technische universität
dortmund

mpi
dortmund

Molecular insights into *Bacillus anthracis* anthrax lethal complex

Dissertation

zur Erlangung des akademischen Grades eines Doktors

der Naturwissenschaften

der Fakultät für Chemie und Chemische Biologie

der Technischen Universität Dortmund

angefertigt am

Max-Planck-Institut für Molekulare Physiologie in Dortmund

vorgelegt von

Claudia Antoni

Alla mia famiglia,
alla mia bimbina.

Table of Contents

Abstract

Bacillus anthracis secretes the major virulent factor anthrax complex toxin, which is responsible for high mortality in mammals. This tripartite complex comprises three soluble proteins: the protective antigen (PA) and two enzymes, the lethal factor (LF) and edema factor (EF). After proteolytic activation by host proteases, PA oligomerizes into octamers and/or heptamers on the cell surface and the LF as well as EF bind competitively to the pre-pore oligomer surface. After a receptor-mediated endocytosis, the pre-pore complex rearranges in a narrow pore channel that delivers the two unfolded enzymes into the cytosol. There, the LF and EF lead to the inactivation of the mitogen-activated protein kinases (MAPKKs) and the dysregulation of the level of cellular cAMP, respectively.

So far, only fragmented information about the heptameric PA in complex with LF are available. Particularly, complex organization and subunits interaction, remain unclear. In this work, I apply cryo-electron microscopy (cryo-EM) to solve the structure of the fully loaded complex of PA with LF (PA_7LF_3) in the pre-pore conformation. Our data demonstrate features of LF molecules assembling in the PA_7LF_3 complex, showing for the first time three different complex combinations, namely PA_7LF_{2+1A} $PA_7LF_{2+1A'}$ and PA_7LF_{2+1B}. These complexes vary on the 3D LF organization on the PA. Our structural analysis defines LF-LF and PA-LF interactions revealing a new potential interaction interface at the C-terminus of LFs. Finally, we demonstrate an intermediate conformation of LF which was not previously predicted.

These results show for the first time that the PA_7LF_3 can generate three different complex states. This work adds another milestone on the intricated mechanism of the anthrax complex translocation, which culminates in the release of LF and EF into the host cytoplasm.

Zusammenfassung

Das Bakterium *Bacillus anthracis* sorgt durch die Sekretierung seines Toxins, dem Anthrax Komplex, der den Hauptteil des virulenten Faktors ausmacht, für eine hohe Mortaliät in allen Säugern. Dabei besteht der Komplex aus den folgenden drei löslichen Proteinen: dem schützenden Antigen (*protective antigen*, PA) sowie zwei Enzymen, dem letalen Faktor (*lethal factor*, LF) und dem Edema Faktor (*edema factor*, EF). Nach einer proteolytischen Aktivierung des Komplexes von Wirtsproteasen oligomerisiert das PA an der Zelloberfläche zu Octameren und/oder Heptameren. Dabei binden der LF sowie der EF kompetitiv an die Zelloberfläche und bilden so einen Prä-Porzustand. Die Bindung zu einem Rezeptor initiiert weiter die Endozytose, nach welcher der Prä-Porzustand zu einem schmalen Porenkanal umgelagert wird und die beiden ungefalteten Enzyme in das Zytosol gelassen werden. Nach der Rückfaltung der Enzyme in deren nativen und enzymatisch aktiven Zustand im Zytosol sorgt der LF für die Inaktivierung der mitogen-activierenden Protein Kinasen (MAPKKs) und der EF zu einem unregelmäßigen Level an zellulärem cAMP.

Bis Heute gibt es nur limitierende Informationen zu dem kompletten PA_7 Komplex mit dem LF (PA_7LF_3), wobei insbesondere die Komplexformierung sowie die intramolekularen Interaktionen der einzelnen Untereinheiten noch unklar sind. In dieser Arbeit habe ich mithilfe der Kryo-Elektronenmikroskopie (cryo-EM) die Struktur des gesamten Komplexes im Prä-Porenzustand bestehend aus PA und LF (PA_7LF_3) gelöst, wodurch die intermolekularen Interaktionen zwischen LF-LF sowie LF-PA näher untersucht werden konnten und noch unbekannte Komplexformierungen beobachtet werden konnten. Hier konnten insgesamt drei neue Kombinationsmöglichkeiten des Komplexes beobachtet werden: $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1A')}$ und $PA_7LF_{(2+1B)}$. Die Strukturanalyse dieser neuen Komplexformierungen zeigten zudem noch unbekannte Konformationen von LF im gebunden Zustand des heptameren PA, die entlang der gesamten Kette Interaktionen aufzeigte. Zudem wies ausschließlich eine der drei LF Moleküle in dem PA α-*clamp*-gebunden Zustand die offene Konformation auf, in der eine erweiterete Flexibilität der C-terminalen Domäne beobachtet wurde. Die beiden anderen LFs hingegen, zeigen eine neue Konformation und befinden sich in einem sogenannten Intermediären Zustand zwischen der Toxinassemblierung und der

Translokation. Durch die Präsenz beider Konformationszustände der LFs in dem PA_7LF_3 Komplex kann von einem neuen Modell der Translokation ausgegangen werden, in dem die LFs in Abhängigkeit ihrer Konformation durch den PA_7 Kanal in die Wrtszelle transloziert werden.

List of abbreviation

˚C	degree Celsius
%	percent
Å	Ångström
CP3	Cryoplunge 3
CV	column volume
cryo-EM	cryo-electron microscope
cryo-TEM	cryo-electron tomography
CTF	contrast transfer function
EDTA	ehylenediaminetetraacetic acid
EM	electron microscopy
FSC	Fourier-Shell-Correlation
ISAC	iterative stable alignment and clustering approach
kDa	kilo Dalton
l	liter
mg	milligram
min	minutes
ml	milliliter
mM	millimolar
MSP	membrane scaffold protein
MW	molecular weight

MWCO	molecular weight cut off
NaCl	sodium chloride
NaOH	sodium hydroxide
nm	nanometer
nM	nanomolar
PEG	polyethylene glycol
pI	isoelectric point
POPC	phosphatidylcholine
RT	Room temperature
sec	seconds
SDS-PAGE	sodium dodecylsulfate-polyacrylamid electrophoresis
SPHIRE	SPARX for High Resolution Electron Microscopy
Tc	Toxin complex
VPP	Volta Phase Plate
YOLO	You only look once
µg	microgram
µl	microliter
2D	Two-dimentional
3D	Three-dimensional

1. Introduction

1.1. History of anthrax

The term anthrax derives from the Greek word “anthrakis” that means coal and it is associated to the black coal skin lesions that are typical in the cutaneous form of anthrax infections. According to historical texts, the first cases that could be assigned to anthrax happened in Egypt and Mesopotamia. During Moses’ time, anthrax could have been responsible of what we know as the fifth and sixth plague, the sickness affecting livestock and humans respectively. During the age of the ancient Greece and Rome, texts of Homer in The Iliad (700 BC) and in Virgil (28-27 BC) referred to what could be associated to anthrax symptoms (Fig.1.1).

In 19th century, Robert Koch defined the etiology of the bacteria, injecting a pure culture of *B. anthracis* in animals and reporting the effects leading to the death of the animals. He discovered the ability of the bacteria to survive for long time in unfavorable environmental conditions and explained it by the ability of bacteria to form spores. Those studies brought the scientists to what is known as Koch’s postulates, which established the causative relationship between a bacterium and a disease. In the same century, Louis Pasteur brought Koch’s results to a step further. He created the first vaccine against *B. anthracis* by generating weakened anthrax bacteria (Fig. 1.1). Later, in the 20th century, Max Sterne created an attenuated vaccine for animals composed by anthrax living spores. Today, derivatives of his strain count as veterinary anthrax vaccine (Goel, A. et al 2015) that together with the antibiotics (Gilfoyle, D. et al. 2006) control and limit the propagation of anthrax. In the past centuries, this infection represented the most uncontrolled cause of death in pigs, cattle, sheep horses and goat worldwide. According to the World Organization for Animal Health (OIE), only sporadic cases of anthrax are recorded worldwide. Anthrax remains enzootic in some countries of Asia and Africa, in some areas of Europe as well as in some countries/areas of North and South America (Anthrax in Humans and Animals, 2008). More recently, the notoriety of anthrax is also associated to bioterrorism and biological warfare (Fig. 1.1).

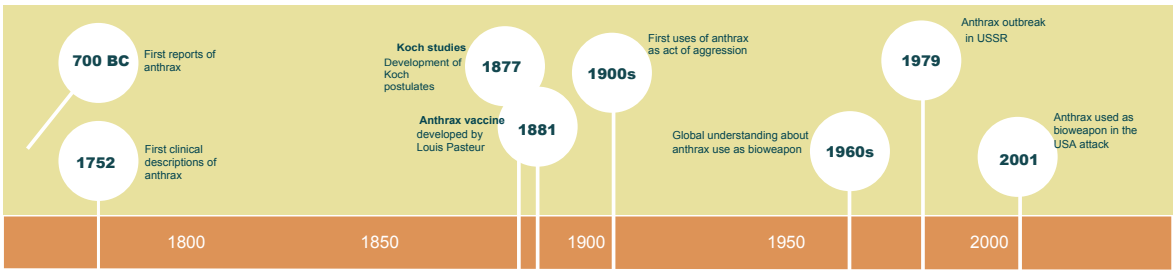


Figure 1.1 Anthrax timeline. From biblical times to today, over the years many sicknesses and symptoms have been associated to anthrax infection. The anthrax timeline reports the first clinical description of anthrax to the latest terroristic attacks in which the anthrax spores were used as biological weapon. Figure adapted from https://www.cdc.gov/anthrax/basics/anthrax-history.html.

