, polyalanine chains are traced in the map improved by density modification. This phasing software

improves its efficiency by iteration of cycles of density modification and autotracing, as better phases coming from improved trace generates a map with more features that are better interpreted by autotracing. At high resolution, better than 2Å, a good indication that a structure is phased is obtained if the Correlation Coefficient (CC) between the traced model and the data is higher than 25%.

3.3 Crystallography, *ab initio,* and ARCIMBOLDO

The *ab initio* methods to address the "phase problem" constitute an excellent alternative as they rely on measured diffraction intensities alone, to solve the phase problem (SHELDRICK et al., 2012). Unfortunately, with the increase of the number of atoms in a molecule, the probability distribution becomes broad and unreliable and direct methods do not work. The dual-space recycling approach, also known as Shake and Bake algorithm, extended the direct methods phasing from the previous maximum limit of 200 non-H equal atoms to 1000 non-H atoms (EALICK, 1997; MILLER et al., 1993). Using atomicity as restraint, the phases of an initial hypothesis, usually of randomly generated atoms, is calculated and modified following direct methods relationships. The electron density map is calculated from those phases and new set of atoms are selected in this map. Cycling between phase refinement to map interpretation steps, as one helps the other, structure solution may be obtained (EALICK, 1997; SHELDRICK et al., 2012). As in protein crystallography the number of atoms exceeds these numbers and atomic resolution is only available for less than 0.5% of structures, different strategies have to be pursued.

Born as an *ab initio* method in this chemical crystallography scenario, ARCIMBOLDO developed by Isabel Usón relies on the conservation of the building blocks among protein families, α-helices, and β-strands instead of using atomicity as restraint (MILLÁN; SAMMITO; USÓN, 2015). Its name was given in analogy of the portraits made by combination of fruits and vegetables by the 16th-century Italian painter Giuseppe Arcimboldo (**Figure 3.1 A**). Instead of using random atoms, it sets up a large number of hypotheses of secondary or tertiary structure by an initial search with the molecular replacement program, PHASER (MCCOY et al., 2007). These elements are small fragments that may be predicted analyzing the protein sequence, such as secondary structures (α-helix, β-wings, among others) and pharmacological sites. To succeed, these fragments need to be accurately positioned in relation to the final structure (below 0.5 RMSD). The polyala α-helix,

for instance, is ideal since it shows constant geometry, it is rigid and periodic (RODRÍGUEZ et al., 2012). If the hypotheses generated are close to the real structure, i.e. contain at least 10% of protein atoms, density modification and autotracing performed by program SHELXE (SHELDRICK, 2010) reveal the real "portrait" of the protein (**Figure 3.1 B**). Otherwise, the trial leads to a "still life" but the key is that both situations can be unequivocally distinguished by reliable figures of merit at medium resolution.

A B

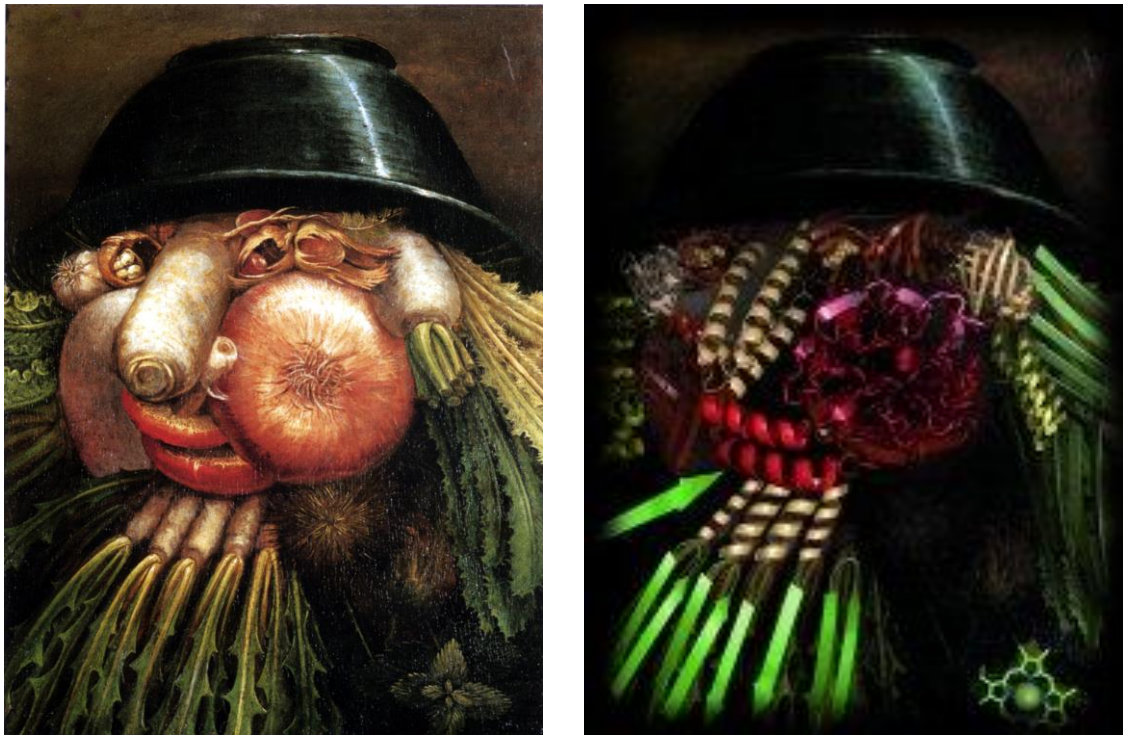


Figure 3.1 - Portrait of a human composed of assembly of vegetables (**A)** and "portrait" of a protein composed of assembly of small fragments on the right (**B**).
Images extracted (ARCIMBOLDO, [s.d.]).

Different modes of ARCIMBOLDO have been designed to tackle different structural scenarios. ARCIMBOLDO_LITE focuses on evaluating different rotation and translation solutions of the same fragment, while ARCIMBOLDO_BORGES evaluates simultaneously a library of different fragments of same local fold (MILLÁN; SAMMITO; USÓN, 2015). Herein, we use the term local fold to determine a set of around 20 to 40 contiguous residues, not necessarily corresponding to a continuous polypeptide stretch. ARCIMBOLDO_LITE has effectively phased new α-helix structures, as low deviations are expected for this highly rigid secondary structure, but alternative fragments may be used as small search models, such as an

iron cluster (SAMMITO et al., 2015). On the other hand, ARCIMBOLDO_BORGES simultaneous evaluation of thousands of similar fragments allows the location of a fragment with higher deviation from expected structure. This is the case for β-strands, as the hydrogen pattern maintaining this secondary structure connects distant regions not covalently bonded allowing higher local variation than a rigid α-helix. The α-helices tend to accumulate local deviation when they are longer than 14 residues, then a library of curved helices or separating the long helix in two smaller helices may suffice to phase the desired structure.

The libraries used in ARCIMBOLDO_BORGES are generated by the algorithm BORGES_MATRIX, which extracts particular folds from proteins available in the whole PDB. The libraries currently available in the CCP4 suite are 3 parallel β-strands, 3 antiparallel β-strands, 2 parallel and 1 antiparallel β-strands, 4 parallel β-strands, 2 parallel and 2 antiparallel β-strands, and 2 parallel helices. The extraction of a common fold by BORGES_MATRIX is done reducing the coordinates of a protein structure into a geometrical description. A protein structure is separated in consecutive tripeptides and their centroids of Cα and O are calculated, then the modulus of the vector defined from the centroid Cα to the centroid O describes the secondary structure while distances and angles between characteristic vectors can be used to capture the structural overall geometry.. Evaluating the whole PDB with BORGES_MATRIX, the α-helix possesses a characteristic vector modulus of 2.2 Å ($\sigma <$ 0.15 Å), while β-strands are usually around 1.4 Å (σ 0.18 Å) and 1.2 Å (σ 0.2 Å) for the fragments of 3 and 9 residues length (SAMMITO, 2015; SAMMITO et al., 2013). Characterization of tertiary structures and partial traces are being developed within the scope of phasing to introduce restriction in autotracing, such as non-crystallographic symmetry.

Another recent development is ARCIMBOLDO_SHREDDER, it requires a distant homolog template from which to derive smaller fragments, in an approach more similar to conventional Molecular Replacement with fragments. Similar to a paper shredder, different hypotheses are generated trimming residues spatially and evaluated against the rotation function of PHASER. A new function based on the LLG was proposed, in which the removal of specific regions that are different from the unknown sequence would improve the LLG, while the removal of highly correct parts (rmsd below 0.5 Å) would decrease LLG. Thus, simple evaluation of the maxima of this function already highlights what is the most probable correct set of coordinates given the remote homolog structure. The following steps are identical to other ARCIMBOLDO approaches, as best hypotheses are submitted to density

modification and autotracing with SHELXE. Thus, the initial coordinates given as input to SHELXE contain a higher number of atoms, which enable ARCIMBOLDO to tackles bigger structures.

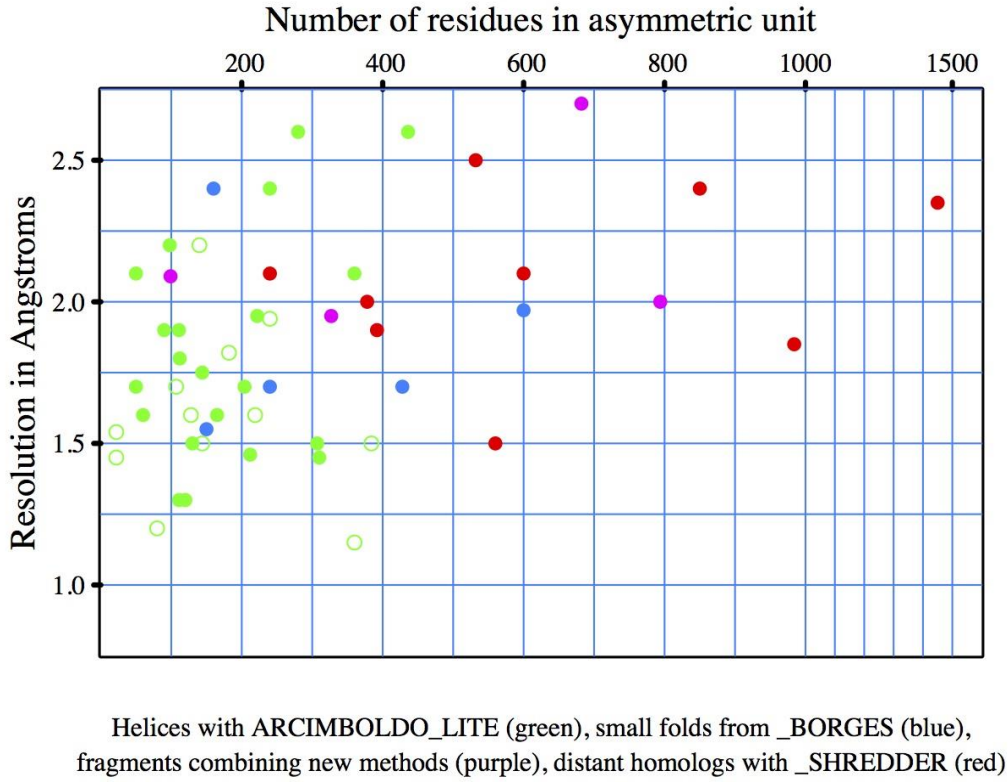


Figure 3.2 - Unknown structures solved with ARCIMBOLDO versions.
Hollow dots represent cases solved by external groups.

Each one of the ARCIMBOLDO versions has been able to phase unknown structures in different scenarios (**Figure 3.2**). ARCIMBOLDO_LITE focuses on datasets with high completeness containing usually up to 300 residues per ASU and diffracting to at least 2.0 Å resolution (green dots in **Figure 3.2**). On the other hand, the fragments or models used in ARCIMBOLDO_BORGES or SHREDDER enable phasing larger structures at lower resolution (purple dots in **Figure 3.2**). As it is seen in **Figure 3.2**, the resolution between 2.0 and 2.5 Å and having more than 400 residues in the ASU are less populated and this is the current challenge for ARCIMBOLDO. As the content of the ASU increases or the resolution lowers, density modification and autotracing becomes less effective in revealing the full structure. Often, recent tests of Usón group show that PHASER has been efficiently finding the correct rotation and translation, but the initial coordinates do not contain enough information to phase the rest of the protein. Moreover, the usual FOM are not so indicative to

filter true from false solutions and brute force is required. One of the ways to overcome this barrier is by improving the initial fragment, an obvious choice is by incorporating its side chains. Since the maps generated by small fragments at such resolution do not possess details for side chain discrimination, multiple hypotheses need to be generated restricted by the possibilities allowed by the known sequence, usually known in the case of recombinant proteins or determined by mass spectrometry for natural products. Possible assignments need to be mutually evaluated and pushed to expansion, the improvement of trace and CC are the indication of structure solution. This method, developed within the present work, has been named SEQUENCE SLIDER (SEQSLIDER).

3.4 Structure elucidation of crystal containing multiple isoforms

The following step of phasing is the refinement and modeling. In the former, errors in the geometry of the molecules are corrected and the atomic coordinates and their B-factors (atomic displacement parameter) are adjusted to minimize differences in the observed and calculated diffraction data. In the latter, the atomic model is manually or automatically fitted into the calculated electron density respecting chemical restrictions. As the phases dominate the reconstruction of the electron density maps over intensities, solutions from MR are biased. Available bias removal algorithms such as simulated annealing and omit maps are often used (PRAŽNIKAR et al., 2009; TERWILLIGER et al., 2008a, 2008b).

Cycles of refinement and modeling are done until the atomic model explains the observed diffraction data, which is measured globally by a low divergence between the $R_{factors}$ (R, R_{free}) and locally by correlation between observed and calculated electron density of each protein residue or ligand molecule (Real Space Correlation Coefficient (RSCC) calculated by **Equation 3**). As a divergence measurement, $R_{factors}$ of a final model should have low values. Its expected value depends on the resolution, for various deposited structures of ~2 Å resolution, their mean R and R_{free} are ~20 and ~25%, respectively (KLEYWEGT; JONES, 2002). As a correlation, RSCC are expected to have higher values, as a rule of thumb, residues/ligands with RSCC below 0.8 and high B-factors are critical. This CC may be calculated by *EDSTATS*.

One of the ways to reduce mode bias of particular atoms is by removing their contribution from the model and reconstruct the electron density. The difference map over this selected region is known as omit map and it will be highly informative. Recently the polder

maps have been proposed as an improved residual OMIT map. It removes bias by calculating a region of 5 Å around chosen coordinates to be omitted and it improves weak density by not including such region in bulk solvent (LIEBSCHNER et al., 2017). RSCC may be calculated with *PHENIX.POLDER* but using synthetic data (F_{obs}=|F_{model}|) and the other is a polder map (LIEBSCHNER et al., 2017). If the atomic model is consistent to known physical and chemical properties, which is calculated by validation software, it is ready to be interpreted.

$$\mathrm{R}SCC = \frac{\sum_r \left(\rho(r)_{obs} - \overline{\rho(r)_{obs}}\right) \cdot \left(\rho(r)_{calc} - \overline{\rho(r)_{calc}}\right)}{\left(\sum_r \left(\rho(r)_{obs} - \overline{(r)_{obs}}\right)^2 \cdot \sum_r \left(\rho(r)_{calc} - \overline{\rho(r)_{calc}}\right)^2\right)^{1/2}} \quad (3)$$

In macromolecular crystallography, previous knowledge is always necessary to elucidate a structure. Knowing the crystal composition from protein sequence to presence of ligands is required to build the model in electron density. To our knowledge, protein sequence is required in all automatic modeling tools. An exception is found for ligands, Arp/wArp has a tool to search for most common compounds in crystallization condition in a unknown blob of electron density (EVRARD et al., 2007). When departing from atomic resolution, modeling side chains *ab initio* is difficult, as the 20 possibilities share many similarities. Moreover, high flexible parts usually have no clear electron density. Mass spectrometry fills this gap determining protein sequence.

In crystallization, the higher the protein purity, the higher chance of success. On the other hand, crystallization constitutes a purification technique, as a particular crystal may grow selecting one component out of a heterogeneous mixture. Crystals are usually composed by a repetition of the same macromolecule(s). Obtaining high purity toxins from natural source, extracted venom, is a challenge due to the co-existence of isoforms that share physicochemical properties. With the function of foraging and protection adaptation, snake venom toxins are characterized by one of the most rapid evolutionary divergence and variability in any category of proteins (CALVETE et al., 2009; CASEWELL et al., 2013). The PLA_2s, per example, share many examples, at least 16 isoforms of β-bungarotoxin and 15 isoforms of crotoxin have been identified and characterized (DOLEY; KINI, 2009). The complexity in protein composition of these purified toxins may be seen in crystallographic and in mass spectrometry data. With the absence of a single sequence, phasing those crystallographic data sets is not the issue, but choosing the nature of specific residues is, since atomic resolution data is not the common scenario. The local RSCC statistics may allow to

establish whether the correct residue can be unequivocally identified from the available data. If not, a more rigorous outcome for the crystallographic analysis would be to acknowledge the sequence indetermination at particular sites. The problem being common to natural products, part of the information can be derived from sequence conservation across toxins of different families, and thus is an approach that can both benefit from and aid venomics in a broad sense.

Therefore, conceived as a side chain evaluator, SEQSLIDER also deals with crystallographic data sets whose crystal composition is not fully known. Previous knowledge, such as partial mass spectrometry data or phylogenetic analysis, may be used to restrict the 20 possible amino acids in each protein residue of the crystallographic model. SEQUENCE SLIDER aims two approaches, phasing unknown structures in the ARCIMBOLDO scope and solving toxin structures with partial known sequence.

3.5 Objective

The objective of this chapter is to propose a general methodology to aid solution of challenging datasets of resolution lower than 2.0 Å or that the sequence is partially known. The first focus on the scope of ARCIMBOLDO and the second in venoms.

3.6 Material and Methods

1. Computer setup used

The massive computation calculations required were executed in a local 180 GFlops Grid of IBMB-CSIC of Barcelona. Smaller computation calculations were done in a desktop computer with an i7 processor 3930K at 3.2 GHz, holding 6 cores sharing 16 GB running Debian 6.0 OS and in a laptop computer with an i7 processor 6700HQ at 2.6 GHz, holding 4 cores sharing 8 GB running Ubuntu 14.04 OS.

2. SEQSLIDER

SEQSLIDER is developed in *python* language. The secondary structure is predicted using the whole sequence with *PSIPRED* (MCGUFFIN; BRYSON; JONES, 2000). The side chains are modeled combining hypotheses assembled with *SCWRL4* (KRIVOV; SHAPOVALOV; DUNBRACK, 2009) and/or *COOT* (EMSLEY et al., 2010) and refinement

through *REFMAC5* (MURSHUDOV et al., 2011, p. 5), *PHENIX.REFINE* (ADAMS et al., 2010) or *BUSTER* (BRICOGNE et al., 2011). Real space correlation coefficient were calculated using *EDSTATS* and *PHENIX.POLDER* (LIEBSCHNER et al., 2017). Prior to *EDSTATS* run, if necessary, map coefficients of the file generated by refinement program are fixed with *mtzfix*, following Ian Tickle recommendation. As *EDSTATS* input is a map extension file, *fft* is used to convert the mtz file into it.

3. Toxin studies

a. Phylogenetic analysis as restriction for SEQSLIDER

We used *COBALT* (PAPADOPOULOS; AGARWALA, 2007) to obtain the local variability in the sequence of the studied proteins. As a protein multiple sequence alignment tool, this software finds a collection of pairwise constraint derived from conserved domain database and local sequence similarity information. We used protein BLAST (ALTSCHUL et al., 1990) to obtain the 100 sequences closer to our query sequence. We chose as query the sequence that has better correlation to the diffraction data, which was obtained with a full search *SEQSLIDER* run.

b. Proteins purification, crystallization and data collection

The BaspTX was purified from *Bothrops asper* venom by our collaborators Maria Laura Fernadez and Bruno Lomonte (Universidad de Costa Rica, Costa Rica). They used a cation-exchange chromatography followed by a reverse-phase high pressure liquid chromatography. It was crystallized and its diffraction data collected by Dr. Juliana I. dos Santos, a former postdoc in our crystallography group in UNESP.

The BbTX was purified from the venom of *Bothrops brazili* by our collaborators Salomón Huancahuire-Vega, Luis Ponce-Soto and Sergio Marangoni (University of Campinas, Brazil) in a single step of reverse-phase high pressure liquid chromatography (HUANCAHUIRE-VEGA et al., 2009). It was crystallized and its diffraction data collected by Edson José Comparetti (FERNANDES et al., 2012), a former undergraduate student in our crystallography group in UNESP.

c. Mass spectrometry analysis

Using LNBio (Campinas - Brazil) equipment and sequencing support, mass spectrometry was used to analyse these two samples. With bottom-up strategy digesting samples with trypsin, peptides were separated by a reverse phase chromatography column (C18 100 μm × 100 mm RP-UPLC from NanoACQUITY UPLC, Waters). The fractions were

measured by nanoelectrospray source on Q-TOF Ultima mass spectrometer, Waters (Q-Tof Ultima API, MicroMass/Waters) at a flow rate of 0.6 µl/min. The gradient was 2–90% acetonitrile in 0.1% formic acid over 45 min. The nanoelectrospray voltage was set to 3.5 kV, a cone voltage of 30 V and the source temperature was 100°C. The instrument was operated in the 'top three' mode, in which one MS spectrum is acquired followed by MS/MS of the top three most-intense peaks detected. After MS/MS fragmentation, the ion was placed on exclusion list for 60 s. The spectra were acquired using software *MassLynx v.4.1* and the raw data files were converted to a peak list format (mgf) without summing the scans by the software *Mascot Distiller v.2.3.2.0*, 2009 (Matrix Science Ldt.) allowing the label-free analysis, and searched against NCBI protein sequence databank filtered in Phylum Chordata using *Mascot engine v.2.3.01* (Matrix Science Ltd.), with carbamidomethylation as fixed modifications, oxidation of methionine as variable modification, one trypsin missed cleavage and a tolerance of 0.1 Da for both precursor and fragment ions.

d. Crystallographic determination

We processed the data with *XDS* (KABSCH, 2010) to a maximum resolution in which low divergence of equivalent measurements was seen (as judged by $CC_{1/2}$ (KARPLUS; DIEDERICHS, 2012)). We manually merged data with *aimless* (EVANS, 2011). We evaluate the crystallographic data with *Xtriage* (ZWART; GROSSE-KUNSTLEVE; ADAMS, 2005), *AIMLESS* (EVANS, 2011) and *XPREP* (written by Sheldrick GM and proprietary of Bruker AXS). We obtained the phases by molecular replacement with *Phaser* (MCCOY et al., 2007) using the native models of the studied toxin. We performed refinement with *Buster* (BRICOGNE et al., 2011), modeling with *Coot* (EMSLEY et al., 2010); model validation with *MolProbity* (DAVIS et al., 2007); and structure analysis, comparison and images with *Coot*, *Pymol* (http://www.pymol.org/).

e. Test datasets

We chose two data sets available in the protein data bank to evaluate the recent version of our SEQSLIDER into venoms. The PDB ids 3G8G and 3G8H. They are two isoforms of the same protein, although one is purified from natural source, while the other is recombinant. We assume that proteins purified from recombinant source will have a known sequence, as the toxin isoforms will not be produced by the bacteria. Instead of using the numbering system proposed by Renetseder *et al.* (1985) as in **chapter 2**, we will use the more intuitive continuous numbering as the models were deposited without gaps.

3.7 SEQUENCE SLIDER into venoms

As the complexity of sample increases in purified toxins from natural source, crystals may contain more than a single isoform. In a complex crystal composition, structure elucidation is not straightforward by the recent methods. *SEQSLIDER* proposes to extract the most of the available data to guide modelling of side chains.

3.7.1 Algorithm description

SEQSLIDER into venoms is separated in the following steps (**Figure 3.3**):

I. generation of evaluated residue types for each residue number in model;
II. modelling the side chain of each trial into the calculated electron density map given in a reflection file (MTZ format);
III. unify Bfactor and occupancy of side chain atoms of evaluated residue;
IV. map calculation using phases from new model;
V. calculate RSCC; and
VI. organize results.

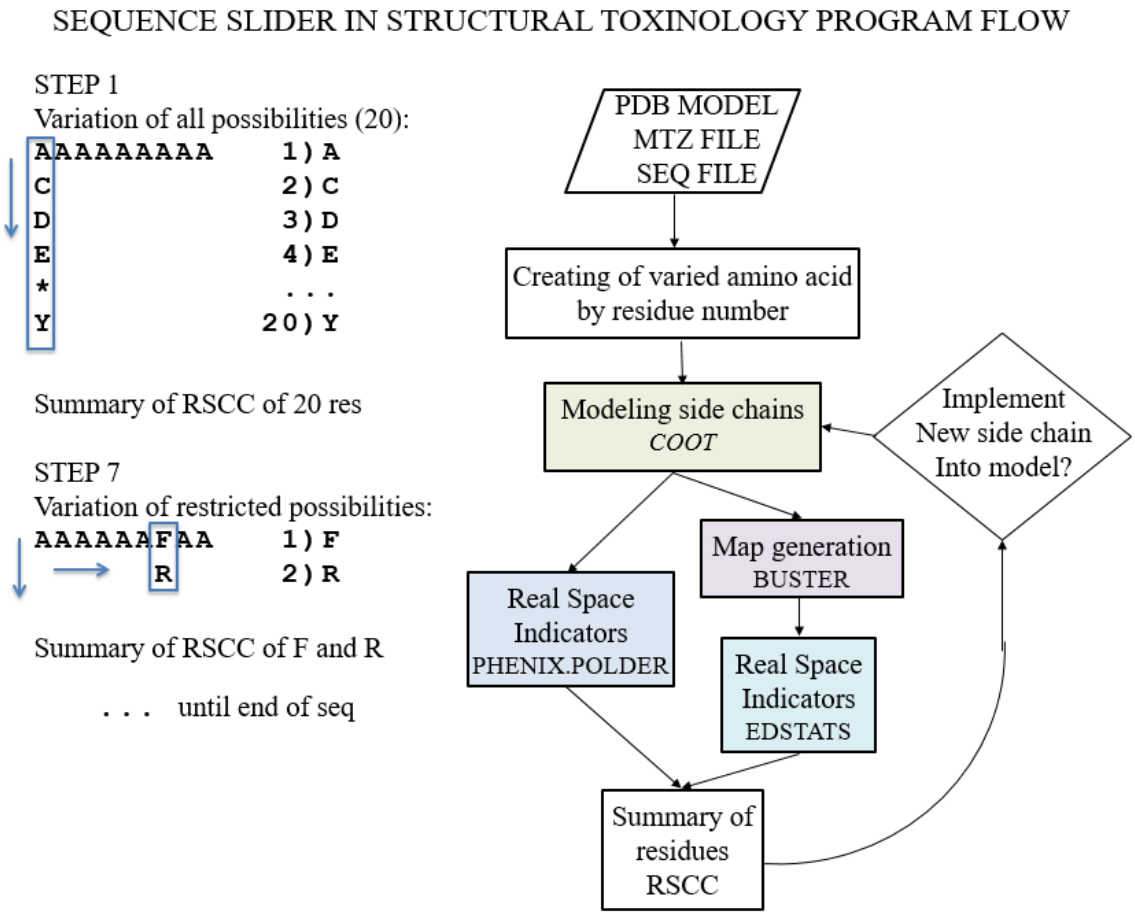


Figure 3.3 - Program flow of SEQSLIDER applied to venoms. Colored background indicates the use of external programs.

The steps I, III and VI are done by our own algorithm. The generation evaluated residues depends on the *SEQSLIDER* mode. It may be run on FULL SEARCH, which evaluates all natural 20 amino acids, or run on restricted by previous known information from phylogenetic or mass spectrometry analysis, per example. For step II, a CONSTRUCTIVE approach may be chosen, if structure is composed by polyala chain(s), after each position is analysed, the amino acid having better agreement with calculated electron density is incorporated in subsequent positions evaluations. The step II is done using the external program functions of *COOT*, mutate to desired residue, automatic rotamer search and sphere refinement of all atoms within radius of 5 Å of modified atoms. The last function ascertains that no clashes will be present among the included atoms, although it does not take symmetry mates into account.

After *COOT* modeling, Bfactor and occupancy are set to a value of 30 and 1 for evaluated side chain atoms, respectively. The main chain atoms values are not changed. As default, steps IV and V are generated using *PHENIX.POLDER*. Alternatively, a real space correlation coefficient (which *EDSTATS* calls CCSa) may be calculated using *EDSTATS* from the calculated map from *BUSTER*. In this thesis, we are considering a distinct amino acid from the others, when its RSCC calculated by *PHENIX.POLDER* (PCC) is at least 1.5% better than other trials.

The input to the *script* comprises:

i. Coordinates file (PDB format)
ii. Reflection file (MTZ format coming from processing software without phases)
iii. Map file (MTZ format from refinement software containing phases)
iv. Chosen output name
v. Type of run:
   a. Using mass spectrometry data (word “MASPEC” should be given), or
   b. Using phylogenetic data (word “ALIGN” should be given), or
   c. Trying all possibilities (word “TRYALL” should be given).
vi. If option “ALIGN” is given, then sequence file needs to be provided in a last argument.

The output of new coordinates, map files and RSCC is organized in folders. The first is a folder with name given in option **iv**, then a folder of the chain letter followed by a folder of residue number. To illustrate *script* output and program flow, if the chosen output name is “test” to evaluate a tyrosine in position 1 of chain A, files would be saved in path “./test/A/1/” as:

- “1Y.pdb” (*COOT* output file)

- “1Y.log” (*POLDER* log)
- “1Y_buster.mtz”, “1Y_buster.pdb” and “1Y_buster.log” (*BUSTER* outputs)
- EDSTATS: “1Y_buster_edstats.out” and “1Y_buster_edst.log”.

The RSCC of evaluated residues in position 1 and chain A would be summarized for *POLDER* in file “A_1_polder.log” and for *EDSTATS* in file “A_1_edstats_fl.log”. An overall summary file would be saved in original path under file “test_all.log” containing sorted *PCC* and CCSa of all atoms in chosen residue. For sake of validation, an additional file “test_final_model.log” is given including only best *PCC* until residue type of original model file is reached. If alignment option was chosen, a last column includes the percentage of appearance in given sequences.

3.7.2 Known test case

Prior to dealing with unknown cases, we applied *SEQSLIDER* to a known structure. We chose Ammodytoxin (Atx) from *Vipera ammodytes ammodytes*. There are two high resolution crystallographic datasets for this toxin, one for isoform C, which was crystallized from samples purified from natural source, and the isoform A produced recombinantly, which would not include presence of other isoforms. These are two PLA_2 isoforms, AtxA (3G8G) and AtxC (3G8H) presenting a difference in only two natural mutations (**Table 3.1**), F113I and K117E, which is enough to alter its toxicity and anticoagulant activities (SAUL et al., 2010). The two datasets are isomorphous and the crystallographic models appear identical within the experimental error (Cα superposition of 0.23 Å). The highest distance between correspondent Cα, after superposition, is on residues 116 and 117 that are 1.6 and 1.1 Å, respectively, being the latter the one of the natural mutation seen for the two isoforms.

Table 3.1 - Overall data statistics of crystallographic models Atx.

Datasets	3G8G	3G8H
Toxin	AtxA	AtxC
Sample source	Synthetic	Natural
Space group	$P3_2$	$P3_2$
Unit cell: abc (Å)	a=b=68.2 c=46.1	a=b=67.7 c=45.3
Resolution (Å)	1.7	1.35
R/R_{free} (%)	17.2 / 21.6	17.1 / 19.6

We applied *SEQSLIDER* to probe a final stage of sequence determination and validation on both datasets in different resolution scenarios: full, 2.0 and 2.3 Å resolution. For sake of reference, we also used the FULL SEARCH mode.

3.7.2.1 Synthetic AtxA

3.7.2.1.1 Full search

In a full search of 122 residues, 2440 possibilities are generated (20*122 (20 is the number possible side and 122 is total number of residues)). Using all diffraction data available (1.7 Å) in 3G8G, *SEQSLIDER* takes ~2 hours and 40 minutes to calculate PCC and ~ 5h20m to calculate CCSa in a 4-core machine. Each *COOT* modeling takes few seconds (1-2 s), *PHENIX.POLDER* run takes ~15 s. EDSTATS is rather fast with few seconds *per* run (1-3 s), although it requires map calculation, for which we are using BUSTER that takes ~28 s. Reducing the resolution to 2.0 or 2.3 Å, SEQSLIDER takes ~1h30m for PCC and ~4h for CCSa calculation.

In **Table 3.2**, we summarize how well the true residues, the ones present in the deposited model, scored in relationship to the rest and if they were distinguished. Lowering the resolution improve the scoring of the true residue, mainly their distinction. The residues whose RSCC were not scored best in 2.3 Å trial were: 18 (L), 23 (F), 30, 46 (V), 97 (R), 53, 70 (N), 33, 65, 77, 116, 117, 121 (K), 69, 87, 118 (E), 112 (D).

3.7.2.1.2 Restricted search

Both sequences of isoform A and C of Atx were used to generate the sequence variability using *COBALT* and protein *BLAST* (**Supplementary Figure 3.27**). Out of the 122 residues, 25 are fully conserved, which are related to calcium binding loop and disulfide bridged residues (shown by only one residue in **Supplementary Figure 3.27**). Moreover, 17, 15, 18, and 14 are seen 2, 3, 4, 5 possibilities, respectively. The other 33 residues are not conserved with more than 5 different seen amino acids. These restrictions reduce to 468 possibilities, which is one fifth that was calculated in full search. The total time spent in each of the resolutions run is one fifth as before. In the resolution to 2.0 or 2.3 Å and with these phylogenetics restrictions, SEQSLIDER took ~45m for PCC and ~48m for CCSa calculation.

In **Table 3.2**, we summarize how well the true residues, the ones present in the deposited model, scored in relationship to the rest and if how they were distinguished. In full

resolution, PCC scored an additional 14 correct residues as best in comparison with the full search, whether the improvement in CCSa was almost 60. In this restricted search mode, lowering resolution stop improvement for PCC, whether for CCSa, it is still seen. Evaluating the PCC results in the 2.3 Å trial, out of the 122 residues, 108 were correctly scored. The ones that were still not distinguished are: 33K, 53N, 65K, 69E, 70N, 77K, 97R, 104N, 112D, 113F, 116K, 117K, 118E, 121K (**Table 3.3** in 3G8G columns).

3.7.2.2 Natural AtxC

3.7.2.2.1 Full search

The crystals of the natural AtxC diffracted to 1.35 Å. The 0.35 Å gained in resolution in comparison to AtxA dataset double the time spent by *PHENIX.POLDER* in each calculation, but increase less than 20% in BUSTER map calculation. *SEQSLIDER* calculation time with *POLDER* was 5h20m and CCSa, 6h20m.

In **Table 3.2**, we summarize how well the true residues, the ones present in the deposited model, scored in relationship to the rest and if they were distinguished. Similar observation is seen for 3G8H data set. Most of the correct residues were best scored by PCC in all evaluated ranges of resolution. This was not the case for CCSa, but lowering resolution improved distinction.

3.7.2.2.2 Restricted search

Following restriction in phylogenetic analysis (**Supplementary Figure 3.27**), we calculated the 468 different possibilities for 3G8H datasets of 1.35, 2.0, and 2.3 Å. Again, restrictions improved distinction of both RSCC statistics, more precisely, ~10 residues for PCC and 20-70 for CSSa depending on resolution. Similar unclear residues are seen for 3G8H data set (**Table 3.3** in 3G8H columns). The true residues whose score was not the best in 2.3 A are similar to those of 3G8G. The exception are the residues 33, 69, and 117 that improved in 3G8H, while 56, 101, and 105 became unclear.

3.7.2.3 Comparison of AtxA and C datasets

PCC distinguishes better correct residues than CCSa (**Table 3.2**). Such difference may be related to different algorithm purpose, while the latter is validations, the former is subtle electron density distinctions to guide modelling. PCC might be better for our case

because the chosen grid size for local CC calculation is more constant, as an intersection of spheres of radius 5 Å is used. Moreover, the use of the less biased polder map may be more sensitive.

Impressively, lowering the resolution improve the scoring and distinction. Electron density calculated from higher resolution datasets have more features, which could change the convergence of the refined side chain position. Local minima in *COOT* auto rotamer search and sphere refine could be found in higher resolution maps, therefore the weaker map features may have been interpreted over the stronger, leading to unrealistic RSCC.

In SEQSLIDER full search, similar residues are mutually evaluated. Despite their chemical property being different, their composition differs only in one single atom with similar number of electrons (C, N or O). This is the case for two groups of residues have electron density almost identical, first L, N, and D, and second T and V (**Figure 3.4**). Almost half of the true residues whose scores were not the best were related to this scenario. The analysis of the different possibilities of local hydrogen networks together with hydrophobic/electrostatic contacts could be a methodology for choice. The validation tool *MOLPROBITY* already evaluates the all-atom contact after adding hydrogen atoms in the model. In our case, the restriction brought by phylogenetic analysis reduces significantly the mutual evaluation of these similar side chains. Therefore, restriction brought by mass spectrometry and by sequence conservation is essential. With exception of four residues, 56, 104, 112, and 118 (**Table 3.3**).

Table 3.2 - Summary of full and restricted SEQSLIDER searches of Atx datasets.

Number of residues	3G8G						3G8H					
	PCC			CCSa			PCC			CCSa		
	Full	2	2.3	Full	2	2.3	Full	2	2.3	Full	2	2.3
Full search												
Distinguished	**49**	**65**	**73**	**17**	**36**	**54**	**44**	**66**	**72**	**10**	**39**	**54**
Within best 1	**91**	**98**	**102**	**30**	**64**	**78**	**90**	**95**	**101**	**16**	**64**	**82**
Within best 2	101	107	107	70	93	95	105	106	109	57	96	104
Within best 3	107	110	111	94	100	101	111	110	113	88	106	109
Within best 4	112	111	111	105	104	106	112	113	114	106	111	111
Within best 5	113	113	113	109	108	108	114	113	114	110	113	114
Restricted search												
Distinguished	**93**	**95**	**95**	**79**	**85**	**93**	**93**	**97**	**97**	**72**	**87**	**91**
Within best 1	**105**	**105**	**108**	**89**	**95**	**103**	**107**	**105**	**108**	**88**	**99**	**99**
Within best 2	112	113	114	108	111	112	113	112	114	111	111	114
Within best 3	118	118	117	116	115	115	116	115	115	116	115	115
Within best 4	119	119	119	118	118	118	118	117	117	119	117	117
Within best 5	122	121	121	121	120	122	119	120	118	120	119	119

The number refers to the residues that were corrected classified in the descripted category.

Table 3.3 - Summary of evaluated residues in 3G8G and 3G8H of SEQSLIDER full search run whose true residue did not score best.

R#	3G8G AA	3G8G CC	3G8G F	3G8H AA	3G8H CC	3G8H F
33	S	92.1	3	-	-	-
	H	91.3	1	-	-	-
	Q	91.1	23	-	-	-
	G	90.8	1	-	-	-
	K!	89.9	53	-	-	-
53	S	88.6	3	S	87.6	3
	N!	86.6	37	N!	85.5	37
56	-	-	-	N	84.5	7
	-	-	-	D!	84	68
65	S	91.7	17	S	92.4	17
	K!	91.4	30	T	92.2	5
	-	-	-	I	91.9	1
	-	-	-	K!	90.3	30
69	K	89.9	36	-	-	-
	E!	88.3	51	-	-	-
70	G	89.1	1	G	85	1
	S	85.7	5	S	82.2	5
	N!	82.5	85	T	80	3
	-	-	-	K	72.9	1
	-	-	-	N!	70	85
77	S	90.9	30	S	89.5	30
	N	88.3	6	G	88.1	18
	D	87.9	1	N	87.9	6
	G	87.4	18	D	87.6	1
	K!	87.4	24	Q	86.5	2
	-	-	-	E	86	13
	-	-	-	K!	84.7	24
97	G	93.8	6	G	94.6	6
	R!	92.3	92	R!	92.7	92
101	-	-	-	P	89.4	5
	-	-	-	A	89.3	2
	-	-	-	S	88.7	1
	-	-	-	T	88	1
	-	-	-	G	86.4	2
	-	-	-	K!	85.5	49

R#	3G8G AA	3G8G CC	3G8G F	3G8H AA	3G8H CC	3G8H F
104	D	94.2	1	D	95	1
	N!	93.5	59	N!	94.6	59
105	-	-	-	A	89	1
	-	-	-	G	88.8	1
	-	-	-	L	88.4	1
	-	-	-	I	87.8	2
	-	-	-	K	87.5	72
	-	-	-	D	87	1
	-	-	-	N	86.5	5
	-	-	-	Y!	85.3	7
112	N	90.9	8	G	92.7	1
	K	88.9	6	D!	88.9	58
	G	88.8	1	-	-	-
	D!	86.8	58	-	-	-
113	S	88.7	10	V	91.7	3
	F!	87.8	74	I!	89.7	10
116	S	83.3	5	S	86.3	5
	I	79.4	1	I	84.8	1
	K!	77.5	77	T	84.6	13
				K!	81.5	77
117	S	83.5	5	-	-	-
	D	78.6	8	-	-	-
	G	76.7	48	-	-	-
	K!	75.4	14	-	-	-
118	S	91.2	1	S	93.9	1
	V	88.3	27	V	90.7	27
	G	87.2	12	K	90.5	11
	K	86.9	11	G	89.4	12
	D	81.1	4	D	84.5	4
	E!	80.6	29	P	82.9	14
	-	-	-	E!	82.6	29
121	V	88.9	1	V	90.2	1
	T	88.4	1	T	86.4	1
	K!	85.3	94	K!	82.8	94

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC), frequency of residue in given position from phylogenetic analysis (F) and true residue (!).

Another reason for the lack of distinction in the restricted search is poor local electron density. Often, exposed residues have side chain positions highly varied within the protein molecules inside the crystal. Also flexible is the polypeptide extremities and regions related to function are also highly flexible. As a consequence of it, unclear electron density is

observed in those regions. In PLA_2s structures, this is the case for C-terminal and calcium binding loop residues. Most of the residues with apparent wrong scoring, i.e., 33, 53, 65, 70, 77, 97, 101, 105, 112, 113, 116, 117, have better RSCC for small side chains than the true one that has more atoms. This is the particular case for large residues K, F, R and Y whose PCC may be not clearly distinct of A, G or S.

Both of these datasets come from the same protein, but 3G8G is isoform A that has 113F and 117K, while 3G8H is isoform C that has 113I and 117E (**Table 3.3**). The SEQSLIDER did not differentiate residues in position 113 residues for both cases. Electron density for 3G8G is weak in 113 position (purple in **Figure 3.5 A**), whether for 3G8H (yellow), it misses one atom. In fact, for the former, PCC scores better a V, which would fit perfectly in this electron density. In the case for 117, in 3G8H SEQSLIDER distinguish well (yellow in **Figure 3.5 B**), but in 3G8G polder map, there is almost no electron density for side chain atoms (purple). Although clear difference is seen in the polder maps coming from the different data sets.

For the large residues, K, F, R, and Y, that may be misjudge with smaller side chains, the difference in their RSCC statistics is rather small and the algorithm could point to those false positives to be checked manually. Depending on the crystallographer, such side chains may be modelled with partial occupancy, omitted or even kept. PCC below 85% of these large residues may be used as a guide to user trying to elucidate the structure.

If we consider a distinct residue if it has at least a PCC 1.5% greater than other evaluated possibilities, SEQSLIDER distinguished in 2.3 Å datasets correctly ~73 and ~95 residues for full and restricted search, respectively. Out of those, 8 would be incorrectly distinguished due to poor electron density, such as 113V for 3G8H. SEQSLIDER has been proven a valuable tool to aid structure solution in the absence of a known single sequence. Moreover, highly flexible regions favor small side chains PCCs and are indicative that occupancy should be refined for large side chain or that their atoms should be omitted.

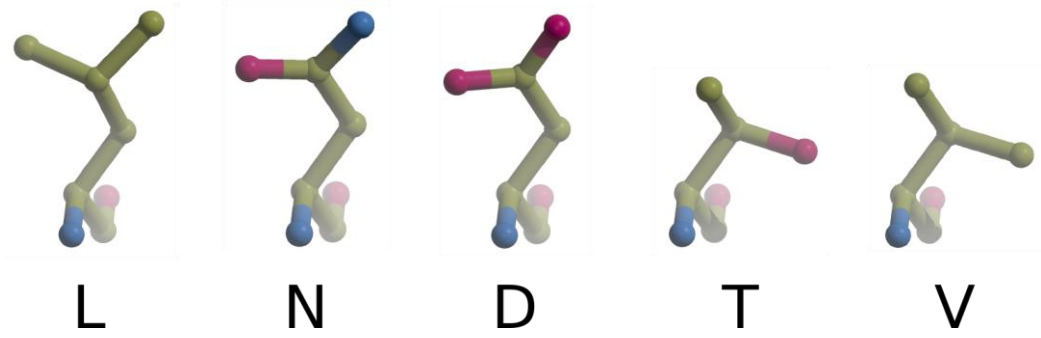


Figure 3.4 - Amino acids with similar atomic composition.

A B

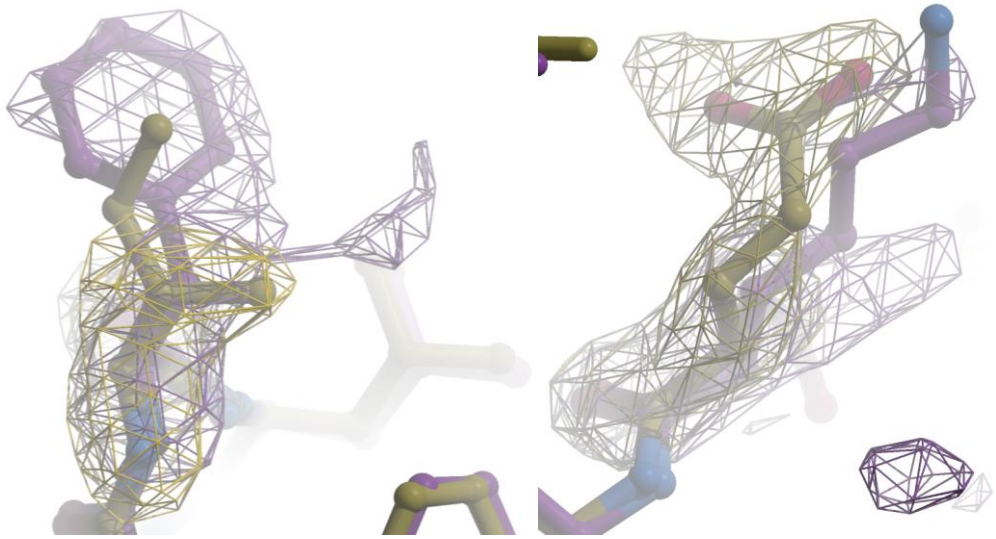


Figure 3.5 - Sticks representation of residues differences in Atx isoforms A (purple) and C (dark yellow) and respective polder maps.
In **A**, weak electron density for 113F of AtxA in purple and I113 and missing electron density for atom CΔ. In **B**, weak electron density for 117K of AtxA in purple and well fitted electron density for 117E of AtxC. The polder maps contour used are 3.5 σ.

3.7.3 Unknown structures

3.7.3.1 BbTX-III

BbTX-III is a basic D49-PLA_2 purified from *Bothrops brazili*. Edson Comparetti crystallized it in the following precipitant solution 2% PEG400, 2 M ammonium sulfate, 0.1 M HEPES Na^+, pH 7.5 (FERNANDES et al., 2012) and diffracted this crystal to 2.0 Å resolution (**Table 3.4**). He successfully obtained the phases from Molecular Replacement using a snake venom PLA_2 model in space group $P6_522$ with a monomer in the ASU. If the crystal content included more than 1 isoform, electron density of different side chain would be

averaged, as we will elucidate the monomer inside the ASU. In modeling and refinement, Edson could not improve much of the model and mainly its side chains. He encountered many disagreements between the electron density and expected sequence (HUANCAHUIRE-VEGA et al., 2009) (sequence E in **Figure 3.6**). These crystals could be composed of an impurity or an isoform in BbTX-III sample, since crystallization is a purification procedure. Or even, these differences could be due to a different lot of lyophilized venom of different snakes but same species despite same purification procedure used.

Thus, we sequenced this protein of the provided sample by our collaborators using mass spectrometry from LNBio facilities (Campinas – Brazil). Evaluating peptide signals against NCBI databank filtered in Phylum Chordata, we found two different proteins in this sample, one is an impurity MTX-II, that is also found in from *Bothrops brazili* venom, and the other assembles D49-PLA$_2$ enzymes. We gathered the results in a single alignment coloring residues according to mass spectrometry results (**Figure 3.6**). Mass spectrometry data covered 106 residues out of 121 (88%). Out of the matched peptides, 17 and 3 residues had 2 and 3 positive matches respectively (colored in beige). Out of the 15 residues whose peptide were not identified in mass spectrometry, 3 are disulfide bridges, reducing to 12 total uncertainty.

```
   1         11         21         31         41         51         60
A) SLVQFETLIMKIAGRSGVWYYGSYGCYCGSGGQGRPQDASDRCCFVHDCCYGKVTDCDPK
B) HLMQFETLIMKIAGRSGIWYYGSYGCYCGKGGQGQPKDASDRCCFVHDCCYGKVTGCDPK
C) DLMQFETLIMKIAGRSGVWFYGYYGCYCGAGGQGRPQDASDRCCFVHDCCYGKVDSCNPK
D) SLVELGKMILQETGKNPAKSYGAYGCNCGVLGRGKPKDATDRCCYVHKCCYKKLTDCDPK
E) SLWEWGQMILKETGKNPFPYYGAYGCYCGWGGRRKPKDATDRCCFVHDCCRYKKLTGCPK

  61         71         81         91        101        111        121
A) ADVYTYSEENGVVVCGGDDPCKKQICECDRVAATCFRDNKDTYDNK-YWFFPAKN-CQEESEPC
B) EDFYTYSEENGAVVCGGDDPCKKEICECDKNAAICFRDNMDTYDSKTYWFLNPKN-CQEESEPC
C) SNTYVYSEENGAIVCGGDDPCKKEVCECDRVAAICFRDNIDTYDSKKYWMLPPRN-CQEESEPC
D) KDRYSYSWKDKTIVCGENNSCLKELCECDKAVAICLRENLDTYN-KKYRNNHLKPFCKKAD-PC
E) TNDRYSYSRLDYTIVCGEDDPCKEICECDKAAAVCFRENLRTYN-KKYMAHLRVL-CKKDK-PC

A: acid PLA2 from Bothrops diporus (UniProt: I2DAL4)
B: PLA2 from Ovophis monticola (UniProt: U5HRG7)
C: PLA2 from Ovophis monticola (UniProt: U5HS18)
D: MTX-II from Bothrops brazili (PDB_id: 4DFC)
E: BbTX-III from Bothrops brazili (HUANCAHUIRE-VEGA et al., 2009)
```

Figure 3.6 - Sequence similarity of BbTX-III against PLA$_2$s.
Matching amino acids from peptide mass spectrometry analysis in bold and blue. Residues in bold and beige mark multiple possibilities; residues in grey and black where no match was found in mass spectrometry.

Table 3.4 - X-ray data collection and refinement with best obtained model of BbTX-III and MT-I

Datasets	BbTX-III	MT-I
Number of monomer per ASU	1	2
Beamlines in LNLS	MX2	MX2
Space group	$P6_522$	$P2_1$
Unit cell: abc (Å) αβγ (º)	a=b=71.8 c=106.4 α=β=90 γ=120	a=39.6 b=71.8 c=45.4 α=γ=90 β=107.6
Resolution (Å)	40-2.05 (2.17-2.05) [a]	40-2.01 (2.44-2.01) [a]
Total reflections	155562	56460
Unique reflections	17539 (3066) [a]	15777 (2404) [a]
Average multiplicity	8.9	3.6
R_{meas} [b] (%)	9.4 (204.7) [a]	12.9 (72.1) [a]
CC (1/2)	99.9 (29.9) [a]	99.3 (48.2) [a]
Completeness (%)	90.6 (98.0) [a]	97.8 (93.4) [a]
Average I/σ (I)	15.0 (0.9) [a]	7.5 (1.9) [a]
R [c] (%)	20.8	18.7
R_{free} [c] (%)	24.2	23.7
RMS deviations from ideal values		
Bond lengths (Å)	0.010	0.008
Bond angles (º)	1.0	1.0
Ramachandran plot (%)		
Residues in most favorable regions	98.3	96.6
Residues in accepted regions	0.8	3.0
Molprobity score	1.13	1.81
Number of molecules/atoms, their averaged B factor (Å^2) and averaged occupancy		
Protein	1 (52.5)	2 (41.9)
Ligand	1 PEG (58.9) 9 EDO (72.2) 1 SO_4 (83.8)	1 PEG (67.3) 1 Ca^{2+} (67.4)
H_2O	87 (61.3)	170 (47.8)

[a] Number in parenthesis are for highest resolution shell. [b] Number in parenthesis and in brackets are B factor and occupation, respectively. $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_t^{n_h} |\hat{I}_h - I_{h,t}|}{\sum_h \sum_t^{n_h} I_{h,t}}$, $\hat{I}_h = \sum_h \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$ redundancy independent R-factor. Calculated for I > -3σ (I). [c] $R_{cryst} = R_{hkl} \frac{||Fobs_{hkl}| - |Fcalc_{hkl}||}{|Fobs_{hkl}|}$. [d] R_{free} is equivalent to R_{crist}, but calculated with reflections (5%) omitted from the refinement process. RMS *root-mean-square*. Anomalous correction: percentage of correlation between random half-sets of anomalous intensity differences. Anomalous signal: mean anomalous difference in units of its estimated standard deviation $R_{hkl} \frac{|F(+) - F(-)| R_{cryst}}{\sigma}$. F(+), F(-) are structure factor estimates obtained from the merged intensity observations in each parity class. EDO is the abbreviation of ethylene glycol.

3.7.3.1.1 Preliminary SEQSLIDER run: CONSTRUCTIVE and FULL SEARCH mode

A crystal grown in a heterogenous sample could correspond to a mixture of different isoforms or to an isolated form through purification. The mass spectrometry results corroborate sample heterogeneity. In a prior step, we will evaluate what information can be extracted from the diffraction data alone as an independent and unbiased experiment.

To remove the bias from a preconceived model, we approximated BbTX-III phases to the polyala chain of monomeric BthTX-II elucidated in this thesis using Molecular Replacement. PHASER clearly found a solution with a Translation Zscore of 18.5. We improved main chain position in relationship to the calculated electron density. And, with all missing side chain atoms, we run *SEQSLIDER* in FULL SEARCH and CONSTRUCTIVE approach to improve the model while side chains were assigned and to introduce no information from mass spectrometry. As accurate RSCCs are dependent on low divergence between model and diffracted data, CONSTRUCTIVE mode is advised when side chain constituency is far from ideal. Polyala model started a *SEQSLIDER* round with R/R_{free} of 43.7/45.6% and identity 16% (in respect to the final model). By end of the run, the more complete model finished with R/R_{free} and identity of 37.7/39.1% and 34%, respectively (**Figure 3.7**).

We performed another round of manual modeling to fit main and side chain atom inside as the increase in the number of atoms in the model improved significantly the electron density map. We modeled, specially, flexible regions, such as C-terminal and loop residues, using a BUSTER map calculated after 10 refinement cycles. We submitted a new round of SEQSLIDER evaluation without the constructive approach but keeping the full search mode. Sequence identity increased to 63% and *BUSTER* refinement adding water improved R/R_{free} to 27.2/32.3%.

In *SEQSLIDER* evaluation in FULL SEARCH mode, 80 residues matched as best scored amino acid the one determined by the mass spectrometry (**Supplementary Table 3.15**). Evaluating the sequences whose positive peptides were found in the protein databank (coloured in blue in **Figure 3.6**), BbTX-II sequence diverge from other PLA_2 (coloured in beige in **Figure 3.6**). Comparing the PCC of the residues in position 8, 16-20, and 31, whose constituency is different for BbTX-II and other PLA_2, the best scores were all constituents of

the former (**Table 3.5**). Therefore, this crystallographic data does not support presence of BbTX-II molecules in the diffracted crystal.

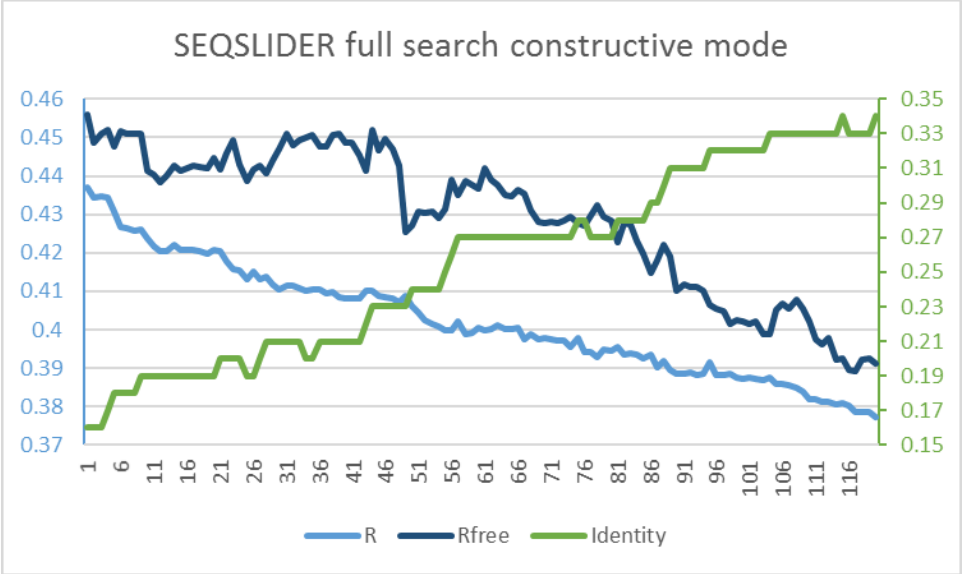


Figure 3.7 - $R_{factors}$ and identity along first round of SEQSLIDER full search and constructive mode with BbTX-III dataset starting from a polyala model.
As sequence is correctly attributed, identity increases and $R_{factors}$ decrease.

Table 3.5 - Correlation coefficient of PLA_2 and BbTX-II residues whose mass spectrometry matched

R#	AA	CC	Bd	M2
8	L!	94.9	L	M
	N	91.9		
	S	91.6		
	D	90.9		
	K	89.8		
	G	88.7		
	A	87.7		
	Q	87		
	E	86.4		
	I	86		
	C	86		
	V	84.6		
	T	82.5		
	H	78.1		
	M	77.7		
16	S!	97.4	S	N
	A	89.3		
	V	87.7		
	C	87.4		
	G	84.5		
	D	82.6		
	N	82.3		
17	G!	95.2	G	P
	A	85.8		
	P	80.4		
18	V!	95.8	V	A
	A	93.7		
19	W!	90.3	W	K
	K	87.2		
20	Y!	96.4	Y	S
	G	90.3		
	A	89.5		
	L	89		
	F	87.9		
	Q	87.8		
	E	87		
	K	86.9		
	H	86.8		
	S	86.3		
31	G!	89	G	L
	A	74.7		
	L	57.1		

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), acid PLA_2 from *Bothrops diporus* (Bd), MT-II, a PLA_2-like protein, from *Bothrops brazili* (M2) and chosen residue (!).

3.7.3.1.2 Introducing mass spectrometry results as restriction

As a next step, we introduced the matched peptides of the PLA_2 sequences as restriction. BbTX-II matched peptides were removed from the generated restriction. Mass spectrometry results lacked definition for 18 residues (excluding 2 cysteines that compose a disulphide bridge) which are represented by black amino acids letters in **Figure 3.6**. Other six and three residues were found two and three possibilities coloured in beige in **Figure 3.6**, respectively. Introducing such restrictions reduced the 2420 calculations in FULL SEARCH to 475.

After *SEQSLIDER* improved the model evaluating these possibilities, we performed two rounds of refinement and manual building obtaining R of 21%. To study the sequence conservation among other snake venom PLA_2, we used this last obtained sequence that generated low divergence with crystallographic data to locate other variations in the PDB using *Protein BLAST* and *COBALT* (**Supplementary Figure 3.28**). Using as restriction the best 100 aligned sequences to our query, we rerun a last *SEQSLIDER* cycle that included 428 calculations.

3.7.3.1.3 Discussion of restricted evaluation of residues

As expected, the PCC calculated on the crystallographic data had a high correlation with the mass spectrometry data. Most of the residues having multiple possibilities from mass spectrometry results were resolved by the crystallographic data.

Few residues had more than one match in mass spectrometry results (**Supplementary Table 3.16**). Most of these, considering the restriction of mass spectrometry or phylogenetic analysis, were resolved by the data itself as a clear difference higher than 1.5% was seen. Again, using the phylogenetic restriction when no data of mass spectrometry was available, RSCC indicate best residue for most of the residue (**Supplementary Table 3.17**). The residues 13, 30, 61, 114, 116, and 117 are undetermined by the crystallographic data as there is no electron density for side chain electron atoms (**Supplementary Table 3.18**). Their side chain atoms were omitted. C-terminal residues together with calcium binding loop are known to be highly flexible as seen in test case and the uncertainty seen for those were already expected. In case of no clear evidence for a single residue, we chose to use X as unknown residue.

The opposite happened for residues 1 and 12, as the best calculated RSCC had a difference of at least 3% compared to the residue expected by mass spectrometry data. With a strong indication, we model the residue suggested by RSCC that fitted better the calculated electron density (**Table 3.6** and **Figure 3.8**). The mass spectrometry results could be impartial due to the absence of particular sequences in the protein databank. The phylogenetic analysis support our choice, as 14 and 44 different snake venom proteins had an asparagine and an isoleucine in position 1 and 12, respectively (**Table 3.6**).

After all the proper electron density analysis and RSCC calculations were done, we concluded the model with *BUSTER* cycles including TLS and anisotropic refinement. The final model reached R/R_{free} of 20.8/24.2% and molprobity score of 1.13, which is in the 100th best percentile of all structures elucidated from datasets of similar resolution (2.05 ± 0.25 Å). In the Ramachandran plot, 98.3% and 0.8% are in the favored and allowed regions, respectively.

Table 3.6 - SEQSLIDER summary of BbTX-III run with RSCC phylogenetic restriction of residues having divergent results with mass spectrometry results

R#	AA	CC	F	M
1	N!	97.4	14	S
	D	93.4	13	
	S	93.2	33	
	H	84.2	39	
12	I!	97.8	44	E
	V	94	52	
	L	91	2	
	A	83.2	1	
	G	79.1	1	
	E	68.7	0	

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). Residues numbers in bold represent side chain atoms were removed due to lack of electron density.

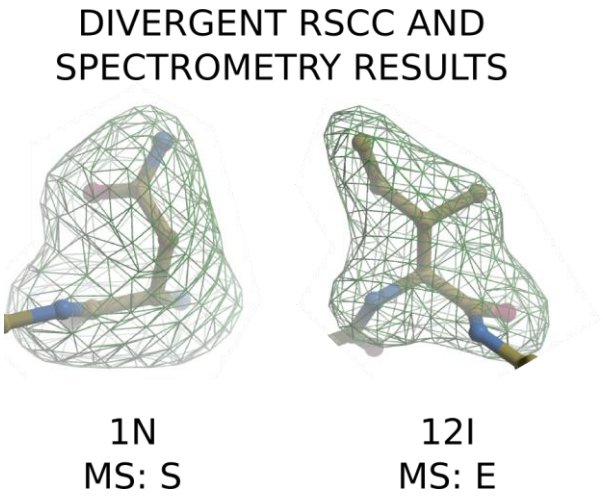


Figure 3.8 – Residue 1 and 12 and their polder map (3.5 σ) correspondent to Asn and Iso, respectively, different than mass spectrometry results.

3.7.3.2 Mt-I

The second unknown model which we submitted SEQSLIDER was Mt-I, also named Mt-III previously, another PLA_2, but purified from *Bothrops asper* venom. Dr. Juliana dos Santos, a former postdoc of our group, successfully phased the dataset using a homologous model with molecular replacement, but she did not manage to improve R factors. Moreover, she noticed a discrepancy between the side chain and the published sequence for this toxin (LIZANO; LAMBEAU; LAZDUNSKI, 2001) (sequence A in **Figure 3.9**). Differently than in BbTX-III, Mt-I contains a dimer in the ASU. If the differences in the sequence of present isoforms are in exposed residues or within dimeric interface, the thermodynamic of each interaction to the symmetry mates among the dimer contact may be different. Thus, a particular isoform could have a preference to compose one of the monomers in the dimer, which could be seen in comparison of side chain electron density of each monomer.

```
   1         11         21         31         41         51         60
A) SLIEFAKMILEETKRLPF-PYYTTYGCYCGWGGQGQPKDATDRCCFVHDCCYGKLSNCKPK
B) GHLLQFNKMIKFETNKNAIPFYAFYGCYCGWGGRGRPKDATDRCCFVHDCCYGKLPNCDTK
C) GSLVQFETLIMKIAGRSGVWYYGSYGCYCGSGGQGRPQDASDRCCFVHDCCYGKVTDCDPK
S) SL?QLGKMILQETGKNPA-KSYGAYGCNCG?GGKG?PKDATDRCCFVH!CCYA?L?GCDP?
MA)SLVELGKMILQETGKNPA-KSYGAYGCNCGVGGKGXPKDATDRCCFVHKCCYAkLTGCDPk
MB)SLvELGKMILQETGKNPA-KSYGAYGCNCGVGGKGXPKDATDRCCFVHDCCYAkLtGCDPk

  61         71         81         91        101        111        121
A1)TDRYSYSRKSGVIICGEGTPCEKQICECDKAAAVCFRENLRTYKKRYMAYPDLLCKKPAEKC
A2)-------------------------------------------------F------DP-
B) WDIYSYSLKSGFITCGGGTWCKKQICECDKAAAICLRENLDTYNKEYQFYGHFHCKEGPKKC
C) ADVYTYSEENGVVVCGGDDPCKKQICECDRVAATCFRDNKDTYDNKYWFFPAKNCQEESEPC
S) ???YSY?W??K?I?C?????CLK?LC?CDKA?AICLRENLNTYNK?YR?Y??P-C??????C
MA)ADRYTYSWEnKVIVCGGdNpCLKQLCECDKAVAICLRENLNTYNKKYRYYLAPKCALXX-a
MB)aDVYtYSWEnKVIVCGEDnpCLKQLCECDKAVAICLRENLNTYNKSYRYYLGPKCKAXX-Ac

A: Mt-III from Bothrops asper (UnitProt: P20474)
B: PLA2 from Sistrurus catenatus catenatus (UniProt: B0LSF5)
C: PLA2 from Bothrops diporus (UniProt: I2DAL4)
S: Seed, consensus in PCC
MA:final sequence of monomer A of elucidated model
MB:final sequence of monomer B of elucidated model
```

Figure 3.9 - Sequence similarity of Mt-I against PLA_2s.
Matching amino acids from peptide mass spectrometry analysis in bold and blue. Residues in grey and black where no match was found in mass spectrometry. The residues whose side chain atoms were omitted from model are shown in lowercase.

We sequenced Mt-I using mass spectrometry from LNBio facilities (Campinas – Brazil). Evaluating peptide signals against NCBI databank filtered in Phylum Chordata, we

found matches for 2 different proteins in this sample, with a sequence coverage of 46 residues out of 121 (almost 40%) (sequence B and C in **Figure 3.9**). We tried to extract the most of this crystallographic dataset using *SEQSLIDER* given restriction of partial mass spectrometry data and phylogenetic analysis in order to elucidate its structure.

3.7.3.2.1 Full search mode

For Mt-I dataset, we performed same strategy as previous, joining rounds of SEQSLIDER in full search mode, manual modeling and *BUSTER* refinement, but starting from the model given by Dr. dos Santos. Considering the two chains, we calculated the RSCC for each chain independently and together to evaluate convergence or possibility of heterogeneity within the crystal molecules. Therefore, *SEQSLIDER* summed 7260 calculations.

Out of the 121 residues, 65 residues had clear distinction in both chains independently or not (**Supplementary Table 3.19**). The residues 65, 68, 71, 73, 82, and 85 diverge to the mass spectrometry data. We used these 65 residues to predict the local variability within observed snake venom PLA_2s and PLA_2-like sequences (S in green in **Figure 3.9**).

3.7.3.2.2 Results and discussion of restricted search mode

The number of calculations given the restriction of phylogenetic analysis decreased to 2172 (**Supplementary Figure 3.29**). Essential to the interpretation of the *SEQSLIDER* run is categorize the residues in whose: i) mass spectrometry and/or phylogenetic analysis converge with RSCC calculations; ii) results diverge; iii) RSCC do not indicate a single answer, i.e., undetermined; and last; iv) RSCC calculate from electron density of chain A and B are different.

Out of the 121 residues, 86 of the residues composed the first group in which results of different techniques corroborate each other (**Supplementary Table 3.20**). Seven residues composed second group, mass spectrometry data indicated a different residue than the one observed in the polder map (**Table 3.7**). Residues 52, 65, 73, and 85 had residues with difference of one atom, whether longer side chains fitted better the polder map for 68, 71, and 82 (few examples in **Figure 3.10**). When a strong evidence RSCC higher than 90%, we chose the residue indicate by this statistic. This was not the case for 65, the PCC difference between

best score and the one indicate by mass spectrometry was only 1.6%, close to our chosen distinction value. For this residue, we modelled Thr suggested by mass spectrometry in chain A, but omitted its side chain atoms in chain B, as an alanine has a high RSCC of 93.2 against 81.3 for threonine.