1.2. Anthrax as biological weapon

Due to the extreme resistance and relatively easy delivery, anthrax has been listed as potential agent used in bioterrorism and biological warfare together with *Clostridium botulinum* toxin, *Yersinia pestis* and some others. The bioterrorist attack could take place in many forms. The spores could be placed in letters and mailed (Jernigan, J. A. et al. 2001). Anthrax spores could also contaminate food or water. The spores could also be released in the air from a building or a plane, transported with the wind and carried on people clothes or objects. In history, the first case of anthrax as biological weapon refers to the World War I even if it was supposed to be developed as bioweapon in the period between the two World Wars (Pita et al., 2009). Other attacks followed during the next years. Japanese aircraft attacked Chinese cities using anthrax spores in 1932. More recently in 2001 letters filled with white dust containing anthrax spores, were mailed to U.S. Senators' offices. These too were considered as an act of bioterrorism (Goel et al., 2015) (Fig. 1.1).

1.3. Classification

B. anthracis belongs to the family of Bacillaceae, which consists of rod-shape bacteria able to form endospores. The family is characterized by anaerobic spore-forming bacteria of the genus *Clostridium,* and the aerobic or facultatively anaerobic spore-forming bacteria of the genus *Bacillus* (e.g. *Bacillus anthracis*). Bacterial cells of the

genus *Bacillus* are classified as Gram-positive and most of them are saprophytes. A common feature among the different species of *Bacillus* is their ability to withstand harsh environment conditions in their dormant spore state. *Bacillus* endospores are not the unique form able to resist to hostile physical and chemical conditions. Indeed, the genus includes also thermophilic, psychrophilic, acidophilic, alkaliphilic, halotolerant and halophilic representatives, which are capable of growing at quite extreme temperatures, pH values and salt concentrations (Turnbull et al., 1996).

1.4. Sporulation

When the bacilli are released from the body of an animal into the environment, the presence of oxygen and the nutrients deprivation, activate the sporulation process. Every bacteria cell generates yields one endospore. After having completed a multi-step process called sporogenesis, the cell finally releases the emergent spore (Fig. 1.2). The spore represents the resistant and dormant form of the vegetative bacteria which is protected from the unfavorable environmental conditions until they become favorable again for growth and replication. Spores contain less water than their vegetative forms. This low water content confers them heat and UV-radiation resistance, protecting the core components (Gould 1977). During sporulation, the vegetative metabolism is minimized.

During sporulation, the genetic material is replicated and concentrated in a small portion of cytoplasm (Fig.1.2, 1). Plasma membranes surround the newly replicated genetic material (Fig.1.2, 2). Afterwards, the spore septum surrounds the genetic material, generating the forespore (Fig.1.2, 3). Peptidoglycan layers are formed in between membranes and the spore coat surrounds the new generated spore (Fig.1.2, 4 and 5). The endospore finally is released from the cell into the environment (Fig.1.2, 6).

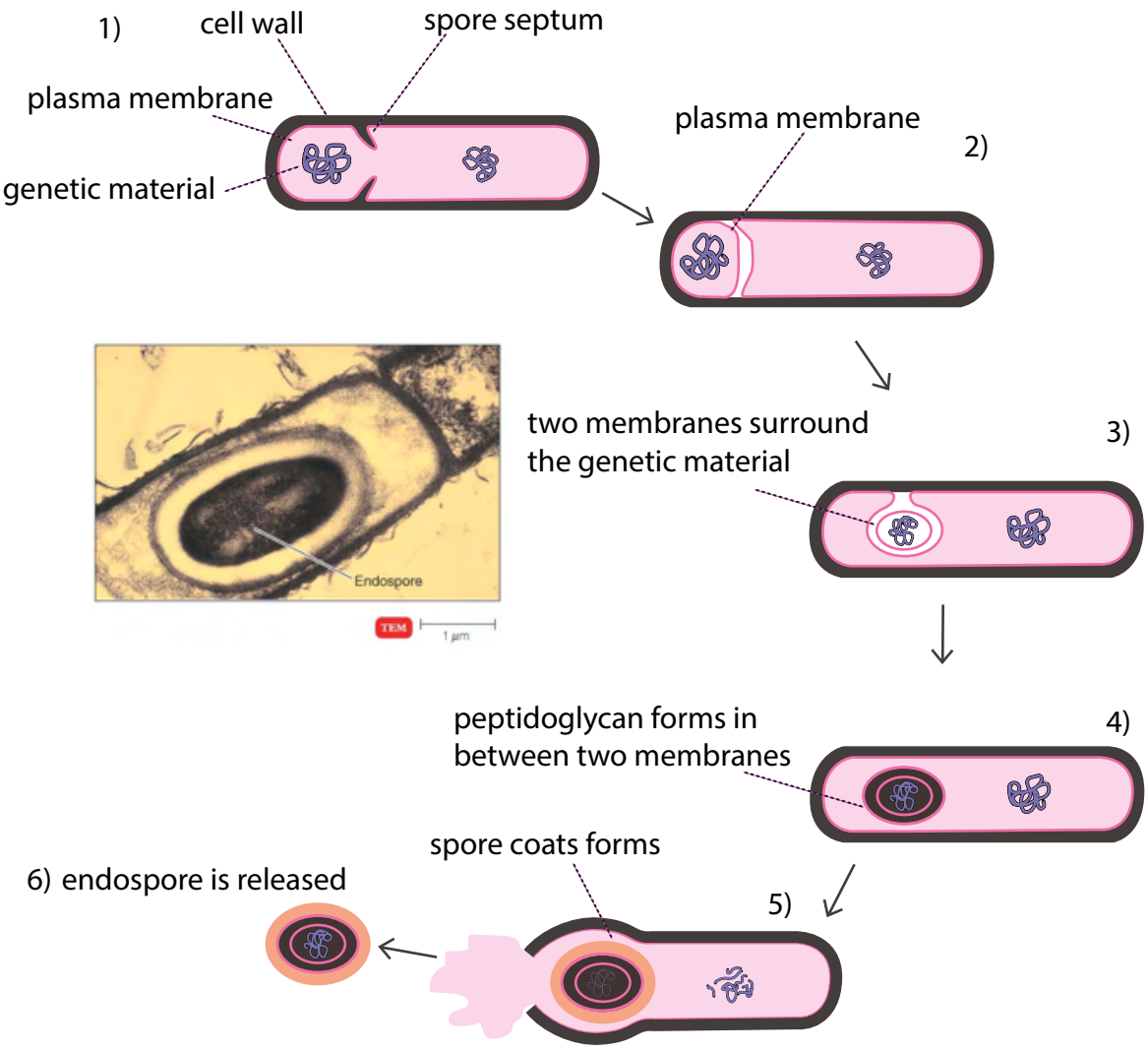


Figure 1.2 : Sporulation process. A drastic change of temperature, oxygen concentration and available nutrients, triggers the anthrax cell through the spore formation. Figure adapted from ako-spice.com/affected-by-nature-pepper/

Nevertheless, the spore interacts with the surrounding environment, most made through the more external layers of the spore, namely exosporium and spore coat. These external layers confer surface hydrophobicity, protecting the inner layers that consist of a thick cortex, a membrane and the core. Inside the core, the bacterial genetic information is stored in a dehydrated form, together with small acid-soluble proteins (SASPs) that bind and protect the nucleic acids, serving also as a source of energy for the later germination (Fig 1.2).

Spores of *B. anthracis* are renowned for being extremely resistant to irradiation, chemicals, desiccation, harsh pH and temperatures, surviving and persisting infectious

even for decades (Fig. 1.2). Notably, the vegetative form is quite fragile to environmental changes. Indeed, *B. anthracis* depends on (i) sporulation to survive against not favorable conditions and on (ii) host infection for multiplication and propagation of the species. These conditions identify it as the only obligate pathogen in the genus of *Bacillus* (Baron, S. et al. 1996).

1.4.1. *B. anthracis* epidemiology

The typical reservoir of *B. anthracis* spores is the soil, in which they can survive for long periods. Herbivores become easily infected during foraging in contaminated areas. Due to this ability to survive in the environment, independent of the presence of host bodies, *B. anthracis* eradication from a certain area is not easy and remains endemic in many countries. Mostly, humans become infected after being in contact with infected animals or animal products. More recently, the human anthrax is classified as industrial or nonindustrial depending if the contamination started from the contact with a contaminated animal product or directly with an animal. The industry form can be caused after handling contaminated hair, wool, bones, with high chances to contract the pulmonary infection as result of the inhalation of anthrax spores. Instead, in the nonindustrial form the contamination can happen for people working with animals or animal carcasses, like farmers, veterinarians, butchers (Turnbull et al., 2008).

1.4.2. *B. anthracis* host infection

Independently from the way of access in the host body, anthrax spores are rapidly phagocyted by antigen presenting cells (APCs, as macrophages), which migrate then to the near lymph nodes.

During this process, anthrax spores are able to resist, survive, germinate into bacilli, multiply and escape the host immune system. This resistance is due to the natural spore composition and by the action of LF (chapter 1.7.2) (Jang et al., 2011). *B. antharcis* cortex is composed by peptidoglycan (murein), which prevents the full

phagocytosis to be completed. In addition, the presence of the capsule, composed of poly-γ-D-glutamic acid (PGA) contributes to that resistance. Furthermore, the weakly immunogenic and antiphagocytic PGA capsule disguises the bacilli from host immune surveillance, letting them replicate, thus invading the bloodstream (Jang et al., 2011).

1.4.3. Host defence

Some species like chickens and dogs are quite resistant to anthrax infection. Others, like herbivores as cattle, sheep and horses are much more susceptible. Humans are defined as being intermediate susceptible.

Immunity protection against anthrax requires the development of antibodies against one of the anthrax toxin components. Both of the so called noncellular human vaccine and the live-spore animal vaccine induce the production of antibodies specific to PA. Instead, the PGA component of the capsule, is poorly immunogenic. Antibodies specific to polysaccharide and other components of the cell wall, are not protective.

1.4.4. Anthrax spores germination

B. anthracis recognizes specific signals coming from the host local environment as the presence of specific amino acids and nucleosides combinations which are recognized by the family of *gerA*-like sensor operons, leading the spore to germinate. The germination process is believed to begin with the activation of the germinat sensor followed by the re-hydration of the core, elimination of cations and dipicolinic acids, cortex breaking as well as metabolism reactivation and toxins synthesis (Liu et al. 2004).

1.5. Pore-forming toxins (PFTs)

Pore forming proteins are involved in the molecular offensive and defensive system of cells, typically present in bacteria and vertebrates (Bischofberger et al. 2012). PFTs are produced and released as virulence factors from a number of pathogenic bacteria. Their role is to hijack the host immune system allowing a free reproduction and colonization of the pathogen. In addition, PFTs represent essential elements used by the innate immune response against pathogens (e.g. the membrane attack complex (MAC) (Dukelberger & Song 2010 and Bischofberger et al. 2012), perforins (Pipkin et al 2007) and regulate signaling pathways (e.g. B-cell lymphoma-2 (Bcl-2) (Garcia-Saez 2012). In comparison, Bcl-2 proteins are involved in the regulation of the inflammasome response during pyroptosis (Shi et al 2019). In the same pathway, the tripartite pore forming toxins (PFT), non-hemolytic enterotoxin (Nhe) and hemolysin BL (Hbl) from the human pathogen *Bacillus cereus,* induce activation of the NLRP3 inflammasome leading to pyroptosis of the host cells. (Xu. H et al., 2014; Wilson et al. 2019) (Fig. 1.3).

Generally, PFTs can enter a cell via two pathways. In a first pathway, PFTs can assemble into a complex able to punch and penetrate the cellular membrane. The broken compartmentalization leads to a disequilibrium of influx and efflux of ions (K^+ and Ca^{2+}), ATP and other molecules that ultimately bring the cell to death. In the other path, the toxin in its pore state translocates an active enzyme through the cellular membrane of the target cell (e.g. anthrax toxin, diphtheria toxin, shiga toxin) (Fig. 1.3). This mechanism is typical for (i) AB-toxins (Odumosu et al. 2010), as well as for the recently characterized (ii) ABC toxins (Meusch et al. 2014; Roderer & Raunser 2019).

PTFs vary in size, like heptamers or octamers as in the PA of anthrax or even bigger like in the cholesterol-dependent cytolysins (CDCs) that form 30-50-meric pore complexes (Peraro et al. 2016).

1.5.1. AB toxins

AB toxins (or type III toxins) are two-component protein complexes released from many bacteria in the extracellular environment, commonly aiming to an intracellular target (e.g. anthrax toxin secreted from *B.anthracis*, chapter 1.7). The name derives from the function of each component. Usually, the active part is defined as A component, whereas the binding part the B component. The AB translocation inside the target cell follows a receptor mediated endocytosis mechanism. The A component has an enzymatic activity and it is transferred to the cytoplasm via a conformational rearrangement of the membrane-bound transport B component. Some members of the AB toxins adopt a syringe-like mechanism as a way to translocate the A moiety through the endosomal membrane, finally reaching their target (e.g. the conformational rearrangement can be triggered by the acidic pH of the endosomes, as in the case of anthrax toxin) (Jiang et al., 2015).

PFTs are classified into two groups, namely α-PFTs or β-PFTs, according to the secondary structural elements of their transmembrane domains in the pore conformation. As the name suggests, the pore forming domain of α-PFTs are characterized by α-helices whereas β-PFTs are composed of β-sheets that form β-barrels (Iacovache et al., 2008). The members belonging to the α-PFTs show high differences at the structural level. Indeed, they include proteins mainly composed of α-helical structure elements (e.g colicins) or β-strand motifs but having a single helix responsible for membrane insertion (actinoporins) (Parker et al. 2005). Commonly, the membrane-spanning domain has a hydrophobic or amphipathic element that initiates the membrane insertion process. One of the few examples of structures resolved in their monomer and oligomer state is the fragaceatoxin C (FraC) (Tanaka et al., 2015). The pore-forming domains of β-PFTs are commonly composed by β-strands in both, monomer and oligomer conformations. The soluble monomers present amphipathic β-strands with small hydrophobic patches. After the oligomerization of monomers, the β-strands form hydrophobic membrane-spanning β-barrels that face the lipid bilayer, which present an essential polar core with a specific transfer of charged molecules (Peraro et al., 2016) (Fig. 1.3).

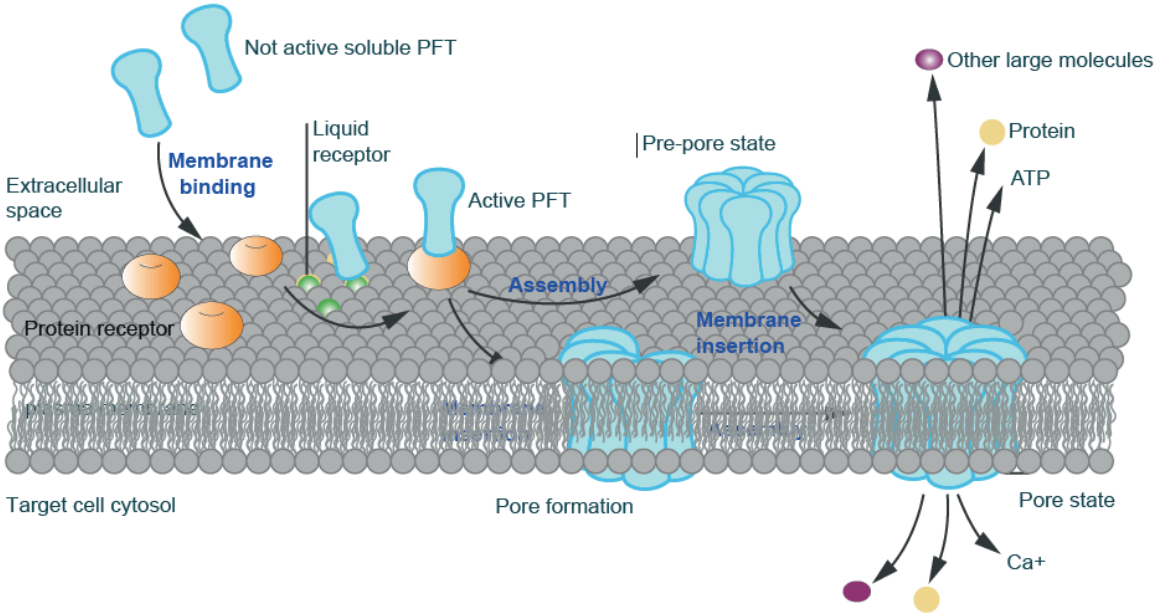


Figure 1.3 Scheme PFTs formation and mode of action. Single monomers bind their specific receptor, lipid or glycan present on the host cellular membrane. This binding promotes the accumulation of other monomers leading to a progressive oligomerization on the cellular membrane. The oligomerization is performed in two ways: (i) the monomers oligomerize in their pre-pore state on the membrane followed by formation of the pore. (ii) Otherwise, membrane insertion is triggered as a consequence of binding to a specific receptor or during oligomerization. Both pathways initiate the formation of a complex in a pore state. Figure adapted from Peraro et al. 2016.

1.5.2. ABC toxins of Photorhabdus luminescens

ABC toxins are multi-subunit toxins composed by A-, B- and C-component which oligomerize into an ABC-complex. The bacteria *P. luminescens* releases the Toxin Complex (Tc), as an ABC-complex with a size of 1.7 MDa (Gastogiannis et al., 2013). Other than TC toxin, *P. luminescens* releases also other three toxin classes: Photorhabdus Virulence Cassettes (PVCs), binary Photorhabdus-toxins (PirAB) and Makes caterpillars floppy (Mcf) toxins (Daborn et al., 2002).

1.5.2.1. Tc toxin complex

Tc toxin complex of *P. luminescens*, also called PTC3 holotoxin, translocates an enzymatic active protein through the transmembrane pore to the cell cytosol. TcA forms a 1.4 MDa pentameric bell-shaped shell which surrounds the translocation

channel. The translocation channel runs over the entire toxin complex, characterized by a long coiled-coil helix with a β-sheet domain at the top, forming the TcB-binding domain of TcA. TcB and TcC form a cocoon-like structure and its β-propeller domain connects with the TcB-binding domain of TcA. The toxin part of this supramolecular structure is represented by the TcB-TcC cocoon, which is autoproteolytically cleaved at the C-terminus of TcC, the so-called hypervariable region (hvr). The actual toxic component has not been resolved due to its highly flexible state the cocoon (Meusch et al., 2014; Gatsogiannis et al., 2018). During intoxication, the TcA component binds via its receptor binding domain to its unknown receptor present on the target membrane, inducing endocytosis of the holotoxin into the target cell. The maturation from early to late endosome state is characterized by a more progressive acidic pH, which triggers a conformational change in the TcdA pore (Meusch et al., 2014). At not acidic pH, an electrostatic lock (neuraminindase-like domain) seals the channel in the pre-pore state.