Two residues with no determination by mass spectrometry had mutual presence of E and Q in aligned sequences and PCC did not differentiate them (**Table 3.8**). For those cases, we chose the one most frequent in sequences, i.e., E and Q for 4 and 11, respectively. Residue 109 had Y and F best scored for chain A and B, respectively. Based on the hydrogen pattern of the Oηs (side chain O) between the Y119A and B (**Figure 3.11**), we model it.

For 13 residues, the RSCC did not indicate a definitive possibility, as no values higher than 90% were obtained (**Table 3.9**). Three of those (35, 119, and 120) were constituent of C-terminal and calcium binding loop, regions known to be highly flexible. We chose to model them as alanines and place a X letter as unknown. The other residues were exposed to the solvent and related to the loops that connect the two anti-parallel β-sheets to the α-helices (**Figure 3.12**). For the residues 53, 69, 70, 78, 79, 80, and 84 that had a mass spectrometry matched a single possibility, we chose to omit its side chain atoms but keep the residue type suggested by mass spectrometry. The residue 79 and the chain B of 78 had a weak density corroborating mass spectrometry result, therefore we model their side chain but place a flag to refine their occupancy in *BUSTER*. The chain A of residue 55 had a generally better RSCC value for Thr that are seen in most of aligned sequences (79), therefore we chose to model its side chain for chain A and omit for B. We performed same approach in both chains of residue 58 modelling Asp. For the last residue 60, we omitted side chain atoms of a Lys, which is the most frequent residue in the aligned sequences.

At last, 9 residues had different RSCC results for each chain (**Table 3.10**) and few of them are shown in **Figure 3.13**. The residue 3 has poor electron density for chain B, therefore we omitted its side chain, keeping Valine as it constituency (**Figure 3.13**). The residue 48 (numbered as 49 following Renetser *et al* (1985)), which is one of the most important residue to differentiate catalytic and non-catalytic PLA_2, are different. Based on the PCC, we modelled Lys in chain A and Asp binding to a calcium ion chain B (**Figure 3.13**), being the latter supported by mass spectrometry results. We modelled Arg, that had a good score, on residue 63 in chain A, while for chain B we chose a Val suggested by mass spectrometry scored second best, both cases are in agreement with polder maps (**Figure 3.13**). The residues

77, 106, 113 and 117 had good scores for larger amino acid for one of the chains, while for the other chain we had to model Ala or Gly due to poor electron density.

Table 3.7 - SEQSLIDER summary of Mt-I run with RSCC residues diverging mass spectrometry match or phylogenetic analysis

R#	AA	CCA	CCB	CCAB	F	MS
52	**A!**	**91.2**	**70.8**	**80.1**	15	G
	G	81.9	65.1	71.8	1	
	K	68.5	69.1	71	78	
	D	63.9	62.8	62.4	4	
	E	60.3	53.3	59.7	1	
65	S	85.9	87.9	87.4	79	T
	T	**83.5!**	81.3!*	81.3	2	
	K		72.8	63.6	4	
	E		57.1	47.1	15	
	A	77.4	**93.2**	87.2	0	
68	**W!**	**95.9**	**95.1**	**95.4**	80	E
	F	85.8	83.7	84.4	1	
	Y	85.7	82.7	83	1	
	E	81.7	82.9	82.2	1	
	R	78.6	84.5	80.2	16	
	L	76.2	74.5	75.1	1	

R#	AA	CCA	CCB	CCAB	F	MS
71	**K!**	**94.8**	**92.9**	**93.6**	78	G
	G	86.7	82.9	80.7	22	
73	**I!**	**95.6**	**96.1**	**95.6**	99	V
	L	85.4	88.7	86.9	1	
	V	89.8	91.9	90.4	0	
82	**L!**	**90.9**	**91.8**	**91**	71	K
	T	77	86.7	82.8	2	
	P	78.5	82.1	81.3	1	
	K	75.5	73.5	75.3	23	
	M	64	69	67.4	3	
85	**L!**	**95.9**	**94.6**	**94.9**	33	I
	I	87.9	86	86.4	21	
	V	86.4	85.7	85.8	10	
	E	86.7	78.9	82.2	1	
	M	78.6	77.5	77.6	35	

The abbreviations represent Residue number (R#), amino acid (AA), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). The real space correlation coefficient of polder map (CC in %) of chain A, of chain B and of both chains are abbreviated as CA, CB and CCAB, respectively. * Residues with side chain omitted. In bold, best CC.

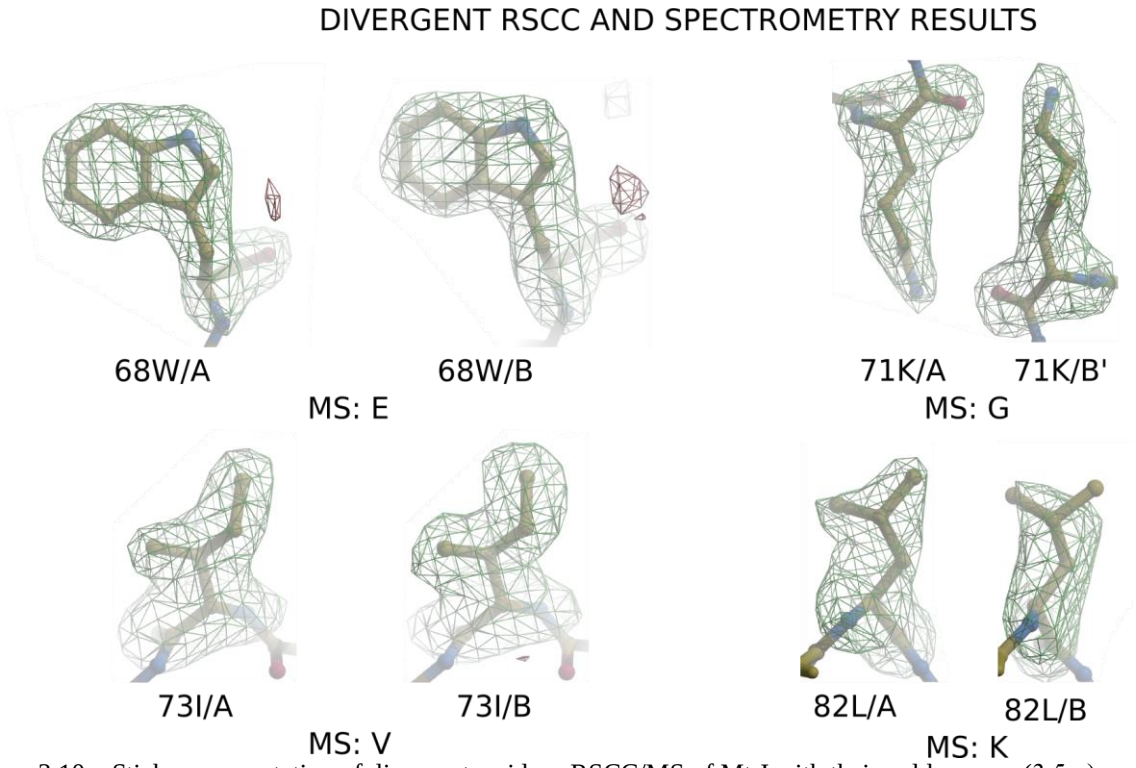


Figure 3.10 – Sticks representation of divergent residues RSCC/MS of Mt-I with their polder maps (3.5 σ).

Table 3.8 - SEQSLIDER summary of Mt-I run with ambiguous RSCC possibilities

R#	A	CC	CC	CCA	F
4	Q	**95.1**	**91.9**	**92.3**	1
	E!	92.3	88.5	90.2	8
11	E	91.3	93.2	**91.7**	1
	Q!	**93.7**	**94**	89.9	8
	L	86.8	82.4	84.4	3

R#	A	CC	CC	CCA	F
109	**Y!**	**93.5**	93.2	93.5	2
	F	89.3	**95.2**	92.8	4
	R	88.3	84.1	86.7	1
	K	86.2	86.4	86.3	1
	N	80.4	85	83	3
	L	82.4	81.9	82	1
	A	82.1	75	78.6	1
	I	73.1	72.7	73.8	2
	V	71.7	70	71.4	2
	P	61.4	60.2	60.8	3

The abbreviations represent Residue number (R#), amino acid (AA), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). The real space correlation coefficient of polder map (CC in %) of chain A, of chain B and of both chains are abbreviated as CA, CB and CCAB, respectively. * Residues with side chain omitted. In bold, best CC.

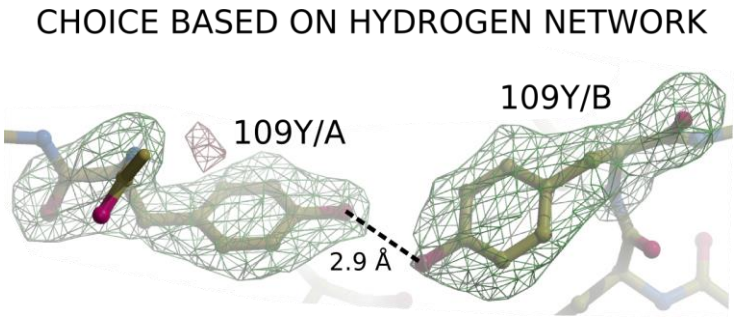


Figure 3.11 – Sticks representation of 109Y of Mt-I and its hydrogen network with its polder maps (3.5 σ).

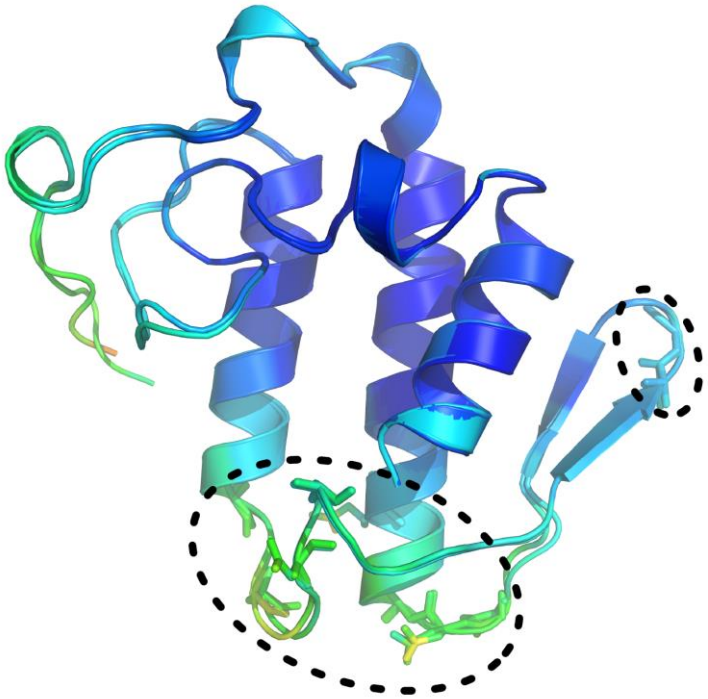


Figure 3.12 – Cartoon representation of Mt-I with uncertain side chain in sticks and circled in dashes. Blue and light green are related to low and high Cα B factors, respectively.

Table 3.9 - SEQSLIDER summary of Mt-I run of residues undetermined

R#	AA	CCA	CCB	CCAB	F	MS
35	G	91.2	89.8	88.6	2	
	T	74.4	84.7	80.9	1	
	P	74.6	83.7	78.5	1	
	K	78.4	83.4	78.4	65	
	R	63.5	77.4	68.7	3	
	Q	62.2	71.4	66.9	24	
	E	59.6	69.4	64.4	3	
	A!	**93**	**94**	**92.1**	0	
53	A	**84.4**	**77.9**	**79.7**	1	K
	K!	64.4*	67.2*	65.9	94	
	T	70.5	56	63	3	
	N	54.3	49.1	53.2	1	
55	T!	86.1	62.4*	74.8	79	
	S	**88.6**	64.9	74.5	14	
	P	72.3	66.5	68.5	5	
	N	79.1	41.9	62	1	
	G	73	**83.8**	74.9	0	
	A	80.2	76.8	**78**	0	
58	S	78.2	77.7	77.6	21	
	D!	84.4	68.8	74.4	45	
	N	79.3	64.5	72.5	33	
	G	**84.4**	**88.2**	**86.7**	0	
	A	80	82.9	81.9	0	
60	I	76.6	80.6	78.8	9	
	L	76.5	74	74.8	1	
	K!	71.5*	71*	72	88	
	Q	62.4	60	62.6	1	
	A	**90.9**	**91.6**	**91.3**	0	
69	V	84.1	84.8	84.7	9	
	K	68	82.2	77.1	58	
	E!	59.6*	73.5*	66.8	32	E
	A	**94.2**	**94.9**	**95.1**	0	
70	N!	64.3*	69.5*	67.3	65	
	D	61.5	67.3	64	34	N
	A	**92.6**	**92.3**	**92.2**	20	

R#	AA	CCA	CCB	CCAB	F	MS
78	**G**	**49.3**	**85.6**	**76.1**	**0**	
	N		82.1	63.2	37	
	S	34.5	80.2	63.1	19	
	D!	*	80.3	61.5	10	D
	E	50.8		35.6	2	
	K		38.8		31	
79	N	**81.7**	**89**	**82.7**	69	
	D!	76	79.4	75.6	7	
	T	61.1	70.6	64	19	D
	K	41.6	67.2	50.7	1	
	Q	26.2	63.8	41.7	3	
80	P!	79.4*	66.9*	73.6	75	
	S	79.1	63	72.4	23	P
	F	54.7	46.3	49.7	1	
	A	**82.9**	84.5	**84.7**	0	
	G	74.6	**86.9**	81.5	0	
84	S	87.1	**92.7**	**89**	0	
	E	86.9	71.8	78.6	68	
	R	74.9	68.6	72.4	4	Q
	Q!	71.3	71.1	72.2	27	
	A	**90.4**	86.6	87.6	0	
119#	S	58.9	63.4	58	3	
	D	48.5	61	49.8	55	
	E	40.7	62	45.5	39	
	K	30.2	**69**	42.3	1	
	R	22.6	63.9	37.8	1	
	G	**79.7**	64.3	**70.3**	0	
	A	76.6	53.3	63.8	0	
120#	A!	61.4*	**76.2**	63.3	11	
	P	51.2	52.7	48.6	28	
	S	26.1	50.2	37.9	5	
	R	24.4	49.1	36.9	1	
	K		43.3	30.8	21	
	T		43.1	30.2	20	
	D	20.2	35.1	28.6	4	
	Q	18.2	37	28.3	7	
	E	17.6	34.4	26.3	3	
	G	**69.9**	75.3	**68.5**	0	

The abbreviations represent Residue number (R#), amino acid (AA), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). The real space correlation coefficient of polder map (CC in %) of chain A, of chain B and of both chains are abbreviated as CA, CB and CCAB, respectively. * Residues with side chain omitted. # Residues that were given X as unknown. In bold, best CC.

Table 3.10 - SEQSLIDER summary of Mt-I residues whose RSCC of chains diverge

R#	AA	CCA	CCB	CCAB	F	MS
3	V!	**92.2**	77.9	84.4	13	
	I	87.7	73.4	81.1	38	
	L	82.1	76.2	79.3	16	
	H	79.7	69.4	76.5	1	
	F	83	68.7	75.9	24	
	Y	76.2	63.1	72.2	4	
	W	65.3	58.6	60.6	2	
	G	88	**92.6***	**88.8**	0	
48	S	86.3	91.4	**88.5**	19	D
	K	**94.6!**	75.4	85.9	73	
	D	73.4	**92.9!**	85.3	2	
	R	87.1	65.8	76.9	5	
63	S	**87.4**	**87.5**	**87**	9	V
	V	85.5	86.3!	86.3	0	
	G	86.9	77.9	83.1	2	
	Q	82.7	68	76.3	3	
	F	81.4	68.4	75.7	1	
	R	87.1!	63.2	71.9	84	
77	G	**72.9!**	86.5	**81.6**	44	G
	E	58.3	**92.2!**	75	68	
	N	63.8	80.8	74	4	
	D	63.7	76.3	71	2	
	S	49.1	86.4	71	15	
	K	47.6	82.6	69.7	4	
	V	54.2	79.4	69.1	2	

R#	AA	CCA	CCB	CCAB	F	MS
106	S	82.7	**89.7!**	**85.4**	33	
	K	**87!**	82.4	84.6	87	
	L	75.6	81.3	79.2	1	
	N	76.2	74.8	75.3	7	
	E	76.7	67.3	73.5	1	
111	L!	86	**95.2**	**90**	52	
	T	87	86.4	85.8	3	
	G	85.5	90.2	85.1	1	
	P	**91.8**	80.9	85.1	17	
	V	84.7	85.9	84.2	4	
	K	77.3	77.9	76.8	1	
	F	64.8	75.5	71.1	15	
	M	61.5	78.7	68.3	6	
112	**G**	84.8	**95.6**	**88.8**	11	
	A	**90.6**	89.7	87.7	11	
	K	73.2	69	69.3	78	
	N	77.7	60.5	65.8	4	
	D	74.8	52.1	62.4	1	
	R	56.2	39	44.8	14	
116	**A**	**93.6**	88.3	**90.9**	0	
	K	71.4	**92.8**	78.2	93	
	D	61.7	79	68.4	3	
	R	59.7	72.4	62.7	3	
117	**A**	77.7	**77.4**	**76.6**	55	
	G	78.2	76	75.1	1	
	V	75.5	69.2	71.6	15	
	T	75.7	63.8	66.9	11	
	L	**82.7**	58.5	65	11	
	D	63.6	58.1	57.8	2	
	K	57.1	66	57.7	60	
	E	56.8	52.8	48.7	3	

The abbreviations represent Residue number (R#), amino acid (AA), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). The real space correlation coefficient of polder map (CC in %) of chain A, of chain B and of both chains are abbreviated as CA, CB and CCAB, respectively. * Residues with side chain omitted. In bold, best CC.

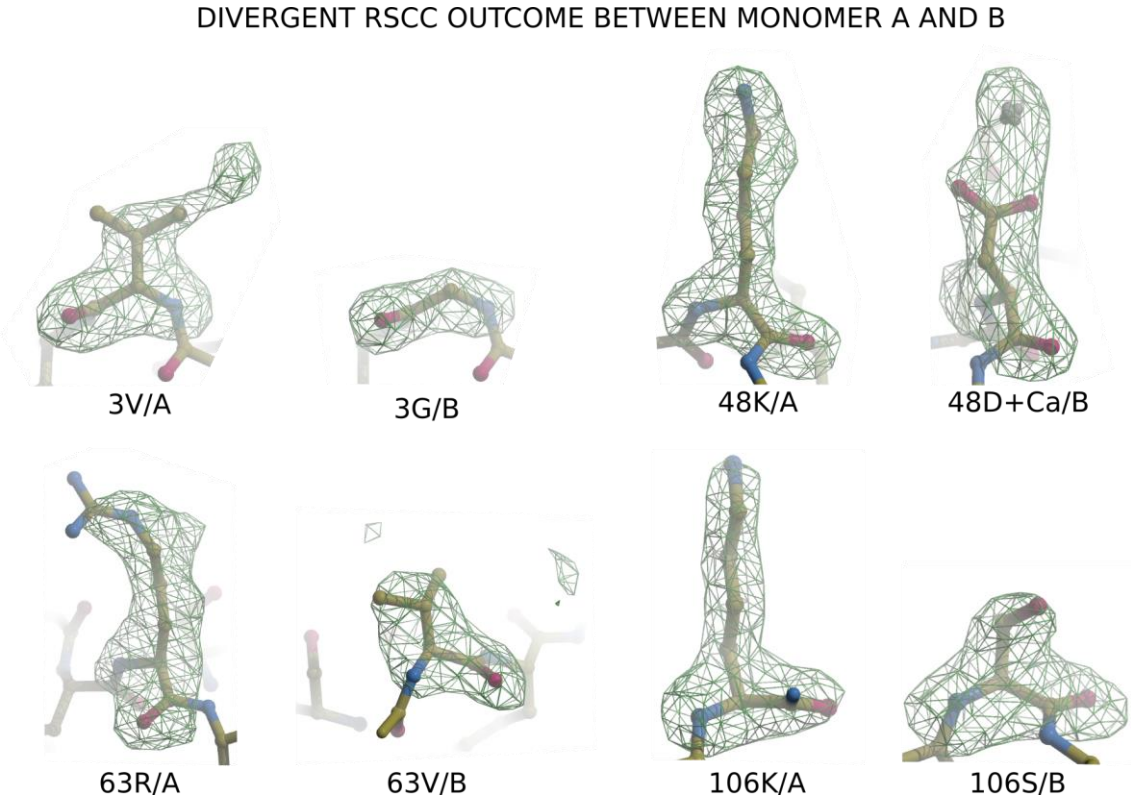


Figure 3.13 - Divergent RSCC between chain A/B of Mt-I and their respective polder maps (3.5 σ).

With the SEQSLIDER restricted search by mass spectrometry and phylogenetic analysis, we could evaluations various side chains and find some differences between the two chains and modelled most of the residues. With such approach, we could refine the model to R/R_{free} of 19.0/24.7 and a molprobity score of 1.86, which is in the 82th best percentile of all structures elucidated from datasets of similar resolution (1.99 ± 0.25 Å). In the Ramachandran plot, 96.6% and 3.0% are in the favored and allowed regions, respectively.

3.8 SEQSLIDER in ARCIMBOLDO scope

As the resolution deviates more from the ideal 2 Å required by the *ARCIMBOLDO* method, density modification and autotracing of partial solutions become less effective. One of the ways to overcome the resolution limitation is to improve the starting map prior to density modification. Modeling and the use of refined maps as the start of density modification have been implemented in the program *SEQSLIDER* (**Figure 3.14**). After a preliminary structure solution with *ARCIMBOLDO* (RODRÍGUEZ et al., 2012, 2009), often only single non-connected fragments like α-helices and β-strands build up the starting structure for refinement. At resolutions between 2 and 3 Å, side chains cannot be identified in

the initial, noisy electron density maps so that it is often unclear how these main chain fragments are connected or to which part of the sequence the fragments correspond (**Figure 3.15**). Starting from a partial *ARCIMBOLDO* solution that usually is composed of polyalanine fragments, SEQSLIDER generates different hypotheses trialing side chain composition and sends to refinement, validation and expansion. This methodology has been successfully used for the solution of two structures, MltC (PDB code: 4C5F) at a resolution of 2.7Å composed by two monomers in the ASU (with 341 residues each) (ARTOLA-RECOLONS et al., 2014); and a complex of DNA (12 nucleotides) and a protein (82 residues) at a resolution of 2.8 Å.

3.8.1 *SEQUENCE SLIDER* program

Building on the experience of a previous prototype made by a former postdoc student of Dr. Usón, Dr. Kathrin Meindl, we have rewritten the *script* of *SEQSLIDER* from scratch following the flow diagram shown in **Figure 3.14**. To launch SEQSLIDER, a single input is needed, an instruction file in *ARCIMBOLDO* usual format. It requires the paths to programs used, the path of the coordinate file, secondary structure prediction (*PSIPRED* output format) file, reflection file in both mtz and hkl formats. *SEQSLIDER* matches the partial model to the predicted secondary structure elements by *PSIPRED* by sliding a frame of sequence over the fragments. *SEQSLIDER* creates several hypotheses written in PDB files (coordinate files), which are scored based on an alignment scoring matrix. The hypotheses with higher scores are evaluated through the modeling of these side chains in the fragments with *SCWRL4* and/or *COOT* and through refinement with *REFMAC*, *BUSTER* or *PHENIX.REFINE*. The different models are scored based on global and local crystallography statistics and submitted to cycles of density modification *SHELXE*. The increase in the CC and in the number of residues traced are an indication of improvement and may recycled within *SEQSLIDER* until the structure is unequivocally solved, shown by an interpretable electron density map and a mainchain trace characterized by a CC over 30%.

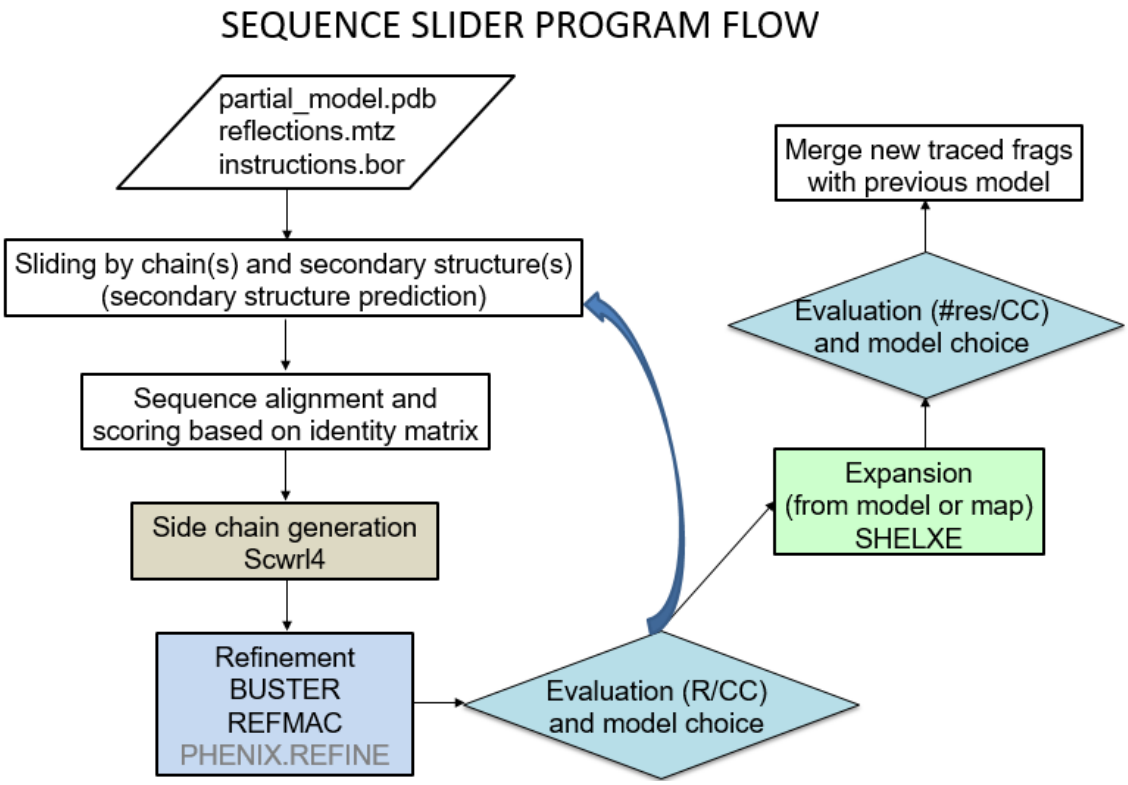


Figure 3.14 - Program flow of SEQSLIDER. Colored background indicates the use of external programs.

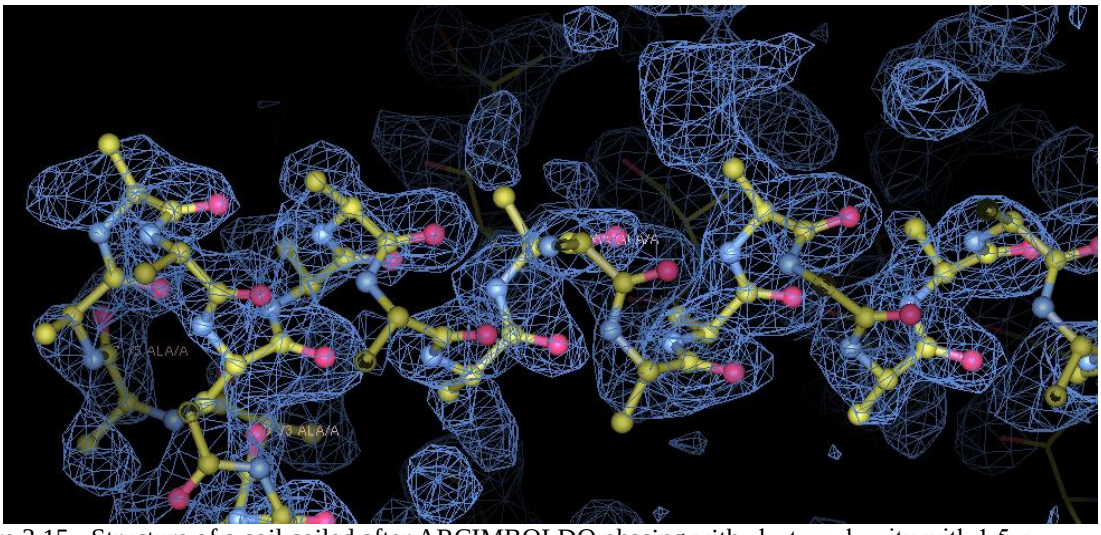


Figure 3.15 - Structure of a coil coiled after ARCIMBOLDO phasing with electron density with 1.5 σ. Most of the side chains cannot be identified in this initial and noisy electron density map.

It should be remarked that while for the use in natural products, the possible sequence assignments are slid over each residue, here a matching sequence is slid over a continuous stretch of main chain trace.

First, *SEQSLIDER* separates the full sequence into smaller pieces classified by their secondary structure prediction (*PSIPRED* file), α-helices, β-strands, or coil. For each fragment available in the partial trace (coordinate file), *SEQSLIDER* reads its secondary structure assignment by *BORGES_MATRIX* and generates all the possible hypotheses within the sequences in the corresponding group of secondary structure prediction. By default, coil fragments and *PSIPRED* confidence level are ignored. If a fragment is large and contains more than one secondary structure element, *SEQSLIDER* generates hypotheses independently for each element and the ones impossible will be filtered by the alignment scoring matrix.

The purpose of the alignment is to score the hypotheses where a higher number of residues are introduced into each chain and, if they are composed of more than one element, what is their expected distance in relationship to the true sequence. This is accomplished by aligning each hypothesis against the full sequence using a scoring matrix that gives a higher weight to assigned residues than the ones not assigned and a low penalty for opening and extending gaps (**Table 3.11**). When we refer to assigned residues, we mean the ones in which side chain will be modeled, that in the default configuration are either α-helix or β-strand residues. This matrix is based on the identity of residues aligned, i.e., a positive value is given if a pair of residues is identical respect to their secondary structure prediction and assignment, if not, a negative value is given (**Table 3.11**). The only exception is when an alanine of the known sequence matches an unassigned alanine, then a neutral value (0) is given, as this is the original composition of polyalanine fragments and we do not want it to drive the alignment. For the local alignment, we used the function *pairwise2.align.locald*, which is based on Smith-Waterman algorithm (dynamic programming) and available in the library *biopython*.

As an example of a coil coiled dataset, its predicted secondary structure (SS) based on the known sequence is composed of three helices (H1, H2, and H3 composed of 26, 32 and 29 residues long, respectively) connected by 3-4 residues loop (**Figure 3.16 A**). *ARCIMBOLDO* successfully obtained a partial model. The SS of this model is composed of a 3-residue (3res) loop, a helix of 27 residues (Ha), connecting 3res loop, a 32res helix (Hb) and a 1res loop (**Figure 3.16 B**). Intuitively, matching the SS of predicted sequence and of the partial model fragments, the Ha and Hb should be either H1 and H2 or H2 and H3. This is exactly what is obtained by the scoring matrix. The hypothesis which shows better agreement between predicted SS and the model SS is matching the last 27 residues out of H2 sequence to Ha and the beginning of H3 to the 11 residues of Hb (Hypothesis 1 in **Figure 3.16 C**). Another good

hypothesis is matching the whole H1 sequence to Ha fragment and the beginning of H2 to Hb (Hypothesis 2 in **Figure 3.16 C**). Increasing the distance between H1-Ha and H2-Hb (hypothesis 3) or H2-Ha and H3-Hb (hypothesis 4) reduces the score as number of assigned side chains decrease and gap penalty increases. An unlikely hypothesis is H1 matching Ha and H3-Hb, the big distance between H1 and H3 are heavily penalized, giving a low score. Therefore, by such score, it is possible to prioritize probable hypotheses and the ones with high number of assigned residues.

SEQSLIDER is run given an instruction file in the same manner as other programs developed by the group of Dra Usón. The essential files are given, such as reflection files in both MTZ and HKL format, coordinate file in PDB format and the prediction file *PSIPRED* format. The model SS is evaluated by *BORGES_MATRIX* (SAMMITO et al., 2013), that classifies α-helices, β-strands and loops based on the characteristic vectors of tripeptides and their geometrical characterization. All the internal *SEQSLIDER* configuration may be set overriding the default values provided by the program. The minimum size for evaluated fragment is 5 residues. In SEQSLIDER for phasing, we define fragment as a set of coordinates of residues, continuous or not sharing the same secondary structure. *PSIPRED* confidence level is set to 0. A tolerance of 2 is set to allow errors in the SS prediction and/or in model coordinates. The 100 best hypotheses from the alignment scoring matrix of each available chain are evaluated by refinement and autotracing independently. This value 100 may be increased automatically if the 100^{th} hypothesis has equal score as the following ones on the list. These are scored according to a Figure of Merit (**Equation 4**) using both local (RSCC) and global statistics ($R_{factors}$). The RSCC are calculated only for the evaluated chain for both main (CC_{mc}) and side chain atoms (CC_{sc}). Afterwards, the best 10 hypotheses of each chain are combined together in pairs, until all fragments with SS were evaluated.

$$FOM = ((R_{work}max - R_{work}) + (R_{free}max - R_{free}) + (CC_{mc} – CC_{mc}min) + (CC_{sc} – CC_{sc}min))*10 \quad (4)$$

Table 3.11 - SEQSLIDER scoring function of the local alignment

Category	Score
Match of assigned residues	6
Mismatch of assigned residues	-4
Match of unassigned residues	4
Mismatch of unassigned residues	-1
Mismatch of unassigned alanine	0
Mismatch of unassigned alanine	0

A)

SECONDARY STRUCTURE PREDICTION

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
CCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHH
C4 H26 C4 H32 C5 H19
```

B)

SECONDARY STRUCTURE COMPOSITION FROM PARCIAL MODEL

C3 H27 C3 H11 C1

C)

HYPOTHESIS 1)

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
                                 ||||||||||||||||||||||||||||||||||||||||||||
-----------------------------------aaaQQIAAVRGASNGLMSEMVEIHLKDELVaaa--PDQRAVRMAEIa-----
```

Score=222

HYPOTHESIS 2)

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
    ||||||||||||||||||||||||||||||||||||||
aaaaPEDKKRILTRVRRIRGQVEALERALEaaa-CLAILQQIAAVa-----------------------------------------
```

Score=217

HYPOTHESIS 3)

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
aaaaPEDKKRILTRVRRIRGQVEALERALE------------aaaRGASNGLMSEMa-------------------------------
```

Score=208

HYPOTHESIS 4)

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
                                  ||||||||||||||||||||||||||||||||||||||||||||||
------------------------------aaaCLAILQQIAAVRGASNGLMSEMVEIHLaaaaa-----PDQRAVRMAa-------
```

Score=205

HYPOTHESIS 5)