Each TcA-protomer has a linker domain which is stretched in the pre-pore state but gets folded in the pore-state. The pH shift leads the lock opening which induces the relaxation of a stretched linker and the subsequent insertion of the translocation channel into the membrane, followed by translocation of TcB-TcC cocoon (Fig. 1.4).

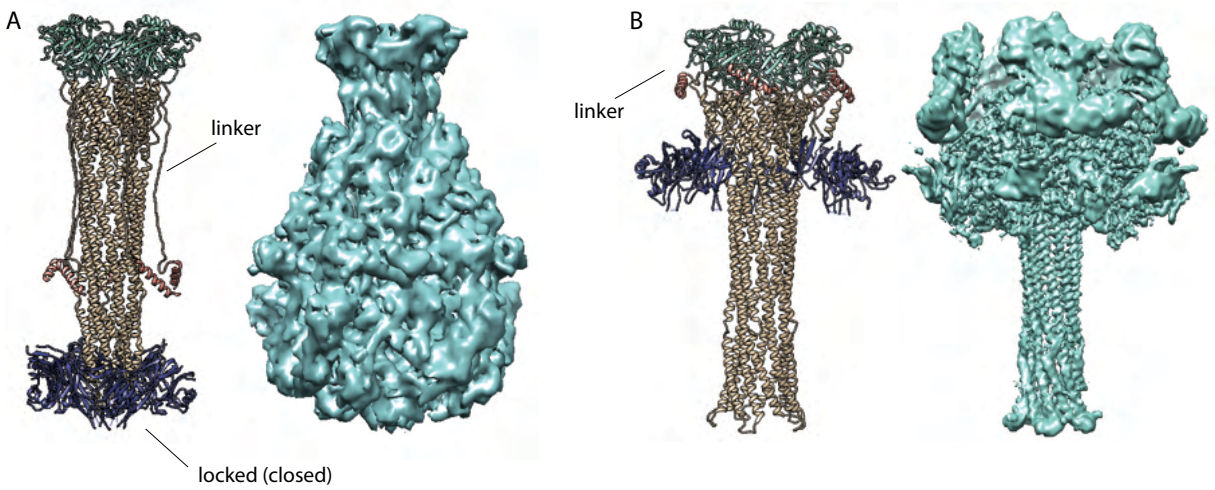


Figure 1.4 : Pentameric structure of TcdA in the pre-pore and pore state. (A) TcdA from *P. luminescens* in its bell- shaped prepore state (EMD-2297). The five different protomers interweaves with each other. In the center of this supramolecular structure the translocation channel is depicted in gold. The linker is stretched, whereas the lock region is closed (dark purple). (B) Pore-state TcdA after conformational

change. TcdA adopts a syringe-like structure with a vuvuzela-like translocation channel. The linker is folded up and the locked region open (dark purple) (EMD: 4068).

1.5.3.Mechanism of pore formation

The secretion of the soluble monomeric toxins (Bischofberger et al., 2012) or alternatively the secretion of the pre-pore complex in an outer membrane vesicle (OMV) represents the first step of the pore formation process (WAI et al. 2003). In order to enter the target cell, the monomeric (or oligomeric) toxins have to pass the hydrophobic or amphipathic environment of the plasma membrane bilayer of the target cell. For this purpose, the monomer assembles in higher oligomers forms to shield the host plasma membrane. Initially, PFTs bind to specific receptors such as proteins, glycans or lipids on the target cellular surface. The binding event of the soluble form to the relative cellular receptor can cause a conformational rearrangement of the toxin (Blewitt et al., 1985). In other cases, the monomeric structure needs to be activated by external factors (as a pH change) (Blewitt et al. 2005), by a catalytic cleavage at the N- or C- terminus or other factors (Parker et al., 2005), which mediate the membrane binding or the transition from monomeric to protomeric pore state. Afterwards, the pore formation happens following two possible pathways (Fig 1.4). The bound and activated protomer intermediates oligomerize on the plasma membrane surface generating the pre-pore complex which after a conformational rearrangement generates the pore state. The membrane insertion follows in a concerted fashion (first pathway fig. 1.4, scheme followed mainly from the β-PFTs, as the anthrax toxin). Alternatively, the membrane insertion can happen just after membrane binding or during the oligomerization process (second pathway fig. 1.4, Cosentino et al., 2016). An example of one α-PFT adopting the second pathway is α-Xenorhabdolysin (Xax) from *Xenorhabdus nematophila* (Cosentino et al. 2016).

1.5.3.1. Xax toxin from *X. nematophila*

The Xax toxins have been first characterized as binary toxins in *X. nemathophila* bacteria (Ribeiro et al., 2003). Afterwards, the toxins were also found in the genomes of *P. luminescens* as well as plant pathogens (*Pseudomonas syringae*) and in some

human pathogen as *Yersinia enterocolitica* (Vigneux et al., 2007). The shortened name is given placing the origin species name in front (e.g. Xax or Pax).

Xax toxins form 1-1.3 MDa large pore complexes able to perforate the host cellular membrane. The proteins are encoded in the genome and released from the bacteria as two single polypeptide chains, named XaxA (45 KDa) and XaxB (40 KDa). The effective toxicity is observed when both toxins are expressed. Xax toxin mechanism lead to granulocyte, phagocytes and erythrocytes lysis (Vigneux et al. 2007). The cryo-EM structure and functional studies of XaxAB toxin revealed that both singular toxins are required to form the pore complex, in which they assemble following a heterodimeric subunit arrangement (Schubert et al. 2018). The heterodimers assemble and form complexes of 12-15 subunits, able to perforate the plasma membrane. Major structural rearrangements in XaxB, as the swinging out of the amphipathic helix are responsible for membrane insertion (Fig. 1.5, Schubert et al., 2018).

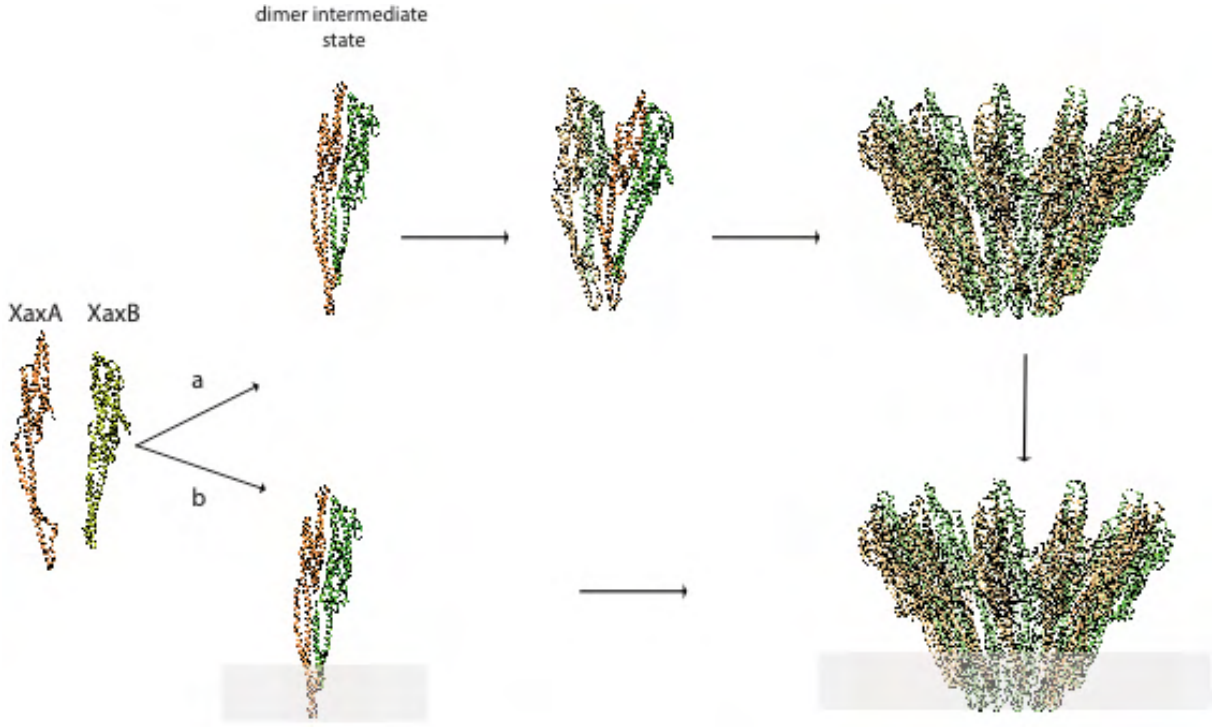


Figure 1.3 : XaxAB complex formation. XaxA and XaxB dimerize in solution followed by (a) either insertion in the plasma membrane and oligomerization or (b) oligomerization in solution and later insertion into the plasma membrane. The grey band mimics the plasma membrane.