```
MPHSPEDKKRILTRVRRIRGQVEALERALESGEPCLAILQQIAAVRGASNGLMSEMVEIHLKDELVSGETTPDQRAVRMAEIGHLLRA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
----aaaKKRILTRVRRIRGQVEALERALE------------aaaaaaa----------------------PDQRAVRMaaaaaa---
```

Score=146

Figure 3.16 - Alignments of predicted secondary structure elements of known sequence (in **A**) against structure fragments from chain in model (in **B**).
Scoring is calculated from our scoring matrix using *pairwise2.align.locald* function available in the library *biopython*. The assigned residues are represented by uppercase letters and non-assigned, by lowercase. The residues colored in green refer to α-helices. In **A** and **B**, letter C and H refer to coil and helix, respectively, and the following number refer to the fragment size.

As proof of principle, a few structures were solved manually by a strategy similar to our algorithm SEQSLIDER, such as MltC (2.7 Å resolution), a protein-DNA complex (2.8 Å resolution) and few coiled coils. Since we aim to distribute SEQSLIDER in phasing to the crystallographic community, we will evaluate recent version and implementations to probe its efficiency. As a *post mortem* analysis, we evaluate the correctness between hypotheses based on the correspondence matching assigned residues in a partial trace to the true residue in the final model. This is accomplished with the *Biopython NeighborSearch*.

3.8.2 MltC

SEQSLIDER originated when phasing the MltC structure (PDB id: 4C5F) by Dr. Kathrin Mendl. The initial available dataset for this protein diffracted to 2.7 Å resolution in $P2_1$ space group with a dimer in its ASU. This was partially phased through molecular replacement with *PHASER* using a common domain of an homologous protein, the MltE (PDB id: 2Y8P), whose identity to the C-terminal domains in MltC is 40%. This C-terminal domain represents 54% of the complete MltC structure and its geometry is similar (0.8 Å of rmsd) (**Figure 3.17**). A clear solution was found in *PHASER* with a Z score of 7.2.

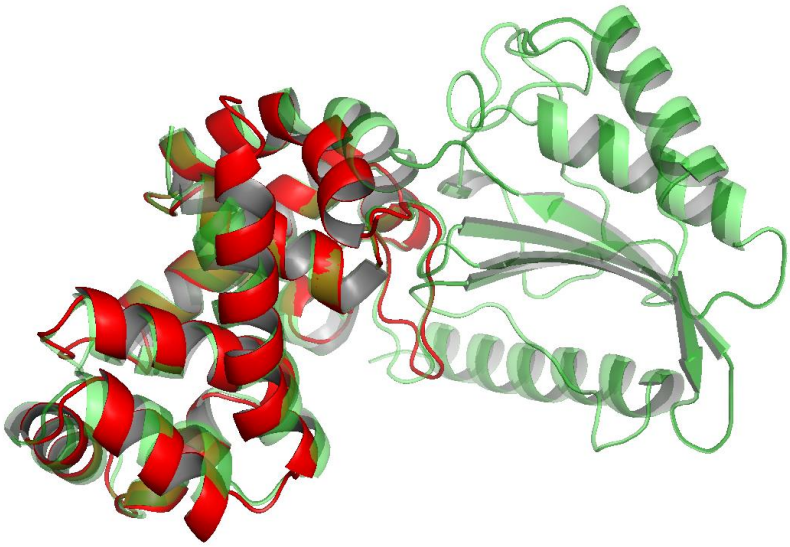


Figure 3.17 - Cartoon representation of MltE (in red PDB id: 2Y8P) and of MltC (transparent in green PDB id: 4C5F).
The structures were superposed using the common C-terminal domain.

The electron density map of the partial solution did not show any additional features in regions where atoms were absent. Preliminary efforts to autotrace missing parts using

SHELXE failed. In all the following expansions, whose results are summarized in **Table 3.12**, we used the *SHELXE* parameterization line: -a5 -t20 -s0.4 -e2.1 -q -m5 -x3. The letters "a" stands for number of autotracing cycles, "t" for time factor spent in probing seeds, "s" for solvent content, "e" for resolution limit of data extrapolation, "q" for helix search, "m" for number of density modification cycles, and "x" to activate *post mortem* analysis. Starting from a 358 residues model with CC_i (initial CC) of 7.03% and $wMPE_i$ (initial weighted mean phase error) of 67.7º, the best trace had only 243 residues, a CC_f (final CC) of 14.86% and $wMPE_f$ (final weighted mean phase error) 76.2º. In a different trial, we performed 100 cycles of BUSTER of the initial model, which gave *SHELXE* CC_i of 31.47%, although with no improvement in wMPE or in trace. Modeling the correct MltC side chains with *SCWRL4* and pushing 100 cycles *BUSTER* refinement improves significantly the model to a CC_i of 43.25% and wMPE of 47.4º, ~20º better than previous models. Using such coordinates as input and the -k option (keeps input coordinates), autotracing finds in the first cycle 36 new residues that increase CC_f to 43.95% and decrease a little bit $wMPE_f$ to 49.4º. The new coordinates are divided in 4 chains, in which 95% are correct. Removing initial fragment in subsequent traces (removing -k) finds 301 residues with a CC_f of 23.12% and $wMPE_f$ 60.9%, corresponding to 90% of correctness. If the BUSTER map is used to provide initial phases instead of deriving them from the refined atomic model, *SHELXE* traces 346 residues divided in 7 chains, including 55 that were not present before. Its CC_f is 27.43% and $wMPE_f$ 56.7. Probably, the wMPE decrease is related to the incorrectness of 26% of the new residues. As *SHELXE* traces polyala residues, it is expected that CC calculated from a model containing side chain atoms are higher than when it is not. Therefore, comparison of CC should be done in two groups, first, the ones containing side chain atoms, i.e. CC_i and CC_f when -k is set; and second, CC_f without -k option.

Table 3.12 - Different strategies to improve initial phases from MltC partial solution and *SHELXE* autotracing.

Strategy	initial		Retaining initial coordinates				Default autotracing			
	CC	wMPE (º)	CC	wMPE (º)	#res	Incorrect (%)	CC	wMPE (º)	#res	Incorrect (%)
After MR	7.0	67.7	19.2	72.4	33	100	18.0	73.9	306	39.1
Refinement	31.5	67.3	-	-	-	-	21.5	71.3	295	25.6
SC modeling + refinement (model)	43.3	47.7	44.0	49.4	36	5	27.0	60.2	337	6.8
SC modeling + refinement (map)	-	46.8	-	-	-	-	27.0	56.3	345	2.9

The abbreviations CC, wMPE and res stands for Correlation Coefficient, weighted mean phase error and residues respectively. The dashes are shown when no values were produced by *SHELXE*. The incorrectness is in respect to the traced residues.

Considering the CC_fs and the number of traced residues, we used both the outputs of the different autotracing of improved SC and refinement. We evaluated convergence of the traced fragments and merged the longer ones. In total, we introduced 42 residues divided in 5 different chains (Structure 0 in **Figure 3.18** and expansion 0 in **Figure 3.19**) into the previous model. New fragments D and J are composed of β-strand(s) of 4 residues long. The other chains E, K, and N are α-helices composed of 5, 9, and 6 residues long, respectively. We set *REFMAC5* to be used within the *SEQSLIDER* run to evaluate hypotheses only these 5 fragments, as the side chains of the C-terminal domain were introduced in previous step.

We organized the results in plots of RSCC and R factors of each evaluated chain. We show chain E with each hypothesis represented by a point in **Figure 3.20**. Evaluating these factors independently, we did not find a consensus on the goodness of the different values to the correct sequence (100% identity). For chain E, for example, the goodness of R and CC_{sc} are correlated to the correct hypothesis, but this is not always the case for the other chains. Therefore, trying different evaluation, we, empirically, chose a FOM (**Equation 4**) that consider all these 4 figures together and better distinguishes correct from false. In **Figure 3.21** we plot the FOM of each hypothesis of all evaluated chains. For chain E and chain K, we did not obtain the correct sequence, probably the number of generated hypotheses would have to be increased, although at least 50% identity was obtained. The objective of our FOM is to prioritize hypotheses that explains better the data, which is measured by the local and global statistics. Differently than the objective SEQSLIDER in venoms targeting the most appropriate sequence, herein the focus is phasing, more specifically, improve the *SHELXE* autotracing by introducing most likely information of side chains. As the SEQSLIDER sends various hypotheses for expansion, the correct side chains will be probably used and the improvement in map may aid phasing.

The FOM of the corrected hypothesis for each chain had good scores (**Figure 3.21**), classifying them at least in the best 10. Chain J was the only exception. Its correct hypothesis had only two residues modeled, as the SS prediction based on the known sequence had the sheet predict smaller than the true structure. The errors in SS prediction are already expected by SEQSLIDER and accounted for by the variable “sliding tolerance” in the instruction file. Specifically, for chain J, the correct hypothesis was tried, but only two residues were assigned, such small differences among hypotheses were not enough for discrimination. Therefore, small number of assigned residues should not be tried, as they offer no distinction.

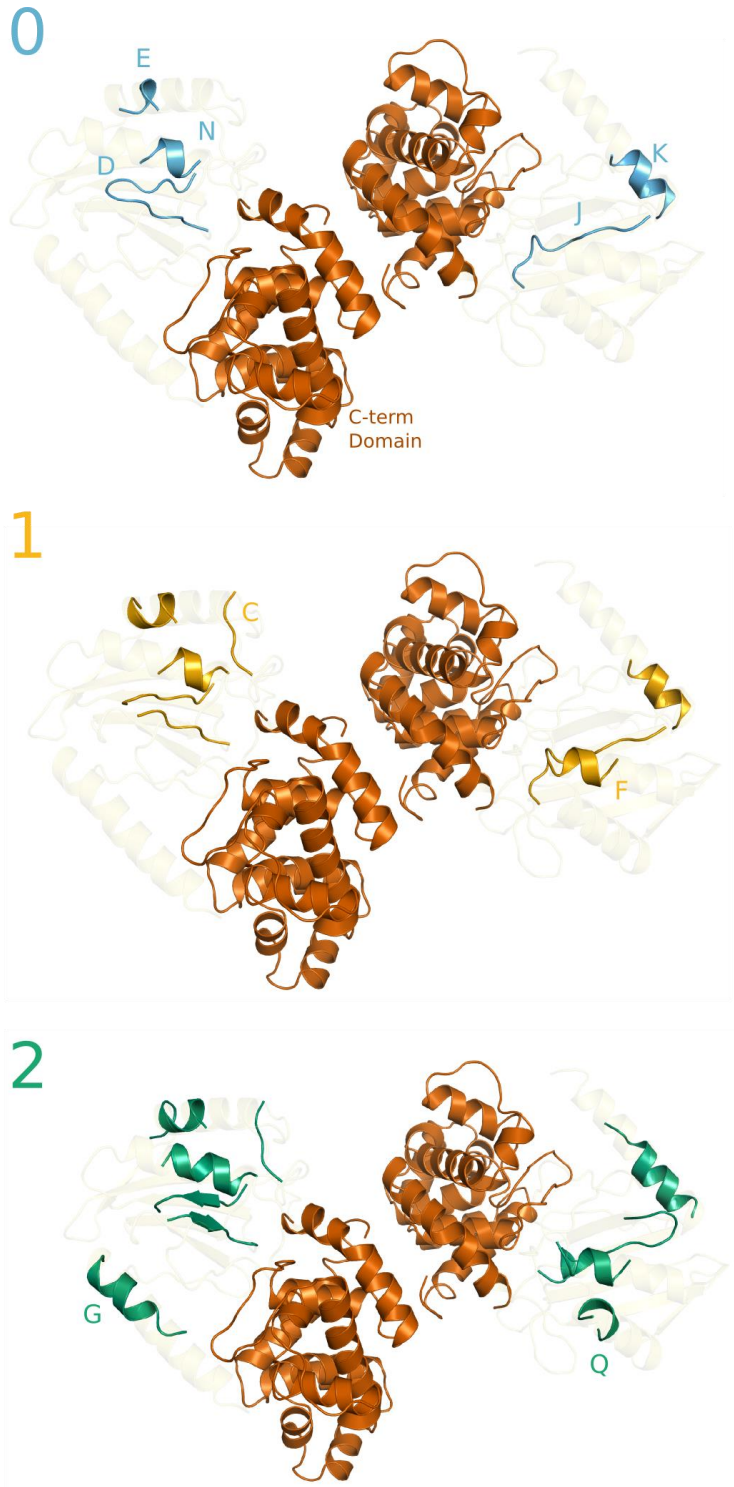


Figure 3.18 - Cartoon representation of main chain expansions of MltC prior and after *SEQSLIDER* runs.
In **0**, fragments in light blue obtained by the first expansion after side chain correction of C-terminal domain (colored orange) and *BUSTER* refinement. In 1, fragments colored in yellow obtained from expansion of 1st SEQSLIDER cycle. In 2, fragments colored in green from expansion of 1st SEQSLIDER cycle. The chain letter of each fragment is followed besides it.

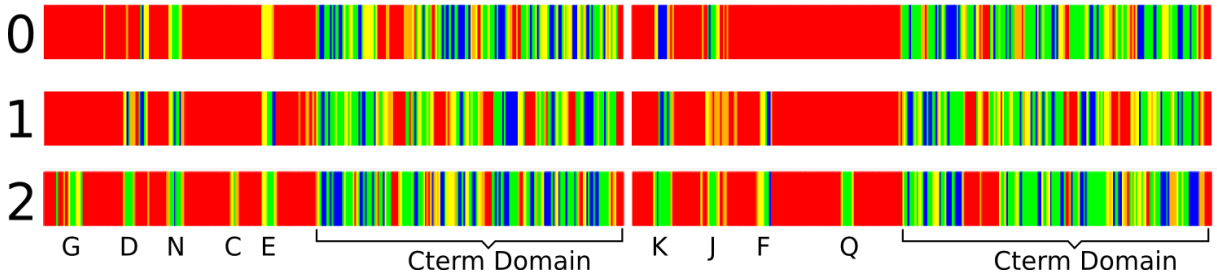


Figure 3.19 - *Post mortem* graph of *SHELXE* of different stages of MltC phasing by SEQSLIDER. 0 is prior to *SEQSLIDER* launch, 1 and 2 are the first and second *SEQSLIDER* runs, respectively.

Table 3.13 - Secondary structure chain composition, number of hypotheses generated by SEQLISDER and maximum observed identity.

Chain	# assigned residues	Secondary Structure	# generated hypotheses	Alignment score (max/min)	Maximum identity (%)
1st ROUND					
D	8	C1 S4 C2 S4 C1	114	43/18	50
E	5	5H	73	30/12	100
J	4	C1 S4 C5	22	24/6	100
K	9	H9	57	54/36	66
N	6	H6	69	36/18	100
2nd ROUND					
C	-	5C	-	-	-
F	5	C1 H5 C1	73	30/12	100

In Secondary Structure, C stands for coil, S for β-sheet and H for α-helix followed by the number of residues composing it.

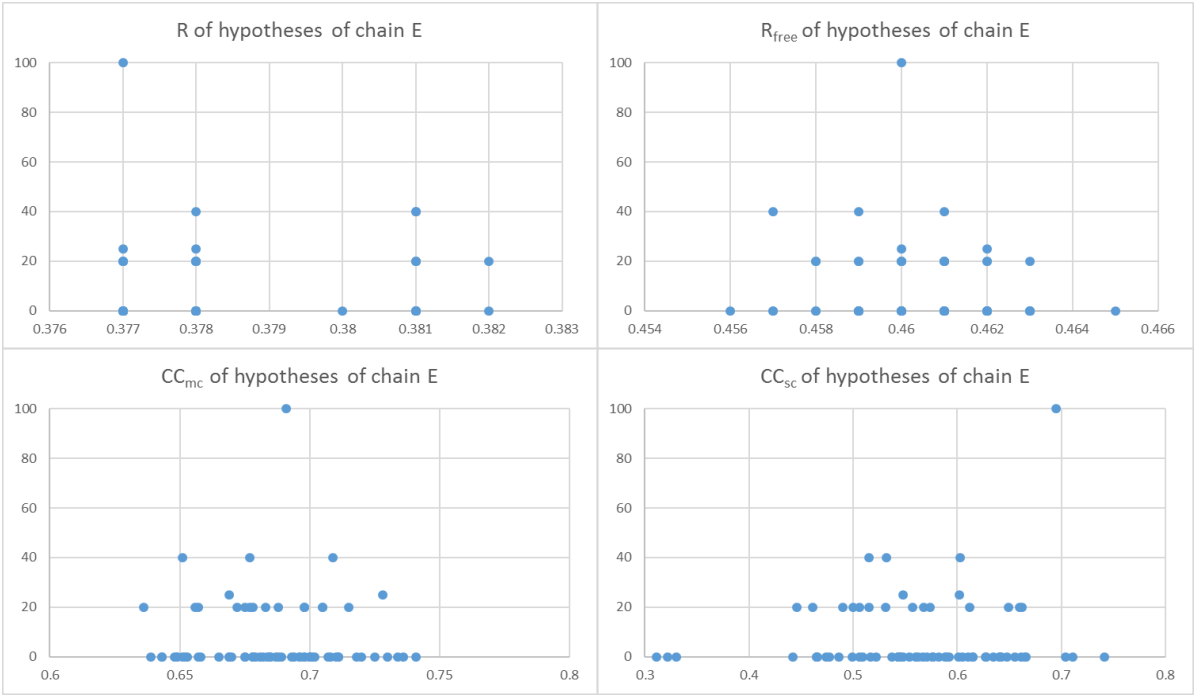


Figure 3.20 - Evaluation of the hypotheses of chain E and its plots of R, R_{free}, CC_{mc}, CC_{sc}.

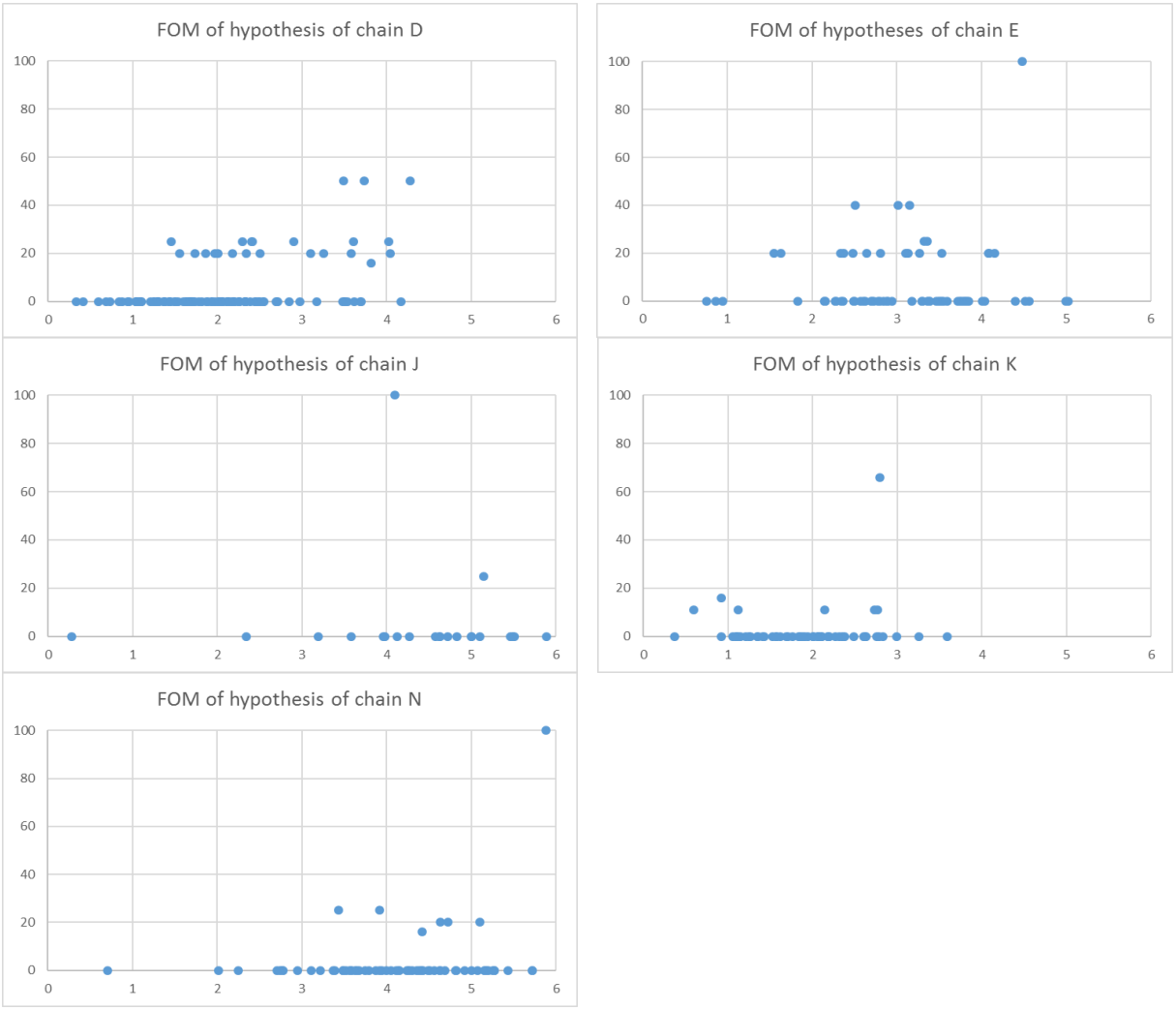


Figure 3.21 - Evaluation of the hypotheses of evaluated chains in SEQSLIDER cycle 1 by FOM and identity.

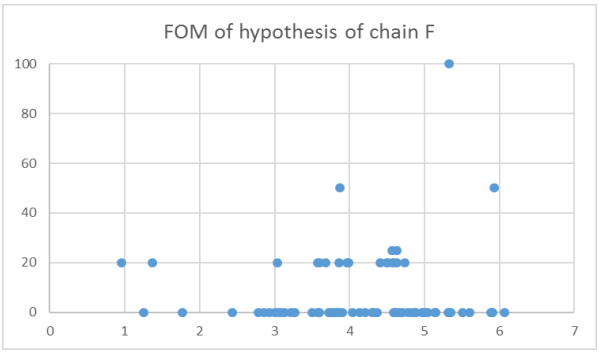


Figure 3.22 - Evaluation of the hypotheses of evaluated chain in SEQSLIDER cycle 2 by FOM and identity.

The next step in *SEQSLIDER* was to combine the best scoring hypotheses from each chain and submit the joint model to refinement and expansion. *SEQSLIDER* is set by default to send the 10 best hypotheses from each evaluation to SHELXE expansion using coordinates and the generated map as input. To extract an improved trace, we select out of the 180

expansion trials, the ones with higher CC and number of residues. The three best traces had: i) CC_f of 30.4% and 374 residues (6.7% incorrect); ii) 30.3% and 382 residues (3.7% incorrect); and iii) 30.0% and 384 residues (6.2% incorrect). Comparing by hand the traced fragments against the initial coordinates, we could see an increase in size of the fragments E and N and the separation of D in two fragments (Structure 1 in **Figure 3.18** and expansion 1 in **Figure 3.19**). Moreover, two additional fragments were found, a coil (C in Structure 1 in **Figure 3.18**) and a helix (F in Structure 1 in **Figure 3.18**). Since we had an increase in the scattering model inside the unit cell, we chose to submit a new round of SEQSLIDER evaluating all chains with secondary structure elements, with the exception of C-terminal domain.

Chains D, E, J, K and N had similar results as before. Using the same methodology as the previous SEQSLIDER run, we selected the best traces. The correct hypothesis was found for the new helix (F) (**Table 3.13**) and had a high FOM score (**Figure 3.22**). The results of traces did not improve much in terms of values, CC_i 30.3, 30.2, and 30.2%, and 378 (8.9% incorrect), 380 (6.0% incorrect) and 387 (7.1% incorrect) residues traced. Although, we identified two new fragments, two helices (G and Q in Structure 2 of **Figure 3.18**). Moreover, the chains E, F, J, K, and N were further extended. With these two new helices, it was possible to find the NCS between the N-terminal domain superimposing the three helices of one monomer to the other (**Figure 3.23**). After NCS discovery, the superposed fragments coming from each monomer were merged, thus completing each other. As the CC is higher than 30, NCS was found, and the generated map is already interpretable, further building can be continued manually or by other programs.

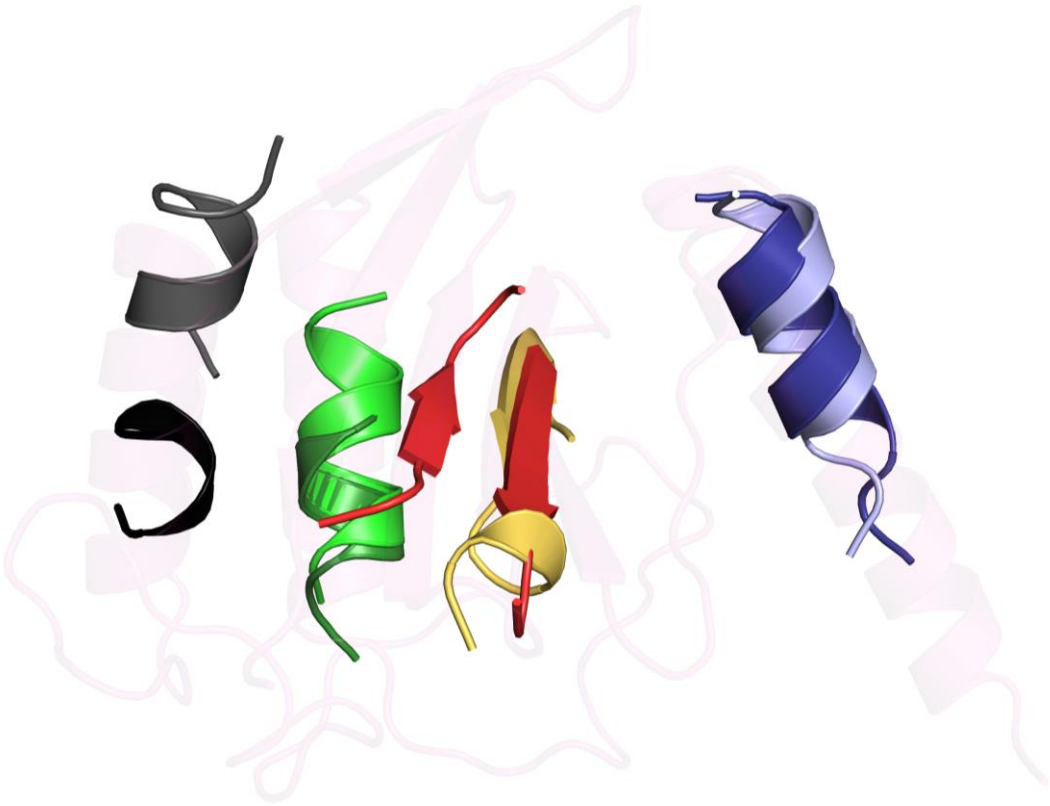


Figure 3.23 - The different fragments of the missing N-terminal domain of MltC.
The fragments from monomer A and B are superposed and are an indication of the NCS.

3.8.3 FrmR E64H de *Salmonella enterica*

Phasing coiled coils structures is a challenge as their diffraction data are usually anisotropic and dominated by the direction of the helices. Distinction of true and false solutions and interpretation of side chains are difficult tasks that we will evaluate using SEQSLIDER. In collaboration with Dr. Ehmke Pohl, we worked in the data of FrmR E64H, a coiled coil with 91 residues, 2.12 Å resolution and space group $P2_1$ (OSMAN et al., 2016). With 4 monomers in the ASU, other methods did not yield a solution. Dr. Usón, using ARCIMBOLDO_LITE version, found a solution composed of 269 residues divided in 6 chains with a CC of 56. Completing the rest of 100 residues and finding correct side chains in electron density were difficult.

We evaluated SEQSLIDER application to the six present chains in the partial trace of FrmR E64H (**Table 3.14**). Chains C, D, E, and F are composed of two helices, whether A and B of one. Due to the low number of assigned residues and hypotheses generated for the single helices A and B, we evaluated them together. Figure 3.16 shows the alignments of few

hypotheses against the chain C among the scoring function. The best hypothesis for chain C has score of 222 and the last score to be included in evaluation was 209. Therefore, the hypotheses 2, 3, and 4 that are unlikely were not included in the 105 evaluations. Four additional hypotheses were included in the default 100 as they had equal scores. In these 105 hypotheses, the top one shows an identity of 86% respect to the final structure.

Table 3.14 - SEQSLIDER overview on FrmR E64H dataset.

Chain	# assigned residues	Secondary structure elements	# generated hypotheses	Alignment score (Max/Min)	Maximum identity (%)
A	20	C5 H20 C1	32	228/228	100
B	18	C3 H18 C2	38		
C	38	C3 H27 C3 H11 C1	105	222/209	86
D	32	C3 H24 C6 H8 C6	100	186/178	93
E	53	C1 H24 C3 H29 C3	56	312/281	94
F	47	C1 H21 C5 H26 C3	100	276/253	89

Chain A and B were evaluated mutually.

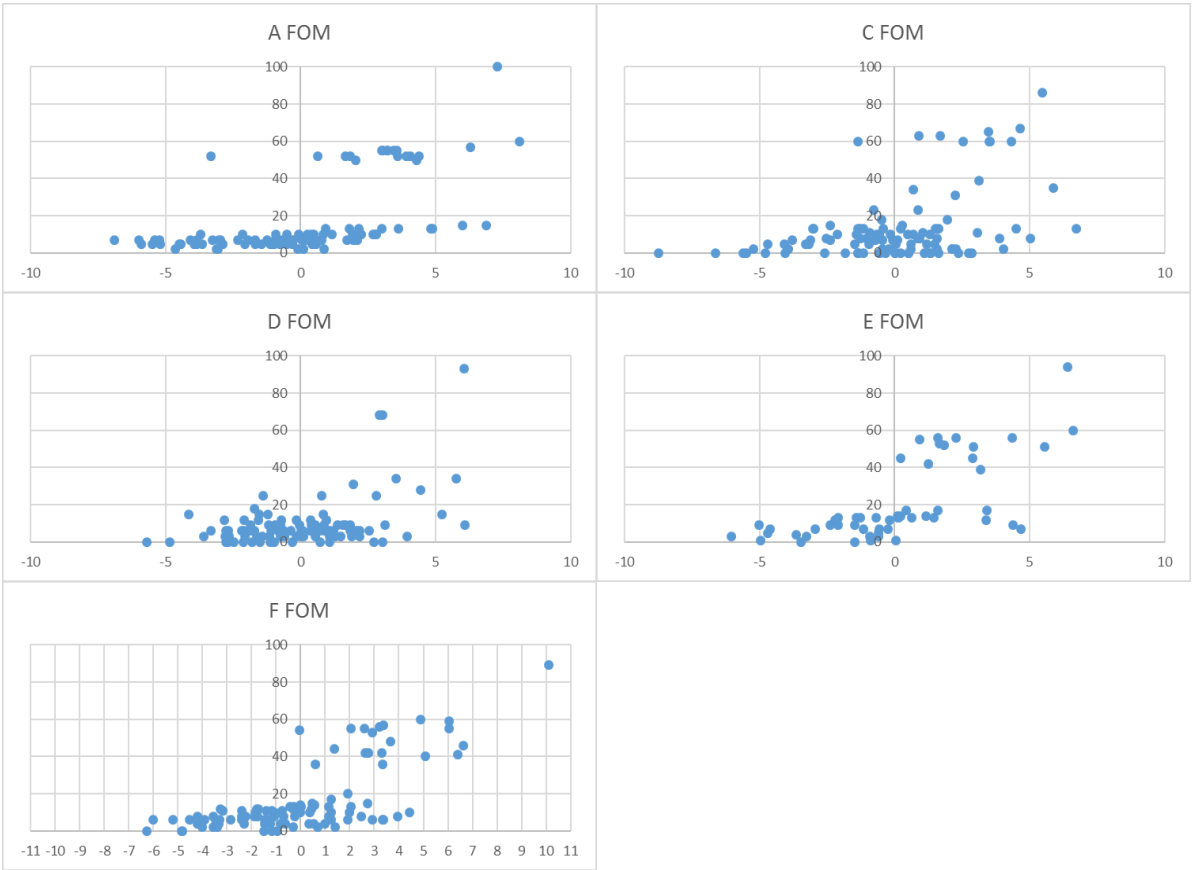


Figure 3.24 - Evaluation of the hypotheses of all chains in SEQSLIDER by FOM and identity (%).

Chains A and B were evaluated simultaneously.

FOM is appropriate to evaluate the correctness as a high correlation is found between FOM and identity of the hypotheses evaluated in all chains (**Figure 3.24**). Once the initial evaluation of independent chains is finished, SEQSLIDER selects the best 10 hypotheses of each evaluation and sorts them by chain and maximum number of assigned residues. In this run, the first chain would be C, followed by D, E, A/B, and C. Independent chain evaluation sorts them in decreasing order by the number of maximum assigned residues. In the present case, this would be first C, then D, E, A/B, and F. In a constructive method, SEQSLIDER combines de best 10 hypotheses of C with the best 10 of D generating 100 new hypotheses, those are evaluated as before. Then, the best 10 hypotheses of this evaluation are combined with the best 10 of E and the new combinations are sent to refinement and expansion. The process finishes when all chains have been combined.

To illustrate the independent and the constructive approach, we ploted in a single graph the $R_{factors}$, RSCC and number of residues assigned of the best hypothesis of each chain or combination (**Figure 3.25**). As more residues are assigned, statistics are improved, getting the R and R_{free} of 39.0% and 42.2%, respectively. From this point on, electron density is improved significantly and manual inspection is possible. Aligning the sequence of the best hypothesis against the known sequence is possible to know what helix corresponds to what stretch in the sequence (**Figure 3.26**). Helices of chain A and B refer to the first helix of predicted SS and they may connect to chains C and D, which only have helices 2 and 3. Therefore, model may be completed in a straightforward way.

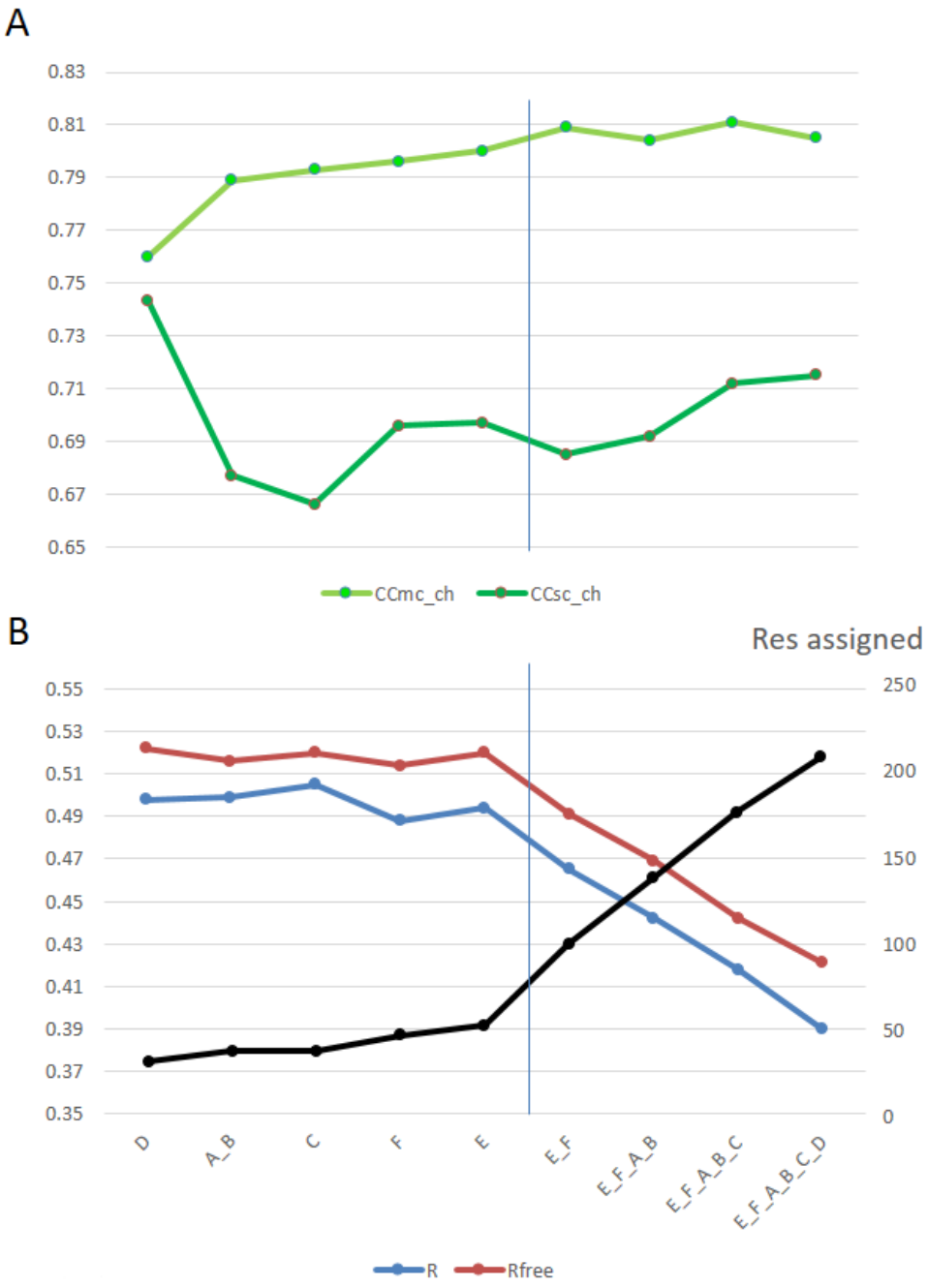


Figure 3.25 – Graphs of the improvement of Real Space Correlation Coefficient (RSCC) and $R_{factors}$ with increase of number of residues assigned by SEQSLIDER run.
Each point represents the best hypothesis based on FOM in the evaluation of the chain(s) in the abscissa. In A, RSCC of main chain atoms (CCmc) and RSCC of side chain atoms (CCsc) are represented in light and dark green, respectively. In B, R and R_{free} are represented in blue and red, respectively with values in left side ordinate. The number of residues assigned is represented by the black line with values corresponding on the right side ordinate.

Figure 3.26 – Alignment of the best hypothesis of each SEQSLIDER evaluation by chain with complete sequence in top.
Green letters refer to predict α-helices by *PSIPRED*. Secondary structure composition of each chain in the right hand side. C stands for coil and H for helix followed by the number of residues composing that element.

3.9 Conclusion

The structure solution software SEQUENCE SLIDER and its underlying algorithms have proven useful to aid elucidation of challenging crystallographic datasets. Focusing on crystals obtained from natural source samples, such as venoms, SEQSLIDER could differentiate and distinguish most of the correct residues in the restricted search. Restrictions may be given by mass spectrometry data or phylogenetic analysis on local variability and conservation. We faced two problematic scenarios. First, the mutual evaluation of residues of similar constituency, such as D/N/L and T/V, which may be addressed in the future looking into hydrogen network. In case where restriction allows more than a single possibility to a residue highly flexible with poor electron density, no conclusion may be extracted. The electron density would best fit to small side chains, such as alanine or serine. This uncertainty is a limitation of the diffraction data and should be addressed experimentally with mass spectrometry. The search for residues insufficiently determined from the crystallographic data and presenting assignments conflicting with experimental or bioinformatic data will provide valuable information. We plan to perform a systematic evaluation on deposited structures

whose crystals were obtained from natural source proteins. In such analysis, we will use the phylogenetic analysis to review side chains and, in case SEQSLIDER has improved the deposited model, we will contact the depositors suggesting review of their structures and sharing the use of our program.

On the second SEQSLIDER approach, increasing the scope of ARCIMBOLDO to lower resolution than usual 2.0 Å, it was possible to solve MltC structure (2.7 Å). This test allowed a series of implementation into the SEQSLIDER previous version. We will apply SEQSLIDER to other test structures to evaluate weakness of the current program and try to push over unknown structures. In the step of mutually evaluating SHELXE expansions, we will automate new traces evaluation and merging steps. Our objective is to distribute an automated version of SEQSLIDER in which the decisions on what fragments to slide and what to incorporate from expansions will be made by the algorithm itself.

4 REFERENCES

REFERENCES

ADAMS, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. **Acta Crystallographica Section D Biological Crystallography**, v. 66, n. 2, p. 213–221, 22 jan. 2010.

ALEXANDROV, V. et al. Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. **Protein Science: A Publication of the Protein Society**, v. 14, n. 3, p. 633–643, mar. 2005.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

AMBROSIO, A. L. B. et al. A molecular mechanism for Lys49-phospholipase A2 activity based on ligand-induced conformational change. **The Journal of biological chemistry**, v. 280, n. 8, p. 7326–7335, 25 fev. 2005.

ANDRIÃO-ESCARSO, S. H. et al. Myotoxic phospholipases A2 in Bothrops snake venoms: Effect of chemical modifications on the enzymatic and pharmacological properties of bothropstoxins from Bothrops jararacussu. **Biochimie**, v. 82, n. 8, p. 755–763, Agosto 2000.

ARCIMBOLDO, G. **The Greengrocer**. Free media repository. Disponível em: <https://upload.wikimedia.org/wikipedia/commons/4/49/Arcimboldo_Vegetables.jpg>. Acesso em: 5 out. 2016.

ARNI, R. K. et al. Structure of a calcium-independent phospholipase-like myotoxic protein from Bothrops asper venom. **Acta crystallographica. Section D, Biological crystallography**, v. 51, n. Pt 3, p. 311–317, 1 maio 1995.

ARTOLA-RECOLONS, C. et al. Structure and cell wall cleavage by modular lytic transglycosylase MltC of Escherichia coli. **ACS chemical biology**, v. 9, n. 9, p. 2058–2066, 19 set. 2014.

BAHNSON, B. J. Structure, function and interfacial allosterism in phospholipase A2: insight from the anion-assisted dimer. **Archives of biochemistry and biophysics**, v. 433, n. 1, p. 96–106, 1 jan. 2005.

BATINA, M. DE F. et al. Inhibition of the lethal and myotoxic activities of Crotalus durissus terrificus venom by Tabernaemontana catharinensis: identification of one of the active components. **Planta medica**, v. 66, n. 5, p. 424–428, jun. 2000.

BATTYE, T. G. G. et al. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. **Acta Crystallographica Section D Biological Crystallography**, v. 67, n. 4, p. 271–281, 18 mar. 2011.

BERG, O. G. et al. Interfacial Enzymology: The Secreted Phospholipase A2-Paradigm. **Chemical Reviews**, v. 101, n. 9, p. 2613–2654, 1 set. 2001.

BERNSTEIN, F. C. et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. **Journal of Molecular Biology**, v. 112, n. 3, p. 535–542, 25 maio 1977.

BORGES, R. J. **Estudos estruturais com fosfolipases A2 homólogas de veneno botrópicos em presença de íons com importância funcional.** Botucatu: Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 2012.

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. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1861, n. 1, p. 3199–3209, jan. 2017.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. In: **Doenças Infecciosas e Parasitárias: Guia de Bolso**. 7. ed. Brasília, DF: Ministério da Saúde, 2008. p. 28–31.

BRASIL. **Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Casos de acidentes por serpentes. Brasil, Grandes Regiões e Unidades Federadas 2000 a 2015**. [s.l: s.n.]. Disponível em: <http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/20/1-Casos-Ofidismo-2000-2015.pdf>.

BRASIL. **Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Óbitos por serpentes. Brasil, Grandes Regiões e Unidades Federadas 2000 a 2015**. [s.l:

s.n.]. Disponível em: <http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/20/3-Obitos-Ofidismo-2000-2015.pdf>.

BRICOGNE, G. et al. **BUSTER**. Cambridge, United Kingdom: Global Phasing Ltd., 2011.

BROOKS, B. R. et al. CHARMM: the biomolecular simulation program. **Journal of computational chemistry**, v. 30, n. 10, p. 1545–1614, 30 jul. 2009.

BULTRÓN, E.; GUTIÉRREZ, J. M.; THELESTAM, M. Effects of Bothrops asper (terciopelo) myotoxin III, a basic phospholipase A2, on liposomes and mouse gastrocnemius muscle. **Toxicon: official journal of the International Society on Toxinology**, v. 31, n. 2, p. 217–222, fev. 1993.

CALVETE, J. J. et al. Venoms, venomics, antivenomics. **FEBS letters**, v. 583, n. 11, p. 1736–1743, 5 jun. 2009.

CALVETE, J. J. et al. Snake venomics of the Central American rattlesnake Crotalus simus and the South American Crotalus durissus complex points to neurotoxicity as an adaptive paedomorphic trend along Crotalus dispersal in South America. **Journal of proteome research**, v. 9, n. 1, p. 528–544, jan. 2010.

CALVETE, J. J.; JUÁREZ, P.; SANZ, L. Snake venomics. Strategy and applications. **Journal of mass spectrometry: JMS**, v. 42, n. 11, p. 1405–1414, nov. 2007.

CARDOSO, F. F. **Estudos estruturais e funcionais da interação entre derivados do ácido cinâmico e fosfolipase A2 homóloga do veneno de Bothrops jararacussu**. Botucatu: Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 2016.

CARVALHO, C. DE. **Ação do extrato bruto da Tabernaemontana catharinensis e de seu alcalóide isolado sobre as atividades de fosfolipases A2 ofídicas em preparação neuromuscular de camundongos**. Botucatu: Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 7 out. 2011.

CASEWELL, N. R. et al. Complex cocktails: the evolutionary novelty of venoms. **Trends in Ecology & Evolution**, v. 28, n. 4, p. 219–229, abr. 2013.

CHANG, C. C.; LEE, C. Y. ISOLATION OF NEUROTOXINS FROM THE VENOM OF BUNGARUS MULTICINCTUS AND THEIR MODES OF NEUROMUSCULAR BLOCKING ACTION. **Archives Internationales De Pharmacodynamie Et De Thérapie**, v. 144, p. 241–257, 1 jul. 1963.

CHIOATO, L. et al. Mapping of the structural determinants of artificial and biological membrane damaging activities of a Lys49 phospholipase A2 by scanning alanine mutagenesis. **Biochimica et biophysica acta**, v. 1768, n. 5, p. 1247–1257, maio 2007.

CHOVANCOVA, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. **PLoS computational biology**, v. 8, n. 10, p. e1002708, 2012.

CORRÊA, L. C. et al. Crystal structure of a myotoxic Asp49-phospholipase A2 with low catalytic activity: Insights into Ca2+-independent catalytic mechanism. **Biochimica et biophysica acta**, v. 1784, n. 4, p. 591–599, abr. 2008.

COSTA, T. R. et al. Myotoxic phospholipases A2 isolated from Bothrops brazili snake venom and synthetic peptides derived from their C-terminal region: Cytotoxic effect on microorganism and tumor cells. **Peptides**, v. 29, n. 10, p. 1645–1656, out. 2008.

CROWFOOT, D. et al. X-ray crystallographic investigation of the structure of penicillin. In: **Chemistry of Penicillin**. Clarke H, Johnson T, Robinson JR ed. Princeton, NJ: University Press, 1949. p. 310–367.

DAVIS, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. **Nucleic acids research**, v. 35, n. Web Server issue, p. W375-383, jul. 2007.

DE AZEVEDO, W. F. et al. Crystal structure of piratoxin-I: a calcium-independent, myotoxic phospholipase A2-homologue from Bothrops pirajai venom. **Toxicon: Official Journal of the International Society on Toxinology**, v. 36, n. 10, p. 1395–1406, out. 1998.

DE OLIVEIRA, A. H. et al. A pH-induced dissociation of the dimeric form of a lysine 49-phospholipase A2 abolishes Ca2+-independent membrane damaging activity. **Biochemistry**, v. 40, n. 23, p. 6912–6920, 12 jun. 2001a.

DE OLIVEIRA, A. H. et al. The effect of resonance energy homotransfer on the intrinsic tryptophan fluorescence emission of the bothropstoxin-I dimer. **Biochemical and biophysical research communications**, v. 284, n. 4, p. 1011–1015, 22 jun. 2001b.

DE OLIVEIRA, A. H. C.; FERREIRA, T. L.; WARD, R. J. Reduced pH induces an inactive non-native conformation of the monomeric bothropstoxin-I (Lys49-PLA2). **Toxicon: official journal of the International Society on Toxinology**, v. 54, n. 3, p. 373–378, 1 set. 2009.

DE OLIVEIRA, M. et al. Antagonism of myotoxic and paralyzing activities of bothropstoxin-I by suramin. **Toxicon: Official Journal of the International Society on Toxinology**, v. 42, n. 4, p. 373–379, 15 set. 2003.

DELATORRE, P. et al. Crystal structure of Bn IV in complex with myristic acid: a Lys49 myotoxic phospholipase A_2 from Bothrops neuwiedi venom. **Biochimie**, v. 93, n. 3, p. 513–518, mar. 2011.

DHAWAN, D. K.; CHADHA, V. D. Zinc: a promising agent in dietary chemoprevention of cancer. **The Indian Journal of Medical Research**, v. 132, p. 676–682, dez. 2010.

DÍAZ, C. et al. The effect of myotoxins isolated from Bothrops snake venoms on multilamellar liposomes: relationship to phospholipase A2, anticoagulant and myotoxic activities. **Biochimica et biophysica acta**, v. 1070, n. 2, p. 455–460, 9 dez. 1991.

DÍAZ, C. et al. Cleavage of the NH2-terminal octapeptide of Bothrops asper myotoxic lysine-49 phospholipase A2 reduces its membrane-destabilizing effect. **Archives of biochemistry and biophysics**, v. 312, n. 2, p. 336–339, 1 ago. 1994.

DÍAZ, C. et al. Modulation of the susceptibility of human erythrocytes to snake venom myotoxic phospholipases A(2): role of negatively charged phospholipids as potential membrane binding sites. **Archives of biochemistry and biophysics**, v. 391, n. 1, p. 56–64, 1 jul. 2001.

DIEBEL, J. **Representing Attitude: Euler Angles, Unit Quaternions, and Rotation Vectors**Stanford, California 94301–9010, , 20 out. 2006. Disponível em: <https://www.astro.rug.nl/software/kapteyn/_downloads/attitude.pdf>

DOLEY, R.; KINI, R. M. Protein complexes in snake venom. **Cellular and molecular life sciences: CMLS**, v. 66, n. 17, p. 2851–2871, set. 2009.

DOS SANTOS, J. I. et al. Structural and functional studies of a bothropic myotoxin complexed to rosmarinic acid: new insights into Lys49-PLA_2 inhibition. **PloS one**, v. 6, n. 12, p. e28521, 2011a.

DOS SANTOS, J. I. et al. Structural, functional, and bioinformatics studies reveal a new snake venom homologue phospholipase A_2 class. **Proteins**, v. 79, n. 1, p. 61–78, jan. 2011b.

DOS SANTOS, J. I.; SOARES, A. M.; FONTES, M. R. M. Comparative structural studies on Lys49-phospholipases A(2) from Bothrops genus reveal their myotoxic site. **Journal of structural biology**, v. 167, n. 2, p. 106–116, ago. 2009.

EALICK, S. E. Now we're cooking: new successes for shake-and-bake. **Structure (London, England: 1993)**, v. 5, n. 4, p. 469–472, 15 abr. 1997.

EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D Biological Crystallography**, v. 66, n. 4, p. 486–501, 24 mar. 2010.

EVANS, P.; MCCOY, A. An introduction to molecular replacement. **Acta Crystallographica Section D Biological Crystallography**, v. 64, n. 1, p. 1–10, 1 jan. 2008.

EVANS, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. **Acta Crystallographica Section D Biological Crystallography**, v. 67, n. 4, p. 282–292, 1 abr. 2011.

EVRARD, G. X. et al. Assessment of automatic ligand building in ARP/wARP. **Acta Crystallographica. Section D, Biological Crystallography**, v. 63, n. Pt 1, p. 108–117, jan. 2007.

FEIGIN, L. A.; SVERGUN, D. I. **Structure Analysis by Small-Angle X-Ray and Neutron Scattering**. Softcover reprint of the original 1st ed. 1987 edition ed. [s.l.] Springer, 2013.

FENTON, A. W. et al. Arthropod venom citrate inhibits phospholipase A2. **Toxicon: Official Journal of the International Society on Toxinology**, v. 33, n. 6, p. 763–770, jun. 1995.

FERNANDES, C. A. H. et al. Comparison between apo and complexed structures of bothropstoxin-I reveals the role of Lys122 and Ca(2+)-binding loop region for the catalytically inactive Lys49-PLA(2)s. **Journal of structural biology**, v. 171, n. 1, p. 31–43, jul. 2010.

FERNANDES, C. A. H. et al. Crystallization and preliminary X-ray diffraction analysis of three myotoxic phospholipases A2 from Bothrops brazili venom. **Acta Crystallographica. Section F, Structural Biology and Crystallization Communications**, v. 68, n. Pt 8, p. 935–938, 1 ago. 2012.

FERNANDES, C. A. H. et al. Structural bases for a complete myotoxic mechanism: Crystal structures of two non-catalytic phospholipases A2-like from Bothrops brazili venom. **Biochimica et biophysica acta**, v. 1834, n. 12, p. 2772–2781, dez. 2013.

FERNANDES, C. A. H. et al. A structure-based proposal for a comprehensive myotoxic mechanism of phospholipase A2-like proteins from viperid snake venoms. **Biochimica Et Biophysica Acta**, 29 set. 2014.

FERNANDES, C. A. H. et al. Structural Basis for the Inhibition of a Phospholipase A2-Like Toxin by Caffeic and Aristolochic Acids. **PloS One**, v. 10, n. 7, p. e0133370, 2015.

FERREIRA, S. H.; ROCHA E SILVA, M. Potentiation of bradykinin by dimercaptopropanol (bal) and other inhibitors of its destroying enzyme in plasma. **Biochemical Pharmacology**, v. 11, p. 1123–1128, dez. 1962.

FRANCIS, B. et al. Myotoxin II from Bothrops asper (Terciopelo) venom is a lysine-49 phospholipase A2. **Archives of biochemistry and biophysics**, v. 284, n. 2, p. 352–359, 1 fev. 1991.

FRANCIS, B.; SEEBART, C.; KAISER, I. I. Citrate is an endogenous inhibitor of snake venom enzymes by metal-ion chelation. **Toxicon: Official Journal of the International Society on Toxinology**, v. 30, n. 10, p. 1239–1246, out. 1992.

FUGLEBAKK, E.; ECHAVE, J.; REUTER, N. Measuring and comparing structural fluctuation patterns in large protein datasets. **Bioinformatics (Oxford, England)**, v. 28, n. 19, p. 2431–2440, 1 out. 2012.

GRATTAN, B. J.; FREAKE, H. C. Zinc and cancer: implications for LIV-1 in breast cancer. **Nutrients**, v. 4, n. 7, p. 648–675, jul. 2012.

GUTIÉRREZ, J. M. et al. Snake venomics and antivenomics: Proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming. **Journal of Proteomics**, Venomics. v. 72, n. 2, p. 165–182, 6 mar. 2009.

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 Medicine**, v. 3, n. 6, jun. 2006.

HELUANY, N. F. et al. Effects induced by bothropstoxin, a component from Bothrops jararacussu snake venom, on mouse and chick muscle preparations. **Toxicon: official journal of the International Society on Toxinology**, v. 30, n. 10, p. 1203–1210, out. 1992.

HENDRICKSON, W. A. Evolution of diffraction methods for solving crystal structures. **Acta Crystallographica Section A Foundations of Crystallography**, v. 69, n. 1, p. 51–59, 1 jan. 2013.

HODGKIN, D. C. et al. The Structure of Vitamin BFormula I. An Outline of the Crystallographic Investigation of Vitamin BFormula. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 242, n. 1229, p. 228–263, 29 out. 1957.

HUANCAHUIRE-VEGA, S. et al. Structural and functional characterization of brazilitoxins II and III (BbTX-II and -III), two myotoxins from the venom of Bothrops brazili snake. **Toxicon: Official Journal of the International Society on Toxinology**, v. 54, n. 6, p. 818–827, nov. 2009.

HÜNEFELD, F. L. **Die Chemismus in der tierescher Organisation**. Leipzig: FA Brockhouse, 1840.

JANOTA, A. et al. Improving the precision and speed of Euler angles computation from low-cost rotation sensor data. **Sensors (Basel, Switzerland)**, v. 15, n. 3, p. 7016–7039, 2015.

JASKOLSKI, M.; DAUTER, Z.; WLODAWER, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. **The FEBS journal**, v. 281, n. 18, p. 3985–4009, set. 2014.

KABSCH, W. A solution for the best rotation to relate two sets of vectors. **Acta Crystallographica Section A**, v. 32, n. 5, p. 922–923, 1 set. 1976.

KABSCH, W. XDS. **Acta crystallographica. Section D, Biological crystallography**, v. 66, n. Pt 2, p. 125–132, fev. 2010.

KARPLUS, P. A.; DIEDERICHS, K. Linking crystallographic model and data quality. **Science (New York, N.Y.)**, v. 336, n. 6084, p. 1030–1033, 25 maio 2012.

KASTURIRATNE, A. et al. The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. **PLoS medicine**, v. 5, n. 11, p. e218, 4 nov. 2008.

KENDREW, J. C. et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. **Nature**, v. 181, n. 4610, p. 662–666, 8 mar. 1958.

KINI, R. M. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. **Toxicon: official journal of the International Society on Toxinology**, v. 42, n. 8, p. 827–840, 15 dez. 2003.

KLEYWEGT, G. J.; JONES, T. A. Homo Crystallographicus—Quo Vadis? **Structure**, v. 10, n. 4, p. 465–472, abr. 2002.

KONAREV, P. V. et al. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. **Journal of Applied Crystallography**, v. 36, n. 5, p. 1277–1282, 1 out. 2003.

KONAREV, P. V. et al. ATSAS 2.1, a program package for small-angle scattering data analysis. **Journal of Applied Crystallography**, v. 39, n. 2, p. 277–286, 12 mar. 2006.

KRISSINEL, E.; HENRICK, K. Inference of macromolecular assemblies from crystalline state. **Journal of molecular biology**, v. 372, n. 3, p. 774–797, 21 set. 2007.

KRIVOV, G. G.; SHAPOVALOV, M. V.; DUNBRACK, R. L. Improved prediction of protein side-chain conformations with SCWRL4. **Proteins**, v. 77, n. 4, p. 778–795, dez. 2009.

LEE, W. H. et al. Structural basis for low catalytic activity in Lys49 phospholipases A2--a hypothesis: the crystal structure of piratoxin II complexed to fatty acid. **Biochemistry**, v. 40, n. 1, p. 28–36, 9 jan. 2001.

LIEBSCHNER, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. **Acta Crystallographica Section D Biological Crystallography**, n. (accepted), 2017.

LIN, Y. et al. Docking phospholipase A2 on membranes using electrostatic potential-modulated spin relaxation magnetic resonance. **Science (New York, N.Y.)**, v. 279, n. 5358, p. 1925–1929, 20 mar. 1998.

LITTLE, P. J. et al. Zinc and cardiovascular disease. **Nutrition (Burbank, Los Angeles County, Calif.)**, v. 26, n. 11–12, p. 1050–1057, dez. 2010.

LIZANO, S.; LAMBEAU, G.; LAZDUNSKI, M. Cloning and cDNA sequence analysis of Lys(49) and Asp(49) basic phospholipase A(2) myotoxin isoforms from Bothrops asper. **The International Journal of Biochemistry & Cell Biology**, v. 33, n. 2, p. 127–132, fev. 2001.

LOMONTE, B. et al. The phospholipase A2 homologues of snake venoms: biological activities and their possible adaptive roles. **Protein and peptide letters**, v. 16, n. 8, p. 860–876, 2009.

LOMONTE, B. Identification of linear B-cell epitopes on myotoxin II, a Lys49 phospholipase A_2 homologue from Bothrops asper snake venom. **Toxicon: official journal of the International Society on Toxinology**, v. 60, n. 5, p. 782–790, out. 2012.

LOMONTE, B.; ANGULO, Y.; CALDERÓN, L. An overview of lysine-49 phospholipase A2 myotoxins from crotalid snake venoms and their structural determinants of myotoxic action. **Toxicon: official journal of the International Society on Toxinology**, v. 42, n. 8, p. 885–901, 15 dez. 2003.

LOMONTE, B.; GUTIÉRREZ, J. M. A new muscle damaging toxin, myotoxin II, from the venom of the snake Bothrops asper (terciopelo). **Toxicon: official journal of the International Society on Toxinology**, v. 27, n. 7, p. 725–733, 1989.

LOMONTE, B.; GUTIÉRREZ, J. M. Phospholipases A2 From Viperidae Snake Venoms: How do They Induce Skeletal Muscle Damage? **Acta Chim. Slov.**, v. 58, p. 647–658, 2011.

MAGRO, A. J. et al. Crystal structures of BnSP-7 and BnSP-6, two Lys49-phospholipases A(2): quaternary structure and inhibition mechanism insights. **Biochemical and biophysical research communications**, v. 311, n. 3, p. 713–720, 21 nov. 2003.

MARAGANORE, J. M. et al. A new class of phospholipases A2 with lysine in place of aspartate 49. Functional consequences for calcium and substrate binding. **The Journal of biological chemistry**, v. 259, n. 22, p. 13839–13843, 25 nov. 1984.

MARCHI-SALVADOR, D. P. et al. Crystal structure of a phospholipase A(2) homolog complexed with p-bromophenacyl bromide reveals important structural changes associated with the inhibition of myotoxic activity. **Biochimica et biophysica acta**, v. 1794, n. 11, p. 1583–1590, nov. 2009.

MCCOY, A. J. et al. Likelihood-enhanced fast translation functions. **Acta Crystallographica. Section D, Biological Crystallography**, v. 61, n. Pt 4, p. 458–464, abr. 2005.

MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658–674, 13 jul. 2007.

MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The PSIPRED protein structure prediction server. **Bioinformatics (Oxford, England)**, v. 16, n. 4, p. 404–405, abr. 2000.

MCPHERSON, A. **Introduction to macromolecular crystallography**. 2nd ed ed. Hoboken, N.J: Wiley-Blackwell, 2009.

MILLÁN, C.; SAMMITO, M.; USÓN, I. Macromolecular ab initio phasing enforcing secondary and tertiary structure. **IUCrJ**, v. 2, n. Pt 1, p. 95–105, 1 jan. 2015.

MILLER, R. et al. On the application of the minimal principle to solve unknown structures. **Science (New York, N.Y.)**, v. 259, n. 5100, p. 1430–1433, 5 mar. 1993.

MONTECUCCO, C.; GUTIÉRREZ, J. M.; LOMONTE, B. Cellular pathology induced by snake venom phospholipase A2 myotoxins and neurotoxins: common aspects of their mechanisms of action. **Cellular and molecular life sciences: CMLS**, v. 65, n. 18, p. 2897–2912, set. 2008.

MORA-OBANDO, D. et al. Role of enzymatic activity in muscle damage and cytotoxicity induced by Bothrops asper Asp49 phospholipase A2 myotoxins: are there additional effector mechanisms involved? **PeerJ**, v. 2, p. e569, 2014.

MURAKAMI, M. T. et al. Inhibition of myotoxic activity of Bothrops asper myotoxin II by the anti-trypanosomal drug suramin. **Journal of molecular biology**, v. 350, n. 3, p. 416–426, 15 jul. 2005.

MURAKAMI, M. T. et al. Interfacial surface charge and free accessibility to the PLA2-active site-like region are essential requirements for the activity of Lys49 PLA2 homologues. **Toxicon: official journal of the International Society on Toxinology**, v. 49, n. 3, p. 378–387, 1 mar. 2007.

MURAKAMI, M. T. et al. Biochemical and structural investigations of Bothropstoxin-II, a myotoxic Asp49 phospholipase A2 from Bothrops jararacussu venom. **Protein and peptide letters**, v. 15, n. 9, p. 1002–1008, 2008.

MURSHUDOV, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. **Acta Crystallographica. Section D, Biological Crystallography**, v. 67, n. Pt 4, p. 355–367, abr. 2011.

NÚÑEZ, V. et al. Structural and functional characterization of myotoxin I, a Lys49 phospholipase A2 homologue from the venom of the snake Bothrops atrox. **Toxicon: Official Journal of the International Society on Toxinology**, v. 44, n. 1, p. 91–101, jul. 2004.

OSMAN, D. et al. The Effectors and Sensory Sites of Formaldehyde-responsive Regulator FrmR and Metal-sensing Variant. **The Journal of Biological Chemistry**, v. 291, n. 37, p. 19502–19516, 9 set. 2016.

OTWINOWSKI, Z.; MINOR, W. Processing of X-ray diffraction data collected in oscillation mode. In: [s.l.] Elsevier, 1997. v. 276p. 307–326.

PAPADOPOULOS, J. S.; AGARWALA, R. COBALT: constraint-based alignment tool for multiple protein sequences. **Bioinformatics (Oxford, England)**, v. 23, n. 9, p. 1073–1079, 1 maio 2007.

PATTERSON, A. L. A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. **Physical Review**, v. 46, n. 5, p. 372–376, 1 set. 1934.

PEDERSEN, J. Z. et al. Autocatalytic acylation of phospholipase-like myotoxins. **Biochemistry**, v. 34, n. 14, p. 4670–4675, 11 abr. 1995.

PEREIRA, P. S. et al. Chemical constituents from Tabernaemontana catharinensis root bark: a brief NMR review of indole alkaloids and in vitro cytotoxicity. **Química Nova**, v. 31, n. 1, p. 20–24, 2008.

PERUTZ, M. F. et al. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. **Nature**, v. 185, n. 4711, p. 416–422, 13 fev. 1960.

PETTERSEN, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, out. 2004.

PONCE-SOTO, L. A. et al. Structural and functional properties of BaTX, a new Lys49 phospholipase A2 homologue isolated from the venom of the snake Bothrops alternatus. **Biochimica Et Biophysica Acta**, v. 1770, n. 4, p. 585–593, abr. 2007.

PRASAD, A. S. et al. Zinc in cancer prevention. **Nutrition and Cancer**, v. 61, n. 6, p. 879–887, 2009.

PRASAD, A. S. Discovery of human zinc deficiency: 50 years later. **Journal of trace elements in medicine and biology: organ of the Society for Minerals and Trace Elements (GMS)**, v. 26, n. 2–3, p. 66–69, jun. 2012.

PRAŽNIKAR, J. et al. Averaged kick maps: less noise, more signal... and probably less bias. **Acta Crystallographica. Section D, Biological Crystallography**, v. 65, n. Pt 9, p. 921–931, set. 2009.

RCSB PDB. **Graph of Yearly Growth of Total Structures**. Disponível em: <http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>. Acesso em: 13 nov. 2016.

RENETSEDER, R. et al. A comparison of the crystal structures of phospholipase A2 from bovine pancreas and Crotalus atrox venom. **The Journal of biological chemistry**, v. 260, n. 21, p. 11627–11634, 25 set. 1985.

RIGDEN, D. J. et al. The structure of the D49 phospholipase A2 piratoxin III from Bothrops pirajai reveals unprecedented structural displacement of the calcium-binding loop: possiblerelationship to cooperative substrate binding. **Acta crystallographica. Section D, Biological crystallography**, v. 59, n. Pt 2, p. 255–262, fev. 2003.

ROCHA E SILVA, M.; BERALDO, W. T.; ROSENFELD, G. Bradykinin, a hypotensive and smooth muscle stimulating factor released from plasma globulin by snake venoms and by trypsin. **The American Journal of Physiology**, v. 156, n. 2, p. 261–273, fev. 1949.

RODRIGUES, V. M. et al. Geographic variations in the composition of myotoxins from Bothrops neuwiedi snake venoms: biochemical characterization and biological activity. **Comparative biochemistry and physiology. Part A, Molecular & integrative physiology**, v. 121, n. 3, p. 215–222, nov. 1998.

RODRIGUES-SIMIONI, L. et al. No role for enzymatic activity or dantrolene-sensitive Ca2+ stores in the muscular effects of bothropstoxin, a Lys49 phospholipase A2 myotoxin. **Toxicon: official journal of the International Society on Toxinology**, v. 33, n. 11, p. 1479–1489, nov. 1995.

RODRÍGUEZ, D. et al. Practical structure solution with ARCIMBOLDO. **Acta crystallographica. Section D, Biological crystallography**, v. 68, n. Pt 4, p. 336–343, abr. 2012.

RODRÍGUEZ, D. D. et al. Crystallographic ab initio protein structure solution below atomic resolution. **Nature methods**, v. 6, n. 9, p. 651–653, set. 2009.

ROSSMANN, M. G.; BLOW, D. M. The detection of sub-units within the crystallographic asymmetric unit. **Acta Crystallographica**, v. 15, n. 1, p. 24–31, 1 jan. 1962.

RUSSO, E. Special Report: The birth of biotechnology. **Nature**, v. 421, n. 6921, p. 456–457, 23 jan. 2003.

SALVADOR, G. H. M. et al. Structural and phylogenetic studies with MjTX-I reveal a multi-oligomeric toxin--a novel feature in Lys49-PLA2s protein class. **PloS one**, v. 8, n. 4, p. e60610, 2013a.

SALVADOR, G. H. M. et al. Structural and functional studies with mytoxin II from Bothrops moojeni reveal remarkable similarities and differences compared to other catalytically inactive phospholipases A_2-like. **Toxicon: official journal of the International Society on Toxinology**, v. 72, p. 52–63, set. 2013b.

SALVADOR, G. H. M. et al. Structural and functional evidence for membrane docking and disruption sites on phospholipase A2-like proteins revealed by complexation with the inhibitor suramin. **Acta Crystallographica. Section D, Biological Crystallography**, v. 71, n. Pt 10, p. 2066–2078, out. 2015.

SALVADOR, G. H. M. et al. Crystal structure of a phospholipase A2 from Bothrops asper venom: Insights into a new putative “myotoxic cluster”. **Biochimie**, v. 133, p. 95–102, fev. 2017.

SAMMITO, M. et al. Exploiting tertiary structure through local folds for crystallographic phasing. **Nature methods**, v. 10, n. 11, p. 1099–1101, nov. 2013.

SAMMITO, M. **Enforcing secondary and tertiary structure for crystallographic phasing**. Barcelona: Universitat de Barcelona, Facultat de Farmàcia, 2015.

SAMMITO, M. et al. ARCIMBOLDO_LITE: single-workstation implementation and use. **Acta Crystallographica. Section D, Biological Crystallography**, v. 71, n. Pt 9, p. 1921–1930, 1 set. 2015.

SÁNCHEZ, E. E.; RODRÍGUEZ-ACOSTA, A. Inhibitors of snake venoms and development of new therapeutics. **Immunopharmacology and immunotoxicology**, v. 30, n. 4, p. 647–678, 2008.

SAUL, F. A. et al. Comparative structural studies of two natural isoforms of ammodytoxin, phospholipases A2 from Vipera ammodytes ammodytes which differ in neurotoxicity and anticoagulant activity. **Journal of Structural Biology**, v. 169, n. 3, p. 360–369, mar. 2010.

SCHALOSKE, R. H.; DENNIS, E. A. The phospholipase A2 superfamily and its group numbering system. **Biochimica et biophysica acta**, v. 1761, n. 11, p. 1246–1259, nov. 2006.

SCOTT, D. L. et al. Interfacial catalysis: the mechanism of phospholipase A2. **Science (New York, N.Y.)**, v. 250, n. 4987, p. 1541–1546, 14 dez. 1990a.

SCOTT, D. L. et al. Crystal structure of bee-venom phospholipase A2 in a complex with a transition-state analogue. **Science (New York, N.Y.)**, v. 250, n. 4987, p. 1563–1566, 14 dez. 1990b.

SEHNAL, D. et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. **Journal of Cheminformatics**, v. 5, n. 1, p. 39, 2013.

SHELDRICK, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. **Acta Crystallographica Section D Biological Crystallography**, v. 66, n. 4, p. 479–485, 24 mar. 2010.

SHELDRICK, G. M. et al. Ab initio phasing. **International Tables for Crystallography**, v. F, n. ch. 16.1, p. 413–432, 2012.

SOARES, A. M. et al. Structural and functional characterization of BnSP-7, a Lys49 myotoxic phospholipase A(2) homologue from Bothrops neuwiedi pauloensis venom. **Archives of biochemistry and biophysics**, v. 378, n. 2, p. 201–209, 15 jun. 2000.

SOARES, A. M. et al. Dissociation of enzymatic and pharmacological properties of piratoxins-I and -III, two myotoxic phospholipases A2 from Bothrops pirajai snake venom. **Archives of biochemistry and biophysics**, v. 387, n. 2, p. 188–196, 15 mar. 2001.

SOARES, A. M. et al. Mn(2+) ions reduce the enzymatic and pharmacological activities of bothropstoxin-I, a myotoxic Lys49 phospholipase A(2) homologue from Bothrops jararacussu snake venom. **The international journal of biochemistry & cell biology**, v. 34, n. 6, p. 668–677, jun. 2002.

SOARES, A. M. et al. Medicinal plants with inhibitory properties against snake venoms. **Current medicinal chemistry**, v. 12, n. 22, p. 2625–2641, 2005.

SOARES, A. M.; GIGLIO, J. R. Chemical modifications of phospholipases A2 from snake venoms: effects on catalytic and pharmacological properties. **Toxicon: official journal of the International Society on Toxinology**, v. 42, n. 8, p. 855–868, 15 dez. 2003.

STORONI, L. C.; MCCOY, A. J.; READ, R. J. Likelihood-enhanced fast rotation functions. **Acta Crystallographica. Section D, Biological Crystallography**, v. 60, n. Pt 3, p. 432–438, mar. 2004.

SVERGUN, D.; BARBERATO, C.; KOCH, M. H. J. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. **Journal of Applied Crystallography**, v. 28, n. 6, p. 768–773, 1 dez. 1995.

SVERGUN, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal of Applied Crystallography**, v. 25, n. 4, p. 495–503, 1 ago. 1992.

TAYLOR, G. The phase problem. **Acta Crystallographica. Section D, Biological Crystallography**, v. 59, n. Pt 11, p. 1881–1890, nov. 2003.

TERWILLIGER, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. **Acta Crystallographica. Section D, Biological Crystallography**, v. 64, n. Pt 1, p. 61–69, jan. 2008a.

TERWILLIGER, T. C. et al. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. **Acta Crystallographica. Section D, Biological Crystallography**, v. 64, n. Pt 5, p. 515–524, maio 2008b.

THOMAS, A. et al. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. **Proteins**, v. 34, n. 1, p. 96–112, 1 jan. 1999.

THORN, A.; SHELDRICK, G. M. Extending molecular-replacement solutions with SHELXE. **Acta crystallographica. Section D, Biological crystallography**, v. 69, n. Pt 11, p. 2251–2256, nov. 2013.

TICLI, F. K. et al. Rosmarinic acid, a new snake venom phospholipase A2 inhibitor from Cordia verbenacea (Boraginaceae): antiserum action potentiation and molecular interaction. **Toxicon: official journal of the International Society on Toxinology**, v. 46, n. 3, p. 318–327, 1 set. 2005.

TOYAMA, M. H. et al. A quick procedure for the isolation of dimeric piratoxins-I and II, two myotoxins from Bothrops pirajai snake venom. N-terminal sequencing. **Biochemistry and Molecular Biology International**, v. 37, n. 6, p. 1047–1055, dez. 1995.

TOYAMA, M. H. et al. Purification and Amino Acid Sequence of MP-III 4R D49 Phospholipase A2 from Bothrops pirajai Snake Venom, a Toxin with Moderate PLA2 and Anticoagulant Activities and High Myotoxic Activity. **Journal of Protein Chemistry**, v. 18, n. 3, p. 371–378, 1 abr. 1999.

UTKIN, Y. N. Animal venom studies: Current benefits and future developments. **World Journal of Biological Chemistry**, v. 6, n. 2, p. 28–33, 26 maio 2015.

VERHEIJ, H. M. et al. Methylation of histidine-48 in pancreatic phospholipase A2. Role of histidine and calcium ion in the catalytic mechanism. **Biochemistry**, v. 19, n. 4, p. 743–750, 19 fev. 1980.

VERONESE, E. L. G. et al. Inhibition of the myotoxic activity of Bothrops jararacussu venom and its two major myotoxins, BthTX-I and BthTX-II, by the aqueous extract of Tabernaemontana catharinensis A. DC. (Apocynaceae). **Phytomedicine: international journal of phytotherapy and phytopharmacology**, v. 12, n. 1–2, p. 123–130, jan. 2005.

VONRHEIN, C. et al. Data processing and analysis with the autoPROC toolbox. **Acta Crystallographica. Section D, Biological Crystallography**, v. 67, n. Pt 4, p. 293–302, abr. 2011.

WANG, B. C. Resolution of phase ambiguity in macromolecular crystallography. **Methods in Enzymology**, v. 115, p. 90–112, 1985.

WARRELL, D. A. Snake bite. **Lancet**, v. 375, n. 9708, p. 77–88, 2 jan. 2010.

WATANABE, L. et al. Structural insights for fatty acid binding in a Lys49-phospholipase A2: crystal structure of myotoxin II from Bothrops moojeni complexed with stearic acid. **Biochimie**, v. 87, n. 2, p. 161–167, fev. 2005.

WATSON, J. D.; CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. **Nature**, v. 171, n. 4356, p. 737–738, 25 abr. 1953.

WINN, M. D. et al. Overview of the CCP4 suite and current developments. **Acta Crystallographica Section D Biological Crystallography**, v. 67, n. 4, p. 235–242, 18 mar. 2011.

WORLD HEALTH ORGANIZATION. **Rabies and envenomings: a neglected public health issue.** Geneva: W.H.O., 2007.

XIMENES, R. M. et al. Harpalycin 2 inhibits the enzymatic and platelet aggregation activities of PrTX-III, a D49 phospholipase A2 from Bothrops pirajai venom. **BMC complementary and alternative medicine**, v. 12, p. 139, 2012.

XU, Q. et al. Statistical analysis of interface similarity in crystals of homologous proteins. **Journal of Molecular Biology**, v. 381, n. 2, p. 487–507, 29 ago. 2008.

XU, Q.; DUNBRACK, R. L. The protein common interface database (ProtCID)--a comprehensive database of interactions of homologous proteins in multiple crystal forms. **Nucleic Acids Research**, v. 39, n. Database issue, p. D761-770, jan. 2011.

YUNES QUARTINO, P. J.; BARRA, J. L.; FIDELIO, G. D. Cloning and functional expression of secreted phospholipases A(2) from Bothrops diporus (Yarará Chica). **Biochemical and Biophysical Research Communications**, v. 427, n. 2, p. 321–325, 19 out. 2012.

ZWART, P.; GROSSE-KUNSTLEVE, R.; ADAMS, P. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. **CCP4 newsletter**, v. 43, Winter 2005.

5 SUPPLEMENTARY MATERIAL

Supplementary Table 2.11 - Summary of PLA_2-like crystallographic models and their dissociation calculations

PDBs	ΔG^{diss} (kcal/mol)		BU	Unit Cell						SG	Res	Crystallization condition
	LD	CD		a	b	c	α	β	γ			
1PA0	-7.2	0.1	2	57.3	57.3	130.9					2.20	PEG6k, ammonium sulphate, sodium cacodilate, pH 6.7
2Q2J	-7.6	3.2	2	55.9	55.9	127.9					1.65	30% PEG4k, 0.1 M TRIS HCl, 0.2 M lithium sulfate, pH 7.8
3HZD	-6	4.9	2	55.7	55.7	127.8					1.91	26% PEG4k, 0.1 M TRIS-HCl, 0.2 M lithium sulfate, pH 8.5
3I3H	-6	4.3	2	55.9	55.9	127.8	90	90	120	$P3_121$	2.17	30% PEG4k, 0.1 M TRIC-HCl, 0.18 M lithium sulfate, pH 8.5
4K09	-7.5	1.5	2	56.4	56.4	129.1					2.11	30% PEG4k, 0.1 M TRIS-HCl, 0.25 M lithium sulfate, pH 8.5
4WTB	-6.9	7.2	2	56.6	56.6	129.5					2.16	30% PEG4k, 0.1 M TRIS-HCl, 0.1 M lithium sulfate, pH 8.5
2H8I	-7.1	5	2	56.0	56.0	127.6					1.90	20% PEG400, 0.1 M HEPES, 2 M ammonium sulphate, pH 7.5
3HZW	-8.3	0.9	2	50.7	62.5	87.1					2.28	20% PEG4k, 0.1 M sodium citrate, 20% isopropanol, pH 5.6
4KF3	-8.8	-0.6	1	50.0	62.2	86.0					1.92	20% PEG4k, 0.1 M sodium citrate, 20% isopropanol, pH 5.6
4YV5	-8	-0.2	2	50.8	63.6	87.7	90	90	90	$P2_12_12_1$	1.90	30% PEG4k, 0.1 M TRIS-HCl, 0.2 M lithium sulfate, pH 8.5
1XXS	-8	-1.1	2	61.2	88.7	51.1					1.80	0.6-0.8 M sodium citrate, pH 6.0
1Y4L	-8.2	-1.2	2	49.2	64.0	86.0					1.70	PEG3350, isopropanol, sodium citrate, pH 4.6
3QNL	-8.9	4	2	49.4	67.0	85.5					1.77	20% PEG4k, sodium citrate, 20% isopropanol, pH 5.6
4YZ7	-6.2	2.5'	2	68.3	70.9	44.0	90	90	90	$P2_12_12$	1.95	PEG4k, TRIS-HCl, lithium sulfate, pH 8.1
3CYL	-8.1	1.2	2	38.9	71.4	44.3		102.3			1.87	30% PEG4k, 0.1 M TRIS HCl, 0.12 M lithium sulfate, pH 7.5
3CXI	-8.6	1.2	2	38.4	70.1	43.8		102.2			1.83	30% PEG4k, 0.1 M TRIS HCl, 0.12 M lithium sulfate, pH 7.5
4YU7	-8.4	0.1	2	39.2	72.1	44.6		102.1			1.65	PEG4k, TRIS-HCl, lithium sulfate, pH 8.1
2OK9	-9	0.7	2	38.6	70.0	43.8	90	102.6	90	$P2_1$	2.34	23% PEG4k, 0.1 M sodium citrate, 20% isopropanol, pH 5.8
4K06	-8.7	1.7	2	39.0	71.4	44.4		102.5			2.08	30% PEG8k, 0.1 M sodium cacodylate, 0.25 M ammonium sulfate, pH 6.5
3IQ3	-8.5	1.1	2	38.5	70.8	43.8		102.2			1.55	29% PEG4k, 0.1 M TRIS-HCl, 0.2 M lithium sulfate, pH 8.5
3MLM	-9.7	2.3	1	38.8	70.5	44.0		109.3			2.20	30% PEG4k, 0.1 M TRIC-HCl, 0.2 M lithium sulfate, pH 8.5
B1MMV	-9.7	2.9	2	104.9	104.9	64.4	90	90	120	$P3_221$	2.10	20% PEG4k, sodium citrate, 22% isopropanol, pH 5.6
1QLL	-9.6	6.6	2	46.2	60.4	58.7	90	96.0	90	$P2_1$	2.04	28% PEG3350, 0.1 M TRIS-HCl, 0.25 M lithium sulfate, pH 8.5

The abbreviations are related to the energy necessary to dissociate the dimer (ΔG^{diss}), Biological Unit suggested by the structure depositor (BU) and Space Group (SG). In PDBs column, font colored in magenta, red and dark green are a representation of asymmetric, exception and symmetric PLA_2-like properties, respectively. ' ΔG^{diss} refers to the model 4YZ7 after completing side chain atoms of 44 incomplete residues in the model, as without them, the value was -1.3.

Supplementary Table 2.12 - Summary of region binding of ligands of available bothropic PLA_2-like crystallographic models

PDBs	Ligands	Hydrophobic Channel	iFace	MDiS	Nt|H1|lp (outer)	R34	Bsh	H2	H3	Cterm	Crystal System
ASYMMETRICAL MONOMERS COMPLEXED TO NATURAL MOLECULES											
2Q2J	1 sulfate / 2 TRIS		AB							A	Trig
2H8I	1 PEG	A									
SYMMETRICAL MONOMERS COMPLEXED TO NATURAL MOLECULES											
4KF3	4 PEGs / 6 isopropanol	AB	AB		A	AB	A				Ort
1XXS*	4 stearic acids / 5 sulfate	AB	AB		A	AB	A			A	
3CYL	2 vitamin Es / 1 PEG / 5 sulfates	AB	AB		B	A	B		A		
3CXI	2 vitamin Es / 1 PEG / 4 sulfates	AB	AB		B		B		A		
4K06	3 PEGs / 5 sulfates	AB	AB		B	A	B		A		Mon
3IQ3	3 PEGs / 2 sulfates	AB			B	A	B		A		
3MLM	2 myristic acids / 4 sulfates	AB	AB			AB					
1QLL	2 N-tridecanoic acids	AB									
COMPLEXES WITH INHIBITORS											
4WTB*	3 zinc / 2 sulfates	A B	AB	AB		A		A			Trig
3HZW	2 BPB / 1 isopropanol	AB				B					
4YV5*	2 suramin / 3 PEGs / 7 sulfates	AB	AB AB	AB	AB A	AB	AB A			AB A	
1Y4L*	1 suramin / 2 PEGs / 5 isopropanol	AB AB	AB		AB	AB		AB B			Ort
3QNL	1 rosmarinic acid / 1 PEG / 8 isopropanol	B AB	A	B	A		B	B	A		
4YZ7	1 aristolochic acid / 1 PEG / 5 sulfates	B	B	A	B	AB			A		
4YU7	4 caffeic acids / 3 PEGs / 1 sulfate	AB	AB		A	A	A		A		Mon
2OK9	2 BPBs / 2 isopropanol	AB AB									
MMV	2 MMV	AB[#]			AB		AB	AB	AB		Trig

In PDBs column, font colored in magenta, red and dark green are a representation of asymmetric, exception and symmetric PLA_2-like properties, respectively. In the other column, negative molecules, PLA_2-like protein inhibitors, and natural compounds are colored in orange, light blue and dark green, respectively. * PDBs had their chain letter A and B inverted. The abbreviations refers to protein region that interact to the membrane (iFace), Membrane disruption site (MDiS), exposed atoms of residues of N-terminal α-helix and loop (Nt| H1|lp (outer)), arginine at position 34 (R34), β-sheets (Bsh), helix 2 (H2), helix 3 (H3), C-terminal (Cterm), Trigonal (Trig), Orthorhombic (Ort), and Monoclinic (Mon). For the site calculation, we considered the basic residues 16-17, 20, 115 and 118 for iFace; 120-126 for MDiS; 1-15, 18-19, 21-24 for Nt| H1|lp (outer); 72-88 for Bsh; 35-47, 49-69 for H2; 91-109 for H3; and 110-114, 116-117, 119 and 127-133 for Cterm. The hydrophobic channel herein is described as the proposed tunnel, which is composed of the inner part of N-terminal residues and the calcium binding loops. [#] MMV interacts only to the calcium binding loop and not to the inner part of N-terminal residues.

Supplementary Table 2.13 - List of residues that composes the inner and outer of the BthTX-I/Zn dimer cavity and their chemical properties calculated by MOLE 2.0.

Residues:	Charge		Polarity	Hydropathy
	Positive	Negative		
Outer: G6/A, N17/A, P18/A, K20/A, L32/A, K49/A, Y52/A, K53/A, P68/A, K69/A, Y119/A, L2/B, K20/B, G30/B, V31/B, K49/B, Y52/B, K115/B, R118/B, H120/B, and L121/B	8	0	12.7	-0.93
Inner: S1/A, L2/A, F3/A, L5/A, K7/A, I9/A, A19/A, S21/A, Y22/A, G23/A, A24/A, Y25/A, N28/A, C29/A, G30/A, V31/A, R34/A, C45/A, H48/A, D99/A, V102/A, L106/A, R118/A, H120/A, L121/A, K122/A, P123/A, L5/B, G6/B, I9/B, N17/B, P18/B, A19/B, S21/B, Y22/B, G23/B, A24/B, Y25/B, N28/B, L32/B, C45/B, H48/B, V102/B, L106/B, and Y119/B	4	1	6.6	-0.07

Extracted from (BORGES et al., 2017).

Supplementary Table 2.14 - List of residues that composes the tunnels and the chemical properties of these tunnels calculated by MOLE 2.0.

Residues:	Length	Positive	Polarity	Hydropathy
Tunnel 1: H48A, C45/A_{bb}, K49A, N28A_{bb}, C45A_{bb}, G30A_{bb}, C29A_{bb}, L5A, I9A, Y22A, G23A_{bb}, P18A_{bb}, L2A, P18A, A19A_{bb}, H120B_{bb}, G6A, L121B, L2A_{bb}, H120B, A19A	19.1 Å	1	10.4	0.29
Tunnel 2: H48A, C45A_{bb}, K49A, N28$_{bb}$, C45A, G30A_{bb}, L5A, K69A, L2A, G23A_{bb}, V31A, H120B_{bb}, A24B, Y119B_{bb}, A19B_{bb}, K20B_{bb}, H120B, A19A_{bb}, K20B, Y119A, R118B	22.8 Å	-0.67	17.3	-0.07
Tunnel3: H48B, C45B, N28B_{bb}, C29B_{bb}, G30B_{bb}, K49B, Y52B	6.0 Å	1	20.32	-0.98

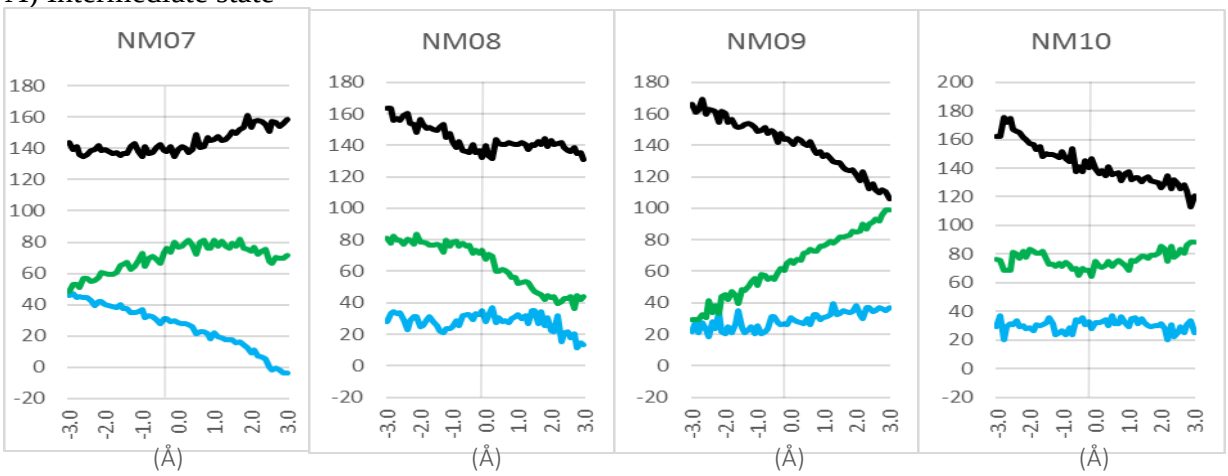
$_{bb}$: backbone. Extracted from (BORGES et al., 2017).

Supplementary Table 2.15 - BthTX-I/MMV comparison with other bothropic PLA_2-like toxin structures in RMSD (Cα).

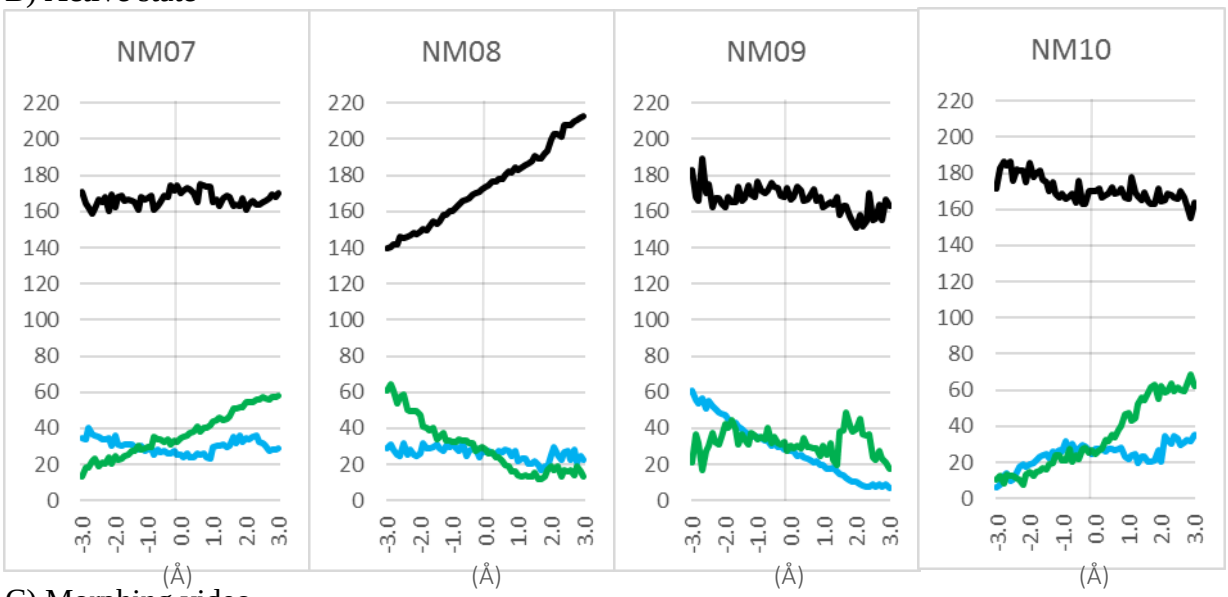
PDB ID monomer	A	B	Toxin	Ligand
4WTB A	1.60	0.74	BthTX-I	Zn^{2+}
4WTB B (non-canonical monomer)	1.67	1.27		
2H8I A (non-canonical monomer)	1.70	1.14	BthTX-I	PEG
2H8I B	1.67	0.70		-
1PA0 A	1.59	0.74	BnSP-VII	-
1PA0 B (non-canonical monomer)	1.62	1.20		
2Q2J A (non-canonical monomer)	1.70	1.18	PrTX-I	-
2Q2J B	1.62	0.75		
3HZD A (non-canonical monomer)	1.68	1.19	BthTX-I	-
3HZD B	1.63	0.73		
3I3H A (non-canonical monomer)	1.71	1.13	BthTX-I	-
3I3H B	1.64	0.94		
4K09 A (non-canonical monomer)	1.64	1.22	BbTX-II	-
4K09 B	1.58	0.72		
4KF3 A	1.93	1.27	MjTX-II	PEG4k
4KF3 B	1.89	1.25		
4YV5 A	1.93	1.27	MjTX-II	suramin/PEG4k
4YV5 B	1.92	1.27		
1XXS A	1.87	1.31	MjTX-II	stearic acid
1XXS B	1.87	1.30		
3HZW A	1.60	0. 96	BthTX-I	BPB
3HZW B	1.65	0. 70		
1Y4L A	1.65	0.47	BaspTX-II	suramin/PEG
1Y4L B	1.57	0.82		
3QNL A	1.49	0.51	BthTX-I	rosmarinic acid/PEG
3QNL B	1.52	0.44		PEG/isopropanol
4YZ7 A	1.66	0.64	PrTX-I	-
4YZ7 B	1.65	0.72		aristolochic acid/PEG4k
3MLM A	1.61	0.70	BnIV	Myristic acid
3MLM B	1.61	0.87		
4K06 A	1.68	0.63	BbTX-II (MTX-II)	PEG
4K06 B	1.59	0.91		
3CYL A	1.72	0.47	PrTX-I	vitE/PEG
3CYL B	1.58	0.92		
3IQ3 A	1.74	0.49	BthTX-I	PEG4k
3IQ3 B	1.61	0.89		
4YU7 A	1.68	0.46	PrTX-I	caffeic acid/PEG4k
4YU7 B	1.62	0.71		
3CXI A	1.72	0.53	BthTX-I	vitE/PEG
3CXI B	1.61	0.91		
2OK9 A	1.64	0.75	PrTX-I	BPB
2OK9 B	1.57	0.88		
MEAN	1.67	0.87		
MINIMUM	1.49	0.44		

Superposition is done with *lsqkab* using all Cα.

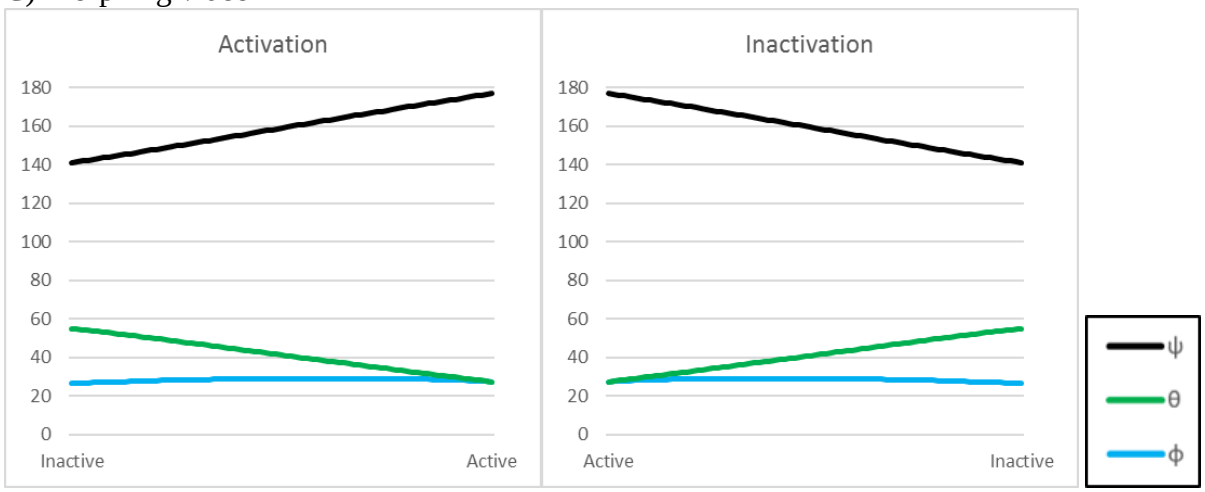
A) Intermediate state



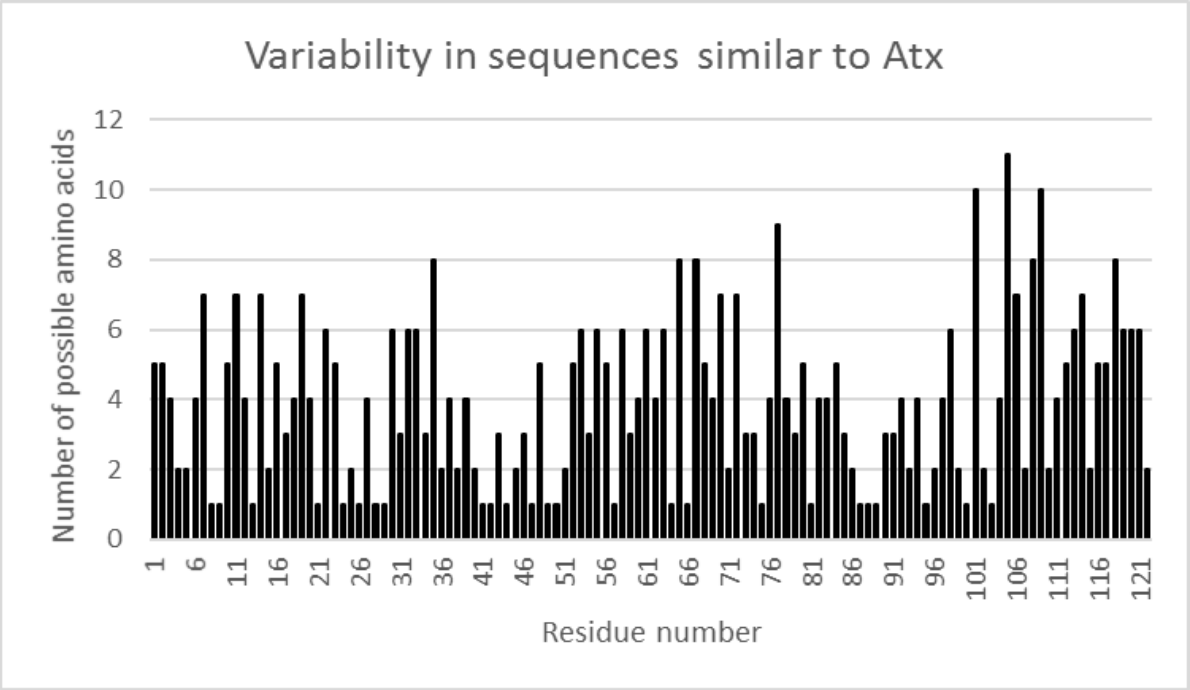
B) Active state



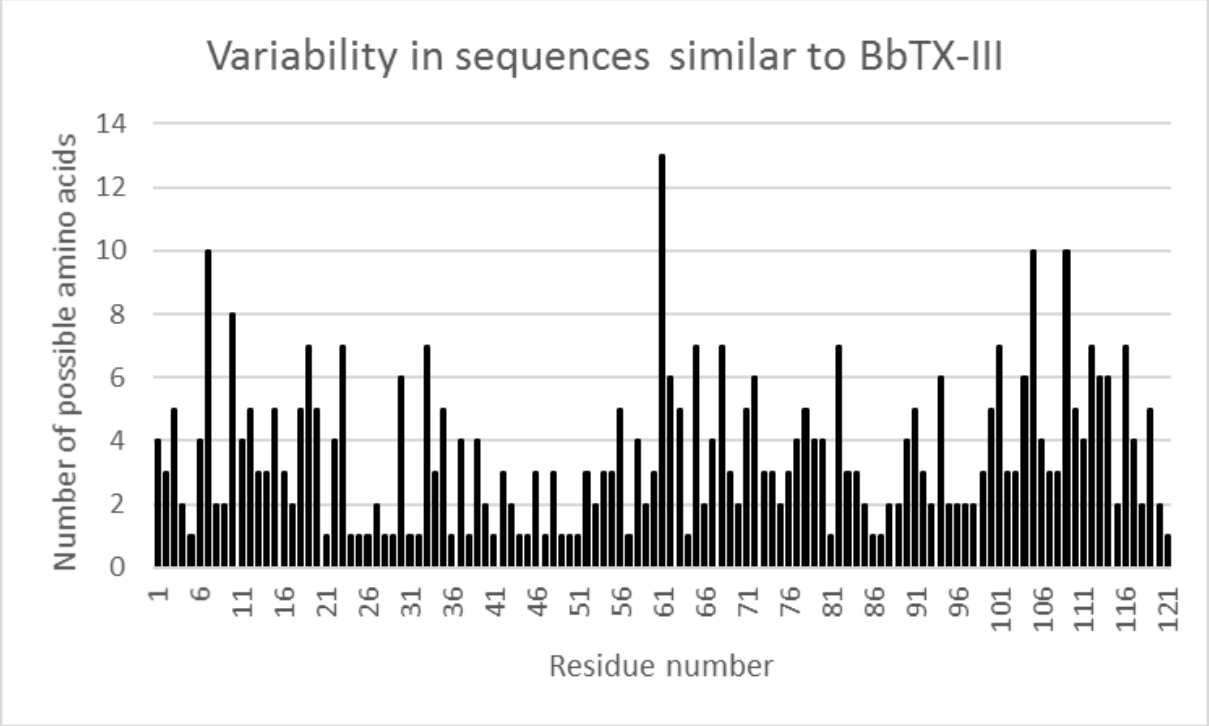
C) Morphing video



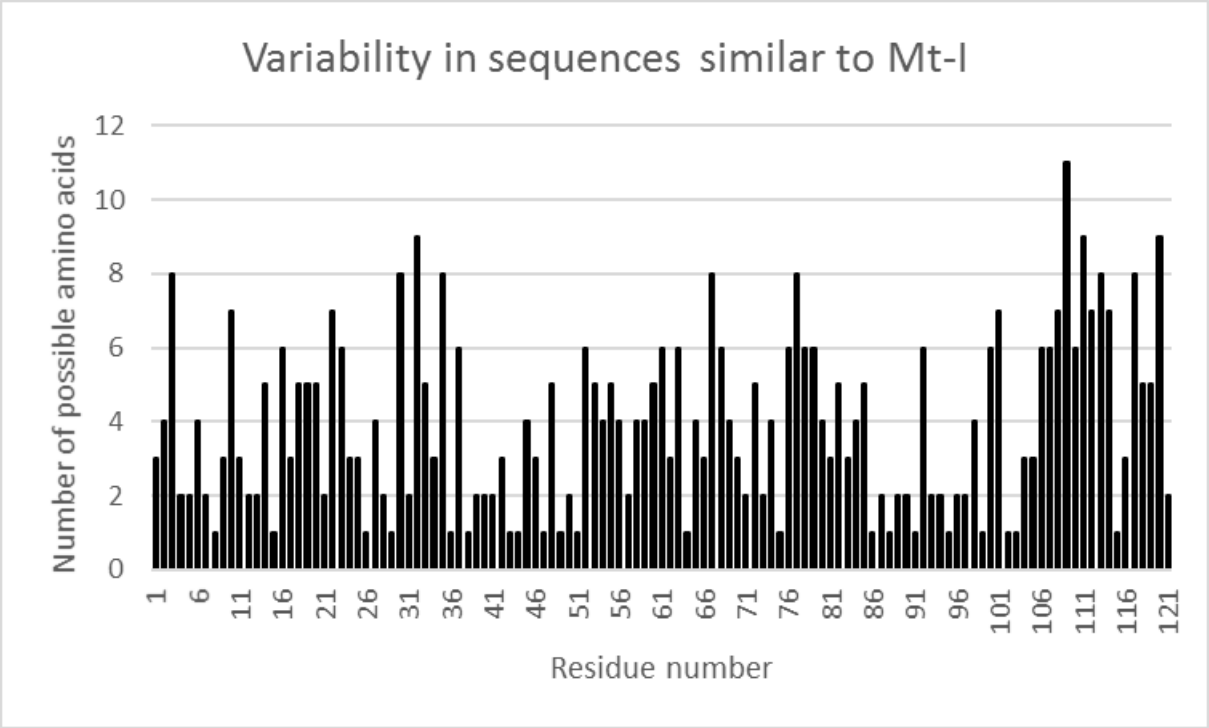
Supplementary Figure 2.26 - Normal Mode (NM) 7 to 10 analysis of structures in intermediate (in **A**) and **active state** (in **B**) and their monomer-monomer angles (ψ, θ, and φ).
In **C**, the angles of morphing video of activation (**inactive** to **active** state) and or inactivation (from **active** to **inactive**)



Supplementary Figure 3.27 - Graph of local variability of similar sequences to Atx.



Supplementary Figure 3.28 - Graph of local variability of similar sequences to BbTX-III.



Supplementary Figure 3.29 - Graph of local variability of similar sequences to Mt-I.

Supplementary Table 3.15 - Correlation coefficient of residues whose scored matched mass spectrometry results of BbTX-III

R#	AA	CC	Dif	F	MS	R#	AA	CC	Dif	F	MS
2	L	96.3	3	97	L	70	N	92.9	5.6	99	N
3	V	96.5	**1.4**	22	V	71	G	93.9	12	94	G
4	Q	96.6	3.9	99	Q	74	V	93.9	5.9	77	V
5	F	97.3	3.3	100	F	75	C	95.9	19	99	C
6	E	96.1	2.2	94	E	76	G	86.9	22	88	G
9	I	96.4	2.3	99	I	77	G	92.3	10	82	G
14	G	91.8	8.5	64	G	79	D	89.4	3	74	D
21	Y	97.4	4.7	100	Y	80	P	88.4	2.8	94	P
22	G	98.4	10.6	48	G	81	C	94.9	14	100	C
24	Y	95.8	3.7	100	Y	82	K	93	**1.1**	62	K
25	G	96.8	8.8	100	G	83	K	94.5	4.8	89	K
26	C	98.1	15.7	100	C	84	Q	94	1.8	39	Q
28	C	97	11.3	100	C	85	I	96.7	4.2	72	I
29	G	97	16.7	100	G	86	C	97.1	9.2	100	C
32	G	93	26	100	G	87	E	96.1	3.4	100	E
34	G	96.6	8.6	98	G	88	C	97.4	12	99	C
36	P	97.4	10	100	P	89	D	96.9	**0.7**	99	D
37	Q	92.5	7.1	93	Q	90	R	95.9	1.5	36	R
38	D	95.4	2.5	100	D	91	V	96.8	4	26	V
39	A	89.6	3	81	A	92	A	97.3	5.1	98	A
40	S	95.5	6.5	51	S	93	A	95.9	1.5	99	A
41	D	97.5	2.2	100	D	95	C	97.5	8.8	99	C
42	R	96.1	1.5	97	R	96	F	96.4	3.1	99	F
43	C	98	9.3	99	C	97	R	93.4	1.8	99	R
44	C	97.4	16.5	100	C	99	N	96.4	2.5	98	N
45	F	95.8	1.9	100	F	101	D	91.2	11	54	D
46	V	96.8	2.1	98	V	102	T	93.1	4.8	90	T
47	H	96.2	1.5	100	H	103	Y	93.3	**1.2**	98	Y
48	D	96.7	**0.9**	98	D	105	S	93	6.6	20	S
49	C	97.6	10.4	100	C	107	Y	95.7	6.6	98	Y
50	C	98.2	10.9	100	C	108	W	89.9	**1.3**	97	W
51	Y	97	4.7	100	Y	109	F	94.5	5.3	37	F
52	G	94.8	7.5	98	G	110	F	95.5	2.6	62	F
53	K	93.7	5.5	90	K	111	P	96.6	5.5	88	P
57	C	94.8	16.2	100	C	112	A	90.7	6.9	53	A
59	P	96.9	7.1	91	P	114	L	94	7.7	0	**?**
60	K	95.8	4.4	98	K	115	C	94.2	20	99	C
64	Y	96.1	4.6	100	Y	119	E	96.5	2.2	91	E
66	Y	94.7	5.2	96	Y	120	P	97.2	9.9	99	P
68	E	94.5	5.8	84	E	121	C	96.7	23	100	C

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), difference of the PCC of the shown residue to the second best score (Dif), frequency of residue in given position from phylogenetic analysis (F) and true residue (!). Differences below 1.5% are in bold.

Supplementary Table 3.16 - SEQSLIDER summary of BbTX-III run with phylogenetic restriction of residues having more than one mass spectrometry match

R#	AA	CC	F	M
62	D!	94	75	D/N
	N	89.1	2	
	I	85.5	1	
	E	84.3	9	
	A	82.8	6	
	V	80.1	7	
63	V!	96.3	22	F/T/V
	T	92.5	10	
	I	90.9	7	
	S	88.9	26	
	F	49	35	
65	T!	93.6	81	T/V
	V	89.9	3	
	S	88.2	8	
	I	86.4	3	
	D	69.4	1	
	K	59.9	3	
	R	37	1	

R#	AA	CC	F	M
72	V!	94.5	6	A/V
	A	92.6	31	
	G	91.6	2	
	N	67.1	6	
	D	63.8	32	
	E	55.1	23	
73	V!	94.5	19	I/V
	I	87.2	80	
	L	85.7	1	
94	T!	97	7	I/T
	V	96.5	7	
	I	89.9	82	
	N	83.4	1	
	D	82.9	1	
	M	75	2	

R#	AA	CC	F	M
98	D!	94.7	99	D/E
	E	88.3	0	
	R	69.9	1	
100	K!	93.7	29	K/M
	I	79.9	33	
	V	79.4	4	
	L	75.4	22	
	M	60.1	12	
105	S	90.8	20	N/S
	G	85.9	1	
	I	81.8	1	
	N	72.4	61	
	K	70.6	1	
	D	69.3	5	
	H	61.8	2	
	Y	54.9	1	
	R	54.3	1	
	W	45.6	6	

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). Residues numbers in bold represent side chain atoms were removed due to lack of electron density. * 10M has a double occupancy, only when the double occupancy is accounted that M is best scored.

Supplementary Table 3.17 - SEQSLIDER summary of BbTX-III run with phylogenetic restriction of unknown residues

R#	AA	CC	F
54	V!	96.5	89
	L	88.4	1
	A	84.6	10
55	T!	92.6	77
	N	76.1	20
	D	74.4	3

R#	AA	CC	F
56	S!	92.4	3
	G	87.6	72
	N	76.8	3
	D	75.6	21
	E	70.5	1
58	D!	95.4	67
	N	92.1	30
	S	87.2	1
	K	84.9	2

R#	AA	CC	F
59	P!	96.9	91
	T	83.7	9
60	K!	95.8	98
	Q	86.4	1
	T	84	1
114	S	86.3	1
	N	83.7	79
	D	82.9	2
	T	78.6	1
	K	70.7	2
	Y	57.4	15

R#	AA	CC	F
118	S!	93	75
	P	67.8	25
119	E!	96.5	91
	Q	94.3	1
	G	87.2	1
	D	81.5	1
	V	71.3	6
120	P!	97.2	99
	Q	80.2	1

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). Residues numbers in bold represent side chain atoms were removed due to lack of electron density.

Supplementary Table 3.18 - SEQSLIDER summary of BbTX-III residues whose electron density was poor

R#	AA	CC	F	M
13	A	96.6	94	T
	V	83.3	3	
	T!	75.9	3	
30	A	92.7	25	S
	V	90	0	
	S	86.4	7	
	L	76.2	5	
	K	54.4	18	
61	A	97.2	5	E/S
	S	86.5	1	
	V	82.2	4	
	T	75.2	21	
	I	74.5	1	
	L	62.5	12	
	K	55.3	11	
	N	48.8	1	
	M	42.5	15	
	D	42.5	23	
	E	40.1	4	
	F	34.1	1	
113	S	70	4	K
	N	57.9	1	
	K!	47	72	
	Q	46.8	6	
	E	44.1	15	
	R	40.4	2	
116	L	60.2	1	-
	P	49.3	5	
	N	44.7	2	
	Q!	44.4	52	
	K	41.5	20	
	R	28.1	9	
117	G	66.7	1	-
	K	59	1	
	Q	48	7	
	E!	45.5	91	

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). Residues numbers in bold represent side chain atoms were removed due to lack of electron density.

Supplementary Table 3.19 - Correlation coefficient of residues with clear distinction of Mt-I

R#	AA	F	A		B		AB		MS
			CC	Dif	CC	Dif	CC	Dif	
1	S	94	93.5	1.4	93.1	1.9	93.1	1.8	
2	L	52	95	1.9	93.5	3.6	94.6	4.2	
4	Q	12	93.4	2.3	93.2	3.1	93.3	3.5	
5	L	84	95.3	1.4	96.2	3.3	95.8	2.1	
6	G	86	95.6	2.8	94.6	2.6	95.4	2.7	
8	M	99	97.2	4	97.6	3	97.2	3.6	
9	I	93	95.8	3.8	95.7	0.8	95.6	2.6	
10	L	57	95.1	2.5	93.9	0.2	94.3	1.3	
12	E	96	94.8	0.4	94.6	2.5	94.7	1.9	
13	T	96	95.2	6.3	94.5	2.5	94.7	3.8	
14	G	81	93	6.9	93.1	8.2	93.2	6.8	
17	P	71	96.3	5.1	95.3	10.3	96.1	9.8	
18	A	39	93.9	3.5	95.7	5.7	94.7	6.2	
19	K	41	96.5	2.4	94.3	3.3	94.5	2.7	
20	S	81	93.7	4.2	95.6	3.6	94.4	4.8	
21	Y	99	95.8	5.6	95.7	5	95.9	6	
22	G	70	96.1	6.7	94.7	8.8	94.9	8.2	
23	A	32	94.2	6.4	95.9	6.7	95.1	6.5	
24	Y	98	96.5	5.7	95.8	3.3	95.8	4.8	
25	G	97	97.2	5	96.8	5.4	97	5.2	
27	N	78	95.9	1.1	97.2	1.4	96.3	1.3	
28	C	99	90.5	2	92.1	1.7	90.8	1.3	
29	G	100	90.4	18.1	90.6	19.6	90.2	20.1	
32	G	46	91.1	9.7	88.9	6.6	90.2	7.8	
34	G	87	93.2	5.2	94.5	10.1	93.7	8.1	
36	P	100	93.9	5.6	93.5	5.9	93.8	8.1	
37	K	84	93.9	3.2	92	3.2	92.7	3.5	
38	D	100	93.5	1.3	96.8	1.9	95.1	1.8	D
39	A	98	94.4	6.8	94.1	5	94.4	6.7	A
40	T	99	97.2	4.9	95.9	0.7	96.5	3.4	T
41	D	99	95.7	2.3	95.3	2.4	95.4	3	D
45	F	58	92.3	1.6	95.3	3.1	92.6	2	F
46	V	91	95.5	3.1	95	3.2	94.3	2.4	V
47	H	100	96.5	2.1	94.9	2.3	95.7	2.1	H
51	Y	100	94.2	4.8	94.5	4.5	94.1	5.4	Y
54	L	86	89.4	4.4	92.6	9.3	89.8	6.7	
56	G	41	84.7	9.7	71.8	21.5	78.2	16.3	
59	P	83	92.4	8.2	94.3	6.9	92.6	6.7	
64	Y	100	93.4	3.9	94.6	5.6	93.7	5.5	Y
65	S	79	88.5	3.3	88.7	0.9	87.7	0.4	T
66	Y	98	96.4	6.5	93.5	4.2	94.7	5.5	Y
68	W	80	94.3	4.1	95.4	10.3	94.9	7.8	E
71	K	78	92.7	5.6	92.9	5.2	92.6	6.6	G
73	I	99	95.7	4.9	95.5	2.5	95.5	3.5	V
74	V	77	95.1	1.7	94.5	1.3	94.8	3.2	V
76	G	72	65.4	8	90.3	15.9	79.5	12.6	G
82	L	71	90.9	4.3	92.2	2.1	91.1	3.1	K
83	K	94	92.1	2	89.2	1.5	90.4	1.9	K
85	L	33	95.6	3.7	96	3.8	95.3	3.3	I

Supplementary Table 3.19 (continued) - Correlation coefficient of residues with clear distinction of Mt-I

R#	AA	F	A		B		AB		MS
			CC	Dif	CC	Dif	CC	Dif	
89	D	99	97.7	1.6	97.5	1.3	97.5	2	D
90	K	85	93.8	0	95.7	5.4	94.7	3.2	K
91	A	100	95.3	6.2	94.1	4.4	95	5.4	
92	V	74	96.6	0.3	96.7	0.7	96.5	0.6	
93	A	99	95.6	5.8	94.4	3.2	94.7	6	
94	I	99	94.3	3.1	94.8	2.3	95	2.5	
96	L	72	95.5	1.4	94.2	0.7	95	1.3	
99	N	100	95.4	1.5	96.9	0.8	96.3	0.9	
100	L	93	94.1	2.2	94.1	2.6	94.4	2.9	
101	N	11	93.6	1.3	92.8	4.1	93.7	1.7	
102	T	100	93.4	4.2	93.2	1.4	93.5	2.3	
103	Y	100	96.3	5.6	94.4	3.9	95.9	5.1	
104	N	85	95	2.1	92.9	3.4	93.7	2.7	
107	Y	80	95.1	4.4	93.3	5.5	93.5	4.5	
110	Y	58	93.8	0.6	94.6	0.7	94.4	1.1	
113	P	37	87.5	6.1	79.6	0.2	82.5	3.6	
118	G	1	82.6	2	73.2	26.6	78.4	14	

The abbreviations represent Residue number (R#), amino acid (AA), frequency of residue in given position from phylogenetic analysis (F), monomer A, B and AB (A, B, and AB, respectively), real space correlation coefficient of polder map (CC in %), difference of the PCC of the shown residue to the second best score (Dif). Residues numbers in bold represent side chain atoms were removed due to lack of electron density.

Supplementary Table 3.20 - SEQSLIDER run on Mt-I and summary of convergence results

R#	AA	CCA	CCB	CCAB	F	MS
1	S!	93.8	92.9	93.2	94	
	G	86.3	84.8	85.1	1	
	N	78.3	74.1	76.6	3	
2	L!	94.5	92.5	91.7	52	
	V	86.4	86.9	87.1	41	
	Q	79.9	74.4	78.2	1	
	M	76.3	74.3	75.9	3	
5	L!	94.5	95.9	94.9	84	
	F	82.4	79.8	81.4	15	
6	G!	97.2	95.1	96.3	86	
	R	80.9	74.3	76.6	1	
	T	76.6	65.7	74.4	4	
	W	65.4	62.1	66.3	8	
7	K!	94.2	93.8	93	98	
	E	83.2	77.9	81.2	1	
8	M!	96.7	96.6	95.9	99	
9	I!	96.2	95.6	95.5	93	
	V	91.1	93.7	91.8	5	
	M	79.5	72.5	76.1	1	
10	L!	95.6	93.9	94	57	
	I	89.5	91.6	90	3	
	V	88.7	92.4	89.6	3	
	K	83.9	77.2	80.3	1	
	Q	77.7	76.1	78.7	15	
	F	62.8	74.5	68.1	17	
	W	62.5	61.2	62.5	3	
12	E!	94.3	93.2	92.9	96	
	L	83.4	87.2	85.3	3	
13	T!	95.2	95.6	93.8	96	
	M	72.9	77.9	66.3	3	
14	G!	93.3	93.5	92.1	81	
	A	86.7	83.4	86.3	1	
	N	57.1	53.1	54.5	3	
	E	48.3	48.8	51	5	
	D	43	50.9	45.7	9	
15	K!	93.7	91.6	91.5	100	
16	N!	94.7	94.6	93.7	88	
	S	89.8	90.3	90.5	1	
	I	88.1	87.5	88.1	2	
	G	90.7	86.1	87.9	1	
	E	79.9	86	82.6	5	
	T	80.5	80.5	81.9	3	
17	P!	96.5	95.9	95.4	71	
	S	89.7	80.6	85.7	1	
	A	87.1	85	85.2	28	
18	A!	92.6	96	92.5	39	
	V	79.8	89.2	85	21	
	I	77.8	81.2	75.7	24	
	L	71.4	80	73.5	11	
	F	57.6	70.7	64.1	2	

R#	AA	CCA	CCB	CCAB	F	MS
19	K!	94.7	95.1	93.7	41	
	I	91.1	90.9	89.9	1	
	A	89.9	91.1	89.9	3	
	T	82.9	88.9	85.2	51	
	P	81	83.5	83	4	
20	S!	95.2	95.7	94.4	81	
	N	86.6	85.9	85.7	11	
	K	80.5	82.5	80.5	1	
	H	74.2	70.1	71.3	2	
	Y	69.6	67.3	68.6	5	
21	Y!	96	95	95.4	99	
	L	83.1	86.5	84.2	1	
22	G!	96.7	95.8	95.7	70	
	A	90.7	88.7	89.3	3	
	S	76.8	72.3	75.6	18	
	I	68.5	70.1	73	1	
	C	76.1	69.3	72.2	1	
	V	69.3	65.8	68.7	3	
	T	68.8	60.7	65.7	4	
23	A!	95	96.5	95.2	32	
	S	90.4	89.6	89.6	17	
	V	87.4	82.5	85.4	5	
	I	75.6	69.5	72.8	1	
	L	71.7	76	71.1	30	
	F	47.6	45.7	47.3	15	
24	Y!	96.5	95	95.1	98	
	F	91.4	88.8	89.5	1	
	H	88.4	85.3	86.2	1	
25	G!	96.6	97.3	96.4	97	
	L	61.1	67.5	64	1	
	Y	53.1				
26	C!	92.6	88.8	90.4	100	
27	N!	96.3	96.8	95.8	78	
	H	89.5	89.6	88.6	15	
	F	87	86.2	86.4	4	
	Y	76.2	75.1	74.1	3	
28	C!	92.7	89.6	90.9	99	
	F	73.4	69.9	71.9	1	
29	G!	92.1	92.5	91	100	
30	V!	86.8	72.2	82	49	
	G	86.5	62.7	78.9	4	
	L	77.9	53.2	75.5	17	
	P	77.5	65.1	71.9	10	
	M	46.9	46.6	55.2	1	
	R	42.6	-	-	1	
	W	47.3	-	28.8	17	
31	G!	79	69.1	72.9	76	
	L	65.3	39.7	52.2	24	

Supplementary Table 3.20 (continued) - SEQSLIDER run on Mt-I and summary of convergence results

R#	AA	CCA	CCB	CCAB	F	MS
32	G!	92.9	84.6	88.1	46	
	A	85.2	79.7	82	2	
	S	63.7	57.2	62.2	10	
	E	59.9	55.1	61.2	1	
	K	61.1	52.5	59.8	3	
	R	61.8	55.9	55.3	10	
	D	62.4	46.5	54.7	3	
	H	58.6	46.6	53.7	13	
	N	50.5	50.4	52.2	12	
33	K!	86.3	86.7	83.5	18	
	P	82	75.2	77.5	1	
	R	78.3	80.8	77.4	76	
	Q	77.3	75	76.2	4	
34	G!	93.3	93.7	91.6	87	
	R	40.3	57.9	47.6	10	
	H	41.2	47.5	45.8	3	
36	P!	93.5	94.6	92.4	100	
37	K!	91.2	91.7	91.4	84	
	L	87	79.8	84.9	7	
	V	82.3	83.9	83.3	6	
	Q	83.5	79.8	80.5	1	
	R	82.1	79.1	79.4	1	
	M	76.1	74.7	74.9	1	
38	D!	94.8	96.4	94.3	100	D
39	A!	95.9	96.6	96	98	A
	G	90.2	88.9	88.8	2	
40	T!	96.5	96.8	96.1	99	T
	V	94	95.7	94.1	1	
41	D!	96.3	95.5	95	99	D
	N	94.7	92.6	93.1	1	
42	R!	87.3	94.2	92.2	94	R
	S	93.9	92	91.9	5	
43	C!	91.8	94.5	92.5	100	C
44	C!	90.2	91.2	89.9	100	C
45	F!	94.6	93.5	92.5	58	
	L	91	90.6	89.8	1	
	S	90.3	89.6	89.2	2	
	Y	89.8	88.3	89	39	
46	V!	96.5	96.1	96.4	91	
	L	85.9	84.3	88	4	
	M	74.3	71.7	73.8	5	
47	H!	96.8	95.2	95	100	H
49	C!	89.2	93.5	90.7	100	C
50	C!	75.6	80.8	78.5	99	C
51	Y!	95.4	94.4	95	100	Y

R#	AA	CCA	CCB	CCAB	F	MS
54	L!	89.7	88.4	89.2	86	
	I	80.8	80.4	79.3	2	
	V	80.4	77.9	77.9	9	
	M	70.8	65.7	69.5	3	
56	G!	83.3	75.5	81.4	41	
	D	46.1		27.9	54	
	N		47.1	24.6	4	
	Y		34.9		1	
57	C!	81.2	81.1	79.3	97	
	S	77.4	73.7	75.1	3	
59	P!	92.2	92.6	92.7	83	
	S	70.6	82.1	76.1	4	
	Q	71.7	74.7	73.2	1	
	H	60.2	60.2	60.4	12	
61	A!	86.9	94	86.3	0	
	T	78.9	69	75.1	38	
	K	69.8	68.4	69.7	55	
	R	67	55.1	61.2	1	A
	M	57.2	58.8	56.7	4	
	D	58.9	55.4	55.7	1	
62	D!	84.3	80.4	82.7	84	D
	N	78.7	70.6	75.4	15	
64	Y!	94.4	94.2	93.7	100	Y
66	Y!	95.9	95.2	95.1	98	
	F	90.4	90.8	90.6	1	Y
	H	88.4	88	88	1	
67	S!	92.2	95	93.6	75	
	G	85.9	85.6	85.5	1	
	K	76.7	76.7	75.5	6	
	D	65.5	77.8	72.5	1	S
	Q	71.3	71.7	71.2	1	
	R	67.3	71.3	69.3	1	
	H	62	60.3	59.7	14	
72	T	93.1	93.9	92.9	37	
	V!	93.6	93.1	92.7	0	
	A	86.5	94.1	91	56	V
	D	79.2	77.5	79	2	
	E	64.7	64.7	65.5	4	
74	V!	94.1	93.1	93.3	77	
	I	92.8	88.2	90	20	V
	L	82	79.4	79.8	2	
75	C!	75.4	64.8	71.8	100	C
76	G!	73	86	75.5	72	
	D	54.2	58	54.2	3	
	K	54.7	33.4	41.8	3	G
	E	27	55.8	39.4	20	
	R	53.3	27.1	36.4	1	
81	C!	71.5	78.9	76.2	95	C

Supplementary Table 3.20 (continued) - SEQSLIDER run on Mt-I and summary of convergence results

R#	AA	CCA	CCB	CCAB	F	MS
83	K!	93	87.9	90.6	94	K
	Q	78.9	83	81.2	5	
86	C!	93.6	94.1	93.3	100	C
87	E!	92.9	92.8	92.3	99	E
88	C!	82.8	87.7	84.9	100	C
89	D!	97.8	96.9	97.2	99	D
	G	88.4	90	88.5	1	
90	K!	94.5	95.4	95	85	K
	R	88.2	84.8	86.6	15	
91	A!	96.9	94.9	95.9	100	
92	V!	97	97.2	97	74	
	A	86.2	89.2	87.9	17	
	L	87.1	86.7	86.7	1	
	M	83.8	82.2	82.7	4	
	F	76.6	73.4	74.6	3	
93	A!	97.6	96.7	97.1	99	
	T	89.5	83.5	86.5	1	
94	I!	95.2	96.2	95.7	99	
	V	93.6	91.4	92.8	1	
95	C!	89	90.7	89.6	100	
96	L!	96.2	95.8	96.1	72	
	F	84.8	84.1	84.5	28	
97	R!	95.9	90.1	93.1	98	
	G	88.6	83.1	86.4	2	
98	E!	94.4	92.2	93.3	89	
	K	91.2	88.3	89.9	4	
	V	87.4	84	86.5	1	
	D	84.5	78.6	81.9	6	
99	N!	96.1	97.3	96.7	100	
100	L!	94.1	94.3	94.4	93	
	K	89.8	89.2	89.7	1	
	I	85.6	89.6	88	1	
	V	82	86.9	84.3	1	
	P	80.7	79.7	80.8	1	
	M	75.9	80.7	79.5	3	
101	N!	91.4	90.8	90.8	11	
	G	88	82.2	86	25	
	E	83.5	85.6	84.9	2	
	D	88.2	70.1	80.8	43	
	Q	84.6	73.2	79.4	2	
	K	70.6	83	78.9	16	
	R		62.6	52.8	1	

R#	AA	CCA	CCB	CCAB	F	MS
102	T!	95.3	94.9	95.2	100	
103	Y!	96.2	95.6	95.5	100	
104	N!	93.7	94.1	93.4	85	
	D	91.1	89.7	89.7	14	
	Q	86.9	85.6	85.7	1	
105	K!	90.5	92	91.1	97	
	N	80.4	79.6	81	2	
	E	76.8	73.8	75.5	1	
107	Y!	95.3	93	93.5	80	
	F	89.9	88.9	88.9	3	
	K	87.7	86.8	87.5	8	
	H	87.5	86.4	86.2	5	
	R	87.8	84.1	85.6	1	
	E	81.5	81.8	81.2	3	
108	R!	95.4	95.1	95.2	43	
	K	90	93.7	91.1	47	
	D	79.9	85.8	82.7	1	
	N	78.6	85.1	81.9	3	
	Q	80.2	84.7	81.7	1	
	T	73.8	81	76.9	4	
	W	66	64.9	65.3	1	
110	Y!	93.2	94.1	93.3	58	
	F	90.5	91.8	90.4	2	
	H	88	89.5	88.1	20	
	K	87.9	85.8	85.6	2	
	N	81.7	86	82	9	
	T	68.9	78.9	72	9	
113	P!	86.9	86.4	86.8	37	
	V	73.4	70.3	72.4	2	
	I	73.9	66.8	70.8	4	
	L	66.6	56.9	62.8	16	
	K	55.7	60	57.8	1	
	F	40.9	50.1	46.8	35	
	Y	38.3	43.8	44.5	3	
	M	32.7	34.1	35	1	
114	K!	91	88.7	89.4	27	
	A	88.1	86.7	86.1	1	
	S	85.4	82.1	83.9	6	
	L	84.6	79.6	82.6	24	
	F	84.7	76	80.4	39	
	Y	79.9	75.1	76.6	1	
	R	72.4	79.9	73.9	1	
115	C!	77.2	77.2	77.7	99	
118	G!	78.8	72.6	75.6	1	
	A	78.5	71.6	73.8	35	
	P	64.3	49	55.9	36	
	S	63.9	30	48.6	24	
	T	52.1		39.2	3	

The abbreviations represent Residue number (R#), amino acid (AA), Real space correlation coefficient of polder map (CC in %), frequency of residue in given position from phylogenetic analysis (F) and chosen residue (!). Residues numbers in bold represent side chain atoms were removed due to lack of electron density. * 10M has a double occupancy, only when the double occupancy is accounted that M is best scored.

6 ATTACHMENTS

8.11 Euler Angle Sequence (3,2,1)

$$R_{321}(\phi,\theta,\psi) = R_3(\phi)R_2(\theta)R_1(\psi) = \begin{bmatrix} c_\phi c_\theta & s_\phi c_\psi + c_\phi s_\theta s_\psi & s_\phi s_\psi - c_\phi s_\theta c_\psi \\ -s_\phi c_\theta & c_\phi c_\psi - s_\phi s_\theta s_\psi & c_\phi s_\psi + s_\phi s_\theta c_\psi \\ s_\theta & -s_\psi c_\theta & c_\psi c_\theta \end{bmatrix} \quad (449)$$

$$L\{R_{321}(\phi,\theta,\psi)\} = \begin{bmatrix} 1 & \phi & -\theta \\ -\phi & 1 & \psi \\ \theta & -\psi & 1 \end{bmatrix} \quad (450)$$

$$\mathbf{u}_{321}(R) = \begin{bmatrix} \phi_{321}(R) \\ \theta_{321}(R) \\ \psi_{321}(R) \end{bmatrix} = \begin{bmatrix} \text{atan2}\left(-r_{21}, r_{11}\right) \\ \text{asin}\left(r_{31}\right) \\ \text{atan2}\left(-r_{32}, r_{33}\right) \end{bmatrix} \quad (451)$$

$$\mathbf{u}_{321}(R_q(\mathbf{q})) = \begin{bmatrix} \text{atan2}\big(-2q_1q_2 + 2q_0q_3, \\ \qquad {q_1}^2 + {q_0}^2 - {q_3}^2 - {q_2}^2\big) \\ \text{asin}\big(2q_1q_3 + 2q_0q_2\big) \\ \text{atan2}\big(-2q_2q_3 + 2q_0q_1, \\ \qquad {q_3}^2 - {q_2}^2 - {q_1}^2 + {q_0}^2\big) \end{bmatrix} \quad (452)$$

$$E_{321}(\phi,\theta,\psi) = \begin{bmatrix} s_\theta & 0 & 1 \\ -s_\psi c_\theta & c_\psi & 0 \\ c_\psi c_\theta & s_\psi & 0 \end{bmatrix} \quad (453)$$

$$L\{E_{321}(\phi,\theta,\psi)\} = \begin{bmatrix} \theta & 0 & 1 \\ -\psi & 1 & 0 \\ 1 & \psi & 0 \end{bmatrix} \quad (454)$$

$$E'_{321}(\phi,\theta,\psi) = \begin{bmatrix} 0 & s_\phi & c_\phi c_\theta \\ 0 & c_\phi & -s_\phi c_\theta \\ 1 & 0 & s_\theta \end{bmatrix} \quad (455)$$

$$L\{E'_{321}(\phi,\theta,\psi)\} = \begin{bmatrix} 0 & \phi & 1 \\ 0 & 1 & -\phi \\ 1 & 0 & \theta \end{bmatrix} \quad (456)$$

$$[E_{321}(\phi,\theta,\psi)]^{-1} = \frac{1}{c_\theta}\begin{bmatrix} 0 & -s_\psi & c_\psi \\ 0 & c_\theta c_\psi & c_\theta s_\psi \\ c_\theta & s_\theta s_\psi & -s_\theta c_\psi \end{bmatrix} \quad (457)$$

$$[E'_{321}(\phi,\theta,\psi)]^{-1} = \frac{1}{c_\theta}\begin{bmatrix} -c_\phi s_\theta & s_\phi s_\theta & c_\theta \\ s_\phi c_\theta & c_\phi c_\theta & 0 \\ c_\phi & -s_\phi & 0 \end{bmatrix} \quad (458)$$

$$\mathbf{q}_{321}(\phi,\theta,\psi) = \begin{bmatrix} c_{\phi/2}c_{\theta/2}c_{\psi/2} - s_{\phi/2}s_{\theta/2}s_{\psi/2} \\ c_{\phi/2}c_{\theta/2}s_{\psi/2} + s_{\phi/2}c_{\psi/2}s_{\theta/2} \\ c_{\phi/2}c_{\psi/2}s_{\theta/2} - s_{\phi/2}c_{\theta/2}s_{\psi/2} \\ c_{\phi/2}s_{\theta/2}s_{\psi/2} + c_{\theta/2}c_{\psi/2}s_{\phi/2} \end{bmatrix} \quad (459)$$

$$\frac{\partial R_{321}}{\partial \phi} = \begin{bmatrix} -s_\phi c_\theta & c_\phi c_\psi - s_\phi s_\theta s_\psi & c_\phi s_\psi + s_\phi s_\theta c_\psi \\ -c_\phi c_\theta & -s_\phi c_\psi - c_\phi s_\theta s_\psi & -s_\phi s_\psi + c_\phi s_\theta c_\psi \\ 0 & 0 & 0 \end{bmatrix} \quad (460)$$

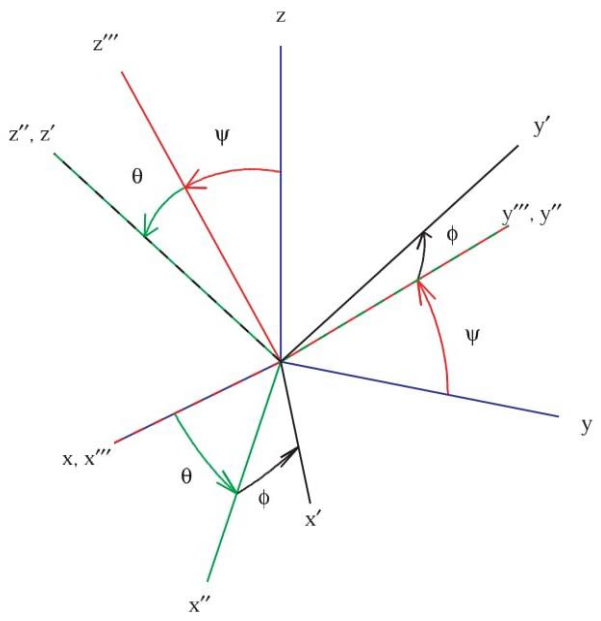


Figure 15: Euler Angle Sequence (3,2,1)

$$\frac{\partial R_{321}}{\partial \theta} = \begin{bmatrix} -c_\phi s_\theta & c_\phi c_\theta s_\psi & -c_\phi c_\theta c_\psi \\ s_\phi s_\theta & -s_\phi c_\theta s_\psi & s_\phi c_\theta c_\psi \\ c_\theta & s_\theta s_\psi & -s_\theta c_\psi \end{bmatrix} \quad (461)$$

$$\frac{\partial R_{321}}{\partial \psi} = \begin{bmatrix} 0 & -s_\phi s_\psi + c_\phi s_\theta c_\psi & s_\phi c_\psi + c_\phi s_\theta s_\psi \\ 0 & -c_\phi s_\psi - s_\phi s_\theta c_\psi & c_\phi c_\psi - s_\phi s_\theta s_\psi \\ 0 & -c_\theta c_\psi & -c_\theta s_\psi \end{bmatrix} \quad (462)$$

$$\frac{\partial E_{321}}{\partial \theta} = \begin{bmatrix} c_\theta & 0 & 0 \\ s_\theta s_\psi & 0 & 0 \\ -s_\theta c_\psi & 0 & 0 \end{bmatrix} \quad (463)$$

$$\frac{\partial E_{321}}{\partial \psi} = \begin{bmatrix} 0 & 0 & 0 \\ -c_\theta c_\psi & -s_\psi & 0 \\ -c_\theta s_\psi & c_\psi & 0 \end{bmatrix} \quad (464)$$

$$\frac{\partial E'_{321}}{\partial \phi} = \begin{bmatrix} 0 & c_\phi & -s_\phi c_\theta \\ 0 & -s_\phi & -c_\phi c_\theta \\ 0 & 0 & 0 \end{bmatrix} \quad (465)$$

$$\frac{\partial E'_{321}}{\partial \theta} = \begin{bmatrix} 0 & 0 & -c_\phi s_\theta \\ 0 & 0 & s_\phi s_\theta \\ 0 & 0 & c_\theta \end{bmatrix} \quad (466)$$

7 APPENDIX

Biochimica et Biophysica Acta 1834 (2013) 2772–2781

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Structural bases for a complete myotoxic mechanism: Crystal structures of two non-catalytic phospholipases A_2-like from *Bothrops brazili* venom

Carlos A.H. Fernandes [a], Edson J. Comparetti [a], Rafael J. Borges [a], Salomón Huancahuire-Vega [b], Luis Alberto Ponce-Soto [b,c], Sergio Marangoni [b], Andreimar M. Soares [d], Marcos R.M. Fontes [a,*]

[a] Dep. de Física e Biofísica, Instituto de Biociências, UNESP — Universidade Estadual Paulista, Botucatu and Instituto Nacional de Ciência e Tecnologia em Toxinas, CNPq, Brazil
[b] Dep. de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil
[c] Dep. de Farmacologia, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brazil
[d] Fundação Oswaldo Cruz, FIOCRUZ Rondônia e Centro de Estudos de Biomoléculas Aplicadas, Universidade Federal de Rondônia, Porto Velho, RO, Brazil

ARTICLE INFO

Article history:
Received 6 September 2013
Received in revised form 7 October 2013
Accepted 12 October 2013
Available online 18 October 2013

Keywords:
Phospholipase A_2-like
Myotoxic mechanism
Snake venom
X-ray crystallography
Lys49-phospholipase A_2
Amazonian snake

ABSTRACT

Bothrops brazili is a snake found in the forests of the Amazonian region whose commercial therapeutic anti-bothropic serum has low efficacy for local myotoxic effects, resulting in an important public health problem in this area. Catalytically inactive phospholipases A_2-like (Lys49-PLA_2s) are among the main components from *Bothrops* genus venoms and are capable of causing drastic myonecrosis. Several studies have shown that the C-terminal region of these toxins, which includes a variable combination of positively charged and hydrophobic residues, is responsible for their activity. In this work we describe the crystal structures of two Lys49-PLA_2s (BbTX-II and MTX-II) from *B. brazili* venom and a comprehensive structural comparison with several Lys49-PLA_2s. Based on these results, two independent sites of interaction were identified between protein and membrane which leads to the proposition of a new myotoxic mechanism for bothropic Lys49-PLA_2s composed of five different steps. This proposition is able to fully explain the action of these toxins and may be useful to develop efficient inhibitors to complement the conventional antivenom administration.

1. Introduction

Bothrops brazili is a snake found in the forests of the Amazonian region particularly in parts of Brazil, Colombia, Peru, Ecuador, Venezuela, Guyana, Suriname and French Guiana [1]. The antivenom access to these remote areas is very limited due to natural geographic barriers, vast territory and lack of road infrastructure. Furthermore, the commercial therapeutic anti-bothropic serum has low efficacy against this and other Amazonian snakes [2] resulting in an important public health problem in this Latin American region. A recent study estimates that at least 421,000 envenomings and 20,000 deaths by ophidian accidents occur each year in the world; however, due to poor statistic data mainly in Asia, Africa and Latin America, these numbers can be as high as 1,841,000 envenomings and 94,000 deaths [3]. It has also been shown that the mortality caused by snakebites is much higher than that of several neglected tropical diseases, including dengue hemorrhagic fever, leishmaniasis, cholera, schistosomiasis and Chagas disease [4]. These studies led the World Health Organization (WHO) to recognize the ophidian accidents as an important neglected tropical disease and the International Society on Toxinology to create the Global Snakebite Initiative [4] aiming to reduce snakebite morbidity and mortality. In Latin America, snakes from the *Bothrops* genus are responsible for approximately 80% of all ophidian accidents [5,6] that cause problems associated with prominent local tissue damage characterized by swelling, blistering, hemorrhaging and necrosis of the skeletal muscle, which are developed rapidly after snakebite. These effects may lead to permanent disability of the victims creating economic and social problems [7,8].

An important component of snake venoms is the phospholipase A_2, an enzyme able to promote Ca^{2+}-dependent hydrolysis of *sn*-2 acyl group of membrane phospholipids, releasing free fatty acids and lysophospholipids [9,10]. A subgroup of these proteins, known as Lys49-phospholipase A_2 (Lys49-PLA_2s), is catalytically inactive due to the lack of Ca^{2+} coordination related to the natural mutations Asp49 → Lys and Tyr28 → Asn [11,12], although it conserves all other catalytic related residues (such as His48, Tyr53, Asp99) [13]. This class of proteins is very common in bothropic venoms and, despite of their catalytic inactivity, they are capable to cause drastic local myonecrosis that is not efficiently neutralized by regular antivenom administration [7]. Several studies have shown that the segment 115–125 of the C-terminal region, which includes a variable combination of positively charged and hydrophobic residues, is responsible for the myotoxic activity [14–17].

Lys49-PLA_2s isolated from *Bothrops* genus snake venoms are dimeric as shown by electrophoresis, spectroscopic [18], small angle X-ray scattering [19], and dynamic light scattering experiments [12]. A

* Corresponding author at: Dep. de Física e Biofísica, IB, UNESP, Cx. Postal 510, CEP 18618-970, Botucatu/SP, Brazil.

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.10.009

dimer is also observed in the asymmetric units for the majority of the crystallographic Lys49-PLA$_2$ structures [20]. Furthermore, the pH-induced dissociation of the dimers showed an abolishment of their ability to disrupt liposomes [21] and reduced approximately 50% of their activity upon skeletal muscle cell and mature muscle tissue [22]. In spite of the experimental consensus that Lys49-PLA$_2$s present a dimeric conformation, the inspection of their unit cells in the crystallographic structures showed two different dimeric conformations as possible "biological dimers". The first possibility is known as "conventional dimer" which is formed by contacts between the tips of β-wing segments and N-terminal α-helices from both monomers being the "choice" for the firsts Lys49-PLA$_2$ structures [13,20]. The second possibility is known as "alternative dimer", being identified initially for Lys49-PLA$_2$s complexed with suramin or α-tocopherol, since this is the unique possible assembly for the binding mode of these molecules [23,24]. Subsequently, bioinformatic analyses with all bothropic Lys49-PLA$_2$s indicate that this conformation is more stable in solution because it presents a larger interfacial area and a smaller free energy compared to the conventional dimer [24]. Furthermore, small angle X-ray scattering experiments [19] also demonstrated that the "alternative conformation" is more likely to occur in solution. This alternative dimer is stabilized by contacts between the putative calcium-binding loop and C-terminal regions, and its dimeric interface is formed by the hydrophobic surfaces surrounding the entrance to the His48 region [24].

In this work we describe the crystal structures of two Lys49-PLA$_2$s (BbTX-II and MTX-II) from *B. brazili* venom and a comprehensive structural comparison with several Lys49-PLA$_2$s. From this comparison, we propose a new myotoxic mechanism for bothropic Lys49-PLA$_2$s that is able to fully explain the action of these toxins.

2. Material and methods

2.1. Protein purification, crystallization and X-ray data collection

Brazilitoxin II (BbTX-II) was isolated from *B. brazili* venom by a single-step reverse phase HPLC as previously described [25]. MTX-II was obtained by the fractionation of *B. brazili* venom on a CM-Sepharose column (2 cm × 20 cm) as previously described [26]. Both proteins have 121 amino acids as determined by mass spectrometry experiments [25,26], and there are five differences between their primary sequences: i) Glu4Gln, Lys20Ala, Gln68Pro, Asn111Gly and Ala130Pro, respectively for BbTX-II and MTX-II. Crystallization experiments were performed using sparse-matrix method [27], as previously described [28]. In this former article, MTX-II was called as MT-II. Basically, the crystallization conditions for each proteins were (i) for BbTX-II, 30% (w/v) polyethylene glycol 4000, 0.25 M lithium sulfate and 0.1 M Tris HCl at pH 8.5 and (ii) for MTX-II, 30% (w/v) polyethylene glycol 8000, 0.25 M ammonium sulfate and 0.1 M sodium cacodylate at pH 6.5. Crystals were grown at 291 K for approximately 3 weeks for both proteins. X-ray diffraction data for all crystals were collected at a wavelength of 1.435 Å using a synchrotron-radiation source (MX1 station, Laboratório Nacional de Luz Síncrotron, LNLS, Campinas, Brazil) and a MAR CCD imaging-plate detector (MAR Research). Crystals were mounted in nylon loops and flash-cooled in a stream of nitrogen gas at 100 K using no cryoprotectant. The data were processed using HKL 2000 program [29].

2.2. Structure determination and refinement

Table 1 shows data-collection and refinement statistics. The crystal structures were determined by molecular-replacement techniques implemented in the program Phaser [30] using the coordinates of piratoxin-I (PrTX-I; PDB ID 2Q2J), a Lys49-PLA$_2$ isolated from *Bothrops pirajai* venom, for BbTX-II and bothropstoxin-I (BthTX-I; PDB ID 3IQ3), a Lys49-PLA$_2$ isolated from *Bothrops jararacussu* venom complexed with polyethylene glycol 4000, for MTX-II. The model choices were based on the best values of final translation function Z score (TFZ) and log-likelihood gain (LGG) followed by an automated model building using AutoBuild in Phenix software suite [31]. Despite the amino acid sequence of MTX-II obtained by mass spectrometry has indicated a Leu in position 125 [26], an electron density compatible to a Phe was observed in our crystallographic data. This residue is also present in the position 125 for other bothropic Lys49-PLA$_2$s and this substitution on the amino acid sequence of MTX-II crystallographic structure does not interfere with the discussion of this work (Sections 4.2 and 4.3) due to maintenance of the hydrophobic characteristic of both residues. The modeling processes were always performed by manual rebuilding with the program Coot [32] using electron density maps calculated with coefficients 2|Fobs|−1|Fcalc|. The models were improved, as judged by the free R-factor [33], through rounds of crystallographic refinement (positional and restrained isotropic individual B-factor refinement, with an overall anisotropic temperature factor and bulk solvent correction) using Phenix [31]. The refinement statistics for the final models are shown in Table 1. In the structure of BbTX-II, due to lack of electron density, side chains of the following residues were excluded: Phe3 (monomer A), Lys16 (monomer B), Lys20 (B), Val31 (B), Leu32 (B), Lys36 (A, B), Lys53 (B), Lys57 (A, B), Gln68 (A, B), Lys69 (A), Lys70 (A, B), Arg72 (A), Asn88 (B), Asn111 (A, B), Lys115 (A, B), Lys116 (A, B), His110 (A), Leu121 (A), Lys122 (A, B), Pro123 (B), Lys127 (A, B), and Lys129 (A). In the structure of MTX-II, the side chains of the following residues were excluded: Val31 (B), Lys36 (A), Lys53 (B), Lys69 (A, B), Lys70 (A), Arg72 (B), Lys78 (A), Asn87 (A), Lys122 (A, B), Lys127 (A, B), and Lys129 (A, B). The stereochemical qualities of the models were checked with the Phenix and MolProbity programs [31,34]. The coordinates were deposited in the Protein Data Bank (PDB) under the identification codes 4K09 (BbTX-II) and 4K06 (MTX-II).

Table 1
X-ray data collection and refinement statistics.

	BbTX-II	MTX-II
Unit cell (Å)	$a = b = 56.4$; $c = 129.1°$	$a = 39.0$; $b = 71.4$; $c = 44.4$; $\beta = 102.5°$
Space group	$P3_121$	$P2_1$
Resolution (Å)	32.3–2.11 (2.19–2.11)[a]	20–2.08 (2.15–2.08)[a]
Unique reflections	13,825 (1363)[a]	13,752 (1365)[a]
Completeness (%)	96.07 (97.64)[a]	95.87 (95.25)[a]
R_{merge}[b]	6.3 (49.0)[a]	12.9 (35.3)[a]
Mean I/σ (I)	14.40 (3.23)[a]	5.29 (1.78)[a]
R_{cryst} (%)[c]	19.36	19.27
R_{free} (%)[d]	24.43	23.62
Number of non-hydrogen atoms[e]		
Protein	1788	1837
PEG molecules	–	53
Sulfates	–	24
Waters	191	251
RMS (bonds)[e]	0.004	0.004
RMS (angles)[e]	0.83	0.86
Average B-factor (Å^2)[e]	47.50	28.90
Ramachandran favored (%)[e]	97	97
Ramachandran outliers (%)[e]	0	0
Clashscore[f]	5.30	8.63
MolProbity overall score[f]	1.54	1.78

[a] Numbers in parenthesis are for the highest resolution shell.
[b] $R_{merge} = \sum_{hkl}(\sum i(|I_{hkl,i} \le I_{hkl} > I)) / \sum_{hkl,i} < I_{hk}>$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices h, k and l, and $<I_{hkl}>$ is the mean intensity of that reflection. Calculated for $I \ge 3 \sum$ (I).
[c] $R_{cryst} = \sum hkl(||Fobs_{hkl}| - |Fcalc_{hkl}||) / |Fobs_{hkl}|$, where $|Fobs_{hkl}|$ and $|Fcalc_{hkl}|$ are the observed and calculated structure factor amplitudes.
[d] R_{free} is equivalent to R_{cryst} but calculated with reflections (5%) omitted from the refinement.
[e] Calculated with Phenix [29].
[f] Calculated with MolProbity [32].

2.3. Comparative analysis

Molecular comparisons between the protomers of both structures were performed using the Coot program [32] with only the C_α coordinates. All figures were generated by PyMOL [35] and Coot [32] programs. Analyses of the quaternary assemblies, interfacial contacts of the crystallographic models and buried surface area of the hydrophobic residues were performed using the online interactive tool PISA [36] available at the European Bioinformatics Institute server (http://www.ebi.ac.uk). Two different angles, θ_A (aperture angle) and θ_T (torsional angle) were used to quantify the oligomeric changes between the apo and complexed forms according to a model previously proposed [24]. In this model, the coordinates (x,y,z) of C_α atoms from the α-helices h2 and h3 were used to define two vectors (A and B) whose scalar product defined the aperture angle (θ_A). The torsional angle (θ_T) is calculated by using the normal plane formed by the vectors A and B from one monomer and the vector B from the other monomer.

2.4. Dynamic light scattering

The dynamic light scattering (DLS) measurements were performed with lyophilized MTX-II at 291 K, dissolved in ultra-pure water at a concentration of 3.5 mg mL^{-1} using the instrument DynaPro TITAN (Wyatt Technology). Data were measured one hundred times and results were analyzed with Dynamics v.6.10 software.

3. Results

3.1. Overall structures

Crystals of both proteins diffracted approximately at 2.1 Å (Table 1) and belonged to $P3_12_1$ and $P2_1$ space groups for BbTX-II and MTX-II, respectively. The refinements converged to final R values of 19.4% (R_{free} = 24.4%) and 19.3% (R_{free} = 23.6%), respectively for BbTX-II and MTX-II. The final models present a stereochemical quality expected for structures with the same resolution (Table 1). Both structures have seven disulfide bridges in each monomer and present the following structural features: (i) an N-terminal α-helix; (ii) a "short" helix; (iii) a Ca^{2+} binding loop; (iv) two anti-parallel α-helices (2 and 3); (v) two short strands of anti-parallel β-sheet (β-wing); and (vi) a C-terminal loop (Fig. 1), similarly to other class II PLA_2s [11,13,20].

BbTX-II and MTX-II structures have two molecules in the asymmetric unit and present similar oligomeric structures, but only MTX-II presents ligand molecules bound to it. The refined MTX-II structure displays three polyethylene glycol (PEG) 4000 ligands and five sulfate ions (Fig. 1). Two PEG molecules establish a large number of hydrophobic interactions along the hydrophobic channel of the molecule (Leu10, Pro17, Val 31, Leu31 and Pro113) while the third one is exposed to solvent and establishes hydrogen bonds with Lys7, Leu10, Glu11 and Gly14 residues. These molecules are positioned similarly to other PEG molecules [12], alpha-tocopherol [24] and fatty acid [37,38] in other Lys49-PLA_2 structures. Five sulfate ions interact through water molecules with Glu12 (A), Ser21 (B), Ile82 (A), Lys100 (A) and Ala130 (A) or directly by hydrogen bonds with Asn17 (A), Arg34 (A), Lys53 (A), Thr81 (A), Lys115 (A) and Arg118 (A). These ions are also positioned similarly to other complexed Lys49-PLA_2 structures, such as BthTX-I and PrTX-I complexed with α-tocopherol [24], BthTX-I complexed with PEG 4000 [12] and Bn-IV, a Lys49-PLA_2 isolated from *Bothrops neuwiedi* venom, complexed with myristic acid [38].

3.2. Biological oligomeric assembly

Results of dynamic light scattering experiments at 291 K (crystallization temperature) with MTX-II indicated a mean hydrodynamic radius (R_H) of 2.6 nm with a polydispersity of 9.3%. This R_H value corresponds to a molecular weight of approximately 32 kDa and is, thus, equivalent to a dimer, being similar to that obtained for other Lys49-PLA_2s previously studied [12].

The examination of unit-cell packing of the BbTX-II and MTX-II structures using the PISA software [36] suggested that their dimeric conformations are stable in solution and are similar to other bothropic Lys49-PLA_2s (Fig. 1A and B). These analyses also demonstrated that the alternative dimer conformation has better values of complexation signification score (CSS), interface area and solvation free energy gain

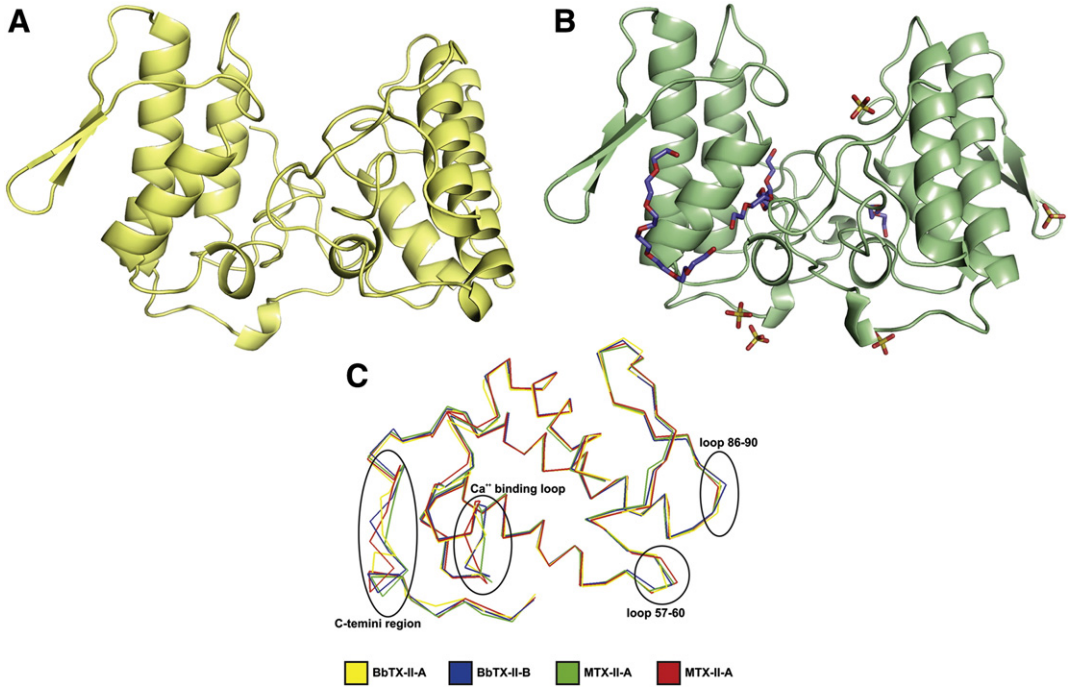


Fig. 1. Dimeric structures of BbTX-II (A) and MTX-II (B) showed as a cartoon representation. The PEG molecules and sulfate ions are showed by sticks. (C) C_α superposition of the BbTX-II and MTX-II structures (monomers A and B) highlighting the most important structural deviations between them.

upon formation of the interface (Δ^iG) than the conventional dimer for both structures. The BbTX-II structure in the alternative dimer conformation presents a complexation significance score (CSS) of 0.1, an interfacial area of 512.5 Å^2 and $\Delta^iG = -9.4$ Kcal/mol, while for the conventional dimer it presents a CSS of zero, interfacial area of 498.4 Å^2 and $\Delta^iG = 0.6$ Kcal/mol. The MTX-II structure presents in the alternative dimer conformation a CSS of 1.0, interfacial area of 524.5 Å^2 and $\Delta^iG = -12.4$ Kcal/mol, while for the conventional dimer conformation it presents a CSS of zero, an interfacial area of the 380.3 Å^2 and $\Delta^iG = -1.3$ Kcal/mol. These results strengthen previously theoretical [24] and experimental studies [19,23,24], indicating that the alternative dimer conformation is the most likely biological dimer for bothropic Lys49-PLA$_2$s.

3.3. Structural comparison of BbTX-II and MTX-II

Superposition between C_α atoms of the monomers A and B from BbTX-II resulted in an RMS deviation of 0.97 Å while the same superposition between MTX-II monomers resulted in 0.66 Å. Taking into account the presence of PEG ligands bound to MTX-II structure and that the BbTX-II structure is unbound to any ligand, these values are comparable to other Lys49-PLA$_2$s [24]. In this previous study a pattern was observed for apo forms (proteins unbound to ligands), which presented an RMSD of around 1.0 Å, while for complexed structures these values are lower. Additionally, when the monomers of BbTX-II are compared with those of the MTX-II (Table 2), it can be observed that one monomer (called herein monomer B) is more similar to the equivalent monomer of the other structure. Interesting, this feature was also observed for BthTX-I bound to different ligands compared to its apo structure [12].

The superposition between BbTX-II and MTX-II monomers (Fig. 1C) shows that the main differences occur in four regions: putative Ca^{2+} binding loop (residues 29–34), loops before and after β-wing region (residues 57–60 and 86–90) and C-termini (117–130); although the major differences are between the C-terminal and the putative Ca^{2+} binding loop regions. These structural deviations were also found in the C_α superposition between several Lys49-PLA$_2$ structures [12,24] and it may be due to higher flexibility of these segments that reflects on their highest B-factors values compared to other regions of the proteins.

Regarding quaternary structure comparisons, it has been shown for other Lys49-PLA$_2$s that when any ligand (PEG, α-tocopherol, BPB, fatty acid) is bound to the hydrophobic channel in at least one monomer, oligomeric changes are induced and the protein adopts an active conformation [24]. In order to characterize these conformational changes observed for Lys49-PLA$_2$s, a tridimensional configuration model has been proposed based on two angles: θ_T (torsional angle) and θ_A (aperture angle) [24]. This model was established for the so-called "alternative dimer" by the observation that the Lys49-PLA$_2$s complexed to any ligand present a rearrangement of the C-terminus of one monomer that changes the symmetry between the monomers aligning them in the same plane. These angles were calculated for several complexed and apo Lys49-PLA$_2$s and their oligomeric conformation could be divided into two classes [24] which complexed structures presented lower torsional and higher aperture angles compared to the apo ones. Subsequently, other Lys49-PLA$_2$s were solved [12,39] and this pattern was also observed. In the case of BbTX-II and MTX-II, their calculated angles are, respectively, $\theta_T = 59°$ and $\theta_A = 7°$ and $\theta_T = 41°$ and $\theta_A = 29°$ being in agreement with the proposed model. In addition, it was previously observed that oligomeric changes induced by ligands in the toxins switch the space groups of the crystals. All apo dimeric Lys49-PLA$_2$s solved to this date belong to $P3_121$ space group while all complexed dimeric Lys49-PLA$_2$ belong to $P2_1$ or $P2_12_12_1$ space groups [12]. These changes were also observed in BbTX-II and MTX-II structures that belong to $P3_121$ and $P2_1$ space groups, respectively (Table 1).

Table 2
Superposition between protomers of BbTX-II and MTX-II crystallographic structures (RMS deviations (Å) of C_α atoms).

	BbTX-II		MTX-II	
	A	B	A	B
BbTX-II (A)	–	0.97	1.03	1.11
BbTX-II (B)	0.97	–	0.45	0.72
MTX-II (A)	1.03	0.45	–	0.66
MTX-II (B)	1.11	0.72	0.66	–

4. Discussion

4.1. Cationic membrane-stabilization site

Sulfate ions are present in crystallization conditions for several PLA$_2$s, and in some cases, electron density maps compatible with sulfate ions were found and incorporated in these crystallographic structures. It has been suggested, based in the crystal structure of porcine pancreatic PLA$_2$, that the interacting region between sulfate ions and PLA$_2$s is the same as where these proteins interact with membranes, since negative bivalent ions could mimic the phosphatidyl group of an anionic lipid bilayer [40,41]. In these structures, sulfates bind with positive charged residues in each monomer and also by other closer-range H-bonding interactions. According to these authors, the interaction of these residues with anionic phosphatidyl groups leads the enzymes to achieve their fully active conformation. These positions are also equivalent to that occupied by the phosphate group of a phospholipid in the structure of an Asp49-PLA$_2$ from cobra-venom complexed with a transition-state analog [42]. Furthermore, forms I and II crystal structures of the ACL myotoxin (a Lys49-PLA$_2$ from *Agkistrodon contortrix laticinctus*) [43] and zhaoermiatoxin structure from *Zhaoermia mangshanensis* [44] presented sulfate ions bound to Arg and Lys residues.

Similarly, the presence of sulfate ions interacting with positive charged residues (Lys20, Arg34, Lys53, Lys115 and Arg118) in other complexed Lys49-PLA$_2$s, such as BthTX-I and PrTX-I complexed to α-tocopherol [24], BthTX-I complexed to PEG 4000 [12] and Bn-IV, a Lys49-PLA$_2$ isolated from *B. neuwiedi* venom, complexed to myristic acid [38] was reported. These structures are in active form because there are ligands bound at their hydrophobic channel leading the alignment of Lys20, Lys115 and Arg118 residues in the same plane (side-by-side) [24]. Sulfate ion interactions with Arg34 and Lys53 seem to be only the result of weak interactions between these ions and superficial positively charges, since they establish at most two hydrogen bonds with only Arg34 or with Lys53 and never are hydrogen bonded with both residues at the same time. Finally, there are no experimental evidences that support any physiological relevance for the sulfate binding in these residues. In contrast, the basic cluster formed by Lys20, Lys115 and Arg118 was proposed to be the "myotoxic site" of bothropic Lys49-PLA$_2$s [24] because scanning alanine site-directed mutagenesis showed a reduction in membrane damaging activity against liposome membranes for Lys115Ala [45] and Arg118Ala [17] mutants of BthTX-I, a Lys49-PLA$_2$ from *B. jararacussu* venom. The structure of the MTX-II complexed to PEG (this work) presents sulfate ions interacting with the residues Lys115 and Arg118 and establishes a large number of interactions (Fig. 2). However, the interaction with Lys20 is absent due to the natural mutation Lys20Ala, indicating that the occurrence of a charged residue in position 20 is not essential for the establishment of the myotoxic activity mechanism previously proposed. The comparison between MTX-II complexed to PEG and BbTX-II structures reveals that the binding of sulfate ions leads to a major extension in solvent direction of Lys115 and Arg118 residues (Fig. 3).

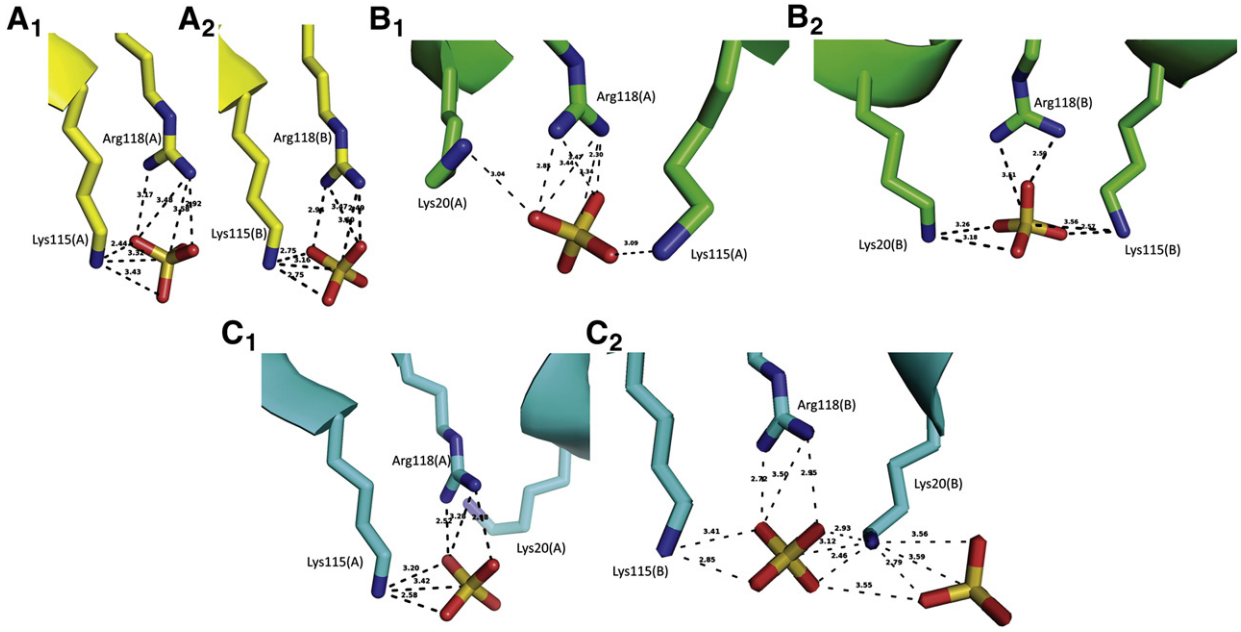


Fig. 2. Interaction of residues Lys20, Lys115 and Arg118 with sulfate ions in different bothropic Lys49-PLA_2 crystallographic structures. (A) MTX-II from *Bothrops brazili* venom complexed with PEG 4000 (this work) (PDB ID 4K06); (B) Bn-IV from *Bothrops neuwiedi* venom complexed with myristic acid [38] (PDB ID 3MLM); (C) PrTX-I from *Bothrops pirajai* venom complexed with α-tocopherol [24] (PDB ID 3CYL). The subscript numbers 1 and 2 refer to monomer A and B, respectively.

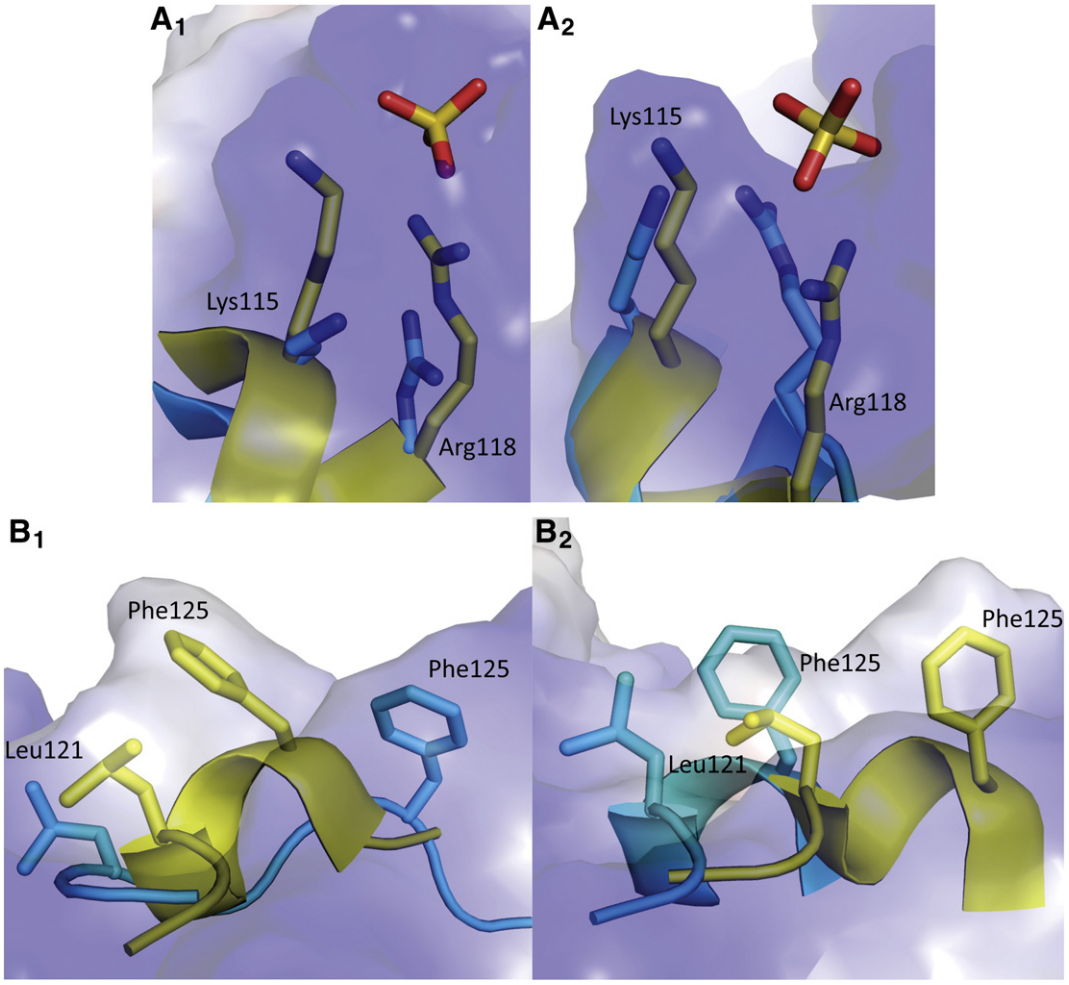


Fig. 3. Structural comparison between membrane-docking site (MDoS) (A) and membrane-disruption site (MDiS) (B) regions from MTX-II complexed to PEG 4000 (Lys49-PLA_2 in active form; yellow structure) and BbTX-II (Lys49-PLA_2 in inactive form; cyan structure) structures. Subscript numbers 1 and 2 refer to monomers A and B, respectively. Lys115, Arg118, Leu121, and Phe125 residues and sulfate ions are represented by sticks.

Taking into account all these structural studies with different PLA_2s, we suggest that this cluster of positive charged residues would be responsible for the toxin "docking" with the phosphatidyl group of anionic lipid bilayer membranes Then, we propose to call this cationic cluster for the Lys49-PLA_2s as "cationic membrane-docking site" (MDoS) which is mainly formed by C-terminus residues (Lys115 and Arg118), but is also aided by other positive and exposed residues such as Lys20, Lys80, Lys122 and Lys127.

4.2. Hydrophobic membrane-disruption site

Based on a crystallographic study with the ACL myotoxin, Ambrosio and colleges [43] proposed that specific hydrophobic residues (Phe121 and Phe124) present in the C-terminal region of this toxin and other Lys49-PLA_2s would be related to their myotoxic activity. In the active form of this structure, Phe121 and Phe124 have their aromatic rings with parallel conformations and salient from the toxin surface forming a contiguous patch resembling a hydrophobic knuckle. This hydrophobic knuckle is only observed when the Lys122 residue interacts with the Cys29 – Gly30 peptide bond and when a ligand is bound to hydrophobic channel of toxin. Similarly, BbTX-II and MTX-II structures also present hydrophobic residues on the C-terminal region as observed in ACL myotoxin structure [43]. However, in the case of BbTX-II and MTX-II, Lys122 side chains were not modeled due the lack of electron densities in their structures, a fact already observed for other Lys49-PLA_2s that, apparently, have random configurations in the majority of Lys49-PLA_2s [12]. Thus, the conformational changes suffered by the C-terminus hydrophobic residues seem to be due to ligand binding in the hydrophobic channel. Sequence alignment of Lys49-PLA_2 C-termini shows a high level of hydrophobic residues in this portion, particularly Leu, Pro and Phe in the positions 121, 124 and 125 (Fig. 4). In the case of bothropic Lys49-PLA_2s, the positions 121 and 125 are always occupied by hydrophobic residues, where Leu and Phe occur for 98% of the sequences. In addition, the side chains of these residues from both monomers are approximately in a symmetric alignment (Fig. 5) after the ligand binding in hydrophobic channel on all complexed bothropic Lys49-PLA_2s.

Chioato and colleagues [17] performed site-directed mutagenesis experiments which demonstrated a strong reduction of myotoxic activity (measured by release of creatine kinase) and membrane damaging on liposome membranes of the Phe125Ala mutant for BthTX-I. These authors also showed that Phe125Trp mutant has the same level of membrane damaging on liposome membranes for BthTX-I [17], reinforcing the importance of a hydrophobic residue in this position for myotoxic activity. Indeed, results of site-directed mutagenesis with aromatic residues performed by the same group [45] suggest that the 115–119 region of protein interacts superficially with the membrane, allowing residues closed to position 125 to be partially inserted into the membrane. Experiments using C-terminal synthetic peptides also demonstrated that a hydrophobic residue in position 121 for the expression of myotoxic activity [16] is essential.

Interestingly, the comparison of the hydrophobic cluster buried surface areas between apo and complexed structures (inactive and active states, respectively) obtained by the analysis with the PISA

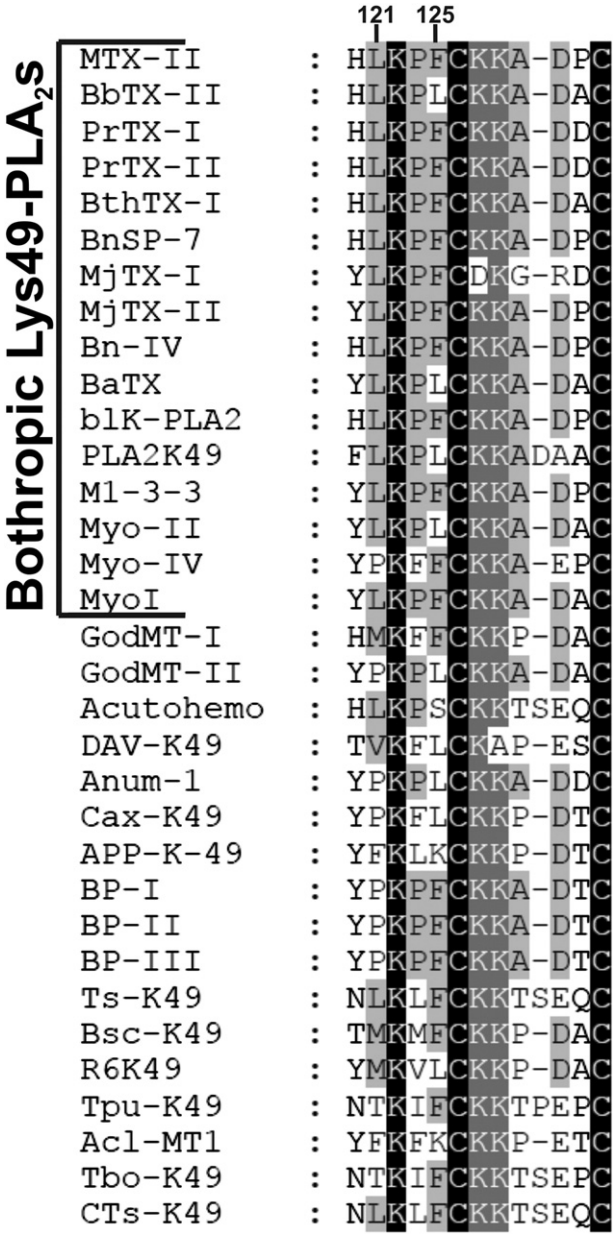


Fig. 4. C-terminal amino acid alignment of Lys49-PLA_2s deposited in NCBI Protein Data Bank. The numbers upside the alignment correspond to positions 121 and 125 (according to numeration proposed by Renetseder et al. [63] from the bovine pancreatic PLA_2, which contains 131 residues). The program AMAP v. 2.2 performed the alignment [64]. The sequences isolated from snakes from *Bothrops* genus are highlighted. MTX-II: Lys49-PLA_2 from *Bothrops brazili* venom (manually extracted from Costa et al. [26]); BbTX-II: Braziliantoxin-II from *B. brazili* venom (manually extracted from Huancahuire-Vega et al. [26]); PrTX-I: piratoxin-I from *Bothrops pirajai* venom (NCBI GI: 17433154); PrTX-II: piratoxin-II from *Bothrops pirajai* venom (NCBI GI: 17368328); BthTX-I: bothropstoxin-I from *Bothrops jararacussu venom* (NCBI GI: 51890398); BnSP-7: Lys49-PLA_2 from *Bothrops pauloensis* venom (NCBI GI: 239938675); MjTX-I: myotoxin-I from *Bothrops moojeni* venom (NCBI GI: 17368325); MjTX-II: myotoxin-II from *Bothrops moojeni* venom (NCBI GI: 62738542); BnIV: Lys49-PLA_2 from *B. pauloensis* venom (NCBI GI: 333361256); BaTX: Lys49-PLA_2 from *Bothrops alternatus* venom (NCBI GI: 292630846); blk-PLA_2: Lys49-PLA_2 from *Bothrops leucurus* venom (NCBI GI: 353678055); PLA_2K49: Lys49-PLA_2 from *Bothrops andianus* venom (NCBI GI: 442738889); M1-3-3: Lys49-PLA_2 from *Bothrops asper* venom (NCBI GI: 6492260); Myo-II: myotoxin-II from *B. asper* venom (BaspTX-II); Myo-IV: Lys49-PLA_2 from *B. asper* venom (NCBI GI: 166216293); MyoI: Lys49-PLA_2 from *Bothrops atrox* venom (NCBI GI: 82201805); GodMT-I: Lys49-PLA_2 from *Cerrophidion godmani* venom (NCBI GI: 4330040); GodMT-II: Lys49-PLA_2 from *C. godmani* venom (NCBI GI: 3122600); Acutohemo: acutohemolysin from *Deinagkistrodon acutus* venom (NCBI GI: 26397573); DAV-K49: Lys49-PLA_2 from *Calloselasma rhodostoma* venom (NCBI GI: 27151658); Anum-1: Lys49-PLA_2 from *Atropoides nummifer* venom (NCBI GI: 17433156); Cax-K49: Lys49-PLA_2 from *Crotalus atrox* venom (NCBI GI: 26397690); APP-K-49: Lys49-PLA_2 from *Agkistrodon piscivorus piscivorus* venom (NCBI GI: 129478); BP-I: Lys49-PLA_2 from *Trimeresurus flavoviridis* venom (NCBI GI: 408407670); BP-II: Lys49-PLA_2 from *T. flavoviridis* venom (NCBI GI: 408407672); BP-III: Lys49-PLA_2 from *T. flavoviridis* venom (NCBI GI: 408407673); Ts-K49: Lys49-PLA_2 from *Viridovipera stejnegeri* venom (NCBI GI: 8220134); Bsc-K49: Lys49-PLA_2 from *Bothriechis schlegelii* venom (NCBI GI: 25453450); R6-K49: Lys49-PLA_2 from *C. godmani* venom (NCBI GI: 59727050); Tpu-K49: Lys49-PLA_2 from *Trimeresurus puniceus* venom (NCBI GI: 123916279); ACL-MT-1: Lys49-PLA_2 from *Agkistrodon contortrix laticinctus* venom (NCBI GI: 1352702); and Tbo-K49: Lys49-PLA_2 from *Trimeresurus borneensis* venom (NCBI GI: 123907684). The *Bothrops pauloensis* species was recently reclassified as *Bothropoides pauloensis* [62].