XaxA acts as an activator of the oligomer complex. In addition, XaxA has a stabilizing effect on XaxB which forms the actual transmembrane pore (Schubert et al. 2018).

1.6. Anthrax spreading

B. anthracis is spread mainly among animals and in particular herbivores. Once the bacilli are released from the carcass to the environment, the increased level of oxygen leads to the conversion from the vegetative to the spore form. The spores might infect other animals at any time, after few hours or even decades (Baron et al., 1996).

Humans get in touch with the bacteria via infected animals or animal products. Generally, humans can contract anthrax via four different ways: (i) wound infection, (ii) injection, (iii) inhalation or (iv) ingestion of anthrax spores. Among these infection ways, the cutaneous one, in which the spores enter the host body through a skin wound, is the most common one and also the least severe. The anthrax infections contracted via this form have the lowest mortality rate, less than 1% when immediately treated with the proper antibiotics, thus representing the 95% of the total registered cases (Turnbull et al., 2008). In contrast, the inhalation and gastrointestinal anthrax infections are more difficult to detect and the mortality is higher. Among the different anthrax forms, the gastrointestinal one attacks the soft tissues and most of the antibiotic treatments have to be associated to surgeries aimed to remove the necrotic tissues. In the last century, fewer cases of anthrax infections for both animals and humans have been registered. In contrast, multiple outbreaks among drug-injection users and the 2001 terrorist attacks brought up the interest for this infection and the associated risks for humans (Friebe et al. 2016).

1.7. Anthrax virulence factors

The major virulence of *B. anthracis* is attributed to two plasmids pXO1 and pXO2. Both of them encode exotoxins and the biosynthetic machinery that is responsible for the biosynthesis of poly-D-γ-glutamic acid (PGA), the major component of the capsule. PGA is a linear polymer of low immunogenicity that prevents the bacteria cells to be

phagocytized thus invading the blood stream and possibly leading to final septicemia (Makino S. et al., 2002). The synthesis of the toxins and the production of the capsule are related to the partial pressure of the CO_2 and the environmental temperature. The carbon dioxide triggers to both transcription of the three genes coding for the toxins (*pagA, lef,* cya) and for the operon *capBCA* which is involved in the biosynthesis of the capsule (Chand et al., 2009). PXO1 encodes for the tripartite toxin, that includes two enzymes (i) LF and EF (A moiety), and the (ii) PA, (B moiety of the AB binary toxin) (Young et al., 2007). The single proteins of the tripartite toxin are not contiguously encoded on plasmid pXO1.

1.7.1.Protective antigen

The protective antigen (PA) is a receptor binding subunit composed of 735 residues. PA is folded in 4 functional domains, mainly organized as antiparallel β-sheets. Each domain is required for a particular step during the intoxication process. Domain 1 (residues 1-258) is characterized by a β-sandwich with a jelly roll topology, several small helices and the coordination residues of a pair of adjacent Ca^{2+} ions. PA activation via the furin like proteases (*in vivo*) or via trypsin (*in vitro*) cut off the so called "activation site" (PA_{20}) represented by a 20 kDa N-terminal fragment. Domain 2 (residues 259-487) shows a β-barrel as well as a large and flexible loop located between the strands $2\beta_2$ and $2\beta_3$ implicated in the membrane insertion. Domain 3 (residues 488-595) is the smallest of the 4 domains and is composed of 4 helices, 4 stranded mixed sheets and 2 smaller sheets. This domain is involved in the oligomerization step. Domain 4 (residues 596-735) begins with a hairpin and helix structure, followed by four β-sandwiches with an immunoglobulin-like fold which mediates the interaction with the receptor (Rainey et al., 2005). Domains 1, 2, and 3 have several deep interaction sites, while domain 4 establishes only few contacts with the other domains (Lacy et al., 2004) (Fig. 1.6).

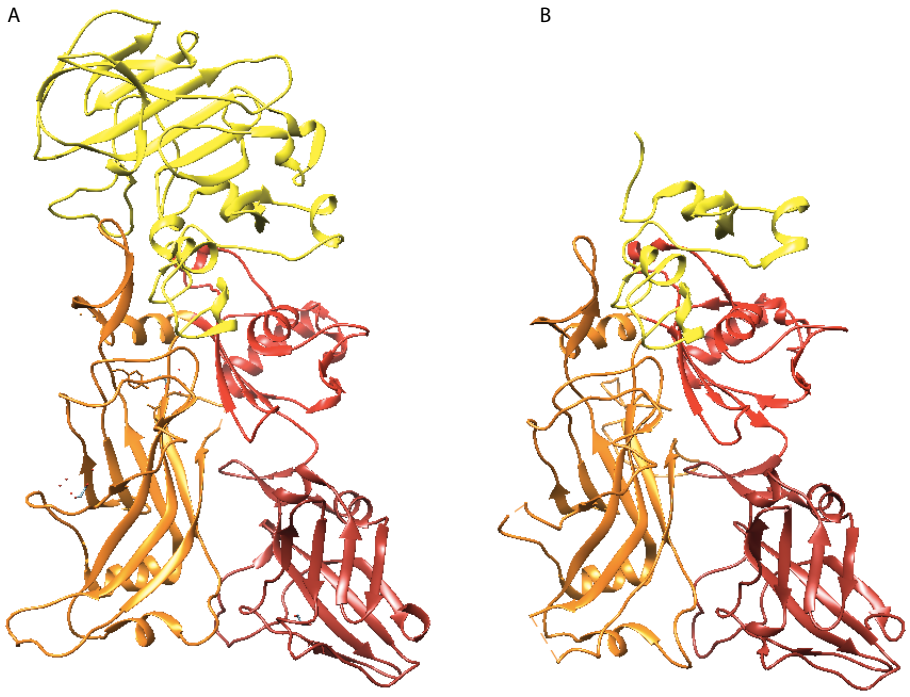


Figure 1.4 : 3D structure of PA_{83} (PDB; 4H2A) and PA_{63} (PDB; 1TZO). (A) PA in its full-length form (PA_{83}, MW: 83 kDa). Domain 1 (residues 1-258) is in yellow, domain 2 (residues 259-487) in orange, domain 3

(residues 488-595) in red and domain 4 (residues 506-735) in dark red. (B) PA after proteolytic cut from furin-like protease, removing 20 kDa from the N-terminal domain (MW: 63 kDa). Domain 1 (residues 174-258) is in yellow, domain 2 (residues 259-487) in orange, domain 3 (residues 488-595) in red and domain 4 (residues 506-735) in dark red. Protomer is extracted from the PA_7 complex.

B. anthracis secretes PA_{83} as preprotein form (MW: 83 kDa). The first check point of toxin host cell entry is the recognition and binding of PA to the host receptor (Santelli et al., 2004). Two human cell-surface receptors are involved in this process: (i) capillary morphogenesis protein 2 (CMG2) and (ii) anthrax toxin receptor/tumor endothelial marker 8 (ATR/TEM8). After receptor binding, proteolytic cleavage of PA induces its oligomerization to heptamers (PA_7) or octamers (PA_8) and leads to exposure of the binding site for EF and LF (Fig 1.7).

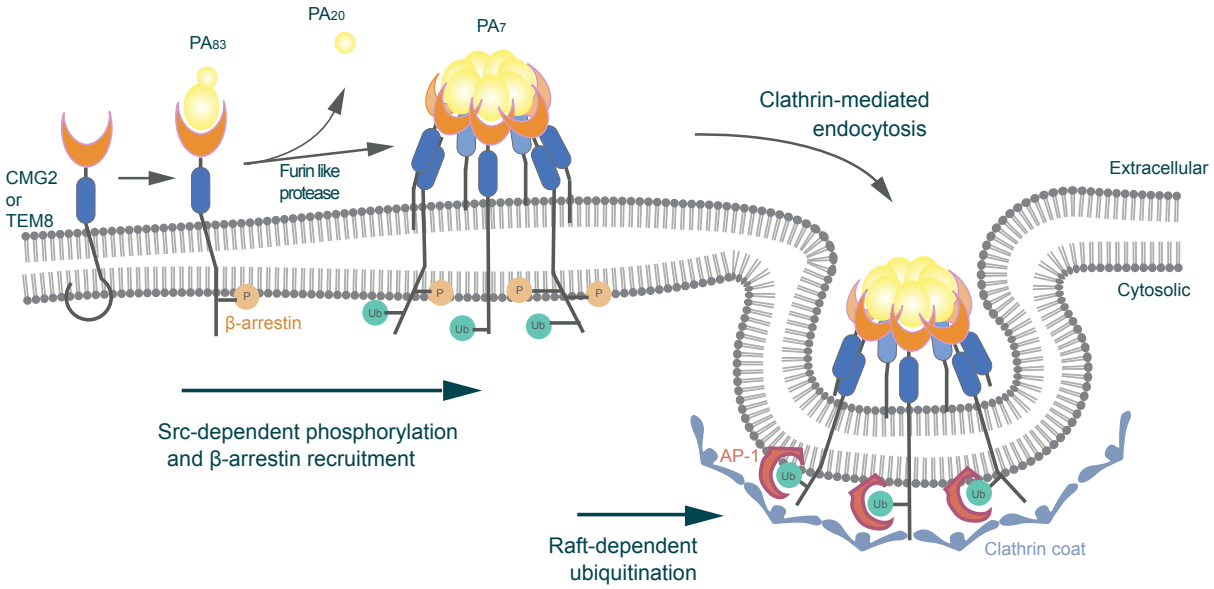


Figure 1.5 : Scheme representing the receptor mediated oligomerization of PA_{63} on the cellular membrane. PA_{63} is the only well characterized ligand of the CMG2 receptor. After the binding between the PA and the vWA domain of the receptor, src-like kinases are activated leading to CMG2 cytosolic tail by phosphorilation. Afterwards, the β-arresting is recruited allowing the E3 ubiquitin ligase to ubiquitinate the CMG2 receptor on its lysin (lys) localized near the cellular membrane. Ubiquitination leads to a clathrin-dependent endocytosis of the CMG2-PA_7, that is dependent on the actin cytoskeleton. Figure adapted from Deuquet et al., 2011.