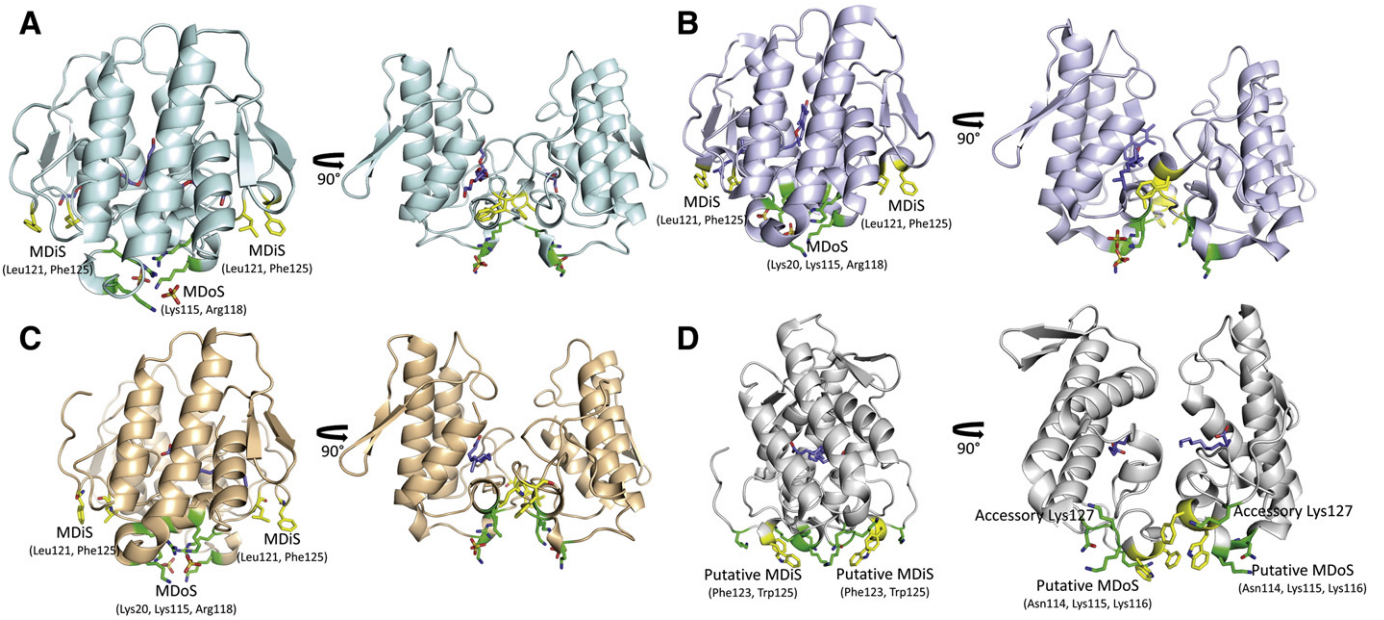


Fig. 5. Localization of MDoS (membrane-docking site) and MDiS (membrane-disruption site) on crystallographic structures of (A) MTX-II from *Bothrops brazili* venom complexed with PEG 4000 (this work) (PDB ID 4K06); (B) Bn-IV from *Bothrops neuwiedi* venom complexed with myristic acid [38] (PDB ID 3MLM); (C) PrTX-I from *Bothrops pirajai* venom complexed with α-tocopherol [24] (PDB ID 3CYL) and (D) ecarpholin S, a Ser49-PLA_2 from *Echis carinatus* venom complexed with lauric acid [51] (PDB ID 2QHD).

software reveals that the presence of a ligand (PEG, α-tocopherol, BPB, fatty acid) reduces the buried surface areas of Leu121 and Phe125 (Table 3). This reduction occurs mainly due to oligomeric changes of the Leu121 side chain after ligand binding, leading to a decrease in approximately 70% of its buried surface area. Structural comparison between MTX-II and BbTX-II Leu121 and Phe125 amino acids reveals that the presence of a ligand in hydrophobic channel (PEG 4000 in MTX-II structure) also leads to a major exposure of these hydrophobic residues (Fig. 3). According to hydrophobicity scales suggested by different authors [46,47] both Leu and Phe are among the residues with higher hydrophobic indices. Furthermore, Leu and Phe have the higher membrane permeability coefficients [48], with their bilayer permeability being up to 100 times higher in relation to charged amino acids [49].

Table 3
Buried surface area (BSA) of residues that constitute the hydrophobic membrane-disruption site (MDiS; Leu-121 and Phe125). BthTX-I: bothropstoxin I from *Bothrops jararacussu* venom [12] (PDB ID: 3HZD); PrTX-I: piratoxin I from *Bothrops pirajai* venom [24] (PDB ID: 2Q2J); BbTX-II: brazilitoxin II from *Bothrops brazili* venom (this work) (PDB ID 4K09); BnSP-7: a Lys49-PLA_2 from *Bothrops pauloensis* venom [60] (PDB ID 1PA0); BthTX-I-PEG4K: BthTX-I complexed with PEG 4000 [12] (PDB ID 3IQ3); BthTX-I-αT: BthTX-I complexed with alpha tocopherol [24] (PDB ID 3CXI); BthTX-I-BPB: BthTX-I chemically modified by p-bromophenacyl bromide (BPB) [12] (PDB ID 3HZW); PrTX-I-αT: PrTX-I complexed with alpha tocopherol [24] (PDB ID 3CYL); PrTX-I-BPB: PrTX-I chemically modified by p-bromophenacyl bromide (BPB) [61] (PDB ID 2OK9); MTX-II-PEG4K: MTX-II from *B. brazili* venom complexed with PEG 4000 (this work) (PDB ID 4K06); BnIV-myristic acid: Bn-IV from *Bothrops neuwiedi* venom complexed with myristic acid [38] (PDB ID 3MLM). The *B. pauloensis* species was recently reclassified as *Bothropoides pauloensis* [62].

		BSA of MDiS (Å^2)
Apo structures	BthTX-I	122.78
	PrTX-I	126.38
	BbTX-II	88.84
	BnSP-7	109.58
Complexed structures	BthTX-I-PEG4K	61.15
	BthTX-I-αT	52.64
	BthTX-I-BPB	47.68
	PrTX-I-αT	42.54
	PrTX-I-BPB	68.49
	MTX-II-PEG4K	63.75
	BnIV-myristic acid	66.49

Consequently, considering the high conservation level, and the structural and functional studies about the hydrophobic cluster (121 and 125 residues) in bothropic Lys49-PLA_2s, we propose that this site is mainly responsible for the membrane disruption after toxin anchorage and we call it herein as "hydrophobic membrane-disruption site" (MDiS). Membrane perturbation would be the key toxic event, allowing an uncontrolled influx of ions (Ca^{2+} and Na^{+}) that initiates a complex series of degenerative effects on muscle fiber [50].

Similarly, structural analysis of other classes of PLA_2-like, Ser49-PLA_2s [51], demonstrated the presence of particular MDoS (Section 4.1) and MDiS. Two hydrophobic residues, Phe123 and Trp125, would form the putative MDiS while Asn114, Lys115 and Lys116, and possibly, Lys127 would form the putative MDoS (Fig. 5). The buried surface area of Phe123 and Trp125 is 38.23 Å^2, indicating a high exposition to the solvent, similarly to MDiS region in Lys49-PLA_2s.

4.3. Structural bases of the myotoxic mechanism for Lys49-PLA_2s

Several hypotheses have been raised to address the membrane damage process produced by Lys49-PLA_2s. Although some propositions point to a direct action of these proteins towards membrane receptors [52–54], the major number of hypotheses is based on the direct action of the Lys49-PLA_2s in the target membrane by electrostatic interactions [14,18,24,43,55,56]. The first proposal was based on the fact that a synthetic peptide formed by C-terminus residues (115–129) of the toxin displayed heparin-binding and cytolytic activities, leading to the hypothesis that this region is mainly responsible for the membrane damage activity produced by Lys49-PLA_2s [14]. The molecular modeling of the monomeric ACL myotoxin from *A. contortrix laticinctus* suggested that partially conserved residues in Lys49-PLA_2s (K7, E12, T13, K16, and N17) associated with some specific lysine residues (K78, K80, K116, and K117) form a site that may be responsible for myotoxic activity in these toxins [57]. Subsequently, spectroscopic and crystallographic studies with the BthTX-I showed that this protein may cause membrane destabilizing activity by an oligomeric triggering process of the toxin (open and closed conformations) using the conventional dimer assembly [18]. During this process, the C-termini change their conformation leading a transition in the membrane bound form that disrupts the packing of the bilayer phospholipids, resulting in the loss of membrane

integrity [18]. Subsequently, Lomonte et al. [55], based on the former model [18], suggested that cationic residues from C and N-terminal regions would establish initially weak electrostatic interactions with the anionic sites of the membrane that would be further strengthened by the contribution of hydrophobic and aromatic residues of the C-termini. These residues may partially penetrate and disorganize the bilayer. Site-directed mutagenesis experiments also highlighted the relevance of the C-terminal region for expression of myotoxic activity, specially the Tyr117–Lys122 segment [45]. Ambrosio et al. [43] based on the crystallographic structure of ACL myotoxin (non-bothropic and monomeric Lys49-PLA_2), proposed that the interaction of the Lys122 with Cys29–Gly30 peptide bond when a ligand is bound to a hydrophobic channel would form a hydrophobic knuckle that would be responsible for membrane perturbation. Posteriorly, Bortoleto-Bugs et al. [56], also using the conventional dimer conformation, hypothesized that a region of the toxin called interfacial recognition site (IRS) binds anionic phospholipids inducing conformational changes in the protein, leading the exposition of hydrophobic surfaces that strengthen the protein/membrane interaction. More recently, dos Santos et al. [24] proposed a myotoxic mechanism based on the alternative dimer conformation for Lys49-PLA_2s, whose myotoxic sites presented in both monomers are formed specifically by the cationic residues Lys20, Lys115 and Arg118 that are aligned side by side in this protein configuration.

Thus, based on the previous propositions and the experimental observations of this work, we suggest a new myotoxic mechanism of bothropic Lys49-PLA_2s where all residues involved in this action are specified. In our hypothesis, the alternative dimer is adopted as the biological oligomeric conformation that would act on the following steps (Fig. 6).

I) Entrance of a hydrophobic molecule (e.g. fatty acid) at the hydrophobic channel leading to a reorientation of a monomer which may be measured by the "two angle model" [24]. This reorientation causes a transition between "inactive" and "active" states, causing exposure to the solvent and alignment of both MDoS and MDiS residues in the same plane, with a symmetric position for both monomers. The fatty acid may come from the lyses process of the membrane by Asp49-PLA_2s, reinforcing the synergism between Asp and Lys49-PLA_2s in snake venom action. Previous studies showed that membrane lesion happens with a small amount of Lys49-PLA_2 when the process occurs in the presence of Asp49-PLA_2s [54]. In addition, this step justifies the conservation of residues from the putative catalytic site and hydrophobic channel for Lys49-PLA_2s regarding to their protein ancestry (catalytic PLA_2s) [58] and, also indicates a functional relationship between the "catalytic" and "myotoxic" sites [24].

II) Stabilization of the toxin on the membrane by interaction of the MDoS from both monomers and the phospholipid head group. As discussed in Section 4.1, sulfate ions observed in several PLA_2 structures would mimic the phosphatidyl group of an anionic lipid bilayer, and then the interaction between MDoS and membrane is a feasible possibility.

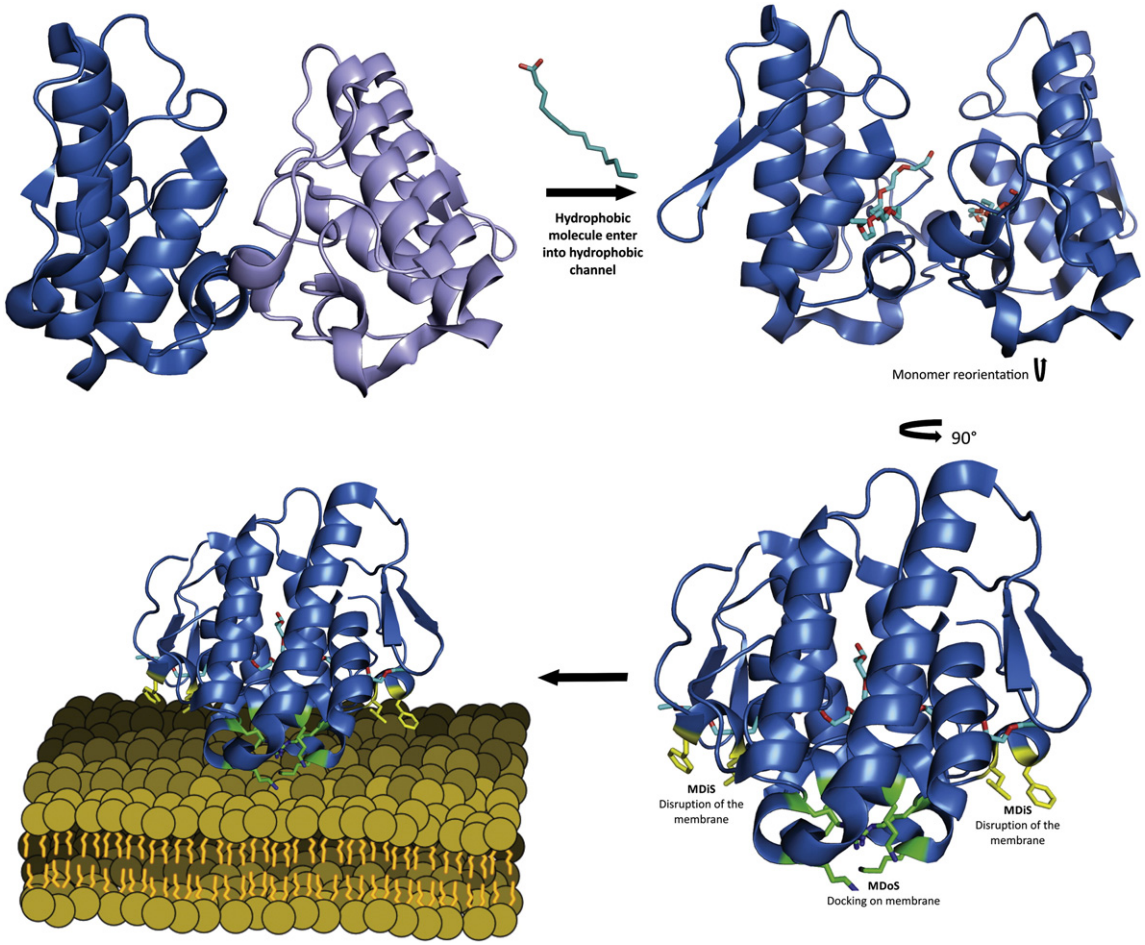


Fig. 6. Myotoxic mechanism of bothropic Lys49-PLA_2s. An apo Lys49-PLA_2 is activated by the entrance of a hydrophobic molecule (e.g. fatty acid) at the hydrophobic channel of the protein leading to a reorientation of a monomer, measured by "two angle model" [24]. This reorientation aligns C-terminal and MDoS regions side-by-side and puts the MDiS region in the same plane, exposed to solvent and in a symmetric position for both monomers. The MDoS region (Lys20; Lys115; Arg118) stabilizes the toxin on membrane by the interaction of charged residues with phospholipid head group. Subsequently, the MDiS region (Leu121; Phe125) destabilizes the membrane with penetration of hydrophobic residues.

III) Membrane destabilization by penetration of the MDiS from both monomers into the membrane. This insertion causes a disorganization of the membrane, allowing an uncontrolled influx of ions (i.e. Ca^{2+} and Na^{+}), and eventually triggering irreversible intracellular alterations and cell death [50]. Furthermore, the cells damaged release ATP into the extracellular environment, which amplifies the effect of these proteins [59]. These ATP molecules bind to muscle P2X purinergic receptors and induce Ca^{2+} and Na^{+} influx and K^{+} efflux in cells that have not been directly damaged by the proteins [59].

5. Concluding remarks

Previously evolutionary studies demonstrated that all Lys49-PLA_2s isolated from *Bothrops* snakes nest in a monophyletic clade [58]. These data support some unique characteristics for the bothropic Lys49-PLA_2s, such as the essential role of the dimer formation for expression of myotoxic activity and a specific mechanism of action. Indeed, the dimer dissociation causes a severe impact in its toxicity [21,22] and the high conservation of the residues that constitute the MDoS [24] and MDiS (Fig. 4) is observed in the bothropic proteins. However, non-bothropic Lys49-PLA_2s have a high level of conservation of Lys115 and other Lys residues (Lys80, Lys122, Lys127 and Lys28) which are exposed to solvent and thus, could constitute an analogous MDoS region. Furthermore, since hydrophobic residues of C-terminal regions have also a high level of conservation among all Lys49-PLA_2s (Fig. 4), then steps II and III may be conserved among all Lys49-PLA_2s.

In conclusion, we performed an extensive comparative structural analysis of the BbTX-II and MTX-II from *B. brazili* snake venom with other bothropic Lys49-PLA_2s available in PDB in order to propose a coherent hypothesis of the molecular events that led to their toxicity. We propose a new myotoxic mechanism composed of five steps: i) fatty acid binding, ii) oligomeric changes, iii) protein–membrane docking, iv) membrane destabilization by penetration of hydrophobic residues and v) uncontrolled influx of ions and cell death. This proposition may be useful to efficiently develop inhibitors that can be used to complement the conventional antivenom administration, thus preventing permanent injuries still caused by these proteins in snakebite victims.

The authors gratefully acknowledge financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil). We also acknowledge the use of the Laboratório Nacional de Luz Síncrotron (LNLS, Brazil).

References

[1] J.A. Campbell, W.W. Lamar, The Venomous Reptiles of the Western Hemisphere, Comstock Press, New York, 2004.
[2] E.G. Muniz, W.S. Maria, M.I. Estevão-Costa, P. Burnheim, C. Chávez-Ortegui, Neutralization potency of horse antibothropic Brazilian antivenom against *Bothrops* snake venoms from the rain forest, Toxicon 38 (2000) 1859–1863.
[3] A. Kasturiratne, A.R. Wickresmasinghe, N. de Silva, N.K. Gunawaderna, A. Pathmeswaran, R. Premaratna, L. Savioli, D.G. Lallo, J. de Silva, The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths, Plos Med. 5 (2008) e218.
[4] D. Williams, J.M. Gutiérrez, R. Harrison, D.A. Warrell, J. White, K.D. Winkel, P. Gopalakrishnakone, The Global Snake Bite Initiative: an antidote for snakebite, Lancet 375 (2010) 89–91.
[5] F.A.A. Araújo, M. Santalúcia, R.F. Cabra, Epidemiologia dos acidentes por animais peçonhentos, in: J.L.C. Cardoso, F.O.S. França, F.H. Wen, C.M.S. Málaque, V. Haddad Jr. (Eds.), Animais Peçonhentos no Brasil, Sarvier, São Paulo, 2003, pp. 06–12.
[6] F.N. Saúde, Manual de Diagnóstico e Tratamento de Acidentes por Animais Peçonhentos, FUNASA, Brasília, 2001.
[7] J.M. Gutiérrez, B. Lomonte, Phospholipases A_2 myotoxins from *Bothrops* snake venoms, Toxicon 33 (1995) 1405–1424.
[8] J.M. Gutiérrez, R.D.G. Theakston, D.A. Warrell, Confronting the neglected problem of snake bite envenoming: the need for a global partnership, Plos Med. 3 (2006) e150.
[9] H.M. Verheij, J.J. Volwerck, E.H. Jansen, W.C. Puyk, B.W. Dijikstra, J. Drenth, G.H. de Haas, Methylation of histidine-48 in pancreatic phospholipase A_2. Role of histidine and calcium ion in the catalytic mechanism, Biochemistry 19 (1980) 743–750.
[10] D.L. Scott, P. White, Z. Otwinowski, W. Yuan, M.H. Gelb, P.B. Sigler, Interfacial catalysis: the mechanism of phospholipase A_2, Science 250 (1990) 1541–1546.
[11] D.R. Holland, L.L. Clancy, S.W. Muchmore, T.J. Ryde, H.M. Eispahr, B.C. Finzel, R.L. Heinrikson, K.D. Watenpaugh, The crystal structure of a lysine 49 phospholipase A_2 from the venom of the cottonmouth snake at 2.0 Å resolution, J. Biol. Chem. 265 (1990) 17649–17656.
[12] C.A. Fernandes, D.P. Marchi-Salvador, G.M. Salvador, M.C. Silva, T.R. Costa, A.M. Soares, M.R.M. Fontes, Comparison between apo and complexed structures of bothropstoxin-I reveals the role of Lys122 and Ca^{2+} binding loop for the catalytically inactive Lys49-PLA_2s, J. Struct. Biol. 171 (2010) 31–43.
[13] R.K. Arni, R.J. Ward, Phospholipase A_2 — a structural review, Toxicon 34 (1996) 827–841.
[14] B. Lomonte, E. Moreno, A. Tarkowski, L.A. Hanson, M. Maccarana, Neutralizing interaction between heparins and myotoxin II, a lysine 49 phospholipase A_2 from *Bothrops asper* snake venom. Identification of a heparin-binding and cytolytic toxin region by the use of synthetic peptides and molecular modeling, J. Biol. Chem. 269 (1994) 29867–29873.
[15] M. Falconi, A. Desideri, S. Rufini, Membrane-perturbing activity of Viperidae myotoxins: an electrostatic surface potential approach to a puzzling problem, J. Mol. Recognit. 13 (2000) 14–19.
[16] C.E. Nunez, Y. Angulo, B. Lomonte, Identification of myotoxic site of the Lys49 phospholipase A(2) from *Agkistrodon piscivorus piscivorus* snake venom: synthetic C-terminal peptides from Lys49, but not from Asp49 myotoxins, exert membrane-damaging activities, Toxicon 39 (2001) 1587–1594.
[17] L. Chioato, E.A. Aragão, T.L. Ferreira, A.I. de Medeiros, L.H. Faccioli, R.J. Ward, Mapping of the structural determinants of artificial and biological membrane damaging activities of a Lys49 phospholipase A(2) by scanning alanine mutagenesis, Biochimica et Biophysica Acta — Biomembranes 1768 (2007) 1247–1257.
[18] M.T. da Silva-Giotto, R.C. Garratt, G. Oliva, Y.P. Mascarenhas, J.R. Giglio, A.C.O. Cintra, W.F. de Azevedo Jr., R.K. Arni, R.J. Ward, Crystallographic and spectroscopic characterization of a molecular hinge: conformational changes in bothropstoxin I, a dimeric Lys49-phospholipase A_2 homologue, Proteins 30 (1998) 442–454.
[19] M.T. Murakami, M.M. Viçoti, J.R. Abrego, M.R. Lourenzoni, A.C. Cintra, E.Z. Arruda, M.A. Tomaz, P.A. Melo, R.K. Arni, Interfacial surface charge and free accessibility to the PLA_2-active site-like region are essential requirements for the activity of Lys49 PLA_2 homologues, Toxicon 49 (2007) 378–387.
[20] J.I. dos Santos, C.A.H. Fernandes, A.J. Magro, M.R.M. Fontes, The intriguing phospholipases A_2 homologues: relevant structural features on myotoxicity and catalytic inactivity, Protein Pept. Lett. 16 (2009) 887–893.
[21] A.H.C. de Oliveira, J.R. Giglio, S.H. Andrião-Escarso, A.S. Ito, R.J. Ward, A pH-induced dissociation of the dimeric form a lysine 49-phospholipase A_2 abolishes Ca^{2+}-independent membrane damaging activity, Biochemistry 40 (2001) 6912–6920.
[22] Y. Angulo, J.M. Gutiérrez, A. Soares, W. Cho, B. Lomonte, Myotoxic and cytolytic activities of dimeric Lys49 phospholipase A_2 homologues are reduced, but not abolished, by a pH-induced dissociation, Toxicon 46 (2005) 291–296.
[23] M.T. Murakami, E.Z. Arruda, P.A. Melo, A.B. Martinez, S. Calil-Elias, M.A. Tomaz, B. Lomonte, J.M. Gutiérrez, R.K. Arni, Inhibition of myotoxic activity of *Bothrops asper* myotoxin II by anti-trypanosomal drug suramin, J. Mol. Biol. 350 (2005) 416–426.
[24] J.I. dos Santos, A.M. Soares, M.R.M. Fontes, Comparative structural studies on Lys49-phospholipases A_2 from *Bothrops* genus reveal their myotoxic site, J. Struct. Biol. 167 (2009) 106–116.
[25] S. Huancahuire-Vega, L.A. Ponce-Soto, D. Martins de Souza, S. Marangoni, Structural and functional characterization of brazilitoxins II and III (BbTX-II and -III), two myotoxins from the venom of *Bothrops brazili* snake, Toxicon 54 (2009) 816–827.
[26] T.R. Costa, D.L. Menaldo, C.Z. Oliveira, N.A. Santos-Filho, S.S. Teixeira, A. Nomizo, A.L. Fuly, M.C. Monteiro, B.M. de Souza, M.S. Palma, R.H. Stábeli, S.V. Sampaio, A.M. Soares, Myotoxic phospholipases A(2) isolated from *Bothops brazili* snake venom and synthetic peptides derived from their C-terminal region: cytotoxic effect on microorganism and tumor cells, Peptides 29 (2008) 1645–1656.
[27] J. Jankarik, S.H. Kim, Sparse-matrix sampling — a screening method for crystallization of proteins, J. Appl. Crystallogr. 24 (1991) 409–411.
[28] C.A.H. Fernandes, E.C.G. Gartuzo, I. Pagotto, S. Huancahuire-Vega, L.A. Ponce-Soto, T.R. Costa, S. Marangoni, A.M. Soares, M.R.M. Fontes, Crystallization and preliminary X-ray diffraction analysis of three myotoxic phospholipases A_2 from *Bothrops brazili* venom, Acta Crystallogr. F 68 (2012) 935–938.
[29] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Macromol. Crystallogr. Pt. A 276 (1997) 307–326.
[30] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
[31] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive python-based system for macromolecular structure solution, Acta Crystallogr. D 66 (2010) 213–221.
[32] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D 60 (2004) 2126–2132.
[33] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system: a new software suite for macromolecular structure determination, Acta Cryst. Sect. D 54 (1998) 905–921.

[34] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormimo, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D 66 (2010) 12–21.
[35] W.S. Delano, The PyMOL Molecular Graphics System, Delano Scientific, San Carlos, 2002.
[36] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.
[37] L. Watanabe, A.M. Soares, R.J. Ward, M.R.M. Fontes, R.K. Arni, Structural insights for fatty acid binding in a Lys49-phospholipase A_2: crystal structure of myotoxin II from *Bothrops moojeni* complexed with stearic acid, Biochimie 87 (2005) 161–167.
[38] P. Delatorre, B.A. Rocha, T. Santi-Gadelha, C.A. Gadelha, M.H. Toyama, B.S. Cavada, Crystal structure of Bn IV in complex with myristic acid: a Lys49 myotoxin phospholipase A_2 from *Bothrops neuwiedi* venom, Biochimie 93 (2011) 513–518.
[39] G.H.M. Salvador, W.L. Cavalcante, J.I. dos Santos, M. Gallacci, A.M. Soares, M.R.M. Fontes, Structural and functional studies with myotoxin II from *Bothrops moojeni* reveal remarkable similarities and differences compared to other catalytically inactive phospholipases A_2-like, Toxicon 72 (2013) 52–63.
[40] Y.H. Pan, T.M. Epstein, M.K. Jain, B.J. Bahnson, Five coplanar anion binding sires on one face of phospholipase A_2: relationships to interface binding, Biochemistry 40 (2001) 609–617.
[41] B.J. Bahnson, Structure, function and interfacial allosterism in phospholipase A_2: insight from the anion-assisted dimer, Arch. Biochem. Biophys. 433 (2005) 96–106.
[42] S.P. White, D.L. Scott, Z. Otwinowski, M.H. Gelb, P.B. Sigler, Crystal structure of a cobra-venom phospholipase A_2 in complex with a transition-state analogue, Science 250 (1990) 1560–1563.
[43] A.L.B. Ambrosio, M.C. Nonato, H.S.S. de Araújo, R.K. Arni, R. Ward, C. Ownby, D.H.F. de Souza, R.C. Garrat, A molecular mechanism for Lys49-phospholipase A_2 activity based on ligand-induced conformational change, J. Biol. Chem. 280 (2005) 7326–7335.
[44] M.T. Murakami, U. Kuch, C. Betzel, D. Mebs, R.K. Arni, Crystal structure of a novel myotoxic Arg49-phospholipase A_2 homologue (zhaoermiatoxin) from *Zhaoermia mangshanensis* snake venom: insights into Arg49 coordination and the role of Lys122 in polarization of the C-terminus, Toxicon 51 (2008) 723–735.
[45] L. Chioato, A.H. de Oliveira, R. Ruller, J.M. Sá, R.J. Ward, Distinct sites for myotoxic and membrane-damaging activities in the C-terminal region of a Lys49-phospholipase A_2, Biochem. J. 366 (2002) 971–976.
[46] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[47] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nat. Struct. Biol. 3 (1996) 842–848.
[48] M.N. Naoi, N. Misako, T. Shimizu, A.N. Malviya, K. Yagi, Permeability of amino acids into liposomes, Biochim. Biophys. Acta 471 (1977) 305–310.
[49] A.C. Chakrabarti, D.W. Deamer, Permeability of lipid bilayers to amino acids and phosphate, Biochim. Biophys. Acta 1111 (1992) 171–177.
[50] J.M. Gutiérrez, C.L. Ownby, Skeletal muscle degeneration induced by venom phospholipases A_2: insights into the mechanisms of local and systemic myotoxicity, Toxicon 42 (2003) 915–931.
[51] X. Zhou, T.C. Tan, S. Valiyaveettil, M.L. Go, R.M. Kini, A. Velazquez-Campoy, J. Sivaraman, Structural characterization of myotoxic ecarpholin S from *Echis carinatus* venom, Biophys. J. 95 (2008) 3366–3380.
[52] Y. Yamazaki, Y. Matsunaga, Y. Nakano, T. Morita, Identification of a vascular endothelial grown factor receptor-binding protein in the venom of Eastern cottonmouth, J. Biol. Chem. 280 (2005) 29989–29992.
[53] D. Fujisawa, Y. Yamazaki, B. Lomonte, T. Morita, Catalytically inactive phospholipase A_2 homologue binds to vascular endothelial growth factor receptor-2 via a C-terminal loop region, Biochem. J. 411 (2008) 515–522.
[54] M. Cintra-Francischinelli, P. Pizzo, L. Rodrigues-Simioni, L.A. Ponce-Soto, O. Rossetto, B. Lomonte, J.M. Gutiérrez, T. Pozzan, C. Montecucco, Calcium imaging of muscle cells treated with snake myotoxins reveals toxin synergism and presence of acceptors, Cell. Mol. Life Sci. 66 (2009) 1718–1728.
[55] B. Lomonte, Y. Angulo, L. Calderon, An overview of lysine-49 phospholipase A_2 myotoxins from crotalid snake venoms and their structural determinants of myotoxic action, Toxicon 42 (2003) 885–901.
[56] R.K. Bortoleto-Bugs, M.R. Bugs, A.A. Neto, R.J. Ward, A micelle nucleation model for the interaction of dodecyl sulphate with Lys49-phospholipases A_2, Biophys. Chem. 125 (2007) 213–220.
[57] H.S. Selistre de Araujo, S.P. White, C.L. Ownby, cDNA cloning and sequence analysis of a lysine-49 phospholipase A_2 myotoxin from *Agkistrodon contortrix laticinctus* snake venom, Arch. Biochem. Biophys. 326 (1996) 21–30.
[58] J.I. dos Santos, M. Cintra-Francischinelli, R.J. Borges, C.A. Fernandes, P. Pizzo, A.C. Cintra, A.S. Braz, A.M. Soares, M.R. Fontes, Structural, functional and bioinformatics studies reveal a new snake venom homologue phospholipase A_2 class, Proteins 79 (2011) 61–78.
[59] M. Cintra Francischinelli, P. Caccin, A. Chiavegato, P. Pizzo, G. Carmignoto, Y. Ângulo, B. Lomonte, J.M. Gutiérrez, C. Montecucco, *Bothrops* snake myotoxins induce a large efflux of ATP and potassium with spreading of cell damage and pain, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14140–14145.
[60] A.J. Magro, A.M. Soares, J.R. Giglio, M.R. Fontes, Crystal structures of BnSP-7 and BnSP-6, two Lys49-phospholipases A(2): quaternary structure and inhibition mechanism insights, Biochem. Biophys. Res. Commun. 311 (2003) 713–720.
[61] D.P. Marchi-Salvador, C.A. Fernandes, L.B. Silveira, A.M. Soares, M.R. Fontes, Crystal structure of a phospholipase A(2) homolog complexed with p-bromophenacyl bromide reveals important structural changes associated with the inhibition of myotoxic activity, Biochim. Biophys. Acta 1794 (2009) 1583–1590.
[62] A.M. Fenwick, R.K. Gutberlet, J.A. Evans, C.K. Parkinson, Morphological and molecular evidence for phylogenetic classification of South American pitvipers, genera *Bothops*, *Bothriopsis* and *Bothrocophias* (Serpentes: Viperidade), Zool. J. Linnean. Soc. 156 (2009) 617–640.
[63] R. Renetseder, S. Brunie, B.W. Dijkstra, J. Drenth, P.B. Sigler, A comparison of the crystal structures of phospholipase A_2 from bovine pancreas and *Crotalus atrox* venom, J. Biol. Chem. 260 (1985) 11627–11634.
[64] A.S. Schwartz, L. Pachter, Multiple alignment by sequence annealing, Bioinformatics 23 (2007) e24–e29.

Biochimica et Biophysica Acta 1844 (2014) 2265–2276

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

A structure-based proposal for a comprehensive myotoxic mechanism of phospholipase A_2-like proteins from viperid snake venoms

Carlos A.H. Fernandes [a], Rafael J. Borges [a], Bruno Lomonte [b], Marcos R.M. Fontes [a,*]

[a] Depto. de Física e Biofísica, Instituto de Biociências, Universidade Estadual Paulista—UNESP, Botucatu, SP, Brazil
[b] Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José 11501, Costa Rica

ARTICLE INFO

Article history:
Received 31 May 2014
Received in revised form 9 September 2014
Accepted 16 September 2014
Available online 30 September 2014

Keywords:
Snake venom
Myotoxin
Phospholipase A_2
Lys49
Inhibitor
Myonecrosis

ABSTRACT

Envenomation via snakebites is an important public health problem in many tropical and subtropical countries that, in addition to mortality, can result in permanent sequelae as a consequence of local tissue damage, which represents a major challenge to antivenom therapy. Venom phospholipases A_2 (PLA_2s) and PLA_2-like proteins play a leading role in the complex pathogenesis of skeletal muscle necrosis, nevertheless their precise mechanism of action is only partially understood. Recently, detailed structural information has been obtained for more than twenty different members of the PLA_2-like myotoxin subfamily. In this review, we integrate the available structural, biochemical and functional data on these toxins and present a comprehensive hypothesis for their myotoxic mechanism. This process involves an allosteric transition and the participation of two independent interaction sites for docking and disruption of the target membrane, respectively, leading to a five-step mechanism of action. Furthermore, recent functional and structural studies of these toxins complexed with ligands reveal diverse neutralization mechanisms that can be classified into at least three different groups. Therefore, the data summarized here for the PLA_2-like myotoxins could provide a useful molecular basis for the search for novel neutralizing strategies to improve the treatment of envenomation by viperid snakes.

1. Introduction

Snakebite envenomings are currently recognized by the World Health Organization as an important neglected tropical disease [1]. Recent estimates indicate that at least 421,000 envenomings and 20,000 deaths by ophidian accidents occur each year in the world; however, these figures may be much higher, considering the under-reporting that occurs in many regions of Asia, Africa and Latin America [2]. The mortality caused by snakebites exceeds that attributed to other neglected tropical diseases such as dengue haemorrhagic fever, leishmaniasis, cholera, schistosomiasis, or Chagas disease; nevertheless, snakebites receive far less attention from governments, health authorities, and development agencies worldwide [3,4].

In addition to mortality, snakebites may result in disability and other permanent sequelae as a consequence of the local tissue damage that can develop in severe envenoming cases [5]. Necrosis and hemorrhage are among the most feared consequences of envenomation, depending on the inflicting snake species. The main toxins involved in the complex pathogenesis of the tissue-damaging activities have been identified as members of the phospholipase A_2 (PLA_2) and metalloproteinase families. Both toxin types are frequently the most abundant components in snake venoms from the Viperidae family [6–8], and their rapid activities often limit the ability of antivenom to prevent tissue damage [9]. Myotoxic PLA_2s target the sarcolemma and induce an acute degeneration of skeletal muscle fibers, whereas hemorrhagic metalloproteinases degrade extracellular matrix components and lead to the loss of microvessel integrity [9].

The targeting of skeletal muscle fibers by specialized PLA_2s, whose genes were recruited, neofunctionalized, and expressed in the venom glands of snakes of the Elapidae and Viperidae families, is unsurprising, considering that muscle represents the largest proportion of the body mass of prey and that its rapid damage contributes to immobilization, capture, and the initiation of digestion of prey [10]. Two related but structurally distinct PLA_2 scaffolds were recruited in the evolution of advanced snakes: group I in the family Elapidae and group II in Viperidae. The independent acquisition of myotoxic activity by PLA_2s from both lineages represents an example of convergent evolution [11].

The ability of snake venom PLA_2s to induce skeletal muscle necrosis depends on their catalytic activity [11], i.e., the hydrolysis of glycerophospholipids at the *sn*-2 position of the glycerol backbone, releasing fatty acids and lysophospholipids [12]. The enzymatic inactivation of myotoxic PLA_2s studied to date hampers their toxicity. However, although phospholipid hydrolysis is a necessary step in their mechanism of toxicity, additional structural and functional aspects of these enzymes must be involved because not all snake venom PLA_2s are myotoxic and a poor correlation between the catalysis and toxicity

* Corresponding author at: Departamento de Física e Biofísica, Instituto de Biociências, UNESP, Botucatu, SP18618-970, Brazil. Tel.: +55 14 3880 0271; fax: +55 14 3815 3744.
E-mail address: fontes@ibb.unesp.br (M.R.M. Fontes).

http://dx.doi.org/10.1016/j.bbapap.2014.09.015
1570-9639/

of these enzymes is well established [13]. Thus, the mechanisms of myotoxicity exerted by venom PLA_2s are still only partially understood [14].

Within the group II PLA_2s of viperids, an intriguing subtype was discovered in 1984 [15] that conserves the basic structural fold of this family of enzymes but lacks catalytic activity. These 'PLA_2 homologues' or 'PLA_2-like' proteins were subsequently found in a large number of viperid species and have the ability to induce a rapid local myonecrosis similarly as their catalytically active PLA_2 counterparts [16–19]. The study of these PLA_2-like myotoxins has attracted growing interest, and several hypotheses have attempted to explain their catalytic-independent mechanisms of myotoxicity. This review presents a summary of the current knowledge about the structure/function relationships of the PLA_2-like myotoxins, focusing on four main topics: (a) the structural basis for their catalytic inactivity; (b) the relevance of the quaternary structure assembly for their myotoxicity; (c) their mechanism of myotoxicity; and (d) the diverse ligand interactions that block their myotoxic activity. It is hoped that insights into these topics, gathered through an increasing number of structural studies, will expand our current view of how these toxins lead to the dramatic myonecrosis that develops in snakebites and pave the way for the identification of novel agents that will improve the medical treatment of this pathology.

2. The 'PLA_2-like myotoxins' or 'PLA_2-homologues': brief history and general characteristics

A seminal paper in 1984 described a new class of PLA_2 found in the venom of the North American Eastern cottonmouth (*Agkistrodon piscivorus piscivorus*), which had a lysine instead of aspartate at position 49, an amino acid residue considered to be an obligate component of the catalytic machinery of these enzymes [15,20]. This protein turned out to be the first of a large subfamily of toxins widely distributed in the venoms of viperids, commonly referred to as PLA_2-like, PLA_2 homologues, or Lys49 myotoxins. By 2012, a review listed 65 proteins of this type [21], and this number continues to grow. In addition to the most frequent Asp49/Lys49 substitution, a few reported variants have Ser, Arg, Asn, or Gln at this position. Originally, the Lys49-PLA_2 from *A. piscivorus piscivorus* (App-K49) was reported to be enzymatically active despite its inability to bind Ca^{2+} [15,20]. Subsequently, other Lys49-PLA_2s were also reported to have low levels of PLA_2 activity in vitro [22–27]. This concept was controversial and conflicted with reports of their inability to bind Ca^{2+} and with site-directed mutagenesis experiments with porcine [28] and bovine [29] pancreatic PLA_2s showing that their corresponding D49K mutants lost catalytic activity. Moreover, the extensive purification of the App-K49 resulted in negligible enzymatic activity that was 4000 times less than that of the Asp49 enzyme isolated from the same source [28]. Due to these conflicting data, most authors studying Lys49-PLA_2s proteins in the following years cautiously referred to them as PLA_2s having "either no, or low, catalytic activity" (for a review, see [16]).

This long-standing uncertainty in the literature was finally settled after recombinant snake venom Lys49 [30] and Ser49 [31] PLA_2-like proteins were produced and demonstrated to be enzymatically inactive. The structural basis for the enzymatic inactivity of the PLA_2-like proteins is discussed in Section 3. A likely explanation for the low levels of catalytic activity initially observed for Lys49-PLA_2s is that they might have contained traces of contaminating catalytic PLA_2s. In further support of the lack of enzymatic activity for Lys49 proteins, even when acting upon their biological targets in vivo, recent studies based on sensitive mass spectrometry techniques have shown that these proteins do not hydrolyze membrane phospholipids of myogenic cells in culture or of mature muscle, unlike their catalytic PLA_2 counterparts [32].

All PLA_2-like toxins from viperid venoms tested to date display myotoxic activity in rodent assays, as evidenced by the histological evaluation of the injected skeletal muscle tissue and by the increase of creatine kinase activity in plasma caused by the release of the CK-MM isozyme from the cytosol of damaged muscle fibers into the bloodstream [33]. The myotoxic effect of these proteins occurs only locally around the site of injection and develops very rapidly. Using intravital microscopy techniques, skeletal muscle fibers were visualized with real-time dynamics to undergo necrosis in as little as 3–4 min after exposure [34]. The localized nature of this effect, as opposed to the systemic myotoxic effect displayed by some PLA_2s or PLA_2 complexes [35,36], suggests that the PLA_2-like myotoxins not only bind to skeletal muscle fibers but also may be sequestered by binding to other cell types or to extracellular matrix, thus precluding their systemic spread and the targeting of distant muscle tissue [37,38]. This view is supported by the fact that PLA_2-like myotoxins are able to lyse a variety of cell types in culture [39,40] and to bind to mouse [41] and human erythrocytes [42], albeit without causing their lysis. Moreover, these myotoxins could not be detected in plasma after their intramuscular or intraperitoneal injection using a sensitive (10 ng/mL) enzyme-immunoassay [41], and conversely, radiolabeled variants were not distributed to skeletal muscle tissue after their intravenous injection [43].

The pathological events induced by the PLA_2-like myotoxins in skeletal muscle tissue in vivo and in myogenic cells in vitro have been studied by a variety of approaches (see recent reviews in [11,37,44]). The available evidence indicates that the sarcolemma is the primary site of action of these myotoxins, and this targeting causes a rapid loss of permeability that leads to a prominent influx of calcium ions [45–47] and an efflux of potassium ions and ATP, among several other intracellular markers [48]. Thus, sarcolemmal damage is currently considered the critical event that triggers the downstream degenerative processes that elicit muscle cell necrosis [32].

Because the PLA_2-like myotoxins are unable to catalyze phospholipid hydrolysis, their toxicity must be explained by a distinct mechanism from that of their catalytically active PLA_2 counterparts. A growing number of amino acid sequences and three-dimensional structures of the PLA_2-like proteins, obtained soon after their discovery but more intensively in recent years, have provided an essential platform for identifying the functional site responsible for their myotoxic activity. The first clue toward mapping the bioactive region of a PLA_2-like protein was obtained in studies on *Bothrops asper* myotoxin II, which identified a segment of thirteen amino acid residues near the C-terminus (^{115}KKYRYYLKPLCKK^{129}; numbering system of Renetseder et al. [49]) in the form of a synthetic peptide capable of causing direct cytolysis [50]. In the same study, this protein site was demonstrated to interact with heparin, explaining the neutralizing action of this glycosaminoglycan toward the effects of myotoxin II [50]. Moreover, a peptide synthesized with the equivalent sequence from the Lys49 PLA_2-like protein from *A. piscivorus piscivorus*, ^{115}KKYKAYFKLKCKK^{129}, induced myonecrosis upon intramuscular injection in mice, unequivocally supporting that this region contains the key structural determinants needed for myotoxicity [51]. Systematic site-directed mutagenesis studies on another PLA_2-like myotoxin, bothropstoxin-I from *Bothrops jararacussu*, concurred in the functional relevance of the cationic/hydrophobic cluster near the C-terminus for its toxic activities and further narrowed the identification of key structural determinants within the sequence 117–122 of this protein [30,52].

A wealth of structural information has accumulated on a considerable number of PLA_2-like myotoxins, of which twenty Lys49-PLA_2s crystallographic structures are currently deposited in the Protein Data Bank (www.rcsb.org/pdb/), including apo forms as well as ligand-complexed forms (Table 1), one crystal structure of Ser49-PLA_2, one of an Arg49-PLA_2, and three structures of a special type of basic myotoxic Asp49-PLA_2s that could represent a new PLA_2-like class (see Section 5). In combination with functional data on these proteins and aided by bioinformatic analyses, this structural information can be integrated into a comprehensive mechanism that would explain their myotoxicity, as proposed in the following sections.

Table 1
Crystallographic models of myotoxic PLA_2-like proteins.

Class/protein	Species	PDB code	Complex	Space group	Reference
Lys49-PLA_2					
PrTX-I	*Bothrops pirajai*	2Q2J	Native	$P3_121$	[70]
		2OK9	BPB	$P2_1$	[105]
		3CYL	α-Tocopherol	$P2_1$	[70]
		3QNL	Rosmarinic acid	$P2_12_12_1$	[100]
PrTX-II BthTX-I		1QLL	N-tridecanoic acid	$P2_1$	[56]
	Bothrops jararacussu	3I3H, 3HZD	Native	$P3_121$	[58]
		2H8I	PEG400	$P3_121$	[68]
		3IQ3	PEG4000	$P2_1$	[58]
		3HZW	BPB	$P2_12_12_1$	[58]
		3CXI	α-Tocopherol	$P2_1$	[70]
MyoII	*Bothrops asper*	1CLP	Native	$P2_12_12_1$	[54]
		1Y4L	Suramin	$P2_12_12_1$	[69]
BnSP-7	*Bothrops neuwiedi pauloensis*	1PA0	Native	$P3_121$	[63]
BnSP-6		1PC9	Native	$P3_121$	[63]
BnIV		3MLM	Myristic acid	$P2_1$	[59]
BbTX-II	*Bothrops brazili*	4K09	Native	$P3_121$	[72]
MTX-II		4K06	PEG	$P2_1$	[72]
		4DCF	PEG	C2	[83]
MjTX-II	*Bothrops moojeni*	4KF3	PEG	$P2_12_12_1$	[60]
		1XXS	Stearic acid	$P2_12_12_1$	[57]
GodMT-II	*Cerrophidion godmani*	1GOD	Native	$P4_32_12$	[57]
MyoII	*Atropoides nummifer*	2AOZ	Native	$P4_32_12$	[75]
ACL myotoxin	*Agkistrodon contortrix laticinctus*	1S8I, 1S8H	Native	$P4_12_12$	[74]
		1S8G	Lauric acid	$P4_12_12$	[74]
AppK49	*Agkistrodon piscivorus piscivorus*	1PPA	Native	$P4_12_12$	[53]
Acutohaemonlysin	*Deinagkistrodon acutus*	1MC2	Native	C2	[76]
Basic Asp49-PLA_2					
PrTX-III	*Bothrops pirajai*	1GMZ	Native	C2	[86]
BthTX-II	*Bothrops jararacussu*	2OQD	Native	C2	[85]
		3JR8	Calcium ions	C2	[87]
Arg49-PLA_2					
Zhaoermiatoxin	*Zhaoermia mangshanensis*	2PH4	PEG	P64	[99]
Ser49-PLA_2					
Ecarpholin	*Echis carinatus*	2QHD	Lauric acid	$P3_121$	[62]
		3BJW	Suramin	$P2_1$	[62]

3. Why are PLA_2-like proteins catalytically inactive?

The inability of Lys49-PLA_2 to bind Ca^{2+} is essential to its catalytic inactivity, and this was attributed to the D49K substitution because the majority of the catalytic network residues are conserved [15]. Crystallographic studies supported this hypothesis by demonstrating that the ε-amino group of Lys49 occupies the position of the Ca^{2+} ion observed in catalytic PLA_2 structures [53–55]. In contrast with this observation, site-directed mutagenesis experiments with a Lys49-PLA_2 showed that the K49D reversion mutant remained catalytically inactive [30], demonstrating that the single D49K replacement is not the only explanation for the catalytic inactivity of these toxins. Since then, two different proposals based on distinct crystallographic structures have attempted to find additional causes for the absence of catalytic activity for Lys49-PLA_2s.

3.1. The role of Lys122

A structural study on piratoxin II (PrTX-II) structure from *Bothrops pirajai* venom [56] found a continuous electron density in its hydrophobic channel and identified it as a putative fatty acid. Based on this observation, the authors proposed a role for the Lys122 residue that is conserved in all Lys49-PLA_2s but uncommon for catalytic PLA_2s. This residue interacts with the carbonyl of Cys29, which hyperpolarizes the peptide bond between Cys29 and Gly30, leading to an increased affinity for the fatty acid head group. According to these authors, the presence of this ligand would impair access to the hydrophobic channel and lead to the catalytic inactivity of Lys49-PLA_2s. However, PrTX-II was not crystallized in the presence of a fatty acid; the electron density found at the hydrophobic channel was only interpreted as a fatty acid. In contrast, MjTX-II from *Bothrops moojeni* venom was co-crystallized with a fatty acid (stearic acid), and its structure presented significant differences [57,58]. In the MjTX-II structure, the carboxyl oxygens of the stearic acid form hydrogen bonds with His48 and Gly30, whereas for the PrTX-II structure, a water molecule makes these interactions. Consequently, the presumed fatty acid in PrTX-II is shifted out one carbon position in relation to the stearic acid of MjTX-II. More recently, in the structure of Bn-IV from *Bothrops neuwiedi* venom co-crystallized with myristic acid [59], the binding of myristic acid was similar to that of stearic acid in MjTX-II.

Several high-resolution structures of Lys49-PLA_2s present unambiguous electron densities for PEG molecules in their hydrophobic channel [58]. Fernandes et al. [58] have suggested that PrTX-II has a PEG molecule instead of a fatty acid molecule in its hydrophobic channel because it was crystallized in the presence of PEG3350 and the electron density map is compatible with this ligand. These authors observed that Lys122 interacts with Cys29 for only 12 of the 30 different monomers for apo and complexed Lys49-PLA_2s structures deposited in the Protein Data Bank. Moreover, B-factor values of Lys122 residues for most Lys49-PLA_2s are very high compared to the average for the proteins, and for some structures, there are no electron density maps for this residue above 1.2σ. Finally, the structure of Bn-IV from *B. neuwiedi* venom co-crystallized with myristic acid [59] showed the presence of this fatty acid in the hydrophobic channel for both monomers, but Lys122 interacts with Cys29 in only one of them. Subsequently, the structure of MjTX-II complexed with PEG molecules [60] demonstrated that Lys122 interacts with Cys29 in both monomers, similarly to the MjTX-II/stearic acid structure. Then, they suggested [60] that this phenomenon may

occur with either fatty acids or PEG molecules in the hydrophobic channel of Lys49-PLA_2s. All these data show that Lys122 is a very flexible residue that can adopt diverse conformations, and consequently, the hyperpolarization of Cys29 is not an obligatory function of Lys122 even in complexed structures. However, Lys122 appears to be an important residue for Lys49-PLA_2s because it is conserved in all Lys49-PLA_2s but is rare in PLA_2s. Accordingly, site-directed mutagenesis showed that the K122A mutant displays a significant reduction in myotoxic activity [30,61]. Therefore, considering the high flexibility of Lys122, it is reasonable to assume that this residue can combine multiple functions related to myotoxic activity rather than to catalytic inactivity. These functions would include assisting fatty acid anchorage in the hydrophobic channel with an eventual hyperpolarization of Cys29, contributing to the activation of the toxin and acting as an auxiliary residue to the membrane docking site (see Section 6).

3.2. The role of Tyr28 and the calcium-binding loop

Another proposal to explain the catalytic inactivity for Lys49 myotoxins suggested that the Y28N substitution plays a key role because it affects the stability of the Ca^{2+}-binding loop [58]. All PLA_2s (catalytic PLA_2s) present a tyrosine residue at position 28, whereas an asparagine occupies this position for all Lys49-PLA_2s. A structural analysis of Tyr28 in PLA_2s reveals an important stability feature for the Ca^{2+}-binding loop region with which it is associated. All snake venom PLA_2 structures display an interaction between the Oγ atom of Tyr28 and the amino group of Gly35 in the range between 3.1 and 3.5 Å (with the exception of a special class of myotoxic basic Asp49-PLA_2s—see Section 5). This interaction also occurs for other PLA_2s, such as pancreatic bovine PLA_2, and provides more structural stability for Ca^{2+}-binding loops. This fact can be verified by the distance between the Cα atoms of Tyr28 and Gly33, which reflects the aperture of this loop. This distance is approximately 4.3 Å in PLA_2s but higher than 6.3 Å for Lys49-PLA_2s, due to their Y28N substitution. This open configuration of the Ca^{2+}-binding loop in Lys49-PLA_2s may be another factor responsible for the inability of Lys49-PLA_2s to bind Ca^{2+} ions in addition to the lysine at position 49. Therefore, the catalytic inactivity of Lys49-PLA_2s can be explained by two replacements, Y28N and D49K [58]. This conclusion is also supported by structural studies of Ecarpholin S, a Ser49-PLA_2 from *Echis carinatus sochureki* venom, which demonstrated that the lack of a Tyr28–Gly35 interaction was also responsible for its Ca^{2+}-binding loop disarrangement [62]. Additionally, site-directed mutagenesis experiments on Ammodytin L, another Ser49-PLA_2 myotoxin, showed that H28Y/S49D mutants have restored enzymatic activity [31]. Finally, the catalytic inactivity of Arg49-PLA_2s can also be explained by Y28N and D49R replacements because these proteins also have Asn at position 28.

4. What is the oligomeric conformation of Lys49 myotoxins?

Many snake venom PLA_2s have a strong tendency to form oligomers. The majority if not all PLA_2-like myotoxins occur as dimers when analyzed by electrophoresis, and accordingly, most structures have been solved as homodimers [17,54,63]. Taking into account that the dissociation of PLA_2-like proteins into monomers has been shown to impair their bioactivities [64,65], their oligomeric assembly appears to be relevant for toxicity.

4.1. The choice between two possible dimeric conformations for bothropic Lys49-PLA_2s

Bothropic Lys49-PLA_2s are the most studied proteins of this group, and the majority of them are dimeric, as shown by electrophoresis, spectroscopic [66,67], small angle X-ray scattering [68], dynamic light scattering [58], and X-ray crystallography experiments (Table 1). The first structure of a Lys49-PLA_2 from the *Bothrops* genus was solved in 1995 [54] and was followed by several others from the same genus [56,57,63,66]. All of these structures adopted a common dimeric conformation. Da Silva-Giotto and colleagues [66] solved the structure of the BthTX-I in this assembly and presented two structures with different aperture angles between the monomers, which they referred to as "open" and "closed" conformations. Subsequently, Magro et al. [63] analyzed eight Lys49-PLA_2s structures and observed that these proteins could adopt many conformations due to the flexibility between the monomers.

In 2005, the crystal structure of BaspTX-II from *Bothrops asper* complexed to suramin showed that this ligand binds simultaneously to both monomers of the protein [69]; these protein/ligand interactions are only possible if a novel Lys49-PLA_2 quaternary assembly is considered (for a review, see [19]). In contrast to the original model of dimerization, referred to as the "conventional dimer", this "alternative dimer" is stabilized by contacts between the putative calcium-binding loop and C-terminal regions, and its dimeric interface is formed by the hydrophobic surfaces surrounding the entrance of the "active sites". If the "conventional dimer" had been used to solve this structure, it would not be energetically favorable because a large hydrophobic portion of this ligand would be exposed to the solvent [69]. Dos Santos and colleagues [70] solved the crystal structure of a Lys49-PLA_2 complexed to alpha-tocopherol and also observed results similar to those obtained for the BaspTX-II/suramin complex. More recently, the crystal structure of MjTX-II complexed to polyethylene glycol 4000 also revealed a binding mode that is only possible for the "alternative dimer" [60]. To address whether this conformation was only induced by binding to these particular ligands or if it can be adopted by all bothropic Lys49-PLA_2s, the structures of these proteins available in the PDB (Table 1) were inspected by crystallographic symmetry application of the "alternative dimer" conformation for all the proteins analyzed [19]. Bioinformatic analyses [70] using the PISA program [71] showed that all Lys49-PLA_2s present a larger interface area and negative solvation free energy values for the "alternative dimer", indicating that this dimer is more stable in solution [19]. Small angle X-ray scattering experiments and molecular dynamics simulations with BthTX-I also indicate that the "alternative dimer" is the most probable assembly in solution [68].

Furthermore, when Lys49-PLA_2s are bound to a ligand, a conformational change occurs that induces the following: i) the reorganization of their C-termini (usually unorganized for unbound proteins), ii) the formation of interchain hydrogen bonding between the Tyr119 residues, and iii) the alignment of basic clusters from both monomers, which may be responsible for the docking of the protein into the membrane (see Section 6) [70]. This allosteric change can be measured by two angles between the monomers: the aperture and torsional angles. Interestingly, this model was able to classify the oligomeric crystal structures into two groups: unbound and bound structures [70]. More recently, a "hydrophobic site" was found in the crystal structures of PLA_2-like proteins and was associated as part of a myotoxic mechanism for these proteins [72]. This site is only exposed to solvent by both monomers when the "alternative" assembly is adopted for bound structures (Section 6).

Differing from the "alternative" oligomeric conformation proposal described above, some studies have raised contrasting explanations: i) Delatorre and colleagues [59], in a study on the structure of BnIV myotoxin complexed with myristic acid, found both "alternative" and "conventional" oligomeric structures in its asymmetric unit and argued that dimer formation would not be relevant to the myotoxic activity; ii) in the crystal structure of a Lys49-PLA_2 from *Bothrops brazili*, only the "conventional dimer" possibility was observed [73]; and iii) spectroscopic studies of BthTX-I showed an interaction between two residues of the "conventional" interface (Trp77) [66]. As discussed by Fernandes et al. [72], the crystal structure of Bn IV/myristic acid has all the indications to be considered a dimer (alternative or conventional), but the authors [59] proposed that the biologically relevant unit would be a monomer and deposited their coordinates in the PDB data bank as a

dimer (PDB ID: 3MLM). The crystal structure of Lys49-PLA_2 from *B. brazili* [73] did not present the "alternative dimer" conformation in its asymmetric unit; however, the high values of refinement statistics may indicate that the space group would be incorrect, consequently affecting the oligomeric possibilities in the asymmetric unit (PDB ID: 4DCF). Finally, spectroscopic studies with BthTX-I indicated that the "conventional dimer" is likely to occur in solution [66]. However, this observation does not exclude the possibility that both forms (conventional and alternative dimer) could be present in solution when the protein is in the unbound (inactive) state because these experiments were not performed in the presence of ligands.

In summary, the mode of oligomeric assembly seems to be a key to understanding the myotoxic function of Lys49-PLA_2s. Despite the extensive experiments performed in the last few years, a consensus has not been reached. Nevertheless, the myotoxic mechanism based on the assembly proposed by Fernandes et al. [72] is the most comprehensive hypothesis for the myotoxic activity of Lys49-PLA_2s presented to date because it takes into account the majority of available data. This hypothesis is based on the "alternative conformation" for bound proteins (active form), and it does not exclude the possible presence of populations of monomeric or "conventional dimer" assemblies when the proteins are in an unbound (or inactive) state.

4.2. Non-bothropic Lys49-PLA_2s

As discussed in Section 4.1, Lys49-PLA_2s isolated from snakes of the *Bothrops* genus exist as dimers, and their conformation appears to be essential to their myotoxic activity. However, this dimeric quaternary structure is not always observed in non-bothropic Lys49-PLA_2s crystallographic studies. The crystal structures of ACL myotoxin from *Agkistrodon contortix laticinctus* [74], AppK-49 from *A. piscivorous piscivorus* [53], myotoxin II from *Atropoides nummifer* [75], godMT-II from *Cerrophidion godmani* [67], and Acutohaemolysin from *Agkistrodon acutus* [76] present only one molecule in an asymmetric unit. The inspection of non-bothropic Lys49-PLA_2 unit cell packing using the PISA program [71] shows that these proteins do not form stable oligomeric complexes with the exception of myotoxin II from *A. nummifer* [75]. However, it is important to note that all non-bothropic Lys49-PLA_2s were crystallized under high concentrations of ammonium sulfate or acidic conditions that could have prevented dimer formation. In Acutohaemolysin [76] and ACL myotoxin [74], the dimer can be generated by the crystallographic 2-fold axes of the unit cell of the space groups, C2 and $P4_12_12$, respectively. Consequently, further biophysical and biochemical experiments are needed to establish their active oligomeric conformations. However, as discussed in Section 6, these proteins conserve several features (e.g., hydrophobic residues in the C-termini) that lead to the conservation of the main steps of the integrative myotoxic mechanism proposed here (Section 6).

4.3. Ser49 and Arg49-PLA_2

The crystal structure of Ecarpholin S, a Ser49-PLA_2 from *Echis carinatus* venom, with lauric acid presented two monomers in its asymmetric unit [62]. The inspection of the cell unit by the PISA program [71] shows that the dimer is stable in solution with a high complexation significance score (CSS; 1.0). Moreover, the dimer interface of Ecarpholin S is formed by hydrophobic surfaces surrounding the entrance to the "active site", which is similar to the alternative dimer assembly of Lys49-PLA_2s. In addition, this protein has similar hydrophobic and charged residue exposure to solvent compared to Lys49-PLA_2s, which suggests a similar myotoxic mechanism for both classes of PLA_2-like proteins (see Section 6). Thus, despite the lack of any other experimental evidence of its dimeric conformation, crystallographic data suggest that Ser49-PLA_2s may assemble as a homodimer.

Zhaoermiatoxin is an Arg49-PLA_2 from *Zhaoermia mangshanensis* venom that migrates in SDS-PAGE at a relative mass of ~14.8 kDa with a weaker band at ~27.6 kDa [77]. Biochemical experiments with Promutoxin (an Arg49-PLA_2 from *Protobothrops mucrosquamatus* venom) also showed the presence of dimers in solution [78]. The crystal structure of Zhaoermiatoxin was obtained with two monomers in its asymmetric unit and displayed a PEG molecule in the hydrophobic channel for both monomers (PDB ID: 2PH4). The inspection of the cell unit by the PISA program [71] shows that the dimer is stable in solution but presents a low CSS (0.112). Despite that the dimer interface of the Zhaoermiatoxin is quite different from Ser49 and bothropic Lys-49-PLA_2s, it is possible to identify similar charged and hydrophobic residues exposed to solvent for both Arg49 and Lys49-PLA_2s, suggesting a similar myotoxic mechanism for all classes of PLA_2-like proteins (see Section 6).

5. Can myotoxic basic Asp49-PLA_2s form a new PLA_2-like protein class?

As mentioned in Section 2, Lys49-PLA_2s are the most studied group of snake venom PLA_2-like toxins, and there are a few reports of Ser49, Arg49, Asn49 and Gln49-PLA_2s. However, more recently, it was reported that two basic Asp49-PLA_2s could form a new PLA_2-like protein class. BthTX-II from *B. jararacussu* [79,80] and PrTX-III from *B. pirajai* [81,82] are basic Asp49-PLA_2s that exhibit high calcium-independent myotoxic activity and display none or lower catalytic activity compared to other snake venom Asp49-PLA_2s [82–84]. To study these particular characteristics observed for BthTX-II and PrTX-III, structural studies have been performed by different research groups [85,86]. Correa and colleagues [85] suggested that the absence of (or low) catalytic activity in BthTX-II and PrTX-III structures is due to Ca^{2+}-binding loop distortion, which impairs co-factor coordination [86]. Subsequently, BthTX-II was co-crystallized using a high concentration of Ca^{2+} [87]. This structure demonstrated that the Ca^{2+}-binding loop conserved the same conformation observed for apo BthTX-II, which strengthens the hypothesis that the absence of catalytic activity is based on the Ca^{2+}-binding loop distortion. Furthermore, based on phylogenetic analysis, these authors [87] also showed that BthTX-II is more closely related to Lys49-PLA_2s than to catalytic PLA_2s, thus allowing PLA_2s from snakes of Crotalinae subfamily to be classified into two main branches. The first branch includes proteins that display Ca^{2+}-independent myotoxic activity (PLA_2-like), whereas the second contains the catalytic PLA_2s. This finding led to the hypothesis that the myotoxic mechanism for these proteins (BthTX-II and PrTX-III) would be analogous to Lys49-PLA_2s.

The dimeric conformation of BthTX-II was observed in solution using non-reducing SDS-PAGE, DLS [87], and SAXS [88], and for PrTX-III, it was observed by non-reducing SDS-PAGE and analytical size-exclusion chromatography [89]. Crystal structures of BthTX-II [85] and PrTX-III [86] revealed a common dimeric structure, which was also observed by dos Santos and colleagues [87]. However, these authors also found a new possibility of a dimeric conformation for BthTX-II using bioinformatic tools based on the more favorable energetic state of the protein. This conformation presents the interaction between Trp31 from both monomers and explains why Ca^{2+}-binding loops are distorted. As a consequence, the Tyr28 is distorted in comparison to catalytic PLA_2s, impairing the hydrogen bond formation between the Oγ atom of Tyr28 and the Gly35 amino group, which is, as noted in Section 3, an important stability feature for the Ca^{2+}-binding loop region [58]. In catalytic Asp49-PLA_2s, the distance between the Cα atoms of Tyr28 and Gly33 residues, which reflects the aperture of this loop, is approximately 4.3 Å, whereas for BthTX-II and PrTX-III structures, this distance is greater than 7 Å. Thus, this Ca^{2+}-binding loop distortion impairs the coordination of Ca^{2+} and may be responsible for the absence of catalytic activity in this special class of myotoxic Asp49-PLA_2s [87]. Furthermore, based on the analysis of the dimeric conformation found for BthTX-II and PrTX-III structures, it can be observed that their C-termini are conveniently located to interact with membranes (side-by-side and in the same face of the toxin) and present charged

and hydrophobic residues exposed to solvent. Although this dimeric conformation is different from that found for Lys49-PLA_2s (Section 4.1), the presence of conveniently positioned charged and hydrophobic residues may lead to a myotoxic mechanism similar to that of Lys49-PLA_2s and other PLA_2-like myotoxins, as discussed in Section 6. All the results presented in this section indicate that BthTX-II and PrTX-III are a special class of Asp49-PLA_2s that display calcium-independent myotoxic activity similar to Lys49-PLA_2s, thus representing a new class of PLA_2-like myotoxins.

6. An integrated hypothesis for the myotoxic mechanism of PLA_2-like myotoxins

As mentioned in previous sections, all PLA_2-like proteins for which crystallographic structures are available (Lys49; Ser49; Arg49 and basic Asp49-PLA_2 myotoxins) have produced biochemical and crystallographic data that support an integrated hypothesis for their mechanism of myotoxicity that will be discussed in this section.

Several hypotheses have been proposed to describe the activity of Lys49-PLA_2s toward membranes. Initially, Lomonte et al. [50] showed that a synthetic peptide composed of the C-terminal region (115–129) from *B. asper* myotoxin II displays cytolytic activities, and subsequently, that the equivalent synthetic peptide of AppK49 myotoxin (from *A. piscivorus piscivorus*) induces myonecrosin in mice [51]. These data led to the conclusion that this region, formed by cationic and hydrophobic residues, is primarily responsible for the toxicity of Lys49-PLA_2s [50,51]. The molecular modeling of ACL myotoxin (from *A. contortix laticinctus*) suggested the relevance of cationic residues in Lys49-PLA_2s and proposed that partially conserved residues (Lys7, Glu12, Thr13, Lys16, and Asn17) and specific basic residues (Lys78, Lys80, Lys116, and Lys117) would constitute a site responsible for their myotoxicity [90]. Subsequently, da Silva Giotto et al. [66] assumed that changes in the oligomeric conformation of Lys49-PLA_2 dimers played a role in their myotoxic mechanism in addition to the role of the C-termini. These authors solved the crystal structure of BthTX-I from *B. jararacussu* in open and closed dimeric states. Consequently, combining crystallographic and spectroscopic data, they proposed that the transition between the open and closed states would occur on the membrane, leading to the disruption of the phospholipid bilayer with a consequent loss of membrane integrity [66].

Subsequently, Lomonte and colleagues [16] combined several aspects of the former models to suggest a hypothetical myotoxic mechanism for Lys49-PLA_2s. In this proposal, positively charged residues from the C- and N-termini would make initially weak electrostatic interactions with negatively charged sites of the membrane, and these would be strengthened by the contribution of aromatic and hydrophobic residues in the C-terminal region. In this model, two additional factors would enhance the membrane perturbation: i) the oligomeric conformation changes through a molecular hinge [66] and; ii) the potential acylation of these toxins, either via autocatalysis [91] or via an interrupted catalytic cycle that fails to release a free fatty acid [56]. Site-directed mutagenesis supported the importance of the C-termini in the myotoxic mechanism and further narrowed this toxic site to the Tyr117–Lys122 segment [52,61]. Additionally, the removal of the N-terminal octapeptide by cyanogen bromide strongly reduced the myotoxic activity of bothropic Lys49-PLA_2s [92,93] with no significant conformational change of the protein [98], also suggesting a relevant role of the N-terminal region in myotoxicity.

Subsequently, Bortoleto-Bugs et al. [94], based on studies of the interaction of sodium dodecyl sulfate molecules to BthTX-I, proposed a micelle nucleation model in which Lys49-PLA_2s extract the phospholipids after electrostatic interactions with the membrane, which would lead to conformational changes in the protein, the reorganization of the phospholipids, and the loss of phospholipid bilayer integrity [94].

Importantly, all these previous hypotheses were based on the conventional dimer assembly as the quaternary structure of bothropic Lys49-PLA_2s. Murakami et al. [69] and dos Santos et al. [70] proposed an alternative dimer assembly in the structure of Lys49-PLA_2s (see Section 4.1). Based on this oligomeric conformation, a new myotoxic mechanism has been hypothesized [70] that is supported by the observation that positively charged residues from the N- and C-termini (Lys20, Lys115 and Arg118) interact with sulfate ions in bothropic Lys49-PLA_2s structures. Indeed, X-ray crystallographic studies with the bovine pancreatic PLA_2s showed that the interactions of cationic ions could indicate the regions of PLA_2s that interact with membranes because these ions can simulate the phospholipid polar head group [95]. Therefore, dos Santos et al. [70] proposed that these three residues (and eventually other basic residues close to this particular region) would constitute the myotoxic site of bothropic Lys49-PLA_2s. More recently, Fernandes et al. [72] suggested that this basic cluster is responsible for toxin docking with the phosphatidyl group of anionic lipid membrane bilayers. Consequently, they called the basic cluster formed by the strictly conserved C-terminal residues (Lys115 and Arg118), eventually aided by other cationic and exposed residues such as Lys20, Lys80, Lys122 and Lys127, the "cationic membrane-docking site" (MDoS) (Fig. 1; Table 2). It was also observed that the presence of a hydrophobic molecule in the hydrophobic channel leads to an oligomeric change, which induces the alignment of the MDoS from both monomers and increases the efficacy of this docking process [70].

Another important functional site was also observed [72] for bothropic Lys49-PLA_2s when the "alternative conformation" was considered. Leu121 and Phe125 are conserved residues in the majority (98%) of bothropic Lys49-PLA_2s, and these hydrophobic residues are more exposed to solvent and have a strong reduction in their buried surface area for complexed structures when the alternative dimer is considered as their quaternary structure. Furthermore, after ligand binding at the hydrophobic channel, the side chains of these residues became aligned for both monomers. Taking into account that these residues (Leu and Phe) have high hydrophobicity and permeability indices [96–98] and the previous structural observations, these authors [72] proposed that this hydrophobic site is responsible for membrane disruption, naming it the "hydrophobic membrane-disruption site" (MDiS). A similar site was also observed for ACL myotoxin by Ambrosio and colleagues [74], who observed an increased hydrophobic residue exposure (Phe121 and Phe124) when a ligand was bound to the hydrophobic channel in this toxin.

Thus, by integrating all previous data, a comprehensive myotoxic mechanism for Lys49-PLA_2s was proposed, which identifies the critical residues involved [72]. This mechanism involves an allosteric transition and the participation of two independent interaction sites with the target membrane and includes the following steps (Fig. 2):

I) Hydrophobic molecule binding at the hydrophobic channel;
II) Allosteric activation;
III) Protein-membrane docking (MDoS);
IV) Protein penetration (MDiS) and membrane disorganization;
V) Cell death.

As previously discussed, the key step for protein activation is the binding of a fatty acid at the hydrophobic channel, which leads to allosteric transition and structure stabilization (see Section 4.1), exposing MDoS and MDiS to the solvent (Fig. 1). In addition, Pedersen et al. [91] showed that Lys49-PLA_2s were able to bind to isolated muscle membranes and to the surface of liposomes with the fatty acid moiety inserted into the lipid bilayer acting as an anchor. Fatty acids may come from membrane phospholipid hydrolysis by catalytic PLA_2s, highlighting the synergism between PLA_2s and PLA_2-like proteins, as shown by Cintra-Francischinelli and colleagues [47]. These authors demonstrated that membrane damage occurs even with low amounts of Lys49-PLA_2s when the process occurs in the presence of PLA_2s. Furthermore, the fatty acid binding event justifies the conservation of fundamental residues from the hydrophobic and catalytic sites of

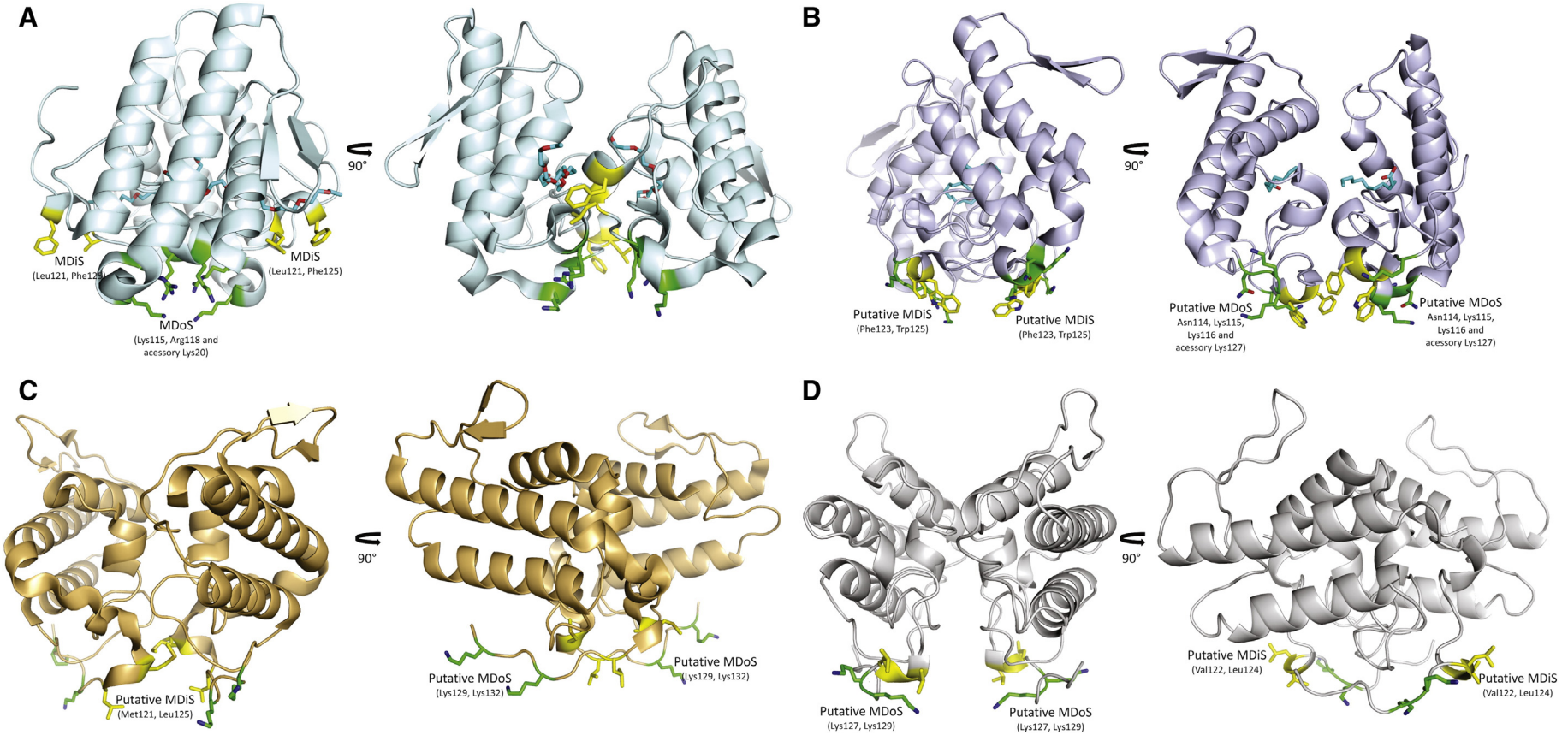


Fig. 1. Localization of MDoS (cationic membrane-docking site) (green sticks) and MDiS (hydrophobic membrane-disruption site) (yellow sticks) on crystallographic structures of PLA_2-like proteins from viperid snake venoms. (A) Bothropstoxin-I (BthTX-I), a Lys49-PLA_2 from *Bothrops jararacussu* venom, complexed with polyethylene glycol 4000 (PDB ID: 3IQ3); (B) Ecarpholin S, a Ser49-PLA_2 from *Echis carinatus* venom, complexed with lauric acid (PDB ID: 2QHD); (C) Zhaoermiatoxin, an Arg49-PLA_2 from *Zhaoermia mangshanensis* venom (PDB ID: 2PH4); (D) Bothropstoxin-II (BthTX-II), a Asp49-PLA_2 basic and myotoxic from *Bothrops jararacussu* venom (PDB ID: 3JR8). PEG and lauric acid molecules are represented as sticks on hydrophobic channel of BthTX-I and Ecarpholin S, respectively. Figure drawn using PyMOL program [106].

Table 2
Cationic membrane-docking site (MDoS) and hydrophobic membrane disruption site (MDiS) of the crystallographic structures of PLA2-like proteins.

Class/protein	Species	MDoS	MDiS
Lys49-PLA_2			
PrTX-I	*Bothrops pirajai*	K20, K115 and R118	L121 and F125
PrTX-II		K20, K115 and R118	L121 and F125
BthTX-I	*Bothrops jararacussu*	K20, K115 and R118	L121 and F125
MyoII	*Bothrops asper*	K19, K115 and R118	L121 and L124
BnSP-7	*Bothrops neuwiedi pauloensis*	K20, K115 and R118	L121 and F125
BnSP-6		K20, K115 and R118	L121 and F125
BnIV		K20, K115 and R118	L121 and F125
BbTX-II	*Bothrops brazili*	K20, K115 and R118	L121 and L125
MTX-II		K115 and R118	L121 and F125
MjTX-II	*Bothrops moojeni*	K20, K115 and R118	L121 and F125
GodMT-II	*Cerrophidion godmani*	K115 and K118	I119 and L125
MyoII	*Atropoides nummifer*	K115 and K118	I119 and L125
ACL myotoxin	*Agkistrodon contortrix laticinctus*	K115 and K118	F121 and F124
AppK49	*Agkistrodon piscivorus piscivorus*	K115 and K118	F121 and L124
Acutohaemonlysin	*Deinagkistrodon acutus*	K20, K115 and R118	L121 and S125
Basic Asp49-PLA_2			
PrTX-III	*Bothrops pirajai*	Putative K127 and K128	L121 and L124
BthTX-II	*Bothrops jararacussu*	Putative K127 and K128	V124 and L125
Arg49-PLA_2			
Zhaoermiatoxin	*Zhaoermia mangshanensis*	Putative K129 and K132	M121 and L125
Ser49-PLA_2			
Ecarpholin	*Echis carinatus*	Putative N114, K115 and K116. Accessory K127	M121 and L124

PLA_2s in PLA_2-like proteins because these enzymes are their ancestral forms [87].

The last two steps are caused by the insertion of the MDiS region from both monomers into the target membrane. This penetration disrupts the lipid bilayer, causing a loss of membrane permeability, a prominent influx of ions (i.e., Ca^{2+} and Na^{+}), and eventually, irreversible intracellular alterations and cell death [37,44]. Moreover, the effects of these toxins can be rapidly amplified in muscle cells because damaged cells release ATP molecules into the extracellular environment that then bind to muscle P2X purinergic receptors and induce Ca^{2+} and Na^{+} influx and K^{+} efflux in cells that have not been directly injured by the proteins [48].

Regarding other PLA_2-like proteins from viperid snake venoms, such as Ser49 and Arg49-PLA_2s, a similar myotoxic mechanism is possible. The crystal structure of Ecarpholin S also presents a hydrophobic molecule (lauric acid) in the hydrophobic channel [62]. Furthermore, in this structure it is possible to identify putative MDoS and MDiS regions, as previously noted [72]. Phe123 and Trp125 may form a putative MDiS, and the basic cluster formed by Asn114, Lys115 and Lys116 (and possibly Lys127) residues may form a putative MDoS (Fig. 1; Table 2). Finally, the buried surface areas of Phe123 and Trp125 are very small, showing high exposure to the solvent, similarly as the MDiS region in Lys49-PLA_2s. Thus, the three steps of the proposed myotoxic mechanism in bothropic Lys49-PLA_2s may also be conserved in Ser49-PLA_2s.

Murakami et al. [99] solved the crystal structure of Arg49-PLA_2 Zhaoermiatoxin, which has PEG molecules in its hydrophobic channels. They suggested that Arg34 and Arg49 residues form the anion-binding site of this protein because these residues interact with sulfate ions in the structure. However, it is difficult to identify this region as a putative MDoS because these two arginine residues are not completely exposed to solvent. Moreover, these residues are involved in the dimer interface in one monomer. A more accurate analysis of the Zhaoermiatoxin crystal structure shows that two hydrophobic (Met121 and Leu125) and two charged residues (Lys129 and Lys132) are exposed to solvent on the same plane of the C-terminal region and may constitute putative MDiS and MDoS regions, respectively (Fig. 1; Table 2). Despite this proposition, it is important to note that the myotoxic mechanism of action of Arg49-PLA_2 remains quite speculative because the dimer of Zhaoermiatoxin crystal structure has a low CCS value according to PISA analysis, and its interface is different from other PLA_2-like dimeric structures.

Finally, in basic myotoxic Asp49-PLA_2s it is possible to identify putative MDoS and MDiS regions by the inspection of hydrophobic and charged residues exposed to solvent in the dimeric conformation suggested by dos Santos et al. [87]. In BthTX-II, Lys127, Lys129 and Lys132 would form a putative MDoS, and Val122 and Leu124 would form a putative MDiS (Fig. 1; Table 2) In PrTX-III, these sites would be formed by Lys127 and Lys128 for MDoS, and Leu121 and Leu124 for MDiS.

7. Complexes of PLA_2-like myotoxins with ligands lead to hypotheses for their activation and inhibition

In several PLA_2-like crystal structures ligands bound to the protein were found; however, the majority of these ligands were present in the crystallization solutions or in the protein sample as impurities. Continuous electron density maps have been found in the hydrophobic channel for different Lys49-PLA_2 structures and have been attributed to fatty acids or polyethylene glycol molecules [56,58,60,72,74,100]. Some authors took advantage of the presence of these ligands and hypothesized functional mechanisms based on their binding regions (refer to Sections 3 and 6) [56,58,60,72,74,100]. Table 3 summarizes the ligands found in PLA_2-like crystal structures and their roles in myotoxic activity.

Salvador and colleagues [60] solved MjTX-II/PEG 4000 in the "alternative dimer" conformation, and its comparison to other Lys49-PLA_2s revealed that this protein presents particular features due to an exclusive insertion of Asn120 residue and Leu32Gly and His121Tyr substitutions. They suggested that these differences compared to other bothropic Lys49-PLA_2 complexed to PEG [58,70] lead to a distinct mechanism of ligand binding at the toxin's hydrophobic channel and also allow the presence of an additional ligand in this region. Consequently, these data suggested that MjTX-II may require different ligands for its complete inhibition, despite functional studies indicating that MjTX-II has similar neuromuscular blockage activities as the majority of other Lys49-PLA_2s [60].

Aiming to understand the myotoxic mechanism of the PLA_2-like proteins and to search for specific inhibitors, structural studies of

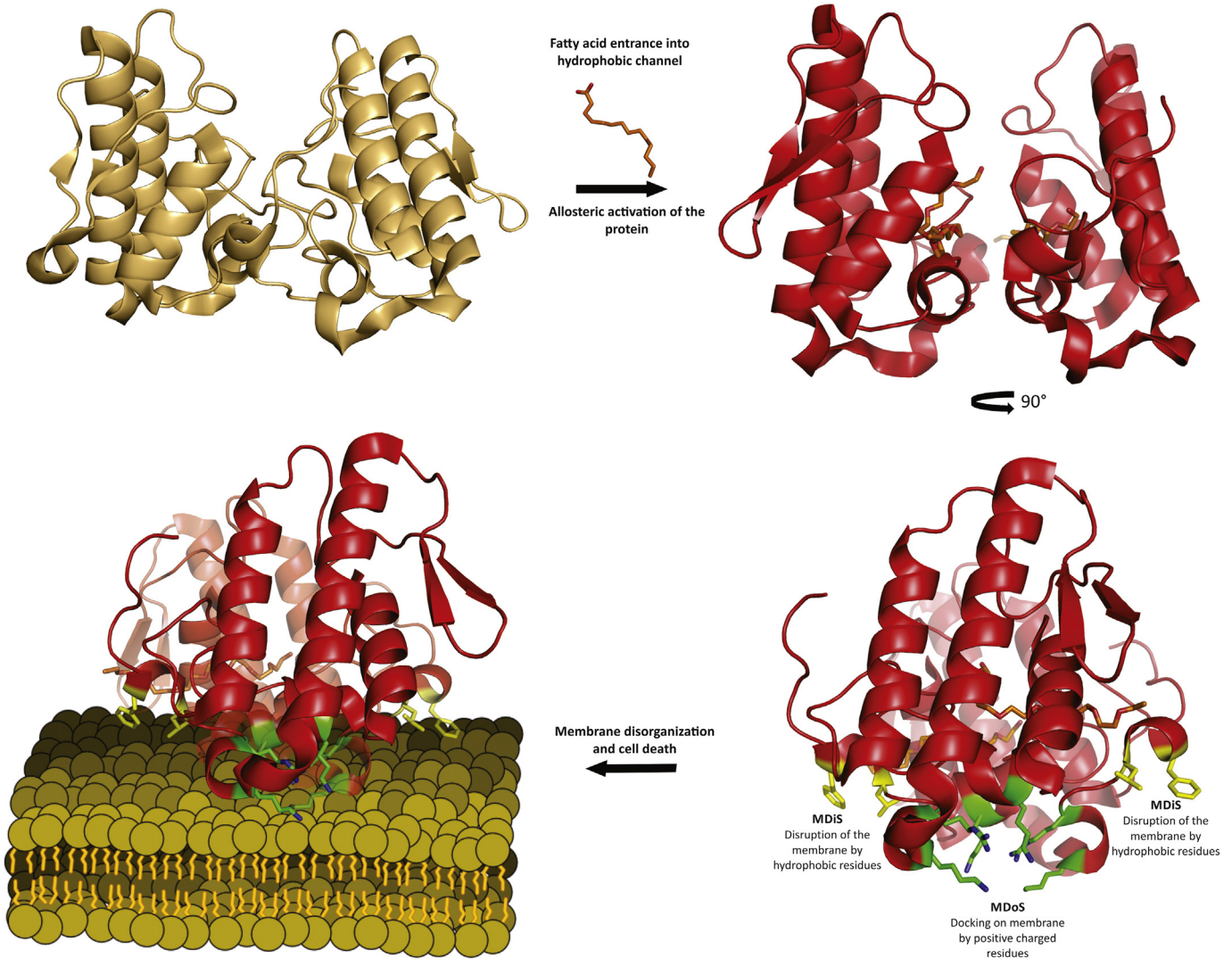


Fig. 2. Myotoxic mechanism of Lys49-PLA_2s. The entrance of a fatty acid at the hydrophobic channel of the protein leads to an allosteric activation by the dimer reorientation. In the active form, the cationic membrane-docking site (MDoS) and the hydrophobic membrane-disruption site (MDiS) became aligned on the same plane, exposed to solvent and in symmetric position on both monomers. The MDoS region stabilizes the protein on membrane by interaction of charged residues with phospholipid head group. Subsequently, the MDiS region by penetration of hydrophobic residues destabilizes the membrane causing cell death. Other PLA_2-like proteins may have similar myotoxic mechanism due to the presence of fatty acids in their hydrophobic channels and putative MDoS and MDiS on their crystallographic structures.

complexes between these toxins and ligands have been performed by co-crystallization assays and by the crystallization of chemically modified proteins. PrTX-I from *B. pirajai* and BthTX-I from *B. jararacussu* were co-crystallized with α-tocopherol [70]. Both PrTX-I/α-tocopherol and BthTX-I/α-tocopherol have very similar structures with ligands bound in the hydrophobic channel. Interestingly, α-tocopherol molecules do not present the same interaction with the residues of the monomers, and the quality of the electron density maps indicates that the ligand occupancy in one monomer is higher compared to the other. This non-symmetric binding behavior of the ligands may be due to the asymmetric structures of the monomers prior to ligand binding, as observed by dos Santos et al. [70].

A comparative structural study between native BthTX-I (apo) and BthTX-I chemically modified by *p*-bromophenacyl bromide (BPB) was published [58]. Interestingly, the authors crystallized each version of the toxin (apo and chemically modified BthTX-I) in two different physicochemical conditions, leading to different crystal packing (three different space groups) with the presence of one or two monomers in the asymmetric units. The analysis of their unit cells revealed that all structures most likely adopt the "alternative dimer" quaternary assembly. BPB molecules were found covalently bound to His48 in BthTX-I/BPB complexes, similarly to catalytic PLA_2s structures [49,101,102]. In the case of catalytically active PLA_2s, the presence of the BPB molecules in their catalytic site abolishes their enzymatic activity. Surprisingly, the binding of BPB ligands to a "catalytic-related" residue (His48) causes the partial inhibition of myotoxic activity for Lys49-PLA_2s, demonstrating that the steric hindrance of this "pseudo-catalytic site" is fundamental to the inhibition of the myotoxic activity of these toxins. An obvious explanation for this phenomenon is that the presence of the BPB molecule prevents the binding of hydrophobic molecules (e.g., fatty acids) at the hydrophobic channel, thus precluding the toxin from acquiring its "active state"—the first step of the proposed myotoxic mechanism (Section 6). In contrast to this possible explanation, the analysis of BthTX-I/BPB structure revealed an oligomeric conformation similar to other complexed Lys49-PLA_2s, i.e., its conformation is in the so-called "active state". Thus, two hypotheses can explain the BPB inhibition process: i) BPB prevents the binding of fatty acids to the toxin, avoiding conventional protein–membrane docking. Due to the presence of the MDoS of the toxin, it could still bind weakly to the membrane, explaining why BPB-complexed toxins have partially inhibited activity; and ii) despite the observation that BPB-complexed proteins have an oligomeric conformation similar to "active-state toxins", the lack of

Table 3
Ligands modeled in crystallographic structures of PLA_2-like proteins involved in activation or inhibition of their myotoxic activity.

Ligands	Structures found	PDB IDs	Main localization site	Function
Sulfate ions	BthTX-I; BnIV; MjTX-II; MTX-II; Myotoxin from *A. contortix laticinctus*; MyoII from *A. nummifer*; PrTX-I; Zhaoermiatoxin	3IQ3; 3CXI; 3MLM; 1XXS; 4K06; 1S8G; 1S8H; 1S8I; 2AOZ; 3CYL; 2PH4	Positively charged residues exposed to solvent	Indication of regions that PLA_2 could interact with membranes [95]
Polyethylene glycol (PEG)	BthTX-I; MjTX-II; MTX-II; Zhaoermiatoxin	3IQ3; 4KF3; 4K06; 2PH4	Hydrophobic channel	Allosteric activator [58,72]
Fatty acids[a]	BnIV; Ecarpholin S; MjTX-II; myotoxin from *A. contortix laticinctus*	3MLM; 2QHD; 1XXS; 1S8G checar	Hydrophobic channel	Allosteric activator [57,62,74]
Alpha-tocopherol	BthTX-I and PrTX-I	3CXI; 3CYL	Hydrophobic channel	Putative allosteric activator [70]
ρ-Bromophenacyl bromide (BPB)	BthTX-I and PrTX-I	3HZW; 3IQ3; 2OK9	Covalent bond with His48	50% inhibition of creatine kinase release in vivo and 50% inhibition liposome disruption activity [58,105]
Suramin	Ecarpholin S and MyoII from *B. asper*	3BJW; 1Y4L	Interactions with N and C-terminal from both monomers; all hydrophobic channel (Ecarpholin S), entrance of hydrophobic channel (MyoII)	85% inhibition of creatine kinase release on in vitro and in vivo assays [62,69]
Rosmarinic acid	PrTX-I	3QNL	Entrance of hydrophobic channel	90% reduction of the blockade of indirectly evoked contractions induced by the protein [103]
Caffeic acid	PrTX-I	Preliminary structure[b]	Lys20, Lys115 and Arg118 (MDoS)	Putative inhibitor

[a] Despite of crystal structure of PrTX-II was modeled with a fatty acid (N-tridecanoic acid) on hydrophobic channel, this structure it was not included in this list due to evidences that this fatty acid could be a PEG molecule, as pointed in Section 3 and by Fernandes et al. [58].
[b] See Shimabuku et al. [104].

interactions between hydrophobic molecules and residues from the hydrophobic channel could create a more unstable conformation that is less likely to act efficiently in the membrane. Both events may occur simultaneously, but these hypotheses need to be tested using other biochemical, biophysical, or functional assays.

The structure of myotoxin II from *Bothrops asper* co-crystallized with suramin (an anti-trypanosomal drug that neutralizes its myotoxic activity) presented a very interesting conformation in which the ligand interacts simultaneously with the two monomers of the dimeric structure [69]. Two important observations could be obtained with this structure: i) only the "alternative dimer" is possible because if the "conventional dimer" is chosen, half of the ligand would be in contact with the solvent, which would be energetically unfavorable; ii) a new inhibition mode is observed by which the ligand restricts access to the hydrophobic channel. The complexation of the Ecarpholin S with suramin was also able to reduce the myotoxic activity of this toxin drastically [62]. In the crystal structure of this complex, the inhibitor induces toxin oligomerization and blocks the access to the hydrophobic channel [62], as observed in the BaspTX-II-suramin complex [69]. However, the binding of suramin to Ecarpholin S establishes a wide range of interactions along the hydrophobic channels that are not observed in the BaspTX-II-suramin structure.

More recently, the crystal structure of PrTX-I from *B. pirajai* co-crystallized with rosmarinic acid (RA—a plant component that neutralizes its neuromuscular blocking activity and its myotoxicity) was solved [103]. This structure revealed that the ligand interacts with the toxin at the entrance of its hydrophobic channel. The authors hypothesized that the inhibition occurs because RA impairs the binding of hydrophobic molecules at the hydrophobic channel. In addition, this structure presents just one RA molecule bound with one monomer, and a PEG molecule interacts with the other monomer. This observation is in agreement with the asymmetry observed for the monomers of Lys49-PLA_2s [70].

Finally, PrTX-I from *B. pirajai* was co-crystallized with caffeic acid [104]. This study showed electron density maps corresponding to caffeic acid molecules that were found in the C-terminal region of the toxin. This result shows that the ligand is bound to a similar region where sulfate ions were found in other Lys49-PLA_2 structures [70] which is proposed region for the "cationic membrane-docking site (MDoS)" (Section 5) [72].

These structural and functional studies with Lys49-PLA_2s and ligands allow us to classify the latter into at least three "classes" (Table 3): i) ligands bound into the hydrophobic channel (e.g., BPB, polyethylene glycol, and α-tocopherol); ii) ligands which block or restrict access to the hydrophobic channel (e.g., rosmarinic acid and suramin); and iii) ligands which bind to C-termini or the "membrane-docking site" (e.g., caffeic acid and sulfate or phosphate ions). All of the structural evidence obtained from the study of Lys49 myotoxin/ligand complexes is in agreement with the integrated mechanism of toxicity proposed here.

8. Conclusions and future perspectives

Recently, an intense data expansion regarding the structural features of PLA_2-like proteins has provided significant insights into the mechanisms by which these proteins induce sarcolemma disorganization. All these data strongly support the alternative dimer assembly as the active quaternary structure of these proteins, and they offer a solid basis for a revised myotoxic mechanism that is able to integrate all the available biochemical and structural data. Structural studies with different inhibitors also corroborate this integrated mechanism of action and show different inhibition modes of PLA_2-like proteins. This knowledge should provide a useful molecular basis for the development of novel neutralizing strategies to improve the treatment of viperid snakebites.

References

[1] World Health Organization, Rabies and envenomings: a neglected public health issue, W.H.O., Geneva, 2007.
[2] A. Kasturiratne, A.R. Wickremasinghe, N. de Silva, N.K. Gunawardena, A. Pathmeswaran, R. Premaratna, et al., The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths, PLoS Med. 5 (2008) e218.
[3] R.A. Harrison, A. Hargreaves, S.C. Wagstaff, B. Faragher, D.G. Lalloo, Snake envenoming: a disease of poverty, PLoS Negl. Trop. Dis. 3 (2009) e569.
[4] D. Williams, J.M. Gutiérrez, R. Harrison, D.A. Warrell, J. White, K.D. Winkel, et al., The global snake bite initiative: an antidote for snake bite, Lancet 375 (2010) 89–91.
[5] D.A. Warrell, Snake bite, Lancet 375 (2010) 77–88, http://dx.doi.org/10.1016/S0140-6736(09)61754-2.
[6] J.J. Calvete, Snake venomics: from the inventory of toxins to biology, Toxicon 75 (2013) 44–62.
[7] L.F. Sousa, C.A. Nicolau, P.S. Peixoto, J.L. Bernardoni, S.S. Oliveira, J.A. Portes-Junior, et al., Comparison of phylogeny, venom composition and neutralization by antivenom in diverse species of *Bothrops* complex, PLoS Negl. Trop. Dis. 7 (2013) e2442.
[8] B. Lomonte, J. Fernández, L. Sanz, Y. Angulo, M. Sasa, J.M. Gutiérrez, et al., Venomous snakes of Costa Rica: biological and medical implications of their venom proteomic profiles analyzed through the strategy of snake venomics, J. Proteomics 105 (2014) 323–339.

[9] J.M. Gutiérrez, B. Lomonte, Efectos locales en el envenenamiento ofídico en América Latina, in: J.L. Costa Cardoso, F.O. de Siqueira França, F.H. Wen, C.M. Sant'Ana Málaque, V. Haddad (Eds.), Animais Peçonhentos no Brasil: Biologia, Clínica e Terapêutica dos Acidentes, Sarvier Editora, São Paulo, Brasil, 2003, pp. 310–323.

[10] J.B. Harris, M.J. Cullen, Muscle necrosis caused by snake venoms and toxins, Electron Microsc. Rev. 3 (1990) 183–211.

[11] B. Lomonte, J.M. Gutiérrez, Phospholipases A2 from viperidae snake venoms: how do they induce skeletal muscle damage? Acta Chim. Slov. 58 (2011) 647–659.

[12] E.A. Dennis, J. Cao, Y.-H. Hsu, V. Magrioti, G. Kokotos, Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention, Chem. Rev. 111 (2011) 6130–6185.

[13] E. Condrea, J.E. Fletcher, B.E. Rapuano, C.C. Yang, P. Rosenberg, Effect of modification of one histidine residue on the enzymatic and pharmacological properties of a toxic phospholipase A2 from *Naja nigricollis* snake venom and less toxic phospholipases A2 from *Hemachatus haemachatus* and *Naja atra* snake venoms, Toxicon 19 (1981) 61–71.

[14] J.M. Gutiérrez, B. Lomonte, Phospholipases A2: unveiling the secrets of a functionally versatile group of snake venom toxins, Toxicon 62 (2013) 27–39.

[15] J.M. Maraganore, G. Merutka, W. Cho, W. Welches, F.J. Kézdy, R.L. Heinrikson, A new class of phospholipases A2 with lysine in place of aspartate 49. Functional consequences for calcium and substrate binding, J. Biol. Chem. 259 (1984) 13839–13843.

[16] B. Lomonte, Y. Angulo, L. Calderón, An overview of lysine-49 phospholipase A2 myotoxins from crotalid snake venoms and their structural determinants of myotoxic action, Toxicon 42 (2003) 885–901, http://dx.doi.org/10.1016/j.toxicon.2003.11.008.

[17] A. Soares, M. Fontes, J. Giglio, Phospholipase A2 myotoxins from *Bothrops* snake venoms: structure–function relationship, Curr. Org. Chem. 8 (2004) 1677–1690.

[18] B. Lomonte, Y. Angulo, M. Sasa, J.M. Gutiérrez, The phospholipase A2 homologues of snake venoms: biological activities and their possible adaptive roles, Protein Pept. Lett. 16 (2009) 860–876.

[19] J. dos Santos, C. Fernandes, A. Magro, M. Fontes, The intriguing phospholipases A2 homologues: relevant structural features on myotoxicity and catalytic inactivity, Protein Pept. Lett. 16 (2009) 887–893.

[20] J.M. Maraganore, R.L. Heinrikson, The lysine-49 phospholipase A2 from the venom of *Agkistrodon piscivorus piscivorus*. Relation of structure and function to other phospholipases A2, J. Biol. Chem. 261 (1986) 4797–4804.

[21] B. Lomonte, J. Rangel, Snake venom Lys49 myotoxins: from phospholipases A(2) to non-enzymatic membrane disruptors, Toxicon 60 (2012) 520–530.

[22] S.Y. Liu, K. Yoshizumi, N. Oda, M. Ohno, F. Tokunaga, S. Iwanaga, et al., Purification and amino acid sequence of basic protein II, a lysine-49-phospholipase A2 with low activity, from *Trimeresurus flavoviridis* venom, J. Biochem. (Tokyo) 107 (1990) 400–408.

[23] Y. Shimohigashi, A. Tani, H. Matsumoto, K. Nakashima, Y. Yamaguchi, Lysine-49-phospholipases A2 from *Trimeresurus flavoviridis* venom are membrane-acting enzymes, J. Biochem. (Tokyo) 118 (1995) 1037–1044.

[24] L. Rodrigues-Simioni, J. Prado-Franceschi, A.C. Cintra, J.R. Giglio, M.S. Jiang, J.E. Fletcher, No role for enzymatic activity or dantrolene-sensitive Ca^{2+} stores in the muscular effects of bothropstoxin, a Lys49 phospholipase A2 myotoxin, Toxicon 33 (1995) 1479–1489.

[25] Y. Yamaguchi, Y. Shimohigashi, T. Chiwata, A. Tani, T. Chijiwa, B. Lomonte, et al., Lys-49-phospholipases A2 as active enzyme for beta-arachidonoyl phospholipid bilayer membranes, Biochem. Mol. Biol. Int. 43 (1997) 19–26.

[26] A.C. Mancin, A.M. Soares, C.A. Giglio, S.H. Andrião-Escarso, C.A. Vieira, J.R. Giglio, The histamine releasers crotamine, protamine and compound 48/80 activate specific proteases and phospholipases A2, Biochem. Mol. Biol. Int. 42 (1997) 1171–1177.

[27] A.M. Soares, Y. Oshima-Franco, C.A. Vieira, G.B. Leite, J.E. Fletcher, M.-S. Jiang, et al., Mn(2+) ions reduce the enzymatic and pharmacological activities of bothropstoxin-I, a myotoxic Lys49 phospholipase A(2) homologue from *Bothrops jararacussu* snake venom, Int. J. Biochem. Cell Biol. 34 (2002) 668–677.

[28] C.J. van den Bergh, A.J. Slotboom, H.M. Verheij, G.H. de Haas, The role of aspartic acid-49 in the active site of phospholipase A2. A site-specific mutagenesis study of porcine pancreatic phospholipase A2 and the rationale of the enzymatic activity of [lysine49]phospholipase A2 from *Agkistrodon piscivorus piscivorus* venom, Eur. J. Biochem. FEBS 176 (1988) 353–357.

[29] Y. Li, B.Z. Yu, H. Zhu, M.K. Jain, M.D. Tsai, Phospholipase A2 engineering. Structural and functional roles of the highly conserved active site residue aspartate-49, Biochemistry (Mosc) 33 (1994) 14714–14722.

[30] R.J. Ward, L. Chioato, A.H.C. de Oliveira, R. Ruller, J.M. Sá, Active-site mutagenesis of a Lys49-phospholipase A2: biological and membrane-disrupting activities in the absence of catalysis, Biochem. J. 362 (2002) 89–96.

[31] T. Petan, I. Krizaj, J. Pungercar, Restoration of enzymatic activity in a Ser-49 phospholipase A2 homologue decreases its Ca(2+)-independent membrane-damaging activity and increases its toxicity, Biochemistry (Mosc) 46 (2007) 12795–12809.

[32] J. Fernández, P. Caccin, G. Koster, B. Lomonte, J.M. Gutiérrez, C. Montecucco, et al., Muscle phospholipid hydrolysis by *Bothrops asper* Asp49 and Lys49 phospholipase A_2 myotoxins—distinct mechanisms of action, FEBS J. 280 (2013) 3878–3886.

[33] J.M. Gutiérrez, B. Lomonte, L. Cerdas, Isolation and partial characterization of a myotoxin from the venom of the snake *Bothrops nummifer*, Toxicon 24 (1986) 885–894.

[34] B. Lomonte, J. Lundgren, B. Johansson, U. Bagge, The dynamics of local tissue damage induced by *Bothrops asper* snake venom and myotoxin II on the mouse cremaster muscle: an intravital and electron microscopic study, Toxicon 32 (1994) 41–55.

[35] M.M. Azevedo-Marques, P. Cupo, T.M. Coimbra, S.E. Hering, M.A. Rossi, C.J. Laure, Myonecrosis, myoglobinuria and acute renal failure induced by South American rattlesnake (*Crotalus durissus terrificus*) envenomation in Brazil, Toxicon 23 (1985) 631–636.

[36] T.F. Salvini, A.C. Amaral, E.H. Miyabara, J.A. Turri, P.M. Danella, H.S. Selistre de Araújo, Systemic skeletal muscle necrosis induced by crotoxin, Toxicon 39 (2001) 1141–1149.

[37] J.M. Gutiérrez, C.L. Ownby, Skeletal muscle degeneration induced by venom phospholipases A2: insights into the mechanisms of local and systemic myotoxicity, Toxicon 42 (2003) 915–931.

[38] J.M. Gutiérrez, L.A. Ponce-Soto, S. Marangoni, B. Lomonte, Systemic and local myotoxicity induced by snake venom group II phospholipases A2: comparison between crotoxin, crotoxin B and a Lys49 PLA2 homologue, Toxicon 51 (2008) 80–92.

[39] B. Lomonte, A. Tarkowski, L.A. Hanson, Broad cytolytic specificity of myotoxin II, a lysine-49 phospholipase A2 of *Bothrops asper* snake venom, Toxicon 32 (1994) 1359–1369.

[40] B. Lomonte, Y. Angulo, S. Rufini, W. Cho, J.R. Giglio, M. Ohno, et al., Comparative study of the cytolytic activity of myotoxic phospholipases A2 on mouse endothelial (tEnd) and skeletal muscle (C2C12) cells in vitro, Toxicon 37 (1999) 145–158.

[41] M.E. Rovira, E. Carmona, B. Lomonte, Immunoenzymatic quantitation of antibodies to *Bothrops asper* myotoxins after polyvalent antivenom administration in mice, Braz. J. Med. Biol. Res. 25 (1992) 23–33.

[42] C. Díaz, G. León, A. Rucavado, N. Rojas, A.J. Schroit, J.M. Gutiérrez, Modulation of the susceptibility of human erythrocytes to snake venom myotoxic phospholipases A(2): role of negatively charged phospholipids as potential membrane binding sites, Arch. Biochem. Biophys. 391 (2001) 56–64.

[43] E. Moreno, J.M. Gutiérrez, Body distribution of *Bothrops asper* (terciopelo) snake venom myotoxin and its relationship to pathological changes, Toxicon 26 (1988) 403–409.

[44] C. Montecucco, J.M. Gutiérrez, B. Lomonte, Cellular pathology induced by snake venom phospholipase A2 myotoxins and neurotoxins: common aspects of their mechanisms of action, Cell. Mol. Life Sci. 65 (2008) 2897–2912.

[45] S. Incerpi, P. de Vito, P. Luly, S. Rufini, Effect of ammodytin L from *Vipera ammodytes* on L-6 cells from rat skeletal muscle, Biochim. Biophys. Acta 1268 (1995) 137–142.

[46] J.C. Villalobos, R. Mora, B. Lomonte, J.M. Gutiérrez, Y. Angulo, Cytotoxicity induced in myotubes by a Lys49 phospholipase A2 homologue from the venom of the snake *Bothrops asper*: evidence of rapid plasma membrane damage and a dual role for extracellular calcium, Toxicol. in Vitro 21 (2007) 1382–1389.

[47] M. Cintra-Francischinelli, P. Pizzo, L. Rodrigues-Simioni, L.A. Ponce-Soto, O. Rossetto, B. Lomonte, et al., Calcium imaging of muscle cells treated with snake myotoxins reveals toxin synergism and presence of acceptors, Cell. Mol. Life Sci. 66 (2009) 1718–1728.

[48] M. Cintra-Francischinelli, P. Caccin, A. Chiavegato, P. Pizzo, G. Carmignoto, Y. Angulo, et al., *Bothrops* snake myotoxins induce a large efflux of ATP and potassium with spreading of cell damage and pain, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14140–14145.

[49] R. Renetseder, S. Brunie, B.W. Dijkstra, J. Drenth, P.B. Sigler, A comparison of the crystal structures of phospholipase A2 from bovine pancreas and *Crotalus atrox* venom, J. Biol. Chem. 260 (1985) 11627–11634.

[50] B. Lomonte, E. Moreno, A. Tarkowski, L.A. Hanson, M. Maccarana, Neutralizing interaction between heparins and myotoxin II, a lysine 49 phospholipase A2 from *Bothrops asper* snake venom. Identification of a heparin-binding and cytolytic toxin region by the use of synthetic peptides and molecular modeling, J. Biol. Chem. 269 (1994) 29867–29873.

[51] C.E. Núñez, Y. Angulo, B. Lomonte, Identification of the myotoxic site of the Lys49 phospholipase A(2) from *Agkistrodon piscivorus piscivorus* snake venom: synthetic C-terminal peptides from Lys49, but not from Asp49 myotoxins, exert membrane-damaging activities, Toxicon 39 (2001) 1587–1594.

[52] L. Chioato, A.H.C. De Oliveira, R. Ruller, J.M. Sá, R.J. Ward, Distinct sites for myotoxic and membrane-damaging activities in the C-terminal region of a Lys49-phospholipase A2, Biochem. J. 366 (2002) 971–976.

[53] D.R. Holland, L.L. Clancy, S.W. Muchmore, T.J. Ryde, H.M. Einspahr, B.C. Finzel, et al., The crystal structure of a lysine 49 phospholipase A2 from the venom of the cottonmouth snake at 2.0-A resolution, J. Biol. Chem. 265 (1990) 17649–17656.

[54] R.K. Arni, R.J. Ward, J.M. Gutierrez, A. Tulinsky, Structure of a calcium-independent phospholipase-like myotoxic protein from *Bothrops asper* venom, Acta Crystallogr. D Biol. Crystallogr. 51 (1995) 311–317.

[55] R.K. Arni, R.J. Ward, Phospholipase A2—a structural review, Toxicon 34 (1996) 827–841.

[56] W.H. Lee, M.T. da Silva Giotto, S. Marangoni, M.H. Toyama, I. Polikarpov, R.C. Garratt, Structural basis for low catalytic activity in Lys49 phospholipases A2—a hypothesis: the crystal structure of piratoxin II complexed to fatty acid, Biochemistry (Mosc) 40 (2001) 28–36.

[57] L. Watanabe, A.M. Soares, R.J. Ward, M.R.M. Fontes, R.K. Arni, Structural insights for fatty acid binding in a Lys49-phospholipase A2: crystal structure of myotoxin II from *Bothrops moojeni* complexed with stearic acid, Biochimie 87 (2005) 161–167.

[58] C.A.H. Fernandes, D.P. Marchi-Salvador, G.M. Salvador, M.C.O. Silva, T.R. Costa, A.M. Soares, et al., Comparison between apo and complexed structures of bothropstoxin-I reveals the role of Lys122 and Ca(2+)-binding loop region for the catalytically inactive Lys49–PLA(2)s, J. Struct. Biol. 171 (2010) 31–43.

[59] P. Delatorre, B.A.M. Rocha, T. Santi-Gadelha, C.A.A. Gadelha, M.H. Toyama, B.S. Cavada, Crystal structure of Bn IV in complex with myristic acid: a Lys49 myotoxic phospholipase A_2 from *Bothrops neuwiedi* venom, Biochimie 93 (2011) 513–518.

[60] G.H.M. Salvador, W.L.G. Cavalcante, J.I. Dos Santos, M. Gallacci, A.M. Soares, M.R.M. Fontes, Structural and functional studies with mytoxin II from *Bothrops moojeni* reveal remarkable similarities and differences compared to other catalytically inactive phospholipases A_2-like, Toxicon 72 (2013) 52–63.

[61] L. Chioato, E.A. Aragão, T. Lopes Ferreira, A.I. de Medeiros, L.H. Faccioli, R.J. Ward, Mapping of the structural determinants of artificial and biological membrane damaging activities of a Lys49 phospholipase A2 by scanning alanine mutagenesis, Biochim. Biophys. Acta 1768 (2007) 1247–1257.

[62] X. Zhou, T.-C. Tan, S. Valiyaveettil, M.L. Go, R.M. Kini, A. Velazquez-Campoy, et al., Structural characterization of myotoxic ecarpholin S from *Echis carinatus* venom, Biophys. J. 95 (2008) 3366–3380.

[63] A.J. Magro, A.M. Soares, J.R. Giglio, M.R.M. Fontes, Crystal structures of BnSP-7 and BnSP-6, two Lys49-phospholipases A(2): quaternary structure and inhibition mechanism insights, Biochem. Biophys. Res. Commun. 311 (2003) 713–720.

[64] Y. Angulo, J.M. Gutiérrez, A.M. Soares, W. Cho, B. Lomonte, Myotoxic and cytolytic activities of dimeric Lys49 phospholipase A2 homologues are reduced, but not abolished, by a pH-induced dissociation, Toxicon 46 (2005) 291–296.

[65] A.H. de Oliveira, J.R. Giglio, S.H. Andrião-Escarso, A.S. Ito, R.J. Ward, A pH-induced dissociation of the dimeric form of a lysine 49-phospholipase A2 abolishes Ca^{2+}-independent membrane damaging activity, Biochemistry 40 (2001) 6912–6920.

[66] M.T. da Silva Giotto, R.C. Garratt, G. Oliva, Y.P. Mascarenhas, J.R. Giglio, A.C. Cintra, et al., Crystallographic and spectroscopic characterization of a molecular hinge: conformational changes in bothropstoxin I, a dimeric Lys49-phospholipase A2 homologue, Proteins 30 (1998) 442–454.

[67] R.K. Arni, M.R. Fontes, C. Barberato, J.M. Gutiérrez, C. Díaz, R.J. Ward, Crystal structure of myotoxin II, a monomeric Lys49-phospholipase A2 homologue isolated from the venom of *Cerrophidion* (*Bothrops*) *godmani*, Arch. Biochem. Biophys. 366 (1999) 177–182.

[68] M.T. Murakami, M.M. Viçoti, J.R.B. Abrego, M.R. Lourenzoni, A.C.O. Cintra, E.Z. Arruda, et al., Interfacial surface charge and free accessibility to the PLA2-active site-like region are essential requirements for the activity of Lys49 PLA2 homologues, Toxicon 49 (2007) 378–387.

[69] M.T. Murakami, E.Z. Arruda, P.A. Melo, A.B. Martinez, S. Calil-Eliás, M.A. Tomaz, et al., Inhibition of myotoxic activity of *Bothrops asper* myotoxin II by the anti-trypanosomal drug suramin, J. Mol. Biol. 350 (2005) 416–426, http://dx.doi.org/10.1016/j.jmb.2005.04.072.

[70] J.I. dos Santos, A.M. Soares, M.R.M. Fontes, Comparative structural studies on Lys49-phospholipases A(2) from *Bothrops* genus reveal their myotoxic site, J. Struct. Biol. 167 (2009) 106–116.

[71] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.

[72] C.A.H. Fernandes, E.J. Comparetti, R.J. Borges, S. Huancahuire-Vega, L.A. Ponce-Soto, S. Marangoni, et al., Structural bases for a complete myotoxic mechanism: crystal structures of two non-catalytic phospholipases A2-like from *Bothrops brazili* venom, Biochim. Biophys. Acta 1834 (2013) 2772–2781.

[73] A. Ullah, T.A.C.B. Souza, C. Betzel, M.T. Murakami, R.K. Arni, Crystallographic portrayal of different conformational states of a Lys49 phospholipase A_2 homologue: insights into structural determinants for myotoxicity and dimeric configuration, Int. J. Biol. Macromol. 51 (2012) 209–214.

[74] A.L.B. Ambrosio, M.C. Nonato, H.S.S. de Araújo, R. Arni, R.J. Ward, C.L. Ownby, et al., A molecular mechanism for Lys49-phospholipase A2 activity based on ligand-induced conformational change, J. Biol. Chem. 280 (2005) 7326–7335.

[75] M.T. Murakami, C.C. Melo, Y. Angulo, B. Lomonte, R.K. Arni, Structure of myotoxin II, a catalytically inactive Lys49 phospholipase A2 homologue from *Atropoides nummifer* venom, Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 62 (2006) 423–426.

[76] Q. Liu, Q. Huang, M. Teng, C.M. Weeks, C. Jelsch, R. Zhang, et al., The crystal structure of a novel, inactive, lysine 49 PLA2 from *Agkistrodon acutus* venom: an ultra-high resolution, AB initio structure determination, J. Biol. Chem. 278 (2003) 41400–41408.

[77] D. Mebs, U. Kuch, F.I.V. Coronas, C.V.F. Batista, A. Gumprecht, L.D. Possani, Biochemical and biological activities of the venom of the Chinese pitviper *Zhaoermia mangshanensis*, with the complete amino acid sequence and phylogenetic analysis of a novel Arg49 phospholipase A2 myotoxin, Toxicon 47 (2006) 797–811.

[78] J.-F. Wei, T. Li, X.-L. Wei, Q.-Y. Sun, F.-M. Yang, Q.-Y. Chen, et al., Purification, characterization and cytokine release function of a novel Arg-49 phospholipase A2 from the venom of *Protobothrops mucrosquamatus*, Biochimie 88 (2006) 1331–1342.

[79] M.I. Homsi-Brandeburgo, L.S. Queiroz, H. Santo-Neto, L. Rodrigues-Simioni, J.R. Giglio, Fractionation of *Bothrops jararacussu* snake venom: partial chemical characterization and biological activity of bothropstoxin, Toxicon 26 (1988) 615–627.

[80] A.M. Moura-da-Silva, M.J. Paine, M.R. Diniz, R.D. Theakston, J.M. Crampton, The molecular cloning of a phospholipase A2 from *Bothrops jararacussu* snake venom: evolution of venom group II phospholipase A2's may imply gene duplications, J. Mol. Evol. 41 (1995) 174–179.

[81] L.C. Mancuso, M.M. Correa, C.A. Vieira, O.A. Cunha, J.J. Lachat, H.S. de Araujo, et al., Fractionation of *Bothrops pirajai* snake venom: isolation and characterization of piratoxin-I, a new myotoxic protein, Toxicon 33 (1995) 615–626.

[82] M.H. Toyama, P.D. Costa, J.C. Novello, B. de Oliveira, J.R. Giglio, M.A. da Cruz-Höfling, et al., Purification and amino acid sequence of MP-III 4R D49 phospholipase A2 from *Bothrops pirajai* snake venom, a toxin with moderate PLA_2 and anticoagulant activities and high myotoxic activity, J. Protein Chem. 18 (1999) 371–378.

[83] S.H. Andrião-Escarso, A.M. Soares, V.M. Rodrigues, Y. Angulo, C. Díaz, B. Lomonte, et al., Myotoxic phospholipases A2 in *Bothrops* snake venoms: effect of chemical modifications on the enzymatic and pharmacological properties of bothropstoxins from *Bothrops jararacussu*, Biochimie 82 (2000) 755–763.

[84] A.M. Soares, S.H. Andrião-Escarso, R.K. Bortoleto, L. Rodrigues-Simioni, R.K. Arni, R. J. Ward, et al., Dissociation of enzymatic and pharmacological properties of piratoxins-I and -III, two myotoxic phospholipases A2 from *Bothrops pirajai* snake venom, Arch. Biochem. Biophys. 387 (2001) 188–196.

[85] L.C. Corrêa, D.P. Marchi-Salvador, A.C.O. Cintra, S.V. Sampaio, A.M. Soares, M.R.M. Fontes, Crystal structure of a myotoxic Asp49-phospholipase A2 with low catalytic activity: insights into Ca^{2+}-independent catalytic mechanism, Biochim. Biophys. Acta 1784 (2008) 591–599.

[86] D.J. Rigden, L.W. Hwa, S. Marangoni, M.H. Toyama, I. Polikarpov, The structure of the D49 phospholipase A2 piratoxin III from *Bothrops pirajai* reveals unprecedented structural displacement of the calcium-binding loop: possible relationship to cooperative substrate binding, Acta Crystallogr. D Biol. Crystallogr. 59 (2003) 255–262.

[87] J.I. dos Santos, M. Cintra-Francischinelli, R.J. Borges, C.A.H. Fernandes, P. Pizzo, A.C. O. Cintra, et al., Structural, functional, and bioinformatics studies reveal a new snake venom homologue phospholipase A_2 class, Proteins 79 (2011) 61–78.

[88] M.T. Murakami, M.R. Lourenzoni, E.Z. Arruda, M.A. Tomaz, M.M. Viçoti, J.R.B. Abrego, et al., Biochemical and structural investigations of Bothropstoxin-II, a myotoxic Asp49 phospholipase A2 from *Bothrops jararacussu* venom, Protein Pept. Lett. 15 (2008) 1002–1008.

[89] R.M. Ximenes, R.S. Alves, T.P. Pereira, R.M. Araújo, E.R. Silveira, M.M. Rabello, et al., Harpalycin 2 inhibits the enzymatic and platelet aggregation activities of PrTX-III, a D49 phospholipase A2 from *Bothrops pirajai* venom, BMC Complement. Altern. Med. 12 (2012) 139.

[90] H.S. Selistre de Araujo, S.P. White, C.L. Ownby, cDNA cloning and sequence analysis of a lysine-49 phospholipase A2 myotoxin from *Agkistrodon contortrix laticinctus* snake venom, Arch. Biochem. Biophys. 326 (1996) 21–30.

[91] J.Z. Pedersen, B. Lomonte, R. Massoud, F. Gubensek, J.M. Gutiérrez, S. Rufini, Autocatalytic acylation of phospholipase-like myotoxins, Biochemistry 34 (1995) 4670–4675.

[92] C. Díaz, A. Alape, B. Lomonte, T. Olamendi, J.M. Gutiérrez, Cleavage of the NH2-terminal octapeptide of *Bothrops asper* myotoxic lysine-49 phospholipase A2 reduces its membrane-destabilizing effect, Arch. Biochem. Biophys. 312 (1994) 336–339.

[93] A.M. Soares, R. Guerra-Sá, C.R. Borja-Oliveira, V.M. Rodrigues, L. Rodrigues-Simioni, V. Rodrigues, et al., Structural and functional characterization of BnSP-7, a Lys49 myotoxic phospholipase A(2) homologue from *Bothrops neuwiedi* pauloensis venom, Arch. Biochem. Biophys. 378 (2000) 201–209.

[94] R.K. Bortoleto-Bugs, M.R. Bugs, A.A. Neto, R.J. Ward, A micelle nucleation model for the interaction of dodecyl sulphate with Lys49-phospholipases A2, Biophys. Chem. 125 (2007) 213–220.

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

[96] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.

[97] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nat. Struct. Biol. 3 (1996) 842–848.

[98] M. Naoi, M. Naoi, T. Shimizu, A.N. Malviya, K. Yagi, Permeability of amino acids into liposomes, Biochim. Biophys. Acta 471 (1977) 305–310.

[99] M.T. Murakami, U. Kuch, C. Betzel, D. Mebs, R.K. Arni, Crystal structure of a novel myotoxic Arg49 phospholipase A2 homolog (zhaoermiatoxin) from *Zhaoermia mangshanensis* snake venom: insights into Arg49 coordination and the role of Lys122 in the polarization of the C-terminus, Toxicon 51 (2008) 723–735.

[100] W.F. de Azevedo Jr, R.J. Ward, J.M. Gutiérrez, R.K. Arni, Structure of a Lys49-phospholipase A2 homologue isolated from the venom of *Bothrops nummifer* (jumping viper), Toxicon 37 (1999) 371–384.

[101] H. Zhao, L. Tang, X. Wang, Y. Zhou, Z. Lin, Structure of a snake venom phospholipase A2 modified by p-bromo-phenacyl-bromide, Toxicon 36 (1998) 875–886.

[102] A.J. Magro, A.A.S. Takeda, A.M. Soares, M.R.M. Fontes, Structure of BthA-I complexed with p-bromophenacyl bromide: possible correlations with lack of pharmacological activity, Acta Crystallogr. D Biol. Crystallogr. 61 (2005) 1670–1677.

[103] J.I. Dos Santos, F.F. Cardoso, A.M. Soares, M. Dal Pai Silva, M. Gallacci, M.R.M. Fontes, Structural and functional studies of a bothropic myotoxin complexed to rosmarinic acid: new insights into Lys49-PLA_2 inhibition, PLoS One 6 (2011) e28521.

[104] P.S. Shimabuku, C.A.H. Fernandes, A.J. Magro, T.R. Costa, A.M. Soares, M.R.M. Fontes, Crystallization and preliminary X-ray diffraction analysis of a Lys49-phospholipase A2 complexed with caffeic acid, a molecule with inhibitory properties against snake venoms, Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 67 (2011) 249–252.

[105] D.P. Marchi-Salvador, C.A.H. Fernandes, L.B. Silveira, A.M. Soares, M.R.M. Fontes, Crystal structure of a phospholipase A_2 homolog complexed with p-bromophenacyl bromide reveals important structural changes associated with the inhibition of myotoxic activity, Biochim. Biophys. Acta 1794 (2009) 1583–1590.

[106] W.L. Delano, The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, 2002..

Biochimica et Biophysica Acta 1861 (2017) 3199–3209

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Functional and structural studies of a Phospholipase A_2-like protein complexed to zinc ions: Insights on its myotoxicity and inhibition mechanism

Rafael J. Borges [a], Fábio F. Cardoso [a], Carlos A.H. Fernandes [a], Thiago R. Dreyer [a], Delkia S. de Moraes [b], Rafael S. Floriano [b], Léa Rodrigues-Simioni [b], Marcos R.M. Fontes [a,*]

[a] Departamento de Física e Biofísica, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Rua Prof. Dr. Antonio Celso Wagner Zanin, s/n, 18618-689 Botucatu, SP, Brazil
[b] Departamento de Farmacologia, Faculdade de Ciências Médicas, Universidade Estadual de Campinas (UNICAMP), Rua Tessália Vieira de Camargo, 126, 13083-862 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 27 April 2016
Received in revised form 5 August 2016
Accepted 10 August 2016
Available online 13 August 2016

Keywords:
PLA2-like proteins
Snake venom
Zinc ions
Inhibition by divalent ions
Myotoxic mechanism
X-ray crystallography

ABSTRACT

Background: One of the main challenges in snakebite envenomation treatment is the development of stable, versatile and efficient *anti*-venom therapies. Local myotoxicity in accidents involving snakes from the *Bothrops* genus is still a consequence of serum therapy inefficient neutralization that may lead to permanent sequelae in their victims. One of the classes of toxins that participate in muscle necrosis is the PLA_2-like proteins. The aim of this work was to investigate the role of zinc ions in the inhibition of PLA_2-like proteins and to advance the current knowledge of their action mechanism.
Methods: Myographic and electrophysiological techniques were used to evaluate the inhibitory effect of zinc ions, isothermal titration calorimetry assays were used to measure the affinity between zinc ions and the toxin and X-ray crystallography was used to reveal details of this interaction.
Results: We demonstrated that zinc ions can effectively inhibit the toxin by the interaction with two different sites, which are related to two different mechanism of inhibition: preventing membrane disruption and impairing the toxin state transition. Furthermore, structural study presented here included an additional step in the current myotoxic mechanism improving the comprehension of the allosteric transition that PLA_2-like proteins undergo to exert their function.
Conclusions: Our findings show that zinc ions are inhibitors of PLA_2-like proteins and suggest two different mechanisms of inhibition for these ions.
General significance: Zinc is a new candidate that can assist in *anti*-venom treatments and can promote the design of new and even more accurate structure-based inhibitors for PLA_2-like proteins.

1. Introduction

Snakebite envenomings are an important global health issue. According to recent estimates, at least 421,000 ophidian accidents occur each year, of which 20,000 cases result in death, although these number may be underestimated by poor reporting in the areas with the highest frequencies of accidents, such as regions in Asia, Africa and Latin America [1]. Specifically, in the Latin America region, the high estimates expect approximately 130,000 accidents, in which 1.8% of cases result in death *per* year, and *Bothrops* genus snakes (American lance heads from Viperidae family) are the most important responsible genus [2]. In addition to the high mortality, the local myotoxicity of bothropic snake bites may cause permanent sequelae and disability, with member amputation in cases where the *anti*-venom is not quickly administered [3]. Most of these accidents happen in rural areas, and these consequences may incapacitate rural workers [4]. This scenario led the World Health Organization to classify snakebite envenomings as a neglected tropical disease [5].