Afterwards, the LFs and/or the EFs bind on top of the heptamer/octamer ring. Host cell receptors do not play any role in this interaction and the binding affinity between

PA and EF or LF is very similar independently if PA is bound or not to the receptor (Tonello et al., 2004). Followed by a more acid pH in the endosome, the structure rearranges its conformation resulting in a flower-like shape, named the pore state (Jiang et al., 2015 and Deuquet et al. 2011) (Fig. 1.7 and 1.8)

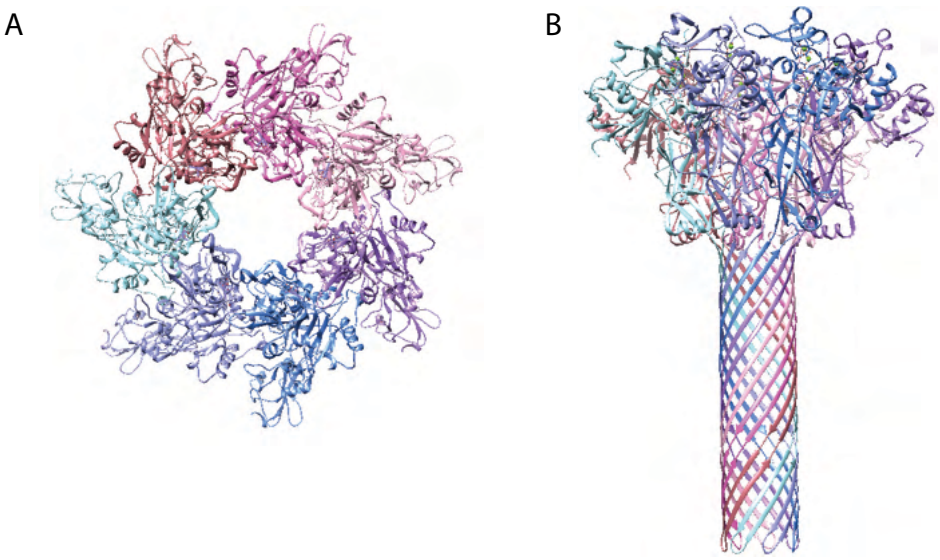


Figure 1.6 : PA_7 pre-pore and pore state. (A) Top view of PA_7 in the pre-pore state. The complex has a disk structure, with a central hole. (B) Side view of PA_7 pore-state, with a flower-like structure. The single protomers are highlighted with different colours.

1.7.2. Lethal Factor

The Lethal Factor (LF) is a 90 kDa protein composed of 776 amino acids organized in four domains, with mainly α-helices. Domain 1 is responsible of the binding to PA, whereas domain 4 (C-terminal) contains the catalytic site. A HExxH zinc-metalloprotease consensus (residues 686-690) is present at this terminus making LF belonging to the family of zinc-dependent metalloproteases (Pannifer et al., 2001). Mutations at this sequence level deprive LF of its activity and ability to bind to Zn^{2+}, proving that, as the clostridial and the tetanus toxins, LF is a zinc protease (Fig. 1.9). LF cleaves the N-terminus of most of the mitogen-activated protein kinases (MAPKKs), leading to their inactivation (Duesbery, N.S. et al 1998). MAPKK plays an intermediate role in the activation of MAPK signaling pathways. Thus, LF promotes macrophage death by disrupting the MAPK-dependent pathways that regulate prosurvival genes and by activating proteasome- and inflammasome-dependent pathways that activate caspase-1 (Jang et al., 2011).

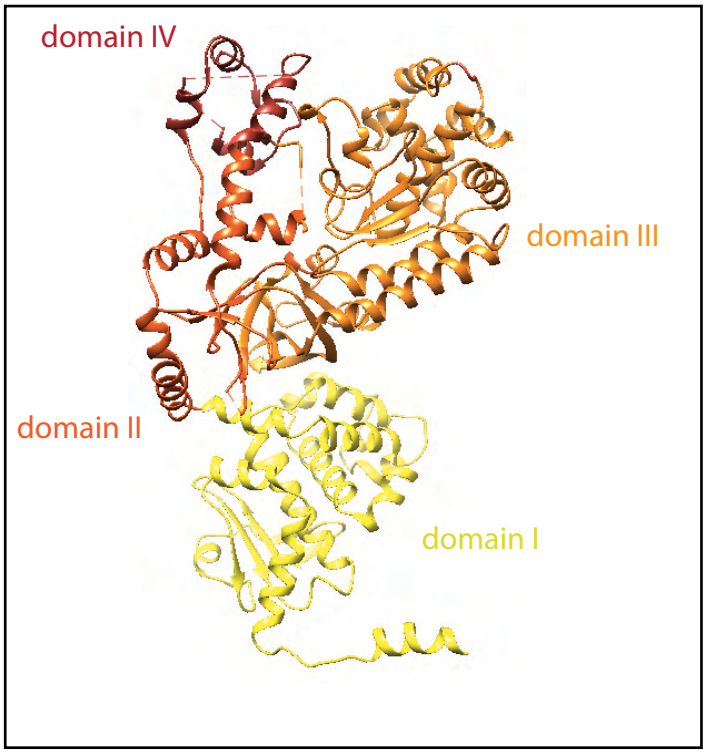


Figure 1.7 : Domain organization of LF (PBD: 1J7N). LF domains are highlighted in different colours.

1.7.3.Edema Factor

EF is a 90 kDa protein with a sequence composed of 767 amino acids. The N-terminal domain 1 is responsible of the binding to the PA showing high similarity to the LF domain 1. The C-terminal part of EF (EF_{58}) represents the catalytic site with three globular domains (1, 2, and 3) and three loops (A, B, and C) (Drum et al., 2002). The active site is located at the interface between domain 1 and 2 and these two domains grouped together are named the catalytic center (Fig. 1.10).

EF is a calmodulin-dependent adenylate cyclase that converts the ATP to cAMP, thus increasing the levels of cellular cAMP, which is the prototypic second messenger that regulates diverse cellular responses. The biological effects of cAMP are mediated with the subsequent binding of cAMP to three families of signal transducers: (i) cAMP-dependent protein kinases, (ii) cyclic nucleotide gated channels and (iii) EPAC (the guanidine exchange factor for Ras GTPase homologs Rap1 and Rap2 (Tang W. et al., 2009).

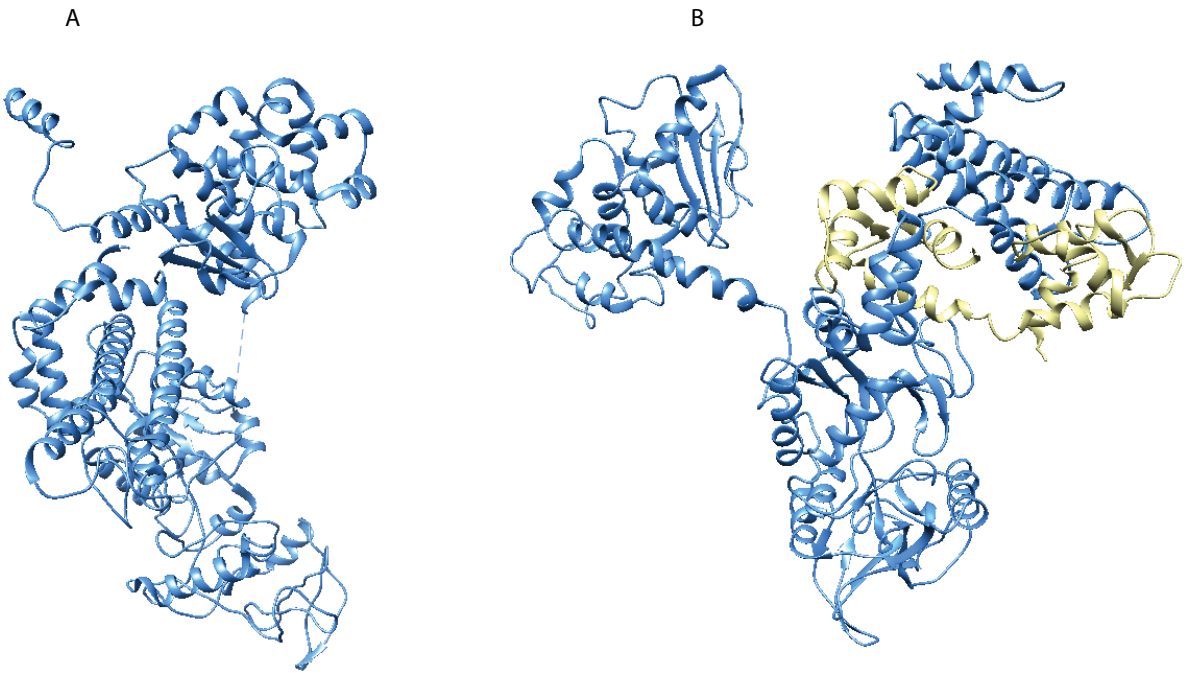


Figure 1.8 : Crystal structures of EF (PDB: 6UZB) and EF-CaM (PDB; 1XFY). (A) Elongated shape of the EF free from any binding. (B) EF in complex with calmodulin (CaM). The structure arrangement of the different domains change, resulting in a Y shape. EF is depicted in blue and CaM in yellow.

Contrary to LF, the 3D structure of EF varies a lot depending if it is bound to CaM (active state) or unbound (inactive state). In order to bind to the CaM, domain 3 rotates and exits from the catalytic center, until it forms a cove between domain 1 and 3, surrounded by CaM. CaM bound to EF assuming an elongated conformation, quite different from the one assumed when bound to the other cellular targets (Shen et al., 2004).

1.7.3.1. CaM and Calcium bound state

CaM is a small soluble protein of 16.5 kDa, largely available in cells and represents 1% of the total protein amount. It is mainly composed of two globular domains, the N-terminal (N-CaM) and the C-terminal (C-CaM) domain which are connected by a flexible α-helix. Each of these two globular domains coordinates two Ca^{2+} ions. The function of CaM is to bind and modulate, in a Ca^{2+}-dependent manner, the activity of a series of enzymes as phosphodiesterases, adenylate cyclases and others (Shen Y. et al. 2005, Ulmer et al 2003). EF is activated by CaM, becoming 1000 more efficient than

mammalian adenylyl cyclase (Friebe S, et al., 2016). The binding to EF and its final activation does not follow exactly the common mechanism of the mammalian adenylate cyclase. Indeed, depending on the presence of Ca^{2+}, CaM adopts a different conformation: more elongated and hydrophobic or compacted and hydrophilic. In the elongated conformation, the structure exposes the residues involved in the binding with the targets (as during the binding with the mammalian adenylate cyclases). However, the binding with the EF takes place when the CaM adopts the compact state, without binding the Ca^{2+} (Shen et al., 2004). This interaction is initially quite weak but launches the opening of the structure of CaM bound to EF. The atomic structure of the CaM-EF complex does not show any Ca^{2+} ion bound to the calcium binding site of the CaM with 1µM Ca^{2+} in the solution. With 1mM Ca^{2+} only one Ca^{2+} binds CaM (Ulmer et al. 2003) (Fig 1.10).

There are two major toxic mechanisms by which bacterial toxins unbalance the intracellular level of cAMP. First, by the action of the bacterial adenylyl cyclase (AC) toxins. These toxins have AC activity and convert ATP in cAMP, and this mechanism becomes active only upon their entrance into host cells and association with specific cellular proteins that act as activators. A high concentration of active AC into the bacterial cell is dangerous similar to mammalian cells. Therefore, these enzymes are released in their inactive state and only after binding with endogenous factors, they become active. Examples of AC toxins are EF from *B. anthracis,* CyaA from *B. pertussis* and ExoY from *P. aeruginosa* (Tang W et al 2009). EF and CyaA share CaM as a common activator while ExoY has F-actin as activator (Belyy et al. 2021).

The second mechanism that leads to an increased level of intracellular cAMP is the ADP-ribosylation of heterotrimeric G proteins by bacterial toxins, which increases the membrane-bound AC (mAC) catalytic activity. An example is the cholera toxin from *Vibrio cholerae* that binds first to the human partner protein ADP-ribosylation factor 6 (Arf6). The binding induces a conformational rearrangement that leads to the exposition of its active site and therefore its catalytic activity. In the presence of NAD^+, the cholera toxin catalyses the ADP-ribosylation of the Gs α-subunit (Gs-α), leading to an increase of the active AC and then of cAMP, manifesting severe diarrhea (Raufman JP et al., 1998).

1.8. Anthrax receptors

So far, two receptors have been identified to bind to the PA oligomer, (i) ATR/TEM8 (Bradley et al., 2001) and (ii) CMG2 (Scobie et al., 2003). The first one, derives its name from its overexpression in colon tumor tissues. The second one is located almost everywhere, and studies showed that it plays a major role in the cellular adhesion on the extracellular matrix (Hanks et al., 2003). Despite their involvement in anthrax toxin infection, their physiological ligands and their exact functions are still not totally uncovered. Due to their similarity to integrins, they could bind proteins of the extracellular matrix (ECM), as the fibronectin (Friebe S. et al., 2016). Other common aspects between the two receptors are their interaction with collagen: TEM8 with collagen VI while CMG2 with collagen VI and laminin. The receptors present the same general architecture composed by: (i) N-terminal region of the two von Willebrand A domain (vWA), characterized by 60% identity with similarity to the domain 1 of the integrins, (ii) an Ig-like domain, (iii) a single transmembrane helix and (iv) a cytoplasmic tail (Frebe S. et al., 2016). The vWA domain has a so-called, metal ion-dependent adhesion site (MIDAS) that mediates the interactions with the domain 2 and 4 of PA (Scobie H.M, et al., 2012). Despite the vWA domain is highly conserved in both receptors, the affinity of TEM8 to the PA is 10- to 1000-fold weaker, in cell cultures and *in vitro*, respectively (Liu S., et al., 2009).

Interestingly, depending on which receptor PA is bound to, the pH at which the conformational rearrangement happens is different. Indeed, when TEM8 is involved in the formation of PA-receptor complex, the conversion to the pore state undergoes at near to neutral pH values. Instead, when PA binds to CMG2, the pore-state is induced at more acidic pH. Consequently, the translocation of the active component of the toxin (EF or LF) takes place at different time points of the endocytic process. Indeed, early endosomes are characterized by pH>6, whereas a pH value of 5.5 indicates late endosomes (Fig. 1.11).

Once the receptor-toxin are assembled, they cluster in “lipid rafts”. These regions of the plasma membrane are essential for the correct internalization of the oligomer into the cell. Indeed, a disruption of the “lipid rafts” is translated in a defect of the oligomer internalization (Abrami L. et al., 2006). The localization of the receptor and its sub-

sequential internalization are in close relation to the palmitoylation of the receptor tail. Indeed, the palmitoylation keeps the receptor out from the lipid rafts, while after palmitoylation the receptor is rapidly internalized (Abrami et al., 2006) (Fig 1.11).

Another important modification that affects the receptor-oligomer complex is the ubiquitination. Indeed, after PA binds to the receptor, the adaptor protein β-Arrestin2 that is involved in the endocytosis, binds to the receptor tail (Moore C.A. et al., 2007). After E3 ubiquitin enzyme is recruited and it promotes ubiquitination of the receptor (Abrami L. et al., 2010), the E3 ligase of TEM8 is Cbl while the one of CMG2 has not been identified yet.

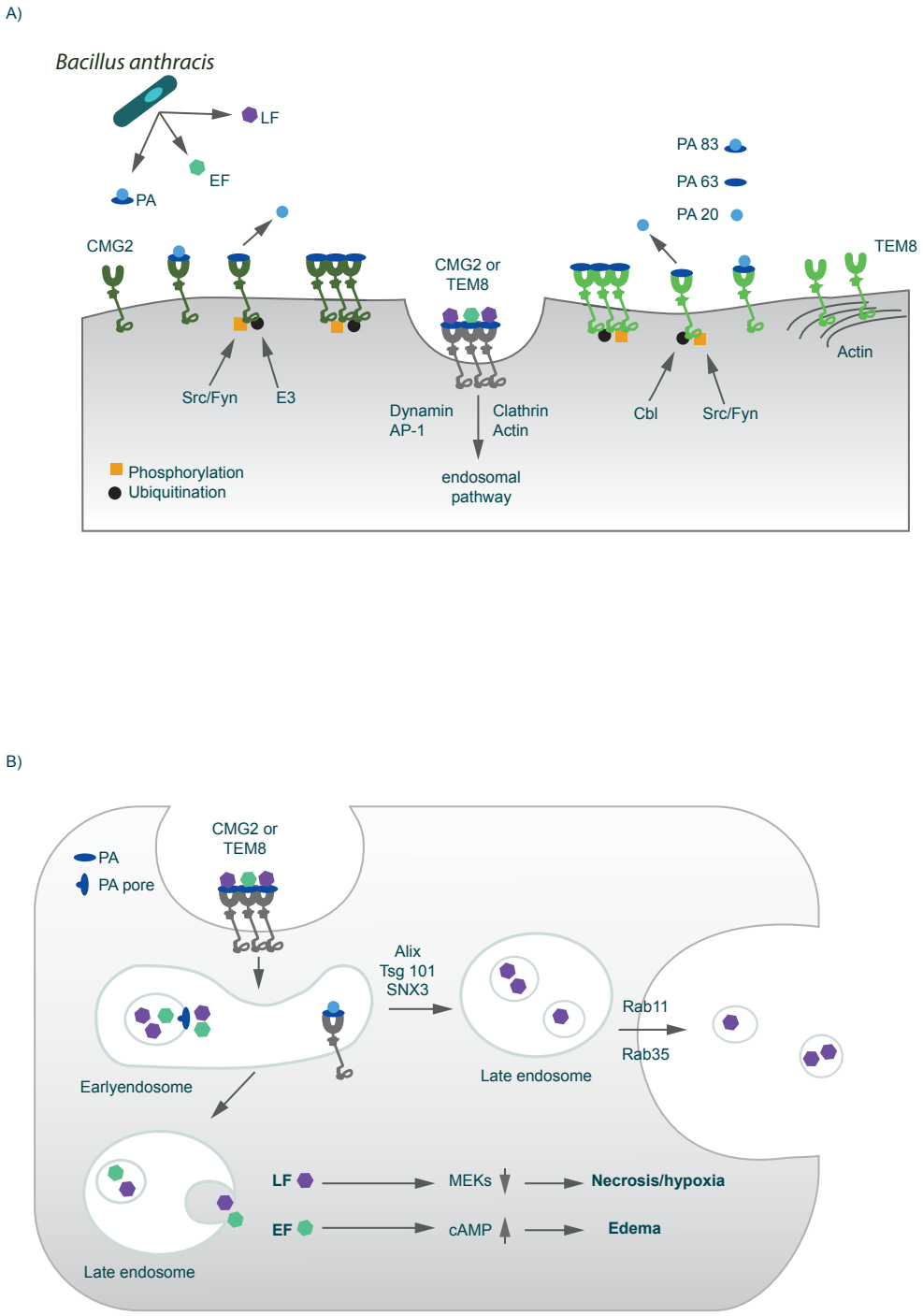


Figure 1.9 : Schematic representation of cellular entry of anthrax toxin, highlighting the endocytic pathway and the final cellular effects. (A) *B. anthracis* releases the components of the tripartite toxin, the PA_{83}, LF and EF. PA_{83} binds to either CMG2 or TEM8 receptor present on the host cellular surface. Subsequently, the furin like protease cleaves 20 kDa off from the PA N-terminus leaving a PA_{63} receptor-associated. After toxin binding, the receptor is phosphorylated on their cytosolic part by Src-like kinases, named Src and Fyn. Furthermore, the receptors are ubiquitinated by the Cbl (for TEM8) or an unknown E3 ligase (for CMG2). The receptor and the related PA_{63} oligomerize in lipid rafts, and the toxin-receptor complex is internalized via receptor mediated endocytosis, which depends on the presence of clathrin, dynamin, AP-1 and actin. (B) The lower pH present in the early endosome leads PA to undergo a

conformational change, that results in a toxin-pore formation. Here, LF and EF are translocated through the membrane into the cytosol via the PA-pore. Once in the cytosol, LF cleaves MEKs and thereby leads the cell to necrosis and hypoxia. EF causes the increase of the cAMP level and causes edema. LF can also be packed into the ILVs, which are released as exosomes, infecting neighboring cells (Abrami et al., 2013). Figure adapted from Friebe S. *et al* 2016, *Toxins*.

1.9. Anthrax complex toxin assembly

Once the vegetative form of *B. anthracis* infects the host, it releases the PA, LF and EF. The toxins reach the target cellular surface, where PA binds to CMG2 or TEM8 receptor. After cleavage of 20kDa by a furin-like protease, PA oligomerizes (PA_7 and/or PA_8). LF and EF competitively bind to the PA_7 or PA_8 surface. Differently than supposed in the past (Tonello et al., 2004), this binding involves and occludes two subunits of the PA oligomer for every LF or EF. Therefore, in the heptameric oligomer not more than three LF and/or EF can bind at once, thus mismatching the complex symmetry (Fabre et al., 2016). Instead, a total of four LF/EF can bind to PA_8. In this case, the symmetry pass from C8 (PA_8) to C4 (PA_8LF_4).

1.10. Anthrax toxin internalization and translocation

As mentioned in the earlier sub-chapters, after anthrax toxin assembling, the complex is internalized by receptor-mediated endocytosis. The progressive decreasing of the pH passing from early to late endosome triggers the conformational rearrangement of the PA oligomer to a pore state, which is defined by an extension of the PA β-barrel and penetration of the endosomal membrane (Fig. 1.8). The funnel-like pore state structure has a narrow channel that allows the translocation of only fully unfolded polypeptides (Jiang et al., 2015). The translocation event is triggered by several crucial aspects: (i) the intrinsic characteristics of the pore which is watertight, with an affinity to neutral and positively charged residues, (ii) low endosomal pH destabilizing the LF and/or EF three-dimensional structure promoting structure unfolding and (iii) the ΔpH from the endosome and the cytosol environment that drives the unidirectional translocation through the channel in accordance with the

electrostatic ratchet generating force. Ion selectivity is of extreme importance during translocation.

PA α-clamp (helix stabilizing cleft) may nucleate the helical structure of the LF/EF into the channel, where the ϕ-clamp grips the amino-terminal leader (Wynia-Smith et al., 2012). A polypeptide chain can pass through the anion-repulsive charge filter of PA channel, once is partially protonated by the lower pH typical of the late endosomes. Such changes in the protonation state may also happen directly into the channel, because the residues at the beginning of the PA channel are also acidic. At this stage, the channel acts like a ratchet holding the chain in a way that limits the retro-translocation (Wynia-Smith et al., 2012). The transition from helix to random coil is highly favorable entropically, and thus should thermodynamically drive the chain translocation from inside to outside the channel (Wynia-Smith et al., 2012). The ϕ clamp site is involved in a coordination of peptide movement or protonation states (Krantz et al., 2006). The deprotonated sequence is unable to re-translocate due to the charge-clamp site. Further cycles complete the translocation of the remaining domains of LF/EF (Wynia-Smith et al., 2012).

1.11. The role of anthrax toxin during infection

Once in the blood stream, the anthrax toxins play fundamental roles during the all infection states. *B. anthracis* could be successfully treated with specific antibiotics (Mayo clinic family health book, 5th edition). Despite the antibiotics therapy, once the bacteria release the toxins those could potentially still cause host death still. LF targets a number of tissues presenting CMG2 receptor (i.e. cardiomyocytes, vascular muscle smooth cells) causing hypoxia-mediated toxicity and ultimately leading to host death. On the other hand, the EF has a direct mode of action on hepatocytes, causing significant liver edema. Additionally, the EF-induced skin edema is not associated to direct targeting of endothelial cells but is probably due to an increase of water influx into the interstitial compartments either through ion channels involving cAMP-dependent process, or through CFTR (Bischofberger et al., 2012).

1.12. Therapy

Ciprofloxacin, doxycycline or levofloxacin are commonly used against human anthrax infection. Depending on the infection way and on the age and other factors (e.g. overall health state of the person) the treatment is given as single antibiotic or combination of antibiotics (Mayo clinic family health book, 5th edition). The main problem of an antibiotic treatment is related to the low efficiency of the antibiotic when administered at the early stages of infections. Once the symptoms are expressed, the bacteria count is high and the toxins have already been released. For that reason, it is necessary to couple an antibiotic treatment with toxin inhibitors. Recently, the recombinant human monoclonal antibody raxibacumab has been introduced as treatment in combination with intravenous antimicrobials providing a meaningful benefit compared to the antibiotic treatment alone. This monoclonal antibody acts on the PA, inhibiting the binding to the natural cellular receptor and follow-up translocation of the LF and/or EF into the target cell (Kummerfeldt, C. E. et al., 2014). A similar mechanism of action was demonstrated for Anthrasil, an immune globulin solution of purified human immune globulin G (IgG) containing polyclonal antibodies that bind the PA. It binds on PA and prevents the LF and/or EF to enter the cell (https://www.fda.gov/vaccines-blood-biologics/approved-blood-products/anthrasil). In addition, many studies successfully indicated other synthetic compounds as LF (Goldeberg et al., 2016 and Yuan et al., 2010) and EF (Jiao et al., 2018) inhibitors.

Aim of the thesis

The anthrax toxin released by the *B. anthracis* has been studied for decades, trying to understand its biological action, more recently contextualized with structural analysis. The structure of the anthrax pre-pore complex (Petosa et al., 1997), bound to the N-terminal domain of LF (Feld et al., 2010) or to its natural receptor (Lacy et al., 2004) have been solved by x-ray crystallography, while the pore state of the PA was recently solved by cryo-EM (Jiang et al., 2015 and Katayama et al., 2010). Even if many structural studies have been recently published, information regarding the real complex arrangement of the anthrax lethal complex (PA_7LF_3) were still missing.

This thesis had as aim to uncover the structural arrangement of this asymmetric complex via cryo-electron microscopy (cryo-EM). Initially, functional analysis were planned to verify the correct assembling and functionality of the homocomplex, PA_7. Therefore, the complex was exposed to lower pH values, thus inducing the conformational rearrangement of the PA_7. With the aim to stabilize the complex pore state, nanodiscs or liposomes at different sizes were planned to be used as substitute of biological membranes. To determine the structure of the fully decorated lethal complex, PA_7LF_3 was planned to be induced testing several ratio between PA_7 and LF. Finally, the structural analysis of the lethal complex was conducted via negatively stained and cryo-EM. The determined structures enabled me to analyze the domain organization of the toxin gaining insight into the molecular mechanism of toxin translocation.

2. Materials and Methods

2.1. Materials

2.1.1. Instruments

Table 2.1. Instruments used in this work

Instrument	Manufacturer
Analytical Balance, ABS 320-4N	Kern & Sohn GmbH (Balingen, DEU)
Analytical Balance, PC 440 deltra range	Mettler Toledo (Columbus, USA)
Centrifuge Allegra X-15R; X-30R	Beckman Coulter (Brea, USA)
Centrifuge Avanti J-26S XP, Rotor: JLA 8.1000	Beckman Coulter (Brea, USA)
ChemiDoc MP Imaging System	Biorad (Munich, DEU)
Cryoplunge Cp