Two groups of toxins, metalloproteinases and phospholipase A_2 (PLA_2) proteins, may cause the local myonecrosis which is the main effect caused by *Bothrops* snake bite. For some snakes of the Viperidae family, PLA_2 protein is the most abundant toxin in their venoms [6,7], and its rapid damage is one of the major difficulties in treating snake venom accidents because *anti*-venom therapy is not able to reverse its effects [4]. PLA_2s are small (approximate molecular weight of 14 kDa) and stable proteins, which frequently have 7 disulfide bridges. These

Abbreviations: Phospholipase A_2, PLA_2; Membrane-Docking Site, MDoS; Membrane-Disruption Site, MDiS; Bothropstoxin I, BthTX-I; Phrenic nerve-diaphragm, PND; Resting potential, RP; Polyethylene glycol, PEG; PEG4000, PEG4k; Protein Data Bank identification code, PDB id; Root-mean square deviation, RMSD; *p*-bromophenacyl bromide, BPB.

* Corresponding author.
E-mail address: fontes@ibb.unesp.br (M.R.M. Fontes).

http://dx.doi.org/10.1016/j.bbagen.2016.08.003
0304-4165/

enzymes are able to hydrolyze the *sn*-2 ester bond of phospholipids in the lipid bilayer of micelles, vesicles and membranes, releasing lysophospholipids and fatty acids through a catalytic mechanism that is dependent on calcium ions [8,9]. A subtype of PLA_2s, the PLA_2-like proteins, also named the PLA_2-homologues due to their similar tertiary structures and common evolutionary ancestor, do not possess catalytic activity, but still present high myotoxicity and a wide range of pharmacological activities [9,10]. The proteins from PLA_2-like group have different residues at 49 positions (Lys, Ser and Arg) and may present similar myotoxic mechanisms of action, although the most studied subgroup is the Lys49-PLA_2 proteins [10].

Since the identification of PLA_2-like toxins in the 1980s, these proteins have been investigated using many different approaches to elucidate their toxic mechanism [10]. The large charge distribution on the toxins' surface and their various pharmacological and toxic effects enhance the complexity and hinder our comprehension of their mechanisms of action. Functional studies highlighted the importance of their dimeric oligomerization [11,12], as well as the C-terminal region, which has a high distribution of basic and hydrophobic residues, for myotoxicity [13,14]. Based on the different dimeric orientations observed in available crystallographic structures, PLA_2-like proteins from the *Bothrops* genus were classified as being in: i) an **inactive state**, whose toxins are native with void hydrophobic channels and ii) an **active state**, whose toxins were complexed to either hydrophobic molecules or inhibitors. Recently, based on structural and functional studies, a myotoxic mechanism has been proposed that involves specific regions in the C-terminal region, the MDoS (Membrane-Docking Site) and the MDiS (Membrane-Disruption Site) [15]. The entrance of a hydrophobic molecule in the toxin hydrophobic channel would induce an allosteric modification of the protein, with dimer reorientation [16] from the **inactive** to **active state,** enabling the MDoS to dock with the cell membrane and the MDiS to perturb cell membrane integrity [15]. After the toxin penetrates the lipid bilayer and disturbs the phospholipids and membrane integrity, ion homeostasis is disrupted [9]. Consequently, the calcium influx increases membrane damage by inducing hypercontraction of myofilaments, mitochondrial damage, calcium-dependent proteases and cytosolic PLA_2s activity, leading to muscle necrosis [9,17]. All of these indirect effects of PLA_2-like activity enhance cytotoxicity [17].

One of the main themes in current toxinology is the development of alternatives to the conventional serum therapy that does not efficiently neutralize snake venom injuries. Currently, the majority of studies with this objective focus on compounds isolated from plants used in folk medicine in developing countries [18,19]. Functional and structural studies of the toxins complexed to these compounds led to the proposal of three different possibilities for the mechanism by which the PLA_2-like proteins are inhibited [20–22]. On the other hand, only a few studies were performed to evaluate the inhibitory effects of divalent cations against snake venoms and, more particularly, PLA_2-like toxins [23–25], despite their role in physiological functions and their use in the treatment of diseases [26,27]. Zinc ions are relatively harmless to humans and have been used in different treatments. It is an effective antioxidant and anti-inflammatory agent, and it may have beneficial effects on myocardial pathologies and atherosclerosis [28]. Zinc has been administered to prevent specific types of cancer [27,29,30] and blindness in patients with age-related macular degeneration, to complement the treatment of diarrhea, to assist the immune system, and to treat the common cold, Wilson's disease, and sickle cell disease [26].

Here, the complex between bothropstoxin I (BthTX-I), a PLA_2-like toxin from *Bothrops jararacussu* venom, and zinc ions was studied using functional, calorimetric and structural methods to obtain new insights into the mechanism by which the PLA_2-like proteins are inhibited. As result of this study, it has been shown that zinc ions are able to inhibit the myotoxic effect through two different ways. Furthermore, by comparing this structure with the crystallographic structures of other PLA_2-like proteins, more details of the current myotoxic mechanism [15] were revealed, particularly those related to the conformational changes of the MDiS region and the hydrophobic channel.

2. Materials and methods

2.1. Animals

Male Swiss mice (25–30 g) obtained from the Multidisciplinary Center for Biological Investigation (CEMIB/Unicamp) were housed 10/cage at 23 °C on a 12 h light/dark cycle, with the lights on at 6 a.m. The animals had free access to food and water. For *in vitro* protocols, the animals were euthanized with isoflurane immediately prior the experiments according to the guidelines of the Brazilian College for Animal Experimentation (COBEA); this study was approved by the institutional Committee for Ethics in Animal Use (CEUA/UNICAMP, protocol no. 3312-1).

2.2. Protein purification

BthTX-I was isolated from *Bothrops jararacussu* snake venom by ion-exchange chromatography on CM-Sepharose, as previously described [31].

2.3. Twitch-tension experiments

Mouse phrenic nerve-diaphragm preparations (PND) were mounted under a resting tension of 1 g in a 5 mL organ bath containing aerated (95% O_2 and 5% CO_2) Tyrode's solution (composition, in mM: NaCl 137, KCl 2.7, $CaCl_2$ 1.8, $MgCl_2$ 0.49, NaH_2PO_4 0.42, $NaHCO_3$ 11.9 and glucose 11.1, pH 7.0) at 37 °C and allowed to stabilize for 10 min prior to use, as described elsewhere [32,33]. Supramaximal stimuli (0.1 Hz and 0.2 ms) were delivered to the nerve from a Grass S88 stimulator (Grass Instrument Co., Quincy, MA, USA), and the muscle twitches were recorded using a TRI201AD force displacement transducer coupled to a Quad Bridge Amp and LabChart 6.0 software (all from AD Instruments Pty Ltd., Bella Vista, Australia). The preparations were incubated with BthTX-I (20 μg/mL), $ZnCl_2$ (0.4 mM) or BthTX-I (20 μg/mL) + $ZnCl_2$ (0.4 mM) for 120 min or until complete neuromuscular blockade, and the changes in twitch-tension were recorded.

2.4. Intracellular recordings

The effects of BthTX-I (20 μg/mL), $ZnCl_2$ (0.4 mM) or BthTX-I (20 μg/mL) + $ZnCl_2$ (0.4 mM) on membrane resting potential (RP) were recorded using a mouse hemidiaphragm muscle mounted in a Lucite chamber containing Tyrode's solution (composition shown above), as previously described [32,34]. The RP was measured using an amplifier (AM 502 Tektronix) and digitized by an A/D converter CAD12/36 12 bits (Lynx, São Paulo, SP, Brazil) coupled to a microcomputer (Microtec, São Paulo, SP, Brazil) loaded with AqDados 5 software (Lynx, São Paulo, SP, Brazil). The RP were monitors in different regions of the muscle considering basal values as t_0 (control); the effects of BthTX-I (20 μg/mL), $ZnCl_2$ (0.4 mM) or BthTX-I (20 μg/mL) + $ZnCl_2$ (0.4 mM) were analyzed in different intervals (t_{15}, t_{30}, t_{60}, t_{90} and t_{120}).

2.5. Crystallographic studies of BthTX-I in the presence of zinc ions

Cocrystallization experiments were performed with lyophilized samples of BthTX-I and $ZnCl_2$. The protein was dissolved in ultrapure water, and the concentration was measured by determining the absorbance at 280 nm with a NanoDrop 2000c (Thermo Scientific™) using the theoretical protein molecular weight and extinction molar coefficient. Crystals were obtained by the hanging drop vapor diffusion method at 18 °C [35]. Hexagonal shape crystals were obtained from drops of 0.8 μL of protein (10.1 mg/mL), 0.09 μL of $ZnCl_2$ (1 M), 0.32 μL of $MnSO_4$ (12.5 mM) and 0.48 μL of reservoir solution composed of 0.1 M TRIS-HCl, pH 8.5,

30% polyethylene glycol 4000 (PEG4k) and 0.1 M $LiSO_4$. Crystals were collected with nylon loops and flash cooled in liquid nitrogen. Datasets from crystals of BthTX-I complexed to Zn^{2+} were collected at the Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, Brazil) using the MX1 beamline on a MAR CCD 165 mm detector, and diffracted up to 2.16 Å resolution (remote dataset in Table 1). To ensure that the metal was present in the crystallographic structure, isomorphous crystals were diffracted by exploiting the zinc anomalous signal in the MX2 beamline and collected on a MarMosaic 225 CCD detector at wavelength of 1.283 Å, in which the zinc peak was determined experimentally using the Fluorescence detector Amptek XR100CR. The datasets were indexed, integrated, and scaled using XDS [36] in the trigonal space group, although the anomalous dataset were processed by specifically distinguishing the Friedel pairs using scaling, strategy and completeness statistics, with strict absorption correction and a 3.24 Å resolution cut, where the anomalous signal was still strong. The structure was solved by the molecular replacement technique using the *apo* (native) BthTX-I structure (PDB id: 3HZD) and the PHASER program [37]. Refinement was performed using Phenix.refine [38] and the model was manually build with Coot [39] using the F_{obs}-F_{calc} and $2F_{obs}$-F_{calc} maps. The side chains of residues K19, L31, K35, K83, K106, and K116 of monomer A and K53, K60, K61, E84, K105, K112 and K116 of monomer B were excluded due to the lack of electron density map. In the final stages of refinement, the optimization of the X-ray/stereochemistry weight and X-ray/ADP weight, occupancy and addition of waters options were included in the refinement. The zinc sites were located by generating the anomalous difference map after the protein phases were obtained by restrained refinement of the high-resolution isomorphous model. The final model is in agreement with the validation results from MOLPROBITY [40], and it was deposited in the PDB under the code 4WTB.

PISA [41], an algorithm that evaluates the interfaces in macromolecular crystallographic structures using thermodynamics, was used to foresee the BthTX-I/Zn biological unit. The MOLE program [42], an algorithm that is proposed to locate and characterize molecular tunnels and pores, was used with the default options to identify the hydrophobic channel and tunnels that access H48, their constituent residues and their physicochemical properties. Structural comparisons were performed using the following structures, with the Protein Data Bank identification code in parentheses (PDB id): *apo* BthTX-I (3HZD and 3I3H), BthTX-I/PEG4k (3IQ3), dimeric BthTX-I/BPB (*p*-bromophenacyl bromide) (3HZW) [43], BthTX-I/PEG400 (2H8I) [44], BthTX-I/α-tocopherol (3CXI), PrTX-I/α-tocopherol (3CYL), *apo* PrTX-I (2Q2J) [16], PrTX-I/BPB (2OK9) [45], PrTX-I/rosmarinic acid/PEG330 (3QNL) [20], PrTX-I/caffeic acid/PEG4k (4YU7), PrTX-I/aristolochic acid/PEG4k (4YZ7) [21], PrTX-II/*n*-tridecanoic acid (1QLL) [46], *apo* BnSP6 (1PC9), *apo* BnSP7 (1PA0) [47], BbTX-II (4K09), MTX-II/PEG4k (4K06) [15], BaspTX-II (1CLP) [48], BaspTX-II/suramin (1Y4L) [49], BnIV/myristic acid (3MLM) [50], MjTX-II/stearic acid (1XXS) [51], MjTX-II/PEG4k (4KF3) [52] and MjTX-II/suramin/PEG4k (4YV5) [22]. The structures complexed with either PEG or fatty acid may be analyzed together because the protein structures are similar; both ligands induce the same effect, and their discrimination by electron density and B-factor analysis alone is difficult [53]. Cartoon and sticks images were generated by PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.) and tunnels images were generated by MOLE local version 2.13.9.6.

Table 1
X-ray data collection and refinement.

Datasets	remote	anomalous
Beamlines in LNLS	MX1	MX2
Wavelength (Å)	1.608	1.283
Space group	$P3_121$	$P3_121$
Unit cell (Å)	a = 56.6, b = 56.6, c = 129.5	a = 57.2, b = 57.2, c = 131.3
Resolution (Å)	45.83–2.16 (2.28–2.16) [a]	49.5–3.24 (3.43–3.24) [a]
Total reflections	52,634	63,982
Unique reflections	13,020 (2051) [a]	7626 (1228) [a]
Average redundancy	3.9	8.4
R_{means} [b] (%)	8.9 (116.2) [a]	9.6 (18.6) [a]
CC (1/2)	99.7 (79.6) [a]	99.7 (98.8)
Completeness (%)	96.5 (96.5) [a]	99.7 (98.4%)
Average I/σ (I)	13.2 (2.0) [a]	19.0 (11.3) [a]
Anomalous correction		29 (2) [a]
Anomalous signal		1.16 (0.82) [a]
R [c] (%)	19.1	
R_{free} [c] (%)	23.4	
RMS deviations from ideal values		
Bond lengths (Å)	0.004	
Bond angles (°)	0.883	
Ramachandran plot (%)		
Residues in most favorable regions	96.7	
Residues in accepted regions	3.3	
Molprobity score	1.20	
Number of molecules/atoms, their averaged B factor (Å^2) and averaged occupancy		
Protein	2 (34.5) [1.0] [b]	
SO_4	2 (48.3) [0.7] [b]	
H_2O	116 (34.7) [1.0] [b]	
Cl^-	2 (41.8) [0.8] [b]	
Zn^{2+}	3 (43.4) [0.6] [b]	

[a]Number in parenthesis are for highest resolution shell. [b]Number in parenthesis and in brackets are B factor and occupation, respectively. $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$, $\hat{I}_h = \sum_h \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$ redundancy independent R-factor. Calculated for $I > -3\sigma$ (I). [c]$R_{cryst} = R_{hkl} \frac{||Fobs_{hkl}| - |Fcalc_{hkl}||}{|Fobs_{hkl}|}$. [d]$R_{free}$ is equivalent to R_{crist}, but calculated with reflections (5%) omitted from the refinement process. RMS *root-mean-square*. Anomalous correction: percentage of correlation between random half-sets of anomalous intensity differences. Anomalous signal: mean anomalous difference in units of its estimated standard deviation $R_{hkl} \frac{|F(+) - F(-)_{R_{cryst}}|}{\sigma}$. F(+), F(−) are structure factor estimates obtained from the merged intensity observations in each parity class. LNLS stands for Brazilian Synchrotron Light Laboratory.

2.6. Calorimetric studies

Calorimetric studies were performed with a microcalorimeter iTC_{200} (GE HealthCare) at 25 °C. $ZnCl_2$ (2 mM) were titrated (40 injections of 1 μL at each 240 s) into the calorimetric cell containing BthTX-I (53 μM) diluted in 100 mM Tris HCl, 100 mM NaCl, pH 7.0. Dilution and stirring effects were determined in control titrations and were subtracted from the BthTX-I - Zn^{2+} titration prior to data analysis, which was performed with Origin v.7.0 software (OriginLab, Northampton, MA, EUA).

3. Results

3.1. Functional experiments

BthTX-I (20 μg/mL) caused irreversible and complete neuromuscular blockade in indirectly stimulated PND preparations from 90 min of incubation. In the preparations incubated with only $ZnCl_2$ (0.4 mM), there was 20 ± 5% reduction on the twitch-tension response after 120 min incubation compared to the control preparations. When the PND preparations were treated with BthTX-I (20 μg/mL) and $ZnCl_2$ (0.4 mM), the BthTX-I-induced neuromuscular blockade was attenuated from 20 min of incubation, with a 19 ± 1% reduction in the twitch amplitude after 120 min of incubation (Fig. 1A). Fig. 1B shows a representative recording of the neuromuscular blockade induced by BthTX-I (20 μg/mL) (Fig. $1B_1$) and the lack of this effect in the presence of $ZnCl_2$ (0.4 mM) (Fig. $1B_2$). In addition, the attenuation of the BthTX-I-induced neuromuscular blockade by $ZnCl_2$ treatment was also corroborated using intracellular recordings to measure the membrane RP of muscle fibers from the PND preparations. BthTX-I (20 μg/mL) induced significant membrane depolarization from 30 min of incubation [from −80 ± 1 mV (t_0) to −27 ± 3 mV (t_{120}), $p < 0.05$, $n = 4$]. Whereas in

the presence of $ZnCl_2$ (0.4 mM), the depolarizing effect of BthTX-I was significantly attenuated [from −79 ± 1 mV (t_0) to −71 ± 2 mV (t_{120}), $p < 0.05$, n = 4]; $ZnCl_2$ (0.4 mM) alone did not induce any membrane depolarization [from −81 ± 1 mV (t_0) to −77 ± 1 mV (t_{120}), $p < 0.05$, n = 4] (Fig. 2).

3.2. Structural experiments

To understand the molecular interactions between BthTX-I and the zinc ions, this protein was crystallized in presence of this ion. Supplementary Fig. 1 shows typical crystals of the BthTX-I/Zn complex, and Table 1 summarizes the data collection and refinement statistics. BthTX-I/Zn preserves the general PLA_2-like fold. The zinc binding sites were confirmed using peaks in the anomalous difference map that were clearly distinguishable from noise (below 3 σ), with the following anomalous signals and interactions: i) Zn1: H48/A (histidine in position 48 of monomer A) with 8.0 σ (Fig. 3A); ii) Zn2: H120/A with 13.2 σ (Fig. 3B); and iii) Zn3: H120/B with 5.8 σ (Fig. 3C). The occupancy and B-factors of the zinc ions are: i) 0.54 and 39.7 Å^2; ii) 0.69 and 38.8 Å^2; iii) 0.58 and 51.8 Å^2, respectively, for Zn1, Zn2 and Zn3. All zinc ions interact with one histidine imidazole nitrogen, and its coordination is completed by water molecules in a tetrahedral geometry with the expected distances and angles, as confirmed by the CheckMyMetal web server [54]. An anomalous signal above noise is observed close to Glu84/A, although the zinc refinement trial at this position generated a B-factor higher than 100 Å^2 and an occupation lower than 0.5 using the non-anomalous high resolution dataset; thus, it was excluded from the final model. The crystallographic model quality was evaluated with validation tools, and it is within all stereochemical requirements, as verified by a MolProbity score of 1.2 (More details in Table 1).

In addition to the zinc-binding regions, a chloride (partial occupation of 0.56 and B-factor of 35.9 Å^2) was shown to interact with Arg34/A. This last residue, together with L31 and G32 from the putative calcium binding loop, also partially occupied two possible conformations, one with the Arg34 side chain pointed toward the end of the C-terminal region, and the other pointed toward the MDiS. This last alternative conformation is different from all other published structures, with exception of an *apo* BthTX-I structure (PDB id: 3I3H/B).

BthTX-I/Zn presents a homo-dimeric structure in the asymmetric unit, but as often found in other PLA_2-like structures (Fig. 4A); thus, two possible structures can be chosen: i) a compact dimer (also termed alternative) and ii) a large dimer (termed conventional). Dimer descriptions and discussion may be found elsewhere [10]. BthTX-I/Zn was solved as a compact dimer because its interface area is greater than the area of the larger dimer (770 *versus* 496.3 Å^2) with a lower free energy (−17.1 against −1.7 kcal/mol); all values were calculated using the PISA program. In addition, previous small-angle X-ray scattering experiments [44] also indicated that the compact dimer is more likely to occur in solution.

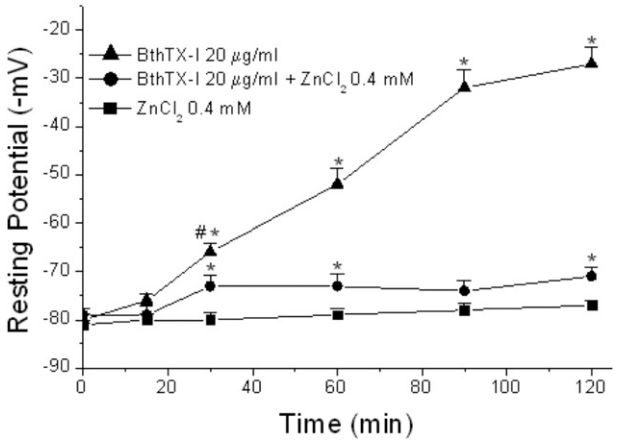


Fig. 2. Effects of BthTX-I and $ZnCl_2$ in membrane resting potential measurements of isolated mouse phrenic nerve-diaphragm preparations. $ZnCl_2$ (0.4 mM) markedly attenuated the known depolarization induced by BthTX-I (20 μg/mL) (▲) from 30 min (●). There were no important changes on membrane resting potential measurements in $ZnCl_2$-incubated preparations. Results are expressed as mean ± S.E.M. (n = 4–6). * $p < 0.05$ indicates significant differences relative to control group ($ZnCl_2$) and # $p < 0.05$ was relative to BthTX-I + $ZnCl_2$ treated group.

3.2.1. Structural comparisons of the BthTX-I/Zn monomers

The monomers of the BthTX-I/Zn are different from each other; their RMSD (root-mean square deviation) is 1.08 Å, mainly due to the variability in the residues from Y119 to F125 in the C-terminal regions (Fig. 4). Among these residues, the hydrophobic ones were described by Ambrosio et al. as able to form a hydrophobic knuckle, conserved in PLA_2-like proteins, and related to toxicity [53]. Subsequently, two residues of this region (L121 and F125) were suggested to be responsible for toxin-induced membrane disruption and called the MDiS [15]. The MDiS is stabilized by one hydrogen bond between the main chain oxygen of L121 and main chain nitrogen of F125, two hydrophobic and exposed residues and by the rigid P123 that, together with K122, form a short 3_{10} helix. BthTX-I/Zn possesses these residues in an asymmetrical configuration: monomer A is in the canonical conformation, where

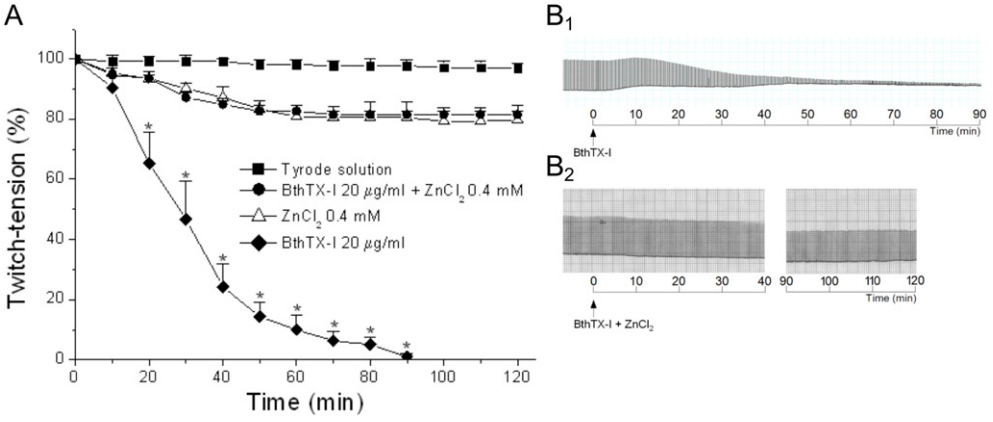


Fig. 1. Effects of BthTX-I and $ZnCl_2$ on indirectly evoked twitches in isolated mouse phrenic nerve-diaphragm preparations. (A) $ZnCl_2$ (0.4 mM) markedly attenuated BthTX-I (20 μg/mL)-induced blockade. There were no significant differences in twitch-tension responses in $ZnCl_2$-incubated preparations. Results are expressed as mean ± S.E.M. ($n = 4–6$); *$p < 0.05$ indicates significant differences relative to BthTX-I + $ZnCl_2$ treated group. (B_1) Neuromuscular blockade induced by BthTX-I in PND preparations. (B_2) Representative record shows the lack of blockade by BthTX-I (20 μg/mL) when incubated with $ZnCl_2$ (0.4 mM) after 120 min.

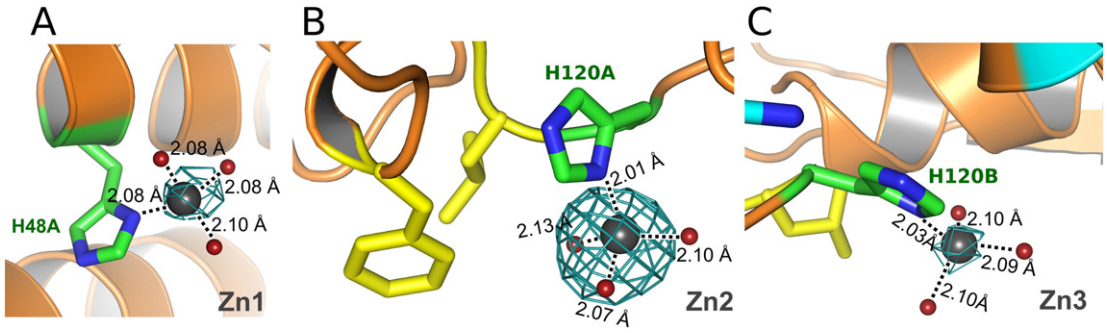


Fig. 3. BthTX-I and zinc ions interacting regions and coordination distances (black dashes). Anomalous difference map is contoured at 5σ and shown in cyan. Zinc spheres are represented in grey, water spheres in red, side chain of interacting histidine in green sticks, MDoS and MDiS side chains in cyan and yellow, respectively. (A) Zn1 and H48/A (monomer A); (B) Zn2, C-terminal H120/A and the MDiS/A; and (C) Zn3 and H120/B.

MDiS is present (Fig. 4 in orange) and will be referred to as the **canonical monomer**; monomer B possesses residues 121 to 125 far from each other in a loop secondary structure and will be referred to as the **non-canonical monomer**. In the BthTX-I/Zn structure, the distances between the L121 and F125 Cβs are 4.7 Å for the **canonical monomer** and 10.3 Å for the **non-canonical monomer**.

3.2.2. Hydrophobic channel accessibility analyses

The BthTX-I/Zn dimer has a large cavity of 2757 $Å^3$ surrounding the interior and interface of both monomers, according to the calculations from the MOLE2 algorithm (Supplementary Table 1). Inside this cavity, hydrophobic channel accessibility was assessed through tunnel computations using each H48 residue as the starting point. This amino acid is the deepest residue that interacts with different ligands found in the hydrophobic channel, such as zinc ions, *p*-bromophenacyl bromide (BPB), fatty acids and polyethylene glycol (PEG) molecules (Supplementary Fig. 2A). Three different tunnels were identified (Supplementary Table 2) in this region: **tunnel 1** (Fig. 5A, green surface and Fig. 5B) and **tunnel 2** (Fig. 5A, dark red surface and Fig. 5B) in the hydrophobic channel of the **canonical monomer** and **tunnel 3** (Fig. 5A, purple surface) in the **non-canonical monomer**. Both **tunnels 1** and **2** are long, 19.1 and 22.8 Å long, respectively, and are accessible to the solvent through the N-terminal region of the **canonical monomer** and C-terminal region of the **non-canonical monomer**. **Tunnel 3** in the **non-canonical monomer** is rather short, 6.0 Å long, and accesses exterior through residues that are essential only for catalytic PLA_2s (calcium binding loop, K49 and Y52). The hydropathy values of **tunnels 1**, **2** and **3** are 0.29, −0.07 and −0.98, respectively; thus, only **tunnel 1** is hydrophobic. A superimposition between the monomers of the BnIV/myristic acid and the BthTX-I/Zn structures shows that the myristic acid lays exactly inside **tunnel 1** of the **canonical monomer** of the BthTX-I/Zn structure (zoomed square in Fig. $5B_1$), which is in contrast to the **non-canonical monomer**, where the fatty acid clashes with the protein residues. Thus, BthTX-I/Zn has only one accessible hydrophobic channel in the **canonical monomer** (**tunnel 1**).

3.3. Calorimetric experiments

As shown in Fig. 6, the titration presented just a monotonic exothermic behavior and a binding isotherm characteristic of low affinity [55] between BthTX-I and Zn ions. In order to obtain the most likely interaction stoichiometry (n) for this assay, it was tested the following possibilities: $n = 1$, 2 or 3. According to previous studies [56], the lowest χ^2 value for the curve adjust to the experimental data represent the most likely stoichimometry to the system. The following χ^2 values were obtained: 7683, 3792 and 2691, respectively for $n = 1$, 2 or 3, suggesting that there is the binding of three Zn ions for each BthTX-I. F-test ($p < 0.05$) also confirmed that the $n = 3$ is the best fit to the isotherm. Therefore, considering three Zn binding sites at BthTX-I, the obtained dissociation constants (K_d) are 83.3 ± 25.4, 62.1 ± 25.1 and 300 ± 104.3 μM.

4. Discussion

4.1. Zinc ions and neuromuscular preparations

Classically, zinc ions have been described to be essential to two main roles in proteomics: as a cofactor for enzymes and as the essential constituent of zinc fingers and related structural motifs found in numerous transcription factors [57,58]. Moreover, this ion has recently been

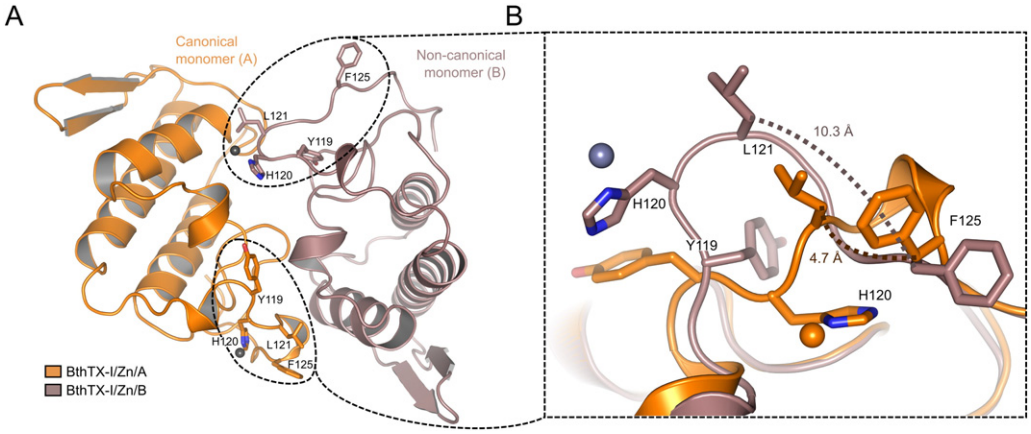


Fig. 4. Cartoon representation of the **canonical** (in orange), **non-canonical monomers** (in brown) of the BthTX-I/Zn and asymmetry between its C-termini. (A) Overall structure dimeric representation highlighting C-terminal regions inside the dashed circle. (B) C_α superposition of the C-terminal monomers displays the high asymmetry between both regions. Distances between MDiS Cβ residues (121 and 124) are showed on brown dots.

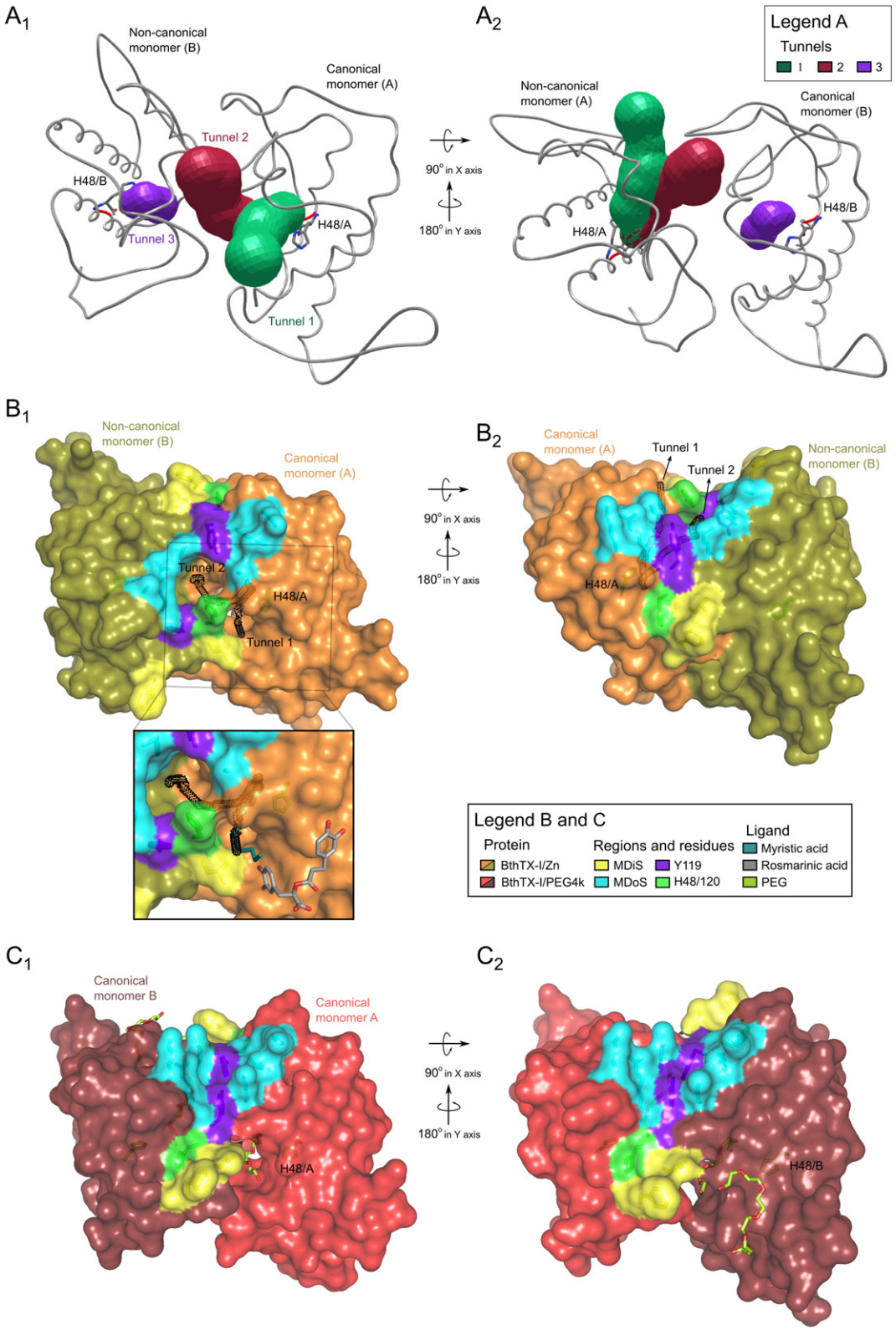


Fig. 5. Tunnels analysis for bothropic PLA_2s-like crystal structures in the **inactive** (A and B) and **active states** (C). **Tunnel 1**, which is only tunnel with hydrophobic feature, exists in **canonical monomers** (A_1, B_1, C_1 and C_2) and is accessible to the solvent through the N-terminal region of the **canonical monomer** and C-terminal region of the **non-canonical monomer** (A and B_1). (Zoom region in B_1) After superimposition, myristic acid (PDB id: 3MLM) occupies **tunnel 1** and rosmarinic acid (PDB id: 3QNL) is in **tunnel 1** entrance region. (B_2) The hydrophobic channel of the **non-canonical monomer** is inaccessible. (B and C) MDiS is in yellow, MDoS is in cyan, Y119 residue is in purple, H48 and H120 residues are in green. Myristic acid and rosmarinic acid ligands are represented in dark green and grey sticks, respectively.

implicated in a number of novel functions and pathologies of various tissues, organs and systems [59]. Here, zinc ions efficiently inhibit the myotoxicity of BthTX-I in mouse phrenic nerve-diaphragm (PND) preparations (Figs. 1A, B_2 and 2). Thus, the question arises of how zinc ions inhibit BthTX-I myotoxicity: by complexing to the toxin, by protecting PND preparation, or both? To address these questions, the effects of zinc alone were evaluated.

This study showed that zinc chloride (0.4 mM) reduced partially the evoked twitch amplitude in mouse PND preparations. Different effects of this concentration were reported for other nerve-muscle preparations, such as partial inhibition of indirect twitches in chick biventer cervicis [60], full inhibition of twitch amplitudes in rat atria [61], and no effect on the rat diaphragm [62]. The observed twitch reduction could be related to membrane damage, as zinc toxicity has been shown to be concentration- and cell type-dependent [59]. However, this hypothesis is rejected by the maintenance of the resting potential throughout entire PND experiment, at least using the evaluated zinc concentration, because sarcolemma damage would generally increase resting potential. Similar electrophysiological results with $ZnCl_2$ treatment were obtained by other groups in rat and mouse PND preparations [63].

When BthTX-I was pre-incubated with zinc ions, the PND preparations were protected from the paralyzing and depolarizing effects initiated by this PLA_2-like protein. Thus, the observed decreased in twitch amplitudes in the myographic assays following this treatment is related to another factor involved in the muscle contraction process. According to the results obtained from the myographic and electrophysiological assays and shown here in the crystallographic and calorimetric experiments, the inhibition of myotoxic activity was due to the interaction between BthTX-I and the zinc ions.

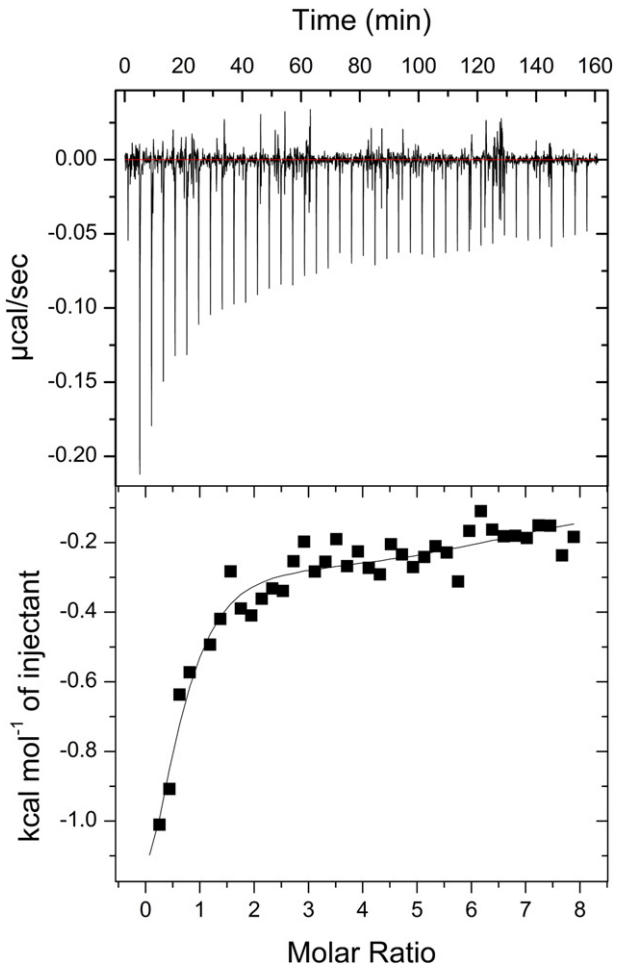


Fig. 6. Calorimetric titration of zinc ions into BthTX-I toxin. The upper panel shows the raw data thermogram (thermal power as a function of time) of the titration of BthTX-I (53 μM) with 2 mM zinc chloride. The lower panel shows the binding isotherm (ligand-normalized integrated heat as a function of the molar ratio). The best fitted binding isotherm with stoichiometry of 3 zinc ions for each BthTX-I molecule is shown as a solid line.

4.2. Zinc interaction sites in BthTX-I and their relationship to the inhibitory process

Based on previous functional and structural experiments with PLA_2-like toxins and inhibitors, three different inhibition mechanisms were suggested for these proteins: i) ligands binding inside the hydrophobic channel or in its entrance, blocking the access of fatty acid molecules; ii) ligands interacting to specific residues of the MDoS or MDiS regions [21], preventing the protein/membrane interaction; or iii) ligands inducing toxin oligomerization [22], which may result in a combination of the two previous processes. BPB and rosmarinic acid are example of the first class, and they inhibit all or a substantial fraction of the myotoxic, paralyzing and edematogenic effects of different toxins [20,64,65]. Caffeic acid and aristolochic acid are examples of the second type of inhibitor, with partial myotoxic inhibition of PrTX-I from *Bothrops pirajai* venom [21]. Curiously, suramin was initially identified as a member of the first group because of its interaction site to BaspTX-I, from *Bothrops asper* venom [49], and it was recently classified into the second and third categories for MjTX-II, from *Bothrops moojeni* venom [22], but it always possesses high inhibition against PLA_2-like proteins [22,49,66]. The crystallographic study identified that the zinc ions interact with the BthTX-I structure at two different sites (Figs. 3 and 4): i) H48 in the hydrophobic channel from the **canonical monomer** (Zn1) and ii) H120 (near the MDiS) from both the **canonical** (Zn2) and **non-canonical monomers** (Zn3). The structural data also corroborated with the calorimetric studies which suggested the stoichiometry of three Zn ions for each BthTX-I.

The conservation of H48 and the other residues from the catalytic network in PLA_2-like toxins has been attributed to the maintenance of the hydrophobic channel [10,53] for entrance of a fatty acid that would lead to an oligomeric change for PLA_2-like proteins, a fundamental step in exerting myotoxicity [10,15]. Supporting this idea, different ligands have been shown to interact with H48, such as the BPB inhibitor, fatty acids and PEG molecules (Supplementary Fig. 2). Although the dimer orientation of the BthTX-I/Zn and dimeric BthTX-I/BPB structures are different, the zinc and C atoms of BPB share almost same position interacting to toxin's H48 Nπ atom (Supplementary Fig. 2); thus, their inhibitory mechanisms may be related. It is also interesting observed that the asymmetrical configuration of the monomers for the BthTX-I/Zn structure led to the non-accessibility of zinc ions by the hydrophobic channel of the non-canonical monomer (tunnels 1 and 2 are not observed in the non-canonical monomer - section 3.2.2); thus, only the zinc ion (Zn1) of the **canonical monomer** bond to H48.

Zn2 and Zn3 interact with the H120 Nπ atom from both monomers (Figs. 3B/C and 5) in the vicinity of the MDiS region. Similarly, the PrTX-I/aristolochic acid/PEG4k and MjTX-II/suramin/PEG4k crystallographic structures show inhibitors bound to the MDiS region, whose interactions were related to their inhibitory abilities [21,22]. The superposition of the BthTX-I/Zn monomers with the latter structures reveals that zinc is located close to aristolochic acid and shares the same position as the suramin urea group (Supplementary Fig. 2 B). Therefore, the binding of Zn ions to H120 may prevent the interaction between the MDiS residues (particularly L121) and the membrane. In addition, the interaction of Zn ions with the H120 in the **non-canonical monomer** may interfere with the oligomeric change of the toxin to its **active state**, as a position inversion between Y119 and H120 is necessary. Although H120 is conserved for many bothropic PLA_2-like proteins, a natural variant of tyrosine is found in a few bothropic snake venoms [48,51,67–72], with which zinc would probably not interact, reducing its inhibitory capability.

Finally, it is interesting to highlight that H48 is a strictly conserved residue for all PLA_2-like proteins, and H120 is conserved for many bothropic PLA_2-like proteins. Thus, the inhibitory mechanism proposed here would be similar for several bothropic PLA_2-like toxins, and zinc could be a potential inhibitory agent for *anti*-bothropic snake venom therapy.

4.3. Canonical and non-canonical monomers and hydrophobic channel accessibility

Bothropic PLA_2-like proteins (compact dimer conformation, also known as the alternative dimer) have a cavity surrounded by positive and hydrophobic residues that are related to the entrance of hydrophobic or negative molecules in the hydrophobic channel [10]. All residues in this cavity are thought to be fundamentally required for the myotoxic activity of PLA_2-like proteins (*i.e.*, MDoS, MDiS, and H48) (Supplementary Table 1), supporting the importance of the hydrophobic channel and C-terminal regions for toxin activity. The entry of hydrophobic molecules into this channel is important to induce dimer reorientation from the **inactive** to **active state**, an essential step of the myotoxic mechanism of action [15,16]. In this context, *apo* PLA_2-like toxins (empty hydrophobic channel) are in the **inactive state**, and their dimers exhibit an asymmetrical conformation (one monomer is canonical and the other is non-canonical). BthTX-I/Zn presents a structure similar to *apo* PLA_2-like structures for both monomers, as can be observed by the low RMSD values (≤ 0.7 Å) of the superposition of correspondent monomers (Table 2). In contrast, the crystallographic structures of the PLA_2-like proteins with ligands bound in the hydrophobic channel have their monomers in a symmetric conformation (canonical), as observed by superposition of the BthTX-I/Zn **canonical** and **non-canonical monomers** (Table 2). **Canonical monomers** are very similar, but the comparison between the non-canonical monomer of the BthTX-I/Zn and **canonical** monomers of the complexed structures present higher RMSD values (≥ 1.1 Å) (Table 2). MjTX-II structures are exceptions because they present specific structural characteristics [15], resulting in higher RMSD compared to other toxins (Table 2).

Table 2
Root mean square deviation (RMSD, Å) of superpositions between monomers of available bothropic PLA_2-like crystallographic dimeric structures and BthTX-I/Zn canonical and non-canonical monomers. It was included only the lower RMSD value of each superposition. RMSD below 0.7 Å are shown in bold.

PDB code	BthTX-I/Zn **Canonical monomer** (chain letter)	BthTX-I/Zn **Non-canonical monomer** (chain letter)	Toxin	Ligand
3HZD	**0.3 (B)**	**0.6 (A)**	BthTX-I	–
2Q2J	**0.3 (B)**	**0.6 (A)**	PrTX-I	–
4K09	**0.3 (B)**	**0.3 (A)**	BbTX-II	–
1PA0	**0.2 (A)**	**0.4 (B)**	BnSP7	–
1PC9	**0.3 (A)**	**0.4 (B)**	BnSP6	–
3I3H	**0.4 (B)**	**0.7 (A)**	BthTX-I	–
2H8I	**0.5 (B)**	**0.5 (A)**	BthTX-I	PEG400
3MLM	**0.5 (A)**	1.1 (B)	BnIV	Myristic acid
2OK9	**0.6 (A)**	1.2 (B)	PrTX-I	BPB
3CYL	**0.6 (A)**	1.2 (B)	PrTX-I	Vitamin E and PEG4k
3IQ3	**0.7 (A)**	1.2 (B)	BthTX-I	PEG4k
1QLL	**0.6 (A)**	1.2 (B)	PrTX-II	N-tridecanoic acid
3QNL	**0.6 (A)**	1.1 (B)	PrTX-I	Rosmarinic acid and PEG330
3CXI	**0.7 (A)**	1.2 (B)	BthTX-I	Vitamin E and PEG4k
4K06	**0.5 (A)**	1.2 (B)	MTX-II	PEG4k
4YU7	**0.6 (A)**	1.1 (B)	PrTX-I	Caffeic acid and PEG4k
1Y4L	**0.6 (A)**	1.1 (B)	BaspTX-II	Suramin and PEG330
4YZ7	**0.5 (A)**	1.2 (B)	PrTX-I	Aristolochic acid and PEG4k
1CLP	**0.6 (A)**	1.2 (B)	BaspTX-II	–
3HZW	0.8 (B)	1.1 (A)	BthTX-I	BPB
4YV5	1.3 (A)	1.5 (B)	MjTX-II	Suramin and PEG4k
1XXS	1.2 (A)	1.4 (B)	MjTX-II	Stearic acid
4KF3	1.3 (A)	1.5 (B)	MjTX-II	PEG4k

Table 3
Distances (Å) between MDiS Cβ residues (121 and 124) for monomers (A and B) of **inactive** (asymmetric monomers) and **active states** (symmetric monomers) of bothropic PLA_2-like crystallographic dimeric structures. Structures are sorted by higher distance.

PDB code	**Canonical monomer** (chain letter)		**Non-canonical monomer** (chain letter)	Toxin	Ligand
3HZD	4.8 (B)		11.2 (A)	BthTX-I	–
2Q2J	4.9 (B)		10.9 (A)	PrTX-I	–
4K09	4.5 (B)		10.9 (A)	BbTX-II	–
1PA0	4.8 (A)		10.8* (B)	Bnsp7	–
1PC9	4.7 (A)		10.3* (B)	Bnsp6	–
4WTB	4.7 (A)		10.3 (B)	BthTX-I	Zn^{2+}
3HZW	6.2 (B)		9.7 (A)	BthTX-I	BPB
3I3H	4.6 (B)		9.3 (A)	BthTX-I	–
2H8I	4.9 (B)		8.5 (A)	BthTX-I	PEG400
3MLM	4.6 (A)	5.4 (B)		BnIV	Myristic acid
2OK9	4.5 (A)	5.3 (B)		PrTX-I	BPB
3CYL	4.5 (A)	5.3 (B)		PrTX-I	Vitamin E and PEG4k
3IQ3	4.7 (A)	5.2 (B)		BthTX-I	PEG4k
1QLL	5.2 (A)	5.2 (B)		PrTX-II	N-tridecanoic acid
3QNL	4.8 (A)	5.2 (B)		PrTX-I	Rosmarinic acid and PEG330
3CXI	4.7 (A)	5.1 (B)		BthTX-I	Vitamin E and PEG4k
4K06	5.1 (A)	4.9 (B)		MTX-II	PEG4k
4YU7	5.0 (A)	4.7 (B)		PrTX-I	Caffeic acid and PEG4k
4YV5	4.9 (A)	5.0 (B)		MjTX-II	Suramin and PEG4k
1Y4L	4.6 (A)	4.8 (B)		BaspTX-II	Suramin and PEG330
4YZ7	4.7 (A)	4.7 (B)		PrTX-I	Aristolochic acid and PEG4k
1XXS	4.4 (A)	4.7 (B)		MjTX-II	Stearic acid
4KF3	4.4 (A)	4.4 (B)		MjTX-II	PEG4k
1CLP	4.2 (A)	3.5* (B)		BaspTX-II	–

* Cβ was added, since this atom was omitted in the crystallographic structure.

In this work, we observed that the MDiS is stabilized by one hydrogen bond between the main chain oxygen of L121 and the main chain nitrogen of F125, two hydrophobic and exposed residues, and by the rigid P123 that, together with K122, form a short 3_{10} helix (Fig. 4B, orange cartoon). Furthermore, the distances between the Cβ atoms from L121 and F125 from all **canonical monomers** are <5.5 Å, whereas the distances for the residues from all **non-canonical monomers** are larger than 10 Å (Table 3). The BthTX-I/PEG400 structure is the only exception, as it is a complex in **inactive state** with a **non-canonical monomer,** with a measured distance that is reduced from the usual 10 Å to 7.5 Å (Table 3). The reduced distance between the MDiS Cβ atoms is probably a consequence of the presence of a short PEG molecule in the **canonical monomer** because that is the main difference compared to *apo* BthTX-I and we observed that the electron density for the other PEG400 molecule in the **non-canonical monomer** is ambiguous. Therefore, BthTX-I/PEG400 structure may represent an intermediate state between the **inactive** and **active states**. Whereas, presence of longer PEG molecules (PEG4k) induces toxin to be in **active state** with symmetrical monomers.

Canonical monomers of bothropic PLA_2-like proteins contain **tunnel 1**, as described in Section 3.2.2, which structurally connects the N-terminal region and H48 with the C-terminal region of the other monomer. As a matter of fact, the N-terminal region has been suggested to be an important region with its hydrophobic residues [73,74], and it may be associated with the membrane interaction, similar to the MDiS. PLA_2-like structures in the **active state** display symmetrical and **canonical monomers**, and **tunnel 1** is occupied by ligands, such as fatty acids or PEG molecules. Interestingly, a rosmarinic acid inhibitor is observed at the entrance of one of the PrTX-I/RA **tunnel 1** regions and interacts with the N-terminal residues of one monomer and the hydrophobic residues of the C-terminal region of the other monomer (gray sticks in the zoom of Fig. $5B_1$), preventing PEG4k entrance.

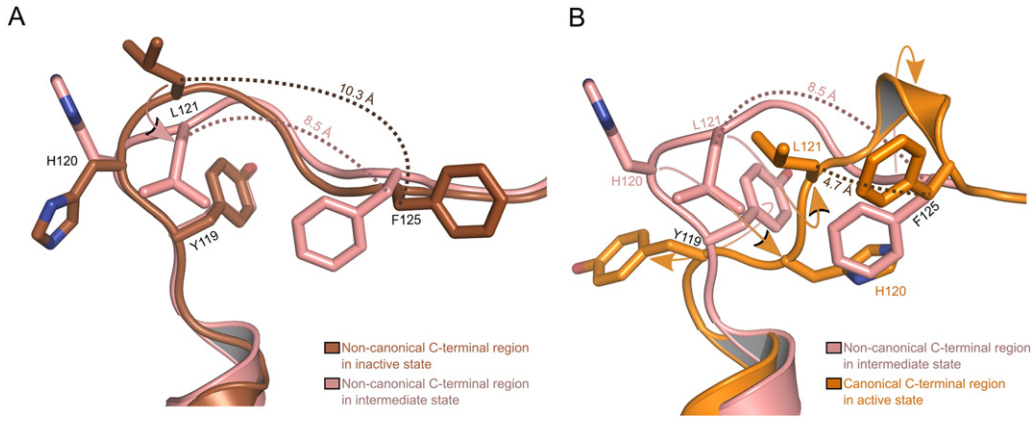


Fig. 7. Structural transitions of the C-termini from the BthTX-I that occurs during hydrophobic molecule entrance. The first state is represented by BthTX-I/Zn **non-canonical monomer** showed as a brown cartoon (A). An intermediate second state is found only in the **non-canonical monomer** of the BthTX-I/PEG400 (PDB id: 2H8I) showed as a pink cartoon (A and B). Finally, a third state is represented by **canonical monomer** of the BthTX-I/PEG4k (PDB id: 2H8I) showed as orange cartoon (B). (A) The transition from first to second state is characterized by a reduction of 10.3 Å to 8.5 Å from L121-F125 Cβ distance resulted from the structural change of L121 caused by hydrophobic molecule entrance in the **canonical monomer** channel. (B) The transition from second to third state is characterized by change of **non-canonical monomer** to **canonical** because Y119 and H120 residues invert their direction (180° rotation), leading to the approximation of the L121 and F125 residues (4.7 Å) and to the formation of a short 3_{10} helix. Side chains of Y119, H120, L121 and F125 residues are shown in sticks. Meaningful conformation changes are pointed in arrows.

PLA_2-like structures in the **inactive state** only exhibit **tunnel 1** for one monomer (**canonical monomer**). The comparison between **tunnels 1** of the structures in the **active** and **inactive states** shows that their cores are identical; however, their entrances are different due to the different dimeric orientations and differences in the C-termini (refer to Section 3.2.1). By comparing the hydrophobic channels of **inactive state** structures, differences in the side chain positions of the H120, K122 and F125 residues can be observed. Such variability is probably related to their flexibility, which is essential for the conformational change from the **non-canonical** to **canonical monomer**, in which Y119 and H120 invert their positions with a 180° rotation (Fig. 7). Curiously, the **hydrophobic tunnels** of BthTX-I/PEG400 and *apo* BthTX-I (3HZD) are closed by the position of the H120 side chain, whereas they are open in the available models of the other bothropic PLA_2-like proteins.

4.4. New steps in the myotoxic mechanism of PLA_2-like toxins

The results presented here identified interesting new structural features related to the oligomeric changes between the **inactive** and **active states** of PLA_2-like proteins. From the detailed structural analysis of the MDiS region, it was possible to identify totally different MDiS conformations in the **inactive** and **active states**, which may be measured by the distance between the L121 and F125 residues (Figs. 5 and 7). Furthermore, from the comprehensive structural bioinformatics study of the protein tunnels, we could identified the relevant tunnels for the PLA_2-like proteins and their relationships to the transition between both states, which is represented by the entrance of hydrophobic molecules in **tunnel 1**of **canonical monomer**, leading to dimer reorientation and a structural change from asymmetrical to symmetrical monomers.

Thus, based on the results presented here, we suggest the following myotoxic mechanism for PLA_2-like proteins, in which item ii was added and additional details were added to items i, iii and iv of the previously proposed myotoxic mechanism [15]:

i) The activation starts with the entrance of **one** hydrophobic molecule/fatty acid through **tunnel 1** of the **canonical monomer**.
ii) Approximation of L121 to F125 residues from the **non-canonical monomer**.
iii) The dimer is reoriented, and the **non-canonical monomer** adopts its canonical conformation by inverting the Y119 and H120 positions and further approximating the L121 and F125 positions.
iv) The inaccessible and vacant hydrophobic channel is opened by dimer reorientation and a hydrophobic molecule/fatty acid is bound, stabilizing the toxin **active state**.
v) Protein-membrane docking through the MDoS.
vi) Membrane destabilization by its interaction with the MDiS.
vii) Uncontrolled influx of ions and, consequently, cell death.

Steps i to iv of the proposed myotoxic mechanism are represented by two morphing videos based on the different identified states of the BthTX-I crystallographic structures (Supplementary Videos 1 and 2).

4.5. Concluding remarks

The structural study presented here and its comparison to other PLA_2-like structures added important elements to the comprehension of the allosteric transition that PLA_2-like proteins undergo to exert their function. The transition between the **inactive state**, which is present as asymmetrical monomers, to the **active states** with symmetrical monomers was observed, and a key tunnel **(Tunnel 1)** was identified for the first time, in which all fundamental residues for the myotoxicity of PLA_2-like proteins are located (MDoS, MDiS and H48). Furthermore, we showed that zinc ions inhibited BthTX-I myotoxicity *in vivo* by interacting to H48 of the accessible hydrophobic channel (**canonical monomer**) and to both H120 residues, as determined in the crystallographic studies. Based on these observations, we can hypothesize that the Zn ions simultaneously inhibit BthTX-I through two different mechanisms: i) preventing fatty acid binding to H48 and thus avoiding the state transition from the **inactive** to **active state**, and ii) by binding directly to the MDiS region, preventing the membrane disruption process. Thus, zinc ions could be a new candidate to aid *anti*-venom treatment against the myonecrotic effects induced by the most abundant toxin in some bothropic snake venoms. Finally, all of the information presented here, regarding the expression of myonecrosis by PLA_2-like proteins and its inhibition, can promote the design of new and even more accurate structure-based inhibitors that can completely inhibit these proteins and provide new insights to complement the conventional serum therapy.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2016.08.003.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgments

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - 2015/17286-0 and 2013/00873-4), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - 300596/2013-8) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - 1592/2011) Brazil). The authors acknowledge Gildo Bernardo Leite for his technical assistance in mouse PND experiments and the Laboratório Nacional de Luz Síncrotron (LNLS, Brazil) for the use and support of the MX beamlines.

References

[1] A. Kasturiratne, A.R. Wickremasinghe, N. de Silva, N.K. Gunawardena, A. Pathmeswaran, R. Premaratna, L. Savioli, D.G. Lalloo, H.J. de Silva, The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths, PLoS Med. 5 (2008) e218, http://dx.doi.org/10.1371/journal.pmed.0050218.
[2] J.M. Gutiérrez, B. Lomonte, G. León, A. Alape-Girón, M. Flores-Díaz, L. Sanz, Y. Angulo, J.J. Calvete, Snake venomics and antivenomics: proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming, J. Proteome 72 (2009) 165–182, http://dx.doi.org/10.1016/j.jprot.2009.01.008.
[3] D.A. Warrell, Snake bite, Lancet 375 (2010) 77–88, http://dx.doi.org/10.1016/S0140-6736(09)61754-2.
[4] J.M. Gutiérrez, R.D.G. Theakston, D.A. Warrell, Confronting the neglected problem of snake bite envenoming: the need for a global partnership, PLoS Med. 3 (2006)http://dx.doi.org/10.1371/journal.pmed.0030150.
[5] World Health Organization, Rabies and Envenomings: A Neglected Public Health Issue, W.H.O., Geneva, 2007 (http://apps.who.int//iris/handle/10665/43858).
[6] J.J. Calvete, P. Juárez, L. Sanz, Snake venomics. Strategy and applications, J. Mass Spectrom. 42 (2007) 1405–1414, http://dx.doi.org/10.1002/jms.1242.
[7] J.J. Calvete, L. Sanz, P. Cid, P. de la Torre, M. Flores-Díaz, M.C. Dos Santos, A. Borges, A. Bremo, Y. Angulo, B. Lomonte, A. Alape-Girón, J.M. Gutiérrez, Snake venomics of the Central American rattlesnake *Crotalus simus* and the South American *Crotalus durissus* complex points to neurotoxicity as an adaptive paedomorphic trend along Crotalus dispersal in South America, J. Proteome Res. 9 (2010) 528–544, http://dx.doi.org/10.1021/pr9008749.
[8] R.H. Schaloske, E.A. Dennis, The phospholipase A2 superfamily and its group numbering system, Biochim. Biophys. Acta 1761 (2006) 1246–1259, http://dx.doi.org/10.1016/j.bbalip.2006.07.011.
[9] B. Lomonte, J.M. Gutiérrez, Phospholipases A2 from Viperidae snake venoms: how do they induce skeletal muscle damage? Acta Chim. Slov. 58 (2011) 647–658.
[10] C.A.H. Fernandes, R.J. Borges, B. Lomonte, M.R.M. Fontes, A structure-based proposal for a comprehensive myotoxic mechanism of phospholipase A2-like proteins from viperid snake venoms, Biochim. Biophys. Acta 1844 (12) (2014) 2265–2276.
[11] A.H. de Oliveira, J.R. Giglio, S.H. Andrião-Escarso, A.S. Ito, R.J. Ward, A pH-induced dissociation of the dimeric form of a lysine 49-phospholipase A2 abolishes Ca2+-independent membrane damaging activity, Biochemistry (Mosc) 40 (2001) 6912–6920.
[12] A.H.C. de Oliveira, T.L. Ferreira, R.J. Ward, Reduced pH induces an inactive non-native conformation of the monomeric bothropstoxin-I (Lys49-PLA2), Toxicon 54 (2009) 373–378, http://dx.doi.org/10.1016/j.toxicon.2009.04.022.
[13] L. Chioato, E.A. Aragão, T. Lopes Ferreira, A.I. de Medeiros, L.H. Faccioli, R.J. Ward, Mapping of the structural determinants of artificial and biological membrane damaging activities of a Lys49 phospholipase A2 by scanning alanine mutagenesis, Biochim. Biophys. Acta 1768 (2007) 1247–1257, http://dx.doi.org/10.1016/j.bbamem.2007.01.023.
[14] B. Lomonte, Y. Angulo, L. Calderón, An overview of lysine-49 phospholipase A2 myotoxins from crotalid snake venoms and their structural determinants of myotoxic action, Toxicon 42 (2003) 885–901, http://dx.doi.org/10.1016/j.toxicon.2003.11.008.
[15] C.A.H. Fernandes, E.J. Comparetti, R.J. Borges, S. Huancahuire-Vega, L.A. Ponce-Soto, S. Marangoni, A.M. Soares, M.R.M. Fontes, Structural bases for a complete myotoxic mechanism: crystal structures of two non-catalytic phospholipases A2-like from *Bothrops brazili* venom, Biochim. Biophys. Acta 1834 (2013) 2772–2781, http://dx.doi.org/10.1016/j.bbapap.2013.10.009.
[16] J.I. dos Santos, A.M. Soares, M.R.M. Fontes, Comparative structural studies on Lys49-phospholipases A(2) from Bothrops genus reveal their myotoxic site, J. Struct. Biol. 167 (2009) 106–116, http://dx.doi.org/10.1016/j.jsb.2009.04.003.
[17] C. Montecucco, J.M. Gutiérrez, B. Lomonte, Cellular pathology induced by snake venom phospholipase A2 myotoxins and neurotoxins: common aspects of their mechanisms of action, Cell. Mol. Life Sci. 65 (2008) 2897–2912, http://dx.doi.org/10.1007/s00018-008-8113-3.
[18] E.E. Sánchez, A. Rodríguez-Acosta, Inhibitors of snake venoms and development of new therapeutics, Immunopharmacol. Immunotoxicol. 30 (2008) 647–678, http://dx.doi.org/10.1080/08923970802279019.
[19] A.M. Soares, F.K. Ticli, S. Marcussi, M.V. Lourenço, A.H. Januário, S.V. Sampaio, J.R. Giglio, B. Lomonte, P.S. Pereira, Medicinal plants with inhibitory properties against snake venoms, Curr. Med. Chem. 12 (2005) 2625–2641.
[20] J.I. Dos Santos, F.F. Cardoso, A.M. Soares, M. Dal Pai Silva, M. Gallacci, M.R.M. Fontes, Structural and functional studies of a bothropic myotoxin complexed to rosmarinic acid: new insights into Lys49-PLA_2 inhibition, PLoS One 6 (2011) e28521, http://dx.doi.org/10.1371/journal.pone.0028521.
[21] C.A.H. Fernandes, F.F. Cardoso, W.G.L. Cavalcante, A.M. Soares, M. Dal-Pai, M. Gallacci, M.R.M. Fontes, Structural basis for the inhibition of a phospholipase A2-like toxin by caffeic and aristolochic acids, PLoS One 10 (2015) e0133370, http://dx.doi.org/10.1371/journal.pone.0133370.
[22] G.H.M. Salvador, T.R. Dreyer, W.L.G. Cavalcante, F.F. Matioli, J.I. Dos Santos, A. Velazquez-Campoy, M. Gallacci, M.R.M. Fontes, Structural and functional evidence for membrane docking and disruption sites on phospholipase A2-like proteins revealed by complexation with the inhibitor suramin, Acta Crystallogr. D Biol. Crystallogr. 71 (2015) 2066–2078, http://dx.doi.org/10.1107/S1399004715014443.
[23] L. Rodrigues-Simioni, J. Prado-Franceschi, A.C. Cintra, J.R. Giglio, M.S. Jiang, J.E. Fletcher, No role for enzymatic activity or dantrolene-sensitive Ca2+ stores in the muscular effects of bothropstoxin, a Lys49 phospholipase A2 myotoxin, Toxicon 33 (1995) 1479–1489.
[24] A.M. Soares, Y. Oshima-Franco, C.A. Vieira, G.B. Leite, J.E. Fletcher, M.-S. Jiang, A.C.O. Cintra, J.R. Giglio, L. Rodrigues-Simioni, Mn(2+) ions reduce the enzymatic and pharmacological activities of bothropstoxin-I, a myotoxic Lys49 phospholipase A(2) homologue from *Bothrops jararacussu* snake venom, Int. J. Biochem. Cell Biol. 34 (2002) 668–677.
[25] N.F. Heluany, M.I. Homsi-Brandeburgo, J.R. Giglio, J. Prado-Franceschi, L. Rodrigues-Simioni, Effects induced by bothropstoxin, a component from *Bothrops jararacussu* snake venom, on mouse and chick muscle preparations, Toxicon 30 (1992) 1203–1210.
[26] A.S. Prasad, Discovery of human zinc deficiency: 50 years later, J. Trace Elem. Med. Biol. 26 (2012) 66–69, http://dx.doi.org/10.1016/j.jtemb.2012.04.004.
[27] A.S. Prasad, F.W.J. Beck, D.C. Snell, O. Kucuk, Zinc in cancer prevention, Nutr. Cancer 61 (2009) 879–887, http://dx.doi.org/10.1080/01635580903285122.
[28] P.J. Little, R. Bhattacharya, A.E. Moreyra, I.L. Korichneva, Zinc and cardiovascular disease, Nutrition 26 (2010) 1050–1057, http://dx.doi.org/10.1016/j.nut.2010.03.007.
[29] D.K. Dhawan, V.D. Chadha, Zinc: a promising agent in dietary chemoprevention of cancer, Indian J. Med. Res. 132 (2010) 676–682.
[30] B.J. Grattan, H.C. Freake, Zinc and cancer: implications for LIV-1 in breast cancer, Nutrients 4 (2012) 648–675, http://dx.doi.org/10.3390/nu4070648.
[31] M.I. Homsi-Brandeburgo, L.S. Queiroz, H. Santo-Neto, L. Rodrigues-Simioni, J.R. Giglio, Fractionation of *Bothrops jararacussu* snake venom: partial chemical characterization and biological activity of bothropstoxin, Toxicon 26 (1988) 615–627.
[32] R.S. Floriano, V.C. Carregari, V.A. de Abreu, B. Kenzo-Kagawa, L.A. Ponce-Soto, M.A. da Cruz-Höfling, S. Hyslop, S. Marangoni, L. Rodrigues-Simioni, Pharmacological study of a new Asp49 phospholipase A(2) (Bbil-TX) isolated from *Bothriopsis bilineata* smargadina (forest viper) venom in vertebrate neuromuscular preparations, Toxicon 69 (2013) 191–199, http://dx.doi.org/10.1016/j.toxicon.2013.03.006.
[33] Y. Oshima-Franco, G.B. Leite, C.A.D. Belo, S. Hyslop, J. Prado-Franceschi, A.C.O. Cintra, J.R. Giglio, M.A. da Cruz-Höfling, L. Rodrigues-Simioni, The presynaptic activity of bothropstoxin-I, a myotoxin from *Bothrops jararacussu* snake venom, Basic Clin. Pharmacol. Toxicol. 95 (2004) 175–182, http://dx.doi.org/10.1111/j.1742-7843.2004.pto_950405.x.
[34] P. Rey-Suárez, R.S. Floriano, S. Rostelato-Ferreira, M. Saldarriaga-Córdoba, V. Núñez, L. Rodrigues-Simioni, B. Lomonte, Mipartoxin-I, a novel three-finger toxin, is the major neurotoxic component in the venom of the redtail coral snake *Micrurus mipartitus* (Elapidae), Toxicon 60 (2012) 851–863, http://dx.doi.org/10.1016/j.toxicon.2012.05.023.
[35] A. McPherson, Introduction to Macromolecular Crystallography, 2nd ed. Wiley-Blackwell, Hoboken, N.J, 2009.
[36] W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132, http://dx.doi.org/10.1107/S0907444909047337.
[37] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674, http://dx.doi.org/10.1107/S0021889807021206.
[38] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221, http://dx.doi.org/10.1107/S0907444909052925.
[39] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501, http://dx.doi.org/10.1107/S0907444910007493.
[40] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall 3rd, J. Snoeyink, J.S. Richardson, D.C. Richardson, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, Nucleic Acids Res. 35 (2007) W375–W383, http://dx.doi.org/10.1093/nar/gkm216.
[41] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797, http://dx.doi.org/10.1016/j.jmb.2007.05.022.
[42] D. Sehnal, R. Svobodová Vařeková, K. Berka, L. Pravda, V. Navrátilová, P. Banáš, C.-M. Ionescu, M. Otyepka, J. Koča, MOLE 2.0: advanced approach for analysis of biomacromolecular channels, J. Cheminform. 5 (2013) 39, http://dx.doi.org/10.1186/1758-2946-5-39.

[43] C.A.H. Fernandes, D.P. Marchi-Salvador, G.M. Salvador, M.C.O. Silva, T.R. Costa, A.M. Soares, M.R.M. Fontes, Comparison between apo and complexed structures of bothropstoxin-I reveals the role of Lys122 and Ca(2+)-binding loop region for the catalytically inactive Lys49-PLA(2)s, J. Struct. Biol. 171 (2010) 31–43, http://dx.doi.org/10.1016/j.jsb.2010.03.019.
[44] M.T. Murakami, M.M. Viçoti, J.R.B. Abrego, M.R. Lourenzoni, A.C.O. Cintra, E.Z. Arruda, M.A. Tomaz, P.A. Melo, R.K. Arni, Interfacial surface charge and free accessibility to the PLA2-active site-like region are essential requirements for the activity of Lys49 PLA2 homologues, Toxicon 49 (2007) 378–387, http://dx.doi.org/10.1016/j.toxicon.2006.10.011.
[45] D.P. Marchi-Salvador, C.A.H. Fernandes, L.B. Silveira, A.M. Soares, M.R.M. Fontes, Crystal structure of a phospholipase A(2) homolog complexed with p-bromophenacyl bromide reveals important structural changes associated with the inhibition of myotoxic activity, Biochim. Biophys. Acta 1794 (2009) 1583–1590, http://dx.doi.org/10.1016/j.bbapap.2009.07.005.
[46] W.H. Lee, M.T. da Silva Giotto, S. Marangoni, M.H. Toyama, I. Polikarpov, R.C. Garratt, Structural basis for low catalytic activity in Lys49 phospholipases A2—a hypothesis: the crystal structure of piratoxin II complexed to fatty acid, Biochemistry (Mosc) 40 (2001) 28–36.
[47] A.J. Magro, A.M. Soares, J.R. Giglio, M.R.M. Fontes, Crystal structures of BnSP-7 and BnSP-6, two Lys49-phospholipases A(2): quaternary structure and inhibition mechanism insights, Biochem. Biophys. Res. Commun. 311 (2003) 713–720.
[48] R.K. Arni, R.J. Ward, J.M. Gutierrez, A. Tulinsky, Structure of a calcium-independent phospholipase-like myotoxic protein from *Bothrops asper* venom, Acta Crystallogr. D Biol. Crystallogr. 51 (1995) 311–317, http://dx.doi.org/10.1107/S0907444994011455.
[49] M.T. Murakami, E.Z. Arruda, P.A. Melo, A.B. Martinez, S. Calil-Eliás, M.A. Tomaz, B. Lomonte, J.M. Gutiérrez, R.K. Arni, Inhibition of myotoxic activity of *Bothrops asper* myotoxin II by the *anti*-trypanosomal drug suramin, J. Mol. Biol. 350 (2005) 416–426, http://dx.doi.org/10.1016/j.jmb.2005.04.072.
[50] P. Delatorre, B.A.M. Rocha, T. Santi-Gadelha, C.A.A. Gadelha, M.H. Toyama, B.S. Cavada, Crystal structure of Bn IV in complex with myristic acid: a Lys49 myotoxic phospholipase A_2 from *Bothrops neuwiedi* venom, Biochimie 93 (2011) 513–518, http://dx.doi.org/10.1016/j.biochi.2010.11.003.
[51] L. Watanabe, A.M. Soares, R.J. Ward, M.R.M. Fontes, R.K. Arni, Structural insights for fatty acid binding in a Lys49-phospholipase A2: crystal structure of myotoxin II from *Bothrops moojeni* complexed with stearic acid, Biochimie 87 (2005) 161–167, http://dx.doi.org/10.1016/j.biochi.2004.11.005.
[52] G.H.M. Salvador, W.L.G. Cavalcante, J.I. Dos Santos, M. Gallacci, A.M. Soares, M.R.M. Fontes, Structural and functional studies with mytoxin II from *Bothrops moojeni* reveal remarkable similarities and differences compared to other catalytically inactive phospholipases A_2-like, Toxicon 72 (2013) 52–63, http://dx.doi.org/10.1016/j.toxicon.2013.06.013.
[53] A.L.B. Ambrosio, M.C. Nonato, H.S.S. de Araújo, R. Arni, R.J. Ward, C.L. Ownby, D.H.F. de Souza, R.C. Garratt, A molecular mechanism for Lys49-phospholipase A2 activity based on ligand-induced conformational change, J. Biol. Chem. 280 (2005) 7326–7335, http://dx.doi.org/10.1074/jbc.M410588200.
[54] H. Zheng, M.D. Chordia, D.R. Cooper, M. Chruszcz, P. Müller, G.M. Sheldrick, W. Minor, Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server, Nat. Protoc. 9 (2014) 156–170, http://dx.doi.org/10.1038/nprot.2013.172.
[55] T. Biswas, O.V. Tsodikov, An easy-to-use for planning and modeling a calorimetric titration, Anal. Biochem. 406 (2010) 91–93, http://dx.doi.org/10.1016/j.ab.2010.06.050.
[56] T. Crépin, A. Dias, A. Palencia, C. Swale, S. Cusack, R.W.H. Ruigrok, Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit, J. Virol. 84 (2010) 9096–9104, http://dx.doi.org/10.1128/JVI.00995-10.
[57] J.M. Berg, Y. Shi, The galvanization of biology: a growing appreciation for the roles of zinc, Science 271 (1996) 1081–1085.
[58] B.L. Vallee, K.H. Falchuk, The biochemical basis of zinc physiology, Physiol. Rev. 73 (1993) 79–118.
[59] R.A. Bozym, F. Chimienti, L.J. Giblin, G.W. Gross, I. Korichneva, Y. Li, S. Libert, W. Maret, M. Parviz, C.J. Frederickson, R.B. Thompson, Free zinc ions outside a narrow concentration range are toxic to a variety of cells in vitro, Exp. Biol. Med. (Maywood) 235 (2010) 741–750, http://dx.doi.org/10.1258/ebm.2010.009258.
[60] S.Y. Lin-Shiau, W.M. Fu, Effects of divalent cations on neuromuscular transmission in the chick, Eur. J. Pharmacol. 64 (1980) 259–269.
[61] F.R. Ciofalo, L.J. Thomas, The effects of zinc on contractility, membrane potentials, and cation content of rat atria, J. Gen. Physiol. 48 (1965) 825–839.
[62] A. Gölgeli, N. Dursun, C. Süer, C. Ozesmi, S. Aydogan, Zinc-induced alterations in contractile properties of rat diaphragm muscle in vitro, Biol. Trace Elem. Res. 60 (1997) 251–260, http://dx.doi.org/10.1007/BF02784445.
[63] E.K. Silbergeld, J.T. Fales, A.M. Goldberg, The effects of inorganic lead on the neuromuscular junction, Neuropharmacology 13 (1974) 795–801.
[64] A.M. Soares, J.R. Giglio, Chemical modifications of phospholipases A2 from snake venoms: effects on catalytic and pharmacological properties, Toxicon 42 (2003) 855–868, http://dx.doi.org/10.1016/j.toxicon.2003.11.004.
[65] F.K. Ticli, L.I.S. Hage, R.S. Cambraia, P.S. Pereira, A.J. Magro, M.R.M. Fontes, R.G. Stábeli, J.R. Giglio, S.C. França, A.M. Soares, S.V. Sampaio, Rosmarinic acid, a new snake venom phospholipase A2 inhibitor from Cordia verbenacea (Boraginaceae): antiserum action potentiation and molecular interaction, Toxicon 46 (2005) 318–327, http://dx.doi.org/10.1016/j.toxicon.2005.04.023.
[66] M. de Oliveira, W.L.G. Cavalcante, E.Z. Arruda, P.A. Melo, M. Dal-Pai Silva, M. Gallacci, Antagonism of myotoxic and paralyzing activities of bothropstoxin-I by suramin, Toxicon 42 (2003) 373–379.
[67] P.J. Yunes Quartino, J.L. Barra, G.D. Fidelio, Cloning and functional expression of secreted phospholipases A(2) from Bothrops diporus (Yarará Chica), Biochem. Biophys. Res. Commun. 427 (2012) 321–325, http://dx.doi.org/10.1016/j.bbrc.2012.09.051.
[68] S. Lizano, G. Lambeau, M. Lazdunski, Cloning and cDNA sequence analysis of Lys(49) and Asp(49) basic phospholipase A(2) myotoxin isoforms from *Bothrops asper*, Int. J. Biochem. Cell Biol. 33 (2001) 127–132.
[69] V. Núñez, V. Arce, J.M. Gutiérrez, B. Lomonte, Structural and functional characterization of myotoxin I, a Lys49 phospholipase A2 homologue from the venom of the snake *Bothrops atrox*, Toxicon 44 (2004) 91–101, http://dx.doi.org/10.1016/j.toxicon.2004.04.013.
[70] L.A. Ponce-Soto, B. Lomonte, J.M. Gutiérrez, L. Rodrigues-Simioni, J.C. Novello, S. Marangoni, Structural and functional properties of BaTX, a new Lys49 phospholipase A2 homologue isolated from the venom of the snake *Bothrops alternatus*, Biochim. Biophys. Acta 1770 (2007) 585–593, http://dx.doi.org/10.1016/j.bbagen.2006.11.015.
[71] G.H.M. Salvador, C.A.H. Fernandes, A.J. Magro, D.P. Marchi-Salvador, W.L.G. Cavalcante, R.M. Fernandez, M. Gallacci, A.M. Soares, C.L.P. Oliveira, M.R.M. Fontes, Structural and phylogenetic studies with MjTX-I reveal a multi-oligomeric toxin—a novel feature in Lys49-PLA2s protein class, PLoS One 8 (2013) e60610, http://dx.doi.org/10.1371/journal.pone.0060610.
[72] P.J. YunesQuartino, J.L. Barra, G.D. Fidelio, Cloning and functional expression of secreted phospholipases A(2) from Bothrops diporus (Yarará Chica), Biochem. Biophys. Res. Commun. 427 (2012) 321–325, http://dx.doi.org/10.1016/j.bbrc.2012.09.051.
[73] C. Díaz, A. Alape, B. Lomonte, T. Olamendi, J.M. Gutiérrez, Cleavage of the NH2-terminal octapeptide of *Bothrops asper* myotoxic lysine-49 phospholipase A2 reduces its membrane-destabilizing effect, Arch. Biochem. Biophys. 312 (1994) 336–339, http://dx.doi.org/10.1006/abbi.1994.1317.
[74] A.M. Soares, R. Guerra-Sá, C.R. Borja-Oliveira, V.M. Rodrigues, L. Rodrigues-Simioni, V. Rodrigues, M.R. Fontes, B. Lomonte, J.M. Gutiérrez, J.R. Giglio, Structural and functional characterization of BnSP-7, a Lys49 myotoxic phospholipase A(2) homologue from *Bothrops neuwiedi pauloensis* venom, Arch. Biochem. Biophys. 378 (2000) 201–209, http://dx.doi.org/10.1006/abbi.2000.1790.

BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Received 24 April 2015
Accepted 4 June 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: phasing; *ARCIMBOLDO_BORGES*; *ab initio*; small fragments.

Supporting information: this article has supporting information at journals.iucr.org/d

ARCIMBOLDO_LITE: single-workstation implementation and use

Massimo Sammito,[a] Claudia Millán,[a] Dawid Frieske,[a] Eloy Rodríguez-Freire,[a] Rafael J. Borges[a] and Isabel Usón[b]*

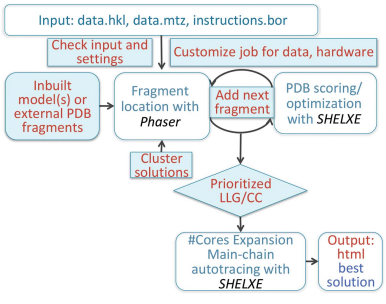
[a]Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB–CSIC), Baldiri Reixac 15, 08028 Barcelona, Spain, and [b]Structural Biology, ICREA at IBMB–CSIC, Baldiri Reixach 13-15, 08028 Barcelona, Spain. *Correspondence e-mail: uson@ibmb.csic.es

ARCIMBOLDO solves the phase problem at resolutions of around 2 Å or better through massive combination of small fragments and density modification. For complex structures, this imposes a need for a powerful grid where calculations can be distributed, but for structures with up to 200 amino acids in the asymmetric unit a single workstation may suffice. The use and performance of the single-workstation implementation, *ARCIMBOLDO_LITE*, on a pool of test structures with 40–120 amino acids and resolutions between 0.54 and 2.2 Å is described. Inbuilt polyalanine helices and iron cofactors are used as search fragments. *ARCIMBOLDO_BORGES* can also run on a single workstation to solve structures in this test set using precomputed libraries of local folds. The results of this study have been incorporated into an automated, resolution- and hardware-dependent parameterization. *ARCIMBOLDO* has been thoroughly rewritten and three binaries are now available: *ARCIMBOLDO_LITE*, *ARCIMBOLDO_SHREDDER* and *ARCIMBOLDO_BORGES*. The programs and libraries can be downloaded from http://chango.ibmb.csic.es/ARCIMBOLDO_LITE.

1. Introduction

The phase problem is the central problem of crystallography and is a bottleneck in the determination of macromolecular crystal structures (Hendrickson, 2013). Several methods exist to provide starting phases for structure factors to approximate the correct phases, which are missed in the diffraction experiment. In the case of macromolecules, experimental phasing through heavy-atom derivatives or anomalous scattering at particular wavelengths (Hendrickson, 1991) is used if no previous structural knowledge from a related structure is available for phasing by molecular replacement (Rossman, 1972; Navaza, 1994).

Ab initio phasing from the native diffraction intensities alone is accomplished by direct methods in the case of molecules with less than 200 atoms that diffract to atomic resolution (Karle & Hauptman, 1956). Dual-space recycling has succeeded in extending the use of direct methods to small proteins of up to 1000 atoms, maintaining the atomicity restriction (Miller *et al.*, 1993; Sheldrick *et al.*, 2012). Approaches to relax this restriction have involved sophisticated use of the Patterson function (Caliandro *et al.*, 2008), data extrapolation to extend the resolution limit beyond the experimentally measured data (Caliandro *et al.*, 2005; Usón *et al.*, 2007) and density-modification algorithms tailored to the high-resolution case, such as low-density elimination (Shiono & Woolfson, 1992; Refaat & Woolfson, 1993), the sphere of influence (Sheldrick, 2002) and *VLD* (Burla *et al.*, 2010, 2011,



2012). Starting phasing from a small but highly accurate substructure in the context of *ACORN* has been reported to be remarkably effective (Yao *et al.*, 2005, 2006) and our own tests have corroborated this finding: as little as 10% of the main-chain atoms suffice to solve a structure at 2 Å resolution through density modification. This provides a way to escape the need for atomic resolution that restricts direct methods and real-space atom selection. The atomicity constraints can be substituted by enforcing secondary- or tertiary-structure stereochemistry. A related proof of principle has been established using small fragments such as α-helices (Glykos & Kokkinidis, 2003) and nucleotides (Robertson & Scott, 2008; Robertson *et al.*, 2010) as search fragments to seed phasing. This is accomplished in *ARCIMBOLDO* (Rodríguez *et al.*, 2009, 2012) by combining a search for small polyalanine model fragments with *Phaser* (McCoy *et al.*, 2007) with density modification and autotracing with *SHELXE* (Thorn & Sheldrick, 2013). Extremely successful approaches based on more complete models of lower accuracy (Rigden *et al.*, 2008) have been developed and include the enhancement of poor search models and *de novo* prediction through modelling with *Rosetta* (Qian *et al.*, 2007) or *QUARK* (Xu & Zhang, 2012) linked to molecular replacement with *Phaser*. This design underlies methods such as *MR-Rosetta* (DiMaio *et al.*, 2011) and *AMPLE* (Bibby *et al.*, 2012, 2013; Keegan *et al.*, 2015), and other implementations (Shrestha *et al.*, 2011; Shrestha & Zhang, 2015).

The original *ARCIMBOLDO* procedure relies on the assumption that the main chain of small secondary-structural elements, such as α-helices, is conserved among unrelated protein structures. Because of the generality of such partial models, which are predictable from the sequence (Cole *et al.*, 2008) or data (Caliandro *et al.*, 2012; Morris *et al.*, 2004), this method is in fact an *ab initio* approach. Recently, it has been shown how unrelated structures can share sufficiently similar local folds comprising the main chain of a few secondary-structure elements. Libraries of such folds can be extracted with *BORGES* (Sammito *et al.*, 2013) from the entire PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000) and then input into *ARCIMBOLDO_BORGES* as search models. If a the structure of a distant homologue is known, fragments from this template can also be extracted and evaluated against the rotation function, as implemented in *ARCIMBOLDO_SHREDDER* (Sammito *et al.*, 2014).

As small models account for a low fraction of the total scattering, the figures of merit characterizing a correct solution may fail to identify it as such, making it indistinguishable from a large pool of incorrect solutions. This has required the implementation of *ARCIMBOLDO* as a multisolution method, in which all possible solutions have to be explored independently up to the step of density modification and autotracing, where a correct density map is unequivocally recognizable at 2 Å resolution. Supercomputing facilities are required to manage, distribute and check this magnitude of computation, and databases to aid in storing and retrieving the amount of data generated. Improvement in the identification of correct solutions allows the pool of solutions to be narrowed in some cases (Oeffner *et al.*, 2013). The accumulated experience in using this method and advances in the software involved (*Phaser* and *SHELXE*) have made it possible to deploy a standalone version which will just require multiprocessing facilities and particular filters to reduce computation depending on the characteristics of the available hardware.

In the present work, we have developed and assessed single-machine implementations of the *ARCIMBOLDO* methods described by Millán *et al.* (2015). *ARCIMBOLDO_LITE* performs *ab initio* phasing using polyalanine helices or other single search fragments. *ARCIMBOLDO_BORGES* performs *ab initio* phasing with nonspecific libraries of small folds. For the sake of completeness, although not discussed in this paper, *ARCIMBOLDO_SHREDDER* uses fragments derived from a distant homologue template. The new software has been thoroughly rewritten and dimensioned to run on a single, multiprocessor machine, thus eliminating the limiting condition of access to a supercomputer or grid cluster. No database is required either. The binary standalone version is distributed for Linux and Mac OS at http://chango.ibmb.csic.es/ARCIMBOLDO_LITE. The performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a pool of test structures and systems is described and the conclusions of these tests have been used to set up data-dependent default parameterization.

Table 1
Performance of *ARCIMBOLDO_LITE* on a pool of 12 workstations for two test structures (PDB entries 2iu1 and 4k82).

Hardware	OS	RT, 2iu1	RT, 4k82
Mean for pool of 12 workstations		7 h 37 min	2 h 1 min
Xeon X5560 at 2.80 GHz, 8 cores sharing 24 GB	Debian 6.0	4 h 26 min	1 h 18 min
i7 2600 at 3.40 GHz, 4 cores sharing 12 GB	Debian 6.0	5 h 34 min	1 h 36 min
i7 3930K at 3.20 GHz, 6 cores sharing 16 GB	Debian 6.0	4 h 10 min	1 h 21 min
Xeon E5410 at 2.33 GHz, 8 cores sharing 32 GB	Debian 6.0	7 h 46 min	1 h 34 min
2 Duo E4500 at 2.20 GHz, 2 cores sharing 512 MB	Debian 6.0	32 h 19 min	8 h 45 min
i7 950 at 3.07 GHz, 4 cores sharing 6 GB	Ubuntu 10.04.2	6 h 41 min	2 h 3 min
Xeon E5440 at 2.83 GHz, 8 cores sharing 16 GB	Ubuntu 10.04.3	6 h 58 min	1 h 56 min
i7 X980 at 3.33 GHz, 6 cores sharing 12 GB	Ubuntu 10.04.4	5 h 1 min	1 h 21 min
Xeon E5-2650 v2 at 2.66–4.00 GHz, 16 cores sharing 125 GB	Centos 7.1.1503	1 h 55 min	32 min
Xeon E5-2680 v3 at 2.50 GHz, 24 cores sharing 16 GB	openSUSE 13.2	4 h 34 min	1 h
I7-3720QM at 2.60 GHz, 8 cores sharing 8 GB	Mac OSX Yosemite 10.10	3 h 16 min	1 h 10 min
I7-3720QM at 2.80 GHz, 8 cores sharing 4 GB	Mac OSX Mavericks 10.9	8 h 54 min	1 h 40 min

2. Materials and methods

2.1. Computing setup

Structure solution was run on the eight identical eight-core machines of an HP ProLiant BL460c blade system, using them as single workstations with dual quad-core Xeon E5440

processors (2.83 GHz and 16 GB RAM). The Linux distribution installed was Ubuntu Server 10.04 LTS.

2.2. Software versions

The *ARCIMBOLDO* binary is deployed for Linux and Macintosh (some examples of compatible distributions and versions can be found in Table 1). It was generated with *PyInstaller* 2.1 and Python 2.6, with version 2014 upwards of *SHELXE* and version 2.5.6 upwards of *Phaser* from the *PHENIX* (Adams *et al.*, 2010) or *CCP*4 (Winn *et al.*, 2011) distributions. Model and maps were examined with *Coot* (Emsley *et al.*, 2010). Figures were prepared with *PyMOL* (v.1.5.0.4; Schrödinger) and *gnuplot* 4.4. *F*-weighted mean phase errors were calculated with *SHELXE* (Sheldrick, 2002) against the final models deposited in the PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000).

2.3. Test set used

In this study, the structure of EIF5 (Bieniossek *et al.*, 2006; PDB entry 2iu1) at 1.7 Å resolution and the structure of Lv-ranaspumin (Hissa *et al.*, 2014; PDB entry 4k82) at 1.7 Å resolution previously solved with *ARCIMBOLDO* were used to benchmark performance on different hardware setups. Furthermore, we used a pool of 294 structures with resolutions between 2.2 and 0.54 Å, sizes of between 44 and 120 residues in the asymmetric unit and 47 different space groups, approximately following the frequency distribution seen in the PDB. In particular, the most frequent 36 space groups are represented, with $P2_12_12_1$ predominating. The PDB codes for these 294 structures are as follows: 1ejg, 1en2, 1ew4, 1ezj, 1f94, 1fk5, 1g2r, 1g6u, 1gk6, 1gmx, 1gvd, 1gxu, 1i2t, 1i71, 1iqz, 1j2l, 1j8b, 1j8e, 1kth, 1kwi, 1l9l, 1lsl, 1lxj, 1m1q, 1mg4, 1mk0, 1ne8, 1njh, 1nnx, 1oap, 1oks, 1ox3, 1p9g, 1pz4, 1q8d, 1r6j, 1r7j, 1riy, 1rlk, 1rw1, 1rwj, 1sbx, 1t07, 1tg0, 1tgr, 1ts9, 1ttz, 1tuk, 1tuw, 1u2h, 1u3y, 1u84, 1u9p, 1ub9, 1ucs, 1uj8, 1uoy, 1use, 1usm, 1v05, 1v2z, 1v70, 1vbw, 1vjk, 1vyi, 1whz, 1wpa, 1wri, 1xak, 1xbi, 1xe1, 1xg8, 1xw3, 1y0n, 1y6x, 1y9l, 1ygt, 1yib, 1yu5, 1yzm, 1z0p, 1z21, 1z96, 1zld, 1zt3, 1zva, 1zzk, 2asc, 2b1y, 2b8i, 2bkf, 2brf, 2c60, 2cbo, 2cg7, 2cmp, 2cwr, 2cwy, 2cyj, 2d3d, 2d9r, 2e3h, 2ea9, 2efv, 2es9, 2ewh, 2ewk, 2ewt, 2f60, 2fb6, 2fht, 2fi0, 2fq3, 2fu2, 2g7o, 2gkr, 2gpi, 2gyz, 2h8e, 2h9u, 2haz, 2hc8, 2hdz, 2hl7, 2hlr, 2hpj, 2i4a, 2i5f, 2i6v, 2iay, 2igp, 2ip6, 2ivy, 2j6b, 2j8b, 2j97, 2jku, 2nml, 2nqw, 2ns0, 2nsc, 2nuh, 2o0q, 2o1k, 2o37, 2o4t, 2o9u, 2od5, 2ooa, 2oqk, 2oqq, 2ouf, 2ovg, 2oxo, 2p5k, 2p6v, 2pk8, 2pnd, 2ppn, 2pst, 2py0, 2q2f, 2q79, 2qff, 2qmt, 2qsb, 2qsk, 2qtd, 2qvo, 2qyw, 2r39, 2r4q, 2rff, 2rh2, 2rhf, 2ril, 2uux, 2v75, 2vc8, 2vkl, 2vq4, 2vsd, 2wbx, 2wj5, 2wkd, 2wuj, 2x3g, 2xfd, 2xnq, 2yv4, 2yvi, 2yxf, 2yzt, 2zqe, 2zqm, 2zxy, 3a0s, 3a38, 3a4c, 3adg, 3agn, 3b64, 3bjo, 3bn0, 3bn7, 3bri, 3bt4, 3bv8, 3c0f, 3ca7, 3ce7, 3cec, 3ci9, 3cq1, 3ctr, 3cw3, 3df8, 3dml, 3dqy, 3e0e, 3e21, 3e56, 3e9v, 3eaz, 3efg, 3emi, 3enu, 3ewg, 3exy, 3f14, 3f2e, 3f40, 3fbl, 3fdr, 3ff2, 3ff5, 3ffy, 3fkc, 3fmy, 3ft7, 3g21, 3g2b, 3ghf, 3goe, 3gv3, 3h01, 3h36, 3h8h, 3h9w, 3hgl, 3hms, 3hnx, 3hqx, 3hrl, 3hro, 3hz7, 3i8z, 3idw, 3im3, 3iv4, 3jsc, 3jsr, 3jtz, 3ju3, 3jvl, 3k0x, 3k3v, 3kkf, 3kp8, 3kw6, 3kzd, 3l32, 3l4h, 3l9a, 3lax, 3lbj, 3ldc, 3le4, 3lwc, 3lyg, 3lyw, 3msh, 3mwz, 3mxz, 3n3f, 3nbm, 3npd, 3nrw, 3nx6, 3nzl, 3oiz, 3oou, 3osh and 3phn.

The program described has also been tested on a pool of 15 structures (Millán *et al.*, 2015) first solved with previous *ARCIMBOLDO* implementations.

2.4. Libraries of local folds

For the *ARCIMBOLDO_BORGES* tests, six precomputed libraries containing geometrically clustered variations of a particular fold were used. Two of them contain pairs of contiguous helices: parallel and antiparallel, respectively. Three correspond to 20 amino acids arranged in three-stranded β-sheets in antiparallel, parallel and parallel–antiparallel dispositions. The last contains a sample of disulfide bridges linking two tetrapeptides in all possible conformations. These libraries of models superimposed to match the generating template were computed with *BORGES* (Sammito *et al.*, 2013) and are available for download from our website, along with libraries of 24 amino acids arranged in four strands with different relative arrangements (http://chango.ibmb.csic.es/download). A library with four antiparallel β-strands has been used redundantly on six cases previously solved with the described library of three-stranded antiparallel sheets to confirm its validity. This and other new libraries will be incorporated into our public repository. Besides the libraries used in this study, a library of DNA-binding motifs is also downloadable (Pröpper *et al.*, 2014).

3. Results and discussion

3.1. Implementation

ARCIMBOLDO_LITE, *ARCIMBOLDO_SHREDDER* and *ARCIMBOLDO_BORGES* have been thoroughly rewritten in the course of the last year. The *ARCIMBOLDO_LITE* implementation is deployed as a binary file and has been designed to run on a single multicore machine. The need for a MySQL database present in previous implementations has been removed, as it is intended to solve easier target structures ranging from 90 to 200 amino acids with complete diffraction data reaching 2.0 Å resolution or beyond. Moreover, particular stress has been placed on simplifying the input, reducing it to minimal requirements and producing helpful output, in particular direct troubleshooting instructions and clear identification of solution files. The program provides default values for all parameters except for the basic specification of experimental data: paths to the reflection data files in `.hkl` (Sheldrick, 2008) and `.mtz` (Winn *et al.*, 2011) format have to be input, as well as the LABELS identifying the data in the `.mtz` file and the contents of the asymmetric unit (molecular weight and number of copies). Paths to *Phaser* and *SHELXE* have to be set unless they are already defined for the user within the system. The default search is exemplified as two copies of a model helix composed of 14 alanines, but the user should select the most appropriate hypothesis according to secondary-structure prediction or previous knowledge about the target structure. For model polyalanine helices, an inbuilt

library is provided and only the length needs to be specified. Model fragments for some cofactors and ligands with high scattering power are also provided internally, in particular for iron–sulfur clusters and the haem group. Information on defaults and usage for all mandatory and optional parameters can be obtained by typing `ARCIMBOLDO_LITE -b` in the terminal window.

Fig. 1 displays a diagram summarizing the flow. The program starts by reading and validating the instruction file. It will substitute default values by user-input choices while checking for formal correctness and an acceptable parameter range. Next, checks are performed to customize the job, adapting it to the experimental data and available hardware, as well as to prevent common errors that cause run termination at a more advanced stage.

(i) The number of physical cores in the computer is read from `/proc/cpuinfo` (in Mac OS this information is extracted with `sysctl -n hw.ncpu`).

(ii) External search models are confirmed to be accessible files in valid PDB format containing at least one atom. Format problems are corrected if possible.

(iii) Paths to files and work directories are verified in terms of existence, accessibility and user permissions.

(iv) If a sequence is provided to compute the molecular weight, its validity is checked.

(v) Python, *Phaser* and *SHELXE* versions are checked for compatibility with the deployed binary.

(vi) A *Phaser* job is launched to validate configuration and data compatibility while generating a corrected data file with anisotropically scaled amplitudes for use in subsequent jobs.

(vii) A *SHELXE* job is launched to test the input line and data format, identifying amplitudes or intensities.

(viii) The space group, if specified, is tested to be a Sohncke space group in a standard setting.

Failure of any of these tests will cause the program to stop, with an error message printed to the standard output indicating the cause. Furthermore, if the resolution of the experimental data does not reach 2 Å a warning will be issued, while below 2.5 Å resolution the program will terminate.

Results are output in xml and html format and are constantly updated for the user to monitor the program flow. The html file (Fig. 2) echoes all of the parameters used in the run and a table of results and figures of merit characterizing final and partial solutions. The table can be sorted by any of its entries. The html page provides links to the files containing the best map and trace.

ARCIMBOLDO_LITE runs in two macrocycles. During the first one *Phaser* sequentially places all requested copies of the search model(s) specified through the configuration file and partial structures are ranked and possibly optimized by trimming against the CC (Fujinaga & Read, 1987) with *SHELXE* (Sheldrick & Gould, 1995). In the second macrocycle, *ARCIMBOLDO* selects from all of the partial solutions produced a number of solutions one less than the number of physical cores in the machine to be expanded through density modification and autotracing with *SHELXE*. To render calculations manageable, hard limits are imposed on the maximum number of solutions after each step. These limits are set in relation to the number of cores available. Therefore, a more powerful machine will by default pursue more partial solutions and may solve a structure that a more limited machine would fail to phase unless a longer job is customized. By default, negative LLG values are discarded and, even though this parameter can be overridden through the `.bor` file, the alternative of revising the expected asymmetric unit composition and/or data resolution appears to be preferable.

Phaser jobs are executed as independent, monitored processes and input is given through keyword files for compatibility with all environments and middleware. Firstly, a rotation search (Storoni *et al.*, 2004) is computed for the input model and solutions within 75% of the top LLG are selected. These are clustered within an angular difference of 15°, taking symmetry into account, after converting the Euler angles to quaternions or rotation matrices. If, exceptionally, the number of rotation peaks exceeds 10 000, numerical *k*-means clustering is used as a filter prior to geometrical clustering. To save time, an ancillary Cantor pairing value is used. The rotation characterized by the highest LLG is used as a reference to determine similar rotations. The process is iterated over the nonclustered rotations until the list has been scanned.

Translation performance (McCoy *et al.*, 2005) is evaluated jointly for each cluster of rotation solutions, limiting the maximum number of solutions to be sent to packing and refinement. In this way, variability is preserved if a given fragment can be placed in several parts of a structure. For the placed fragments, an initial CC is computed with *SHELXE*, optionally eliminating residues to increase this figure of merit. PDB optimization (Sheldrick & Gould, 1995) is usually not

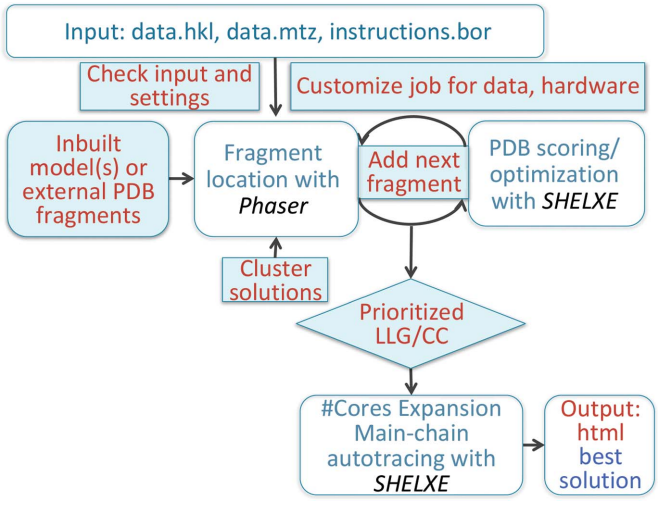


Figure 1
ARCIMBOLDO_LITE flow scheme. *ARCIMBOLDO* operations are labelled in red. Input data and program settings are validated and the run is configured with parameterization defaults depending on data resolution and limiting calculations by the number of cores in the workstation. In a first macrocycle all copies of the search fragments are placed with *Phaser* and optimized with *SHELXE*. Similar solutions are clustered. A number of substructure solutions equal to the number of physical cores minus one is subject to density modification and autotracing with *SHELXE*, prioritizing according to top LLG and CC.

==========Summary of your data==========

SPACEGROUP: P 1 21 1
CELL DIMENSIONS: 23.26, 60.44, 23.59, 90.00, 90.99, 90.00
RESOLUTION: 0.92
NUMBER OF UNIQUE REFLECTIONS: 35634.00

2. STEP: Locating Sequentially the Fragments

Fragment 1																						
Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
0	20	18.99	14.21	4.62	3.73	2	35.51	32.35	6.90	5.71	2	35.51	32.35	6.90	5.71	2	39.20	37.75	7.56			
1	1	16.68	16.68	4.19	4.19	4	25.36	23.54	4.01	3.50	4	25.36	23.54	4.01	3.50	4	34.80	28.17	7.08			
2	3	16.67	14.46	4.19	3.78	7	22.00	21.12	3.62	3.00	7	22.00	21.12	3.62	3.00	7	24.00	22.34	6.20			
3	1	16.33	16.33	4.12	4.12	4	22.49	21.67	3.46	3.18	4	22.49	21.67	3.46	3.18	4	24.60	23.20	6.56			
4	3	13.79	13.54	3.65	3.61	2	26.09	25.99	5.18	5.15	2	26.09	25.99	5.18	5.15	2	28.90	28.80	3.84			
5	1	13.71	13.71	3.64	3.64	1	21.86	21.86	4.04	4.04	1	21.86	21.86	4.04	4.04	1	22.90	22.90	7.59			

Fragment 2																						
Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
(0, 6)	11	56.31	50.27	4.43	3.58	7	108.16	93.75	10.25	9.18	7	108.16	93.75	10.25	9.18	4	124.90	118.85	13.38			
(0, 9)	33	52.94	49.95	3.79	3.51	21	117.06	98.86	12.30	10.05	21	117.06	98.86	12.30	10.05	14	125.00	106.64	12.16			
(1, 9)	18	48.93	40.73	3.85	3.73	92	99.10	52.27	10.29	5.14	3	99.10	91.80	10.29	9.31	3	103.90	98.40	9.99			
(1, 11)	96	54.19	42.45	4.86	3.88	163	119.42	59.00	12.64	5.70	13	119.42	94.96	12.64	9.67	5	124.60	111.36	13.08	4	45.79	81
(4, 6)	2	45.49	45.49	4.42	4.41	2	91.19	88.89	9.53	9.15	2	91.19	88.89	9.53	9.15	2	103.90	101.30	10.00			
(4, 11)	42	47.61	42.52	4.80	3.85	28	99.58	83.12	10.81	8.91	28	99.58	83.12	10.81	8.91	12	105.70	95.93	10.30			

Show All Hide Not Relevant

The current best solution is: th10_0_0xx2FR1_2-1.pdb with FINALCC: 45.79 and n. residues traced 81
file is: /home/dawid/1l9l/8_EXP_LIBRARY/4/0/th10_0_0xx2FR1_2-1.pdb

- FRF: Pos. in Rank: **48** LLG: **52.27** ZSCORE: **4.46** Top LLG in Cluster (**1, 11**): **56.31** Top ZSCORE in Cluster (**1, 11**): **4.94**
- REFINEMENT ROTATION AND MODEL

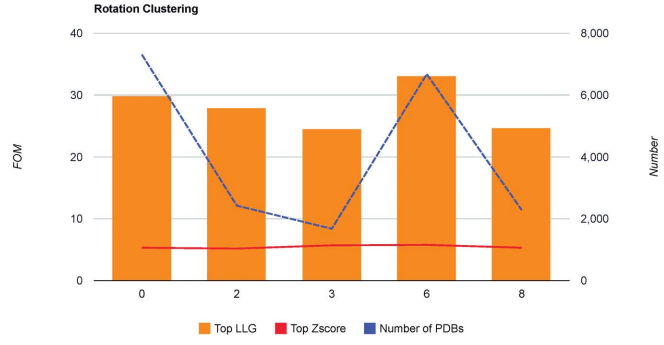
- FTF: Pos. in Rank: **1** LLG: **119.42** ZSCORE: **12.64** Top LLG in Cluster (**1, 11**): **119.42** Top ZSCORE in Cluster (**1, 11**): **12.64**
- PACK: Pos. in Rank: **1** LLG: **119.42** ZSCORE: **12.64** Top LLG in Cluster (**1, 11**): **119.42** Top ZSCORE in Cluster (**1, 11**): **12.64**
- RNP: Pos. in Rank: **19** LLG: **124.60** ZSCORE: **0.00** Top LLG in Cluster (**1, 11**): **125.00** Top ZSCORE in Cluster (**1, 11**): **0.00**
- INITIAL CC
  - After Refinement: Pos. in Rank: **5** INITCC: **12.67** Top INITCC in Cluster (**1, 11**): **13.38**
- EXPANSION

Cycle 4:
Final CC: **45.79%** N. Residues Traced: **81.00**

(*a*)

#Cluster	#Rotations	#Distinct Pdb	Top LLG	LLG Mean	Top Zscore	Zscore Mean
0	7310	4018	29.81	17.99	5.33	3.64
2	2428	1918	27.95	15.44	5.19	3.38
3	1679	1368	24.50	13.73	5.72	3.19

Show All Hide Not Relevant



#Cluster	Rotation and Model Refinement					Translation Function					Packing					Rigid Body Refinement			Initial CC		Best Trace CC/aa			
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Before Refinement CC	After Refinement CC	MODE	Cycle	CC	#Res. traced
0	7310	29.81	17.99	5.33	3.64	700	57.20	39.86	8.07	5.54	267	57.20	40.11	8.07	5.43	267	63.60	42.44	7.49	7.85	FAST	1	45.4	103

Show All Hide Not Relevant

The current best solution is: 1les_0_115.pdb with FINALCC: 45.40 and n. residues traced 103
file is: /home/alfotest/1v70/11_EXP/0/1/0/1les_0_115_rottra.pdb

- FRF: Pos. in Rank: **6758** LLG: **26.50** ZSCORE: **4.04** Top LLG in Cluster **0**: **33.11** Top ZSCORE in Cluster **0**: **5.79**
- REFINEMENT ROTATION AND MODEL

- FTF: Pos. in Rank: **1** LLG: **57.20** ZSCORE: **7.94** Top LLG in Cluster **0**: **57.20** Top ZSCORE in Cluster **0**: **8.07**
- PACK: Pos. in Rank: **1** LLG: **57.20** ZSCORE: **7.94** Top LLG in Cluster **0**: **57.20** Top ZSCORE in Cluster **0**: **8.07**
- RNP: Pos. in Rank: **5** LLG: **58.50** ZSCORE: **0.00** Top LLG in Cluster **0**: **63.60** Top ZSCORE in Cluster **0**: **0.00**
- INITIAL CC
  - Before Refinement: Pos. in Rank: **1** INITCC: **7.49** Top INITCC in Cluster **0**: **7.49**
  - After Refinement: Pos. in Rank: **5** INITCC: **7.25** Top INITCC in Cluster **0**: **7.85**
- EXPANSION

Cycle 1:
Final CC: **45.40%** N. Residues Traced: **103.00**

It seems you have a good solution!
Here you can find the best solution and map for further refinement.
BORGES will end now.

(*b*)

Figure 2
Html output for (*a*) the *ARCIMBOLDO_LITE* solution of PDB entry 1l9l and (*b*) the *ARCIMBOLDO_BORGES* solution of PDB entry 1v70.

Table 2
Performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a pool of 294 test structures.

A detailed table with PDB codes and parameterizations is available as Supporting Information.

Test	Pool of structures	Solved by *ARCIMBOLDO_LITE*	Solved by *ARCIMBOLDO_BORGES*	Total solved
Total	294	143	38	181
Resolution				
1–0.54 Å	24	14	5	19
1.3–1.0 Å	43	25	8	33
1.6–1.3 Å	78	38	12	50
2.2–1.6 Å	149	66	13	79
Fragments				
Helices	230†	139	3	142
Fe clusters, haem	5	4	0	4
β-Sheets	147	0	34	34
Antiparallel (udu)	131	—	31	31
Parallel–antiparallel (uud)	24	—	2	2
Parallel (uuu)	13	—	1	1
Antiparallel (udud)	6 (70)	—	6	6
Disulfide bridges	20	0	1	1

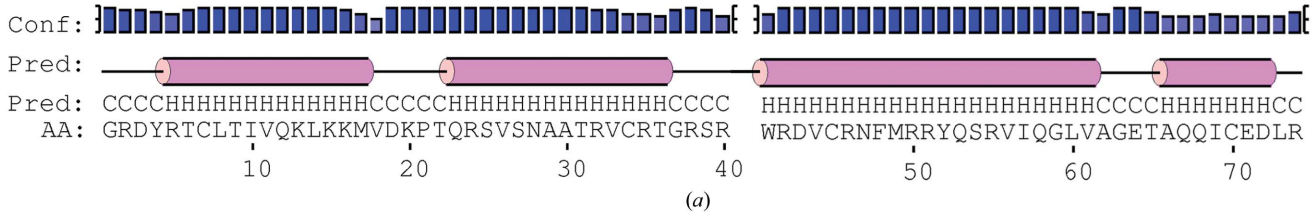
† 1 curved; 2 antiparallel (ud).

critical if the search fragment is a helix, but is the default for libraries of folds.

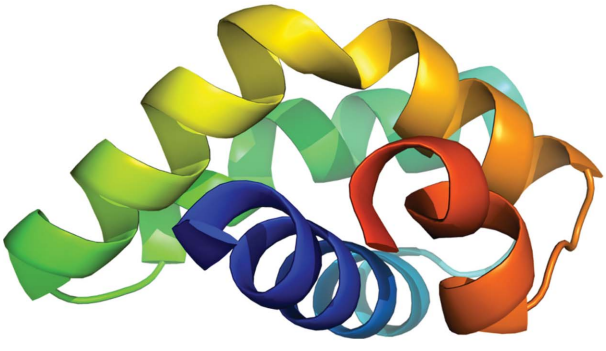
Subsequent searches for additional copies of the same or different fragments are also subjected to clustering of rotation solutions. Clusters are identified by combinations of numbers corresponding to the original and the new rotation peaks. Equivalent combinations are kept apart for simplicity when calculating the translation search but are joined for the refinement stage so that they may be fused if equal. Clustering and evaluation results are written into formatted text files, which can be used by the program to continue an interrupted run. Relaunching an interrupted program will read the present folders and continue from the latest stage completed, which implies that a change in parameterization will only be effective if all previous output for this and subsequent steps has been deleted.

After all fragment-placement operations have been performed, selected solutions are subjected to density modification and autotracing with *SHELXE* (Sheldrick, 2010). Selected solutions may correspond to intermediate fragment-placement rounds as they are chosen to produce the highest CC and LLG. The default number of solutions probed is equal to the number of physical cores in the machine minus one.

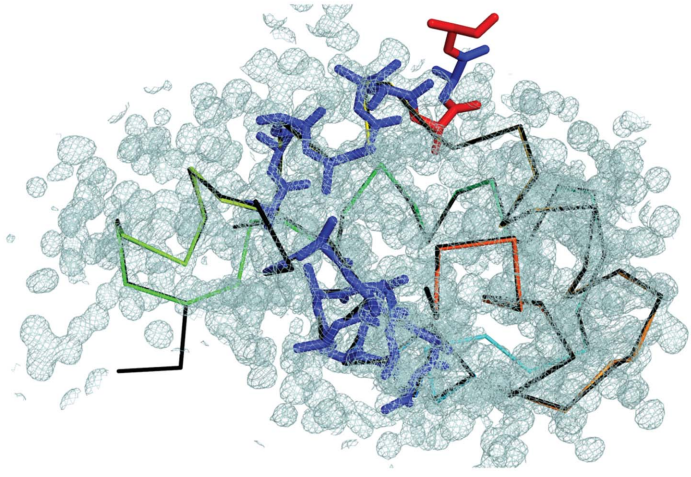
If the user specifies a PDB format file with the extension `.ent` in the configuration file, analysis of results compared with this reference is performed. Goodness of partial solutions is evaluated through the r.m.s.d. of rotations after optimal translation and mean phase errors *versus* the reference for placed fragments and final solutions. This information is not output to the standard `.html` file, but has been used in the present study of test structures to evaluate results and to guide



(*a*)



(*b*)



(*c*)

Figure 3
ARCIMBOLDO_LITE solution of PDB entry 1191. (*a*) Secondary-structure prediction. (*b*) Structure in rainbow-coloured cartoon representation. (*c*) *SHELXE* electron-density map of a solution; the trace is shown as a coil and both placed helical fragments are shown as sticks, with residues eliminated during PDB optimization displayed in red.

default parameterization. The procedure starts by extracting all of the helices (or any other search fragment given as input) found in the `.ent` file. For each one of them a rotation LLG at their current position is computed in *Phaser* through the GYRE mode to provide an estimate of the value range for a correct solution. For the placed fragments, the *F*-weighted mean phase difference to the reference structure (Lunin & Woolfson, 1993) is computed with *SHELXE*.

3.2. Timing benchmarks on various hardware

The structure of the C-terminal domain of human anti-termination factor 5 (PDB entry 2iu1; Bieniossek *et al.*, 2006) was used on 12 machines with hardware ranging from four to 24 physical cores and a minimum of 1 GB of RAM per core. A case is included in which 512 MB of RAM was shared by two cores, showing that the associated increase in run time (RT) makes such a configuration unsuitable. Conversely, a machine endowed with CPU overclocking up to 4 GHz and almost 8 GB of RAM per core rendered an RT that was half that of a typical workstation. Table 1 shows the time taken for structure solution on the various machines. It ranged from one to several hours in the default mode that runs as many processes as there are physical cores detected, reserving one thread for program calculations, mainly clustering and combination of solutions, and running as many *Phaser* and *SHELXE* jobs as there are cores minus one. It is worth remarking that the job computed is tailored to the hardware and thus a greater or lesser number of solutions are pursued depending on the number of cores. Accordingly, the time required by a more powerful machine may not be much less, but it may succeed in solving a structure where a machine with fewer cores fails.

3.3. Performance of *ARCIMBOLDO* tests

Table 2 summarizes the single-workstation performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a set of test structures. An initial baseline was set by running *ARCIMBOLDO_LITE* blindly on the pool of 294 structures with a fragment search configured to find two polyalanine helices of 14 residues and using a common *SHELXE* parameterization. This resulted in the solution of 100 of the 294 structures. *SHELXE* was set to perform five iterations involving 30 cycles of density modification and autotracing, increasing the time for random search 20-fold and taking into account the presence of helices, a solvent content of 55% and fragment optimization against the CC. In many cases, three iterations and a tenfold time increase to seed autotracing led to structure solution. Accounting for the particularities of individual structures opens a way to improve the success rate with minimum intervention.

3.3.1. Ultrahigh resolution (beyond 1 Å). Within the pool of 294 test cases, 24 are atomic resolution structures diffracting to 1 Å resolution or beyond and, of these, 11 contain at least one helix of six amino acids or longer. Such short helices are useful only in this resolution context, where some of the structures are very small. All such helical structures in this group were eventually solved. The behaviour of this group will be illustrated by the case of PDB entry 1l9l in space group $P2_1$ with diffraction to 0.92 Å resolution. Fig. 2(*a*) displays the html output from the *ARCIMBOLDO_LITE* run. The all-helical structure displays three long helices of 18, 15 and 25 residues, as expected from the secondary-structure prediction displayed in Fig. 3(*a*). The long third helix is divided into two stretches by a sharp kink. This scenario would appear to be deceptively simple, as the data quality is excellent, but the blind parameterization that was initially used failed to solve it. The same happened with a number of other structures in this group during the first, blind test. Instead, search trials with different helix lengths and parameterization tests led to a solution when searching for two helices of ten residues and increasing the number of density-modification cycles, turning on low-density elimination (Yao *et al.*, 2005), decreasing the solvent content and increasing the rounds of autotracing but lowering the time seed (*SHELXE* command line `-m200 -v0.5 -a10 -t1 -q -f -s0.25 -o`).

This result suggested the convenience of taking into account particularities anticipated from data resolution to automate default parameterization. Some common settings are adopted, for instance the presence of helices is always assumed for tracing whenever a helical search model is used and default memory allocation is increased to fit the hardware. The model helix of 14 amino acids fits most helical stretches within an r.m.s.d. of 0.3 Å for the C_α atoms. The availability of extremely high resolution data allows smaller fragments to be extended, but suffers more severely from the lack of accuracy in the model. As can be seen in Figs. 3(*b*) and 3(*c*), they contained errors as one of the helices is misplaced. A solution from an imperfect structure is frequently found, and wrong residues may be initially eliminated to increase the CC. Thus, PDB optimization is switched on and, given sufficient cycles and high resolution, the remaining errors are slowly erased. Starting from shorter helices may be more effective as accuracy is enhanced. More density-modification cycles are beneficial and low-density elimination should be turned on, whereas tracing of the atomic map is simpler and can be accelerated with a smaller time parameter. The solvent content is typically extremely low for this resolution group.

3.3.2. High resolution (1.3–1 Å). In the group of 33 structures with helices and resolutions between 1 and 1.3 Å, 25 could be solved when searching for predicted helices and running eight iterations of density modification and autotracing; the number of density-modification cycles was not increased so much this time, setting it to 100, and low-density elimination was given a lower weight of 0.25. A solvent content of 35% was assumed and a time seed of 10 was used. As the resolution becomes lower, more starting information is needed for successful expansion; thus, longer helices of 14 amino acids rather than the extremely small helices of less than ten residues were used as search fragments.

3.3.3. Medium resolution (1.6–1.3 Å). In the pool of 62 structures with helices and resolutions between 1.3 and 1.6 Å, 38 could be solved when searching for predicted helices and running eight autotracing cycles interspersed with 50 density-modification cycles, with low-density elimination set to

0.1 and a time seed of 10 and assuming 45% solvent content.

3.3.4. Low resolution (2.2–1.6 Å). At the lower resolution end, 66 of 149 structures were solved. 124 of the structures contained at least 8% helices, which is the lowest limit found to be accessible at ultrahigh resolution, but in this context such a low percentage is clearly not sufficient to extend helical fragments. More fragments of longer size but tolerating higher deviations provide successful starting substructures. The lower the resolution, the more difficult it becomes to successfully extend the partial structures, even if they are located. The number of density-modification cycles is gradually decreased from 15 at the higher resolution end of this span to ten below 2 Å, a solvent content of 50–60% is assumed and low-density elimination is entirely switched off. The time seed for autotracing is increased.

3.3.5. Alternative search fragments. Five of the structures in the test pool contained cofactors, in particular Fe_4S_4, Fe_2S_2 and a haem group. Despite the presence of helices, solving such structures without first accounting for the electron-rich group failed at the stage of fragment location. As the presence of cofactors is easily predictable, their use as search fragments appeared to be indicated. The search for them may be combined with that for other fragments, but in the cases tested the location of model fragments from the *REFMAC*5 dictionary (Murshudov *et al.*, 2011; Vagin *et al.*, 2004) was sufficient for a successful expansion to the full structure in four of the five cases. The parameterization was set to match the resolution group, but the initial fragment was kept throughout the first autotracing cycles (*SHELXE* `-K` flag) as only polypeptide main chain is built into the electron-density map (Sheldrick, 2010). These fragments may be specified by their three-letter library name in the instruction `.bor` file as they have been included along with the model helix in the internal library.

The straight model helix built into *ARCIMBOLDO* may not be the optimal search fragment for helical structures containing markedly curved or otherwise distorted helices. This is the case for the 90-amino-acid, 1.5 Å resolution structure with PDB code 3l32 in space group $I4_122$. Solution with *ARCIMBOLDO_LITE* failed, whereas using a collection of 162 models for differently distorted helices of 18 amino acids extracted from the PDB and gathered in a library rendered a solution. PDB entry 3l32 is one of the 38 cases solved with libraries of similar models with a single-computer implementation of *ARCIMBOLDO_BORGES*.

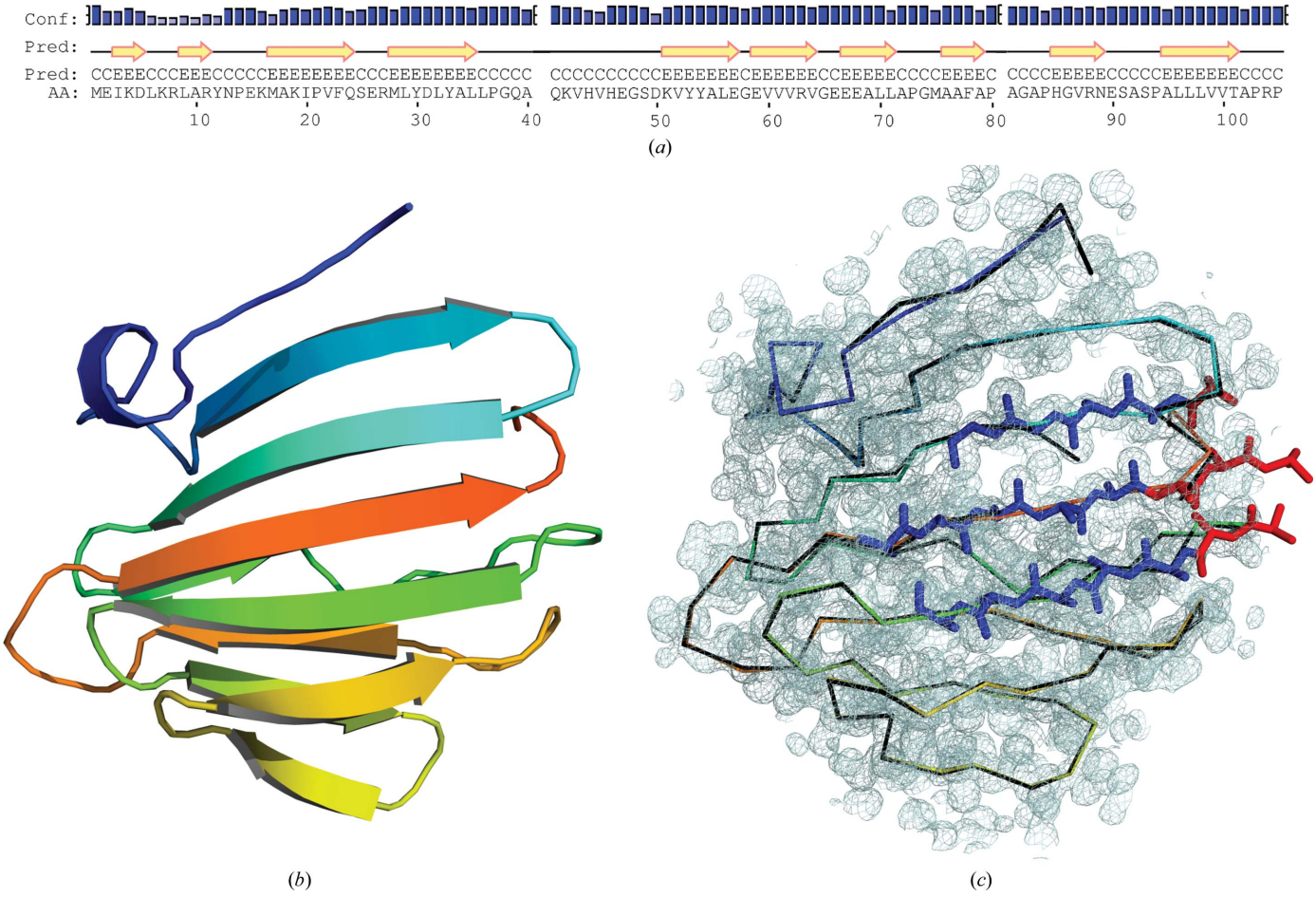


Figure 4
ARCIMBOLDO_BORGES solution of PDB entry 1v70. (*a*) Secondary-structure prediction. (*b*) Structure in rainbow-coloured cartoon representation. (*c*) *SHELXE* electron-density map of a solution; the trace is shown as a coil and the placed antiparallel three-stranded fragments are shown as sticks, with residues eliminated during PDB optimization displayed in red.

3.4. Performance of *BORGES* libraries on test structures

The pool of test structures contains 147 cases in which β-strands are the only secondary-structure fragments or make up most of the fold. In addition, 20 structures feature four or more disulfide bridges anchoring a peculiar fold, as is typical in small toxin inhibitors. Such cases are outside the scope of an *ARCIMBOLDO_LITE* secondary-structure fragment search, but can be solved with libraries of small folds in *ARCIMBOLDO_BORGES* (Sammito *et al.*, 2013). Although the method is better suited to large computing resources, the new implementation can run either connecting to a local or a remote grid or on a single multicore machine. Again, automatic parameterization is related to the available hardware, and only four rotation clusters will be sequentially evaluated by default when the program is run on a single machine. The test structures in this pool are small enough to be computed on the eight-core workstations used for the *ARCIMBOLDO_LITE* tests. The typical run time was 2–5 d. The same precomputed libraries were used in all tests. They are available for download from the *ARCIMBOLDO* web page. Three libraries contain β-sheets with three strands in parallel, parallel–antiparallel and antiparallel arrangements. The latter have been the most successful, being more frequent; 131 of the cases in the group of β-sheet-containing structures display this arrangement of strands. In addition to the library of single distorted helices mentioned, two libraries contain two contiguous helices in parallel and antiparallel dispositions, respectively. They are especially indicated for coiled coils, as single helices tend to be placed in the same position in fragment searches. Finally, a library of two tetrapeptides linked by a disulfide bridge has also been tested.

Two test cases were solved with the library of helices, and 31 with the antiparallel, two with the parallel–antiparallel and one with the parallel three-stranded library, respectively. One structure was solved with the disulfide-linked peptide library. The optimal parameterization has been found to be different for structures mainly composed of β-strands and is accordingly set as default for *ARCIMBOLDO_BORGES* as exemplified in Fig. 4. If the program is launched with one of the nonhelical libraries, the default expansion stage will not search for helices, will use a lower solvent content than in the helical case, will keep the original fragment for *K* cycles and will perform more autotracing iterations. As in *ARCIMBOLDO_LITE*, a line in the `.bor` file setting the parameterization for *SHELXE* will override any of the inbuilt defaults, but leaving this line unset will apply the resolution-tailored and hardware-tailored values described.

3.5. Statistics and performance

Table 2 summarizes the single-workstation performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on the 294 test structures. Beyond the 181 structures (62%) solved as described, some could be solved if run on a computer with more cores or if the parameterization was further adapted to run a more exhaustive attempt or even with less restrictions on a grid of computers. For example, the 114-amino-acid structure with PDB code 3f40 in space group $C222_1$ at 1.27 Å resolution renders correct substructures composed of two helices of 12 and eight alanines, respectively, but these are not prioritized among the seven expanded helices. Running *ARCIMBOLDO_LITE* on a machine with more cores or a grid would solve it. Alternatively, the number of expansions to try may be increased by the user. Regarding *ARCIMBOLDO_BORGES*, the 105-amino-acid structure with PDB code 2haz at 1.7 Å resolution is solved automatically on a double Xeon workstation with similar features to those used for benchmarking performance in this study but endowed with 3 GB rather than 2 GB of RAM per core. This case has not been counted as solved in our results.

High-resolution cases yield a higher success rate: 78% for atomic resolution structures, decreasing to 64% for data in the 1.6–1.3 Å resolution range and to 53% below 1.6 Å resolution. For structures mainly composed of β-strands, phasing with the secondary-structure fragments has not found any practical use within *ARCIMBOLDO*, whereas precomputed libraries of β-sheets have been successful in 22% of the identified cases. Not all have been tested, as structures previously solved with helices were eliminated from the pool. In the same manner, the more frequent antiparallel disposition of three strands was tested first and only structures where this fold would not work were further considered for phasing with parallel–antiparallel or parallel libraries. Six structures were phased both with libraries containing three-stranded and four-stranded models. They are counted only once in the total of 181 structures. The four-stranded libraries are also available for download and might be useful for larger structures. In the same way, two structures were phased with the library of disulfide bridges and one with a library of polyalanine helices extracted from real structures and presenting different deviations from the regular model helix. This structure, PDB entry 3l32, could not be solved with a model helix, and the use of the library illustrates that a starting model with a lower r.m.s.d may succeed where the model fails and that the more exhaustive run needed may be worthwhile to solve an unknown structure.

4. Concluding remarks

ARCIMBOLDO_LITE succeeds in solving 143 out of a pool of 294 test structures on an eight-core workstation, with sizes ranging from 40 to 120 residues and resolutions between 2.2 and 0.54 Å. The fragments placed are predominantly straight polyalanine helices and, in three cases, iron cofactors. *ARCIMBOLDO_BORGES* solves a further 38 structures using precomputed libraries of small folds consisting of contiguous helices, three-stranded β-sheets and disulfide-linked tetrapeptides. The run times for *ARCIMBOLDO_LITE* jobs are typically a couple of hours, whereas *ARCIMBOLDO_BORGES* requires one or two days on a single machine.

The results of this study have been used to incorporate default parameterization as a function of data resolution and search fragment. Leaving the *SHELXE* line unset in the input `.bor` file will activate *ARCIMBOLDO* defaults, which may differ from *SHELXE* defaults.

The binaries for Linux and MAC OS have been validated for the latest distributions and can be downloaded from http://chango.ibmb.csic.es/ARCIMBOLDO_LITE. The same binaries that run on a single workstation can be run to distribute jobs on a local or remote Condor, SGE/Opengrid or Torque/MOAB grid. The precomputed libraries can be downloaded together with the *ARCIMBOLDO_BORGES* binary.

Acknowledgements

We thank George M. Sheldrick, Randy J. Read and Airlie McCoy for helpful discussions. This work was supported by grants BFU2012-35367 and BIO2013-49604-EXP (the Spanish Ministry of Economy and Competitiveness) and Generalitat de Catalunya (2014SGR-997). RJB acknowledges a grant from the Brazilian Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2013). *Acta Cryst.* D**69**, 2194–2201.
Bieniossek, C., Schütz, P., Bumann, M., Limacher, A., Usón, I. & Baumann, U. (2006). *J. Mol. Biol.* **360**, 457–465.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2011). *J. Appl. Cryst.* **44**, 1143–1151.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2012). *J. Appl. Cryst.* **45**, 1287–1294.
Burla, M. C., Giacovazzo, C. & Polidori, G. (2010). *J. Appl. Cryst.* **43**, 825–836.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. & Siliqi, D. (2008). *J. Appl. Cryst.* **41**, 548–553.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005). *Acta Cryst.* D**61**, 1080–1087.
Caliandro, R., Dibenedetto, D., Cascarano, G. L., Mazzone, A. & Nico, G. (2012). *Acta Cryst.* D**68**, 1–12.
Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Glykos, N. M. & Kokkinidis, M. (2003). *Acta Cryst.* D**59**, 709–718.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. (2013). *Acta Cryst.* A**69**, 51–59.
Hissa, D. C., Bezerra, G. A., Birner-Gruenberger, R., Silva, L. P., Usón, I., Gruber, K. & Melo, V. M. M. (2014). *Chembiochem*, **15**, 393–398.
Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
Keegan, R. M., Bibby, J., Thomas, J., Xu, D., Zhang, Y., Mayans, O., Winn, M. D. & Rigden, D. J. (2015). *Acta Cryst.* D**71**, 338–343.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Millán, C., Sammito, M. & Usón, I. (2015). *IUCrJ*, **2**, 95–105.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Morris, R. J., Blanc, E. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 227–240.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oeffner, R. D., Bunkóczi, G., McCoy, A. J. & Read, R. J. (2013). *Acta Cryst.* D**69**, 2209–2215.
Pröpper, K., Meindl, K., Sammito, M., Dittrich, B., Sheldrick, G. M., Pohl, E. & Usón, I. (2014). *Acta Cryst.* D**70**, 1743–1757.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Rigden, D. J., Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 1288–1291.
Robertson, M. P., Chi, Y.-I. & Scott, W. G. (2010). *Methods*, **52**, 168–172.
Robertson, M. P. & Scott, W. G. (2008). *Acta Cryst.* D**64**, 738–744.
Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. (2009). *Nature Methods*, **6**, 651–653.
Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M. & Usón, I. (2012). *Acta Cryst.* D**68**, 336–343.
Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.
Sammito, M. D., Meindl, K., de Ilarduya, I. M., Millán, C., Artola-Recolons, C., Hermoso, J. A. & Usón, I. (2014). *FEBS J.* **281**, 4029–4045.
Sammito, M. D., Millán, C., Rodríguez, D. D., de Ilarduya, I. M., Meindl, K., De Marino, I., Petrillo, G., Buey, R. M., de Pereda, J. M., Zeth, K., Sheldrick, G. M. & Usón, I. (2013). *Nature Methods*, **10**, 1099–1101.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sheldrick, G. M., Gilmore, C. J., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2012). *International Tables for Crystallography*, 2nd online ed., edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 413–429. Chester: International Union of Crystallography.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.
Shrestha, R., Berenger, F. & Zhang, K. Y. J. (2011). *Acta Cryst.* D**67**, 804–812.
Shrestha, R. & Zhang, K. Y. J. (2015). *Acta Cryst.* D**71**, 304–312.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Thorn, A. & Sheldrick, G. M. (2013). *Acta Cryst.* D**69**, 2251–2256.
Usón, I., Stevenson, C. E. M., Lawson, D. M. & Sheldrick, G. M. (2007). *Acta Cryst.* D**63**, 1069–1074.
Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* D**60**, 2184–2195.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, D. & Zhang, Y. (2012). *Proteins*, **80**, 1715–1735.
Yao, J., Dodson, E. J., Wilson, K. S. & Woolfson, M. M. (2006). *Acta Cryst.* D**62**, 901–908.
Yao, J., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2005). *Acta Cryst.* D**61**, 1465–1475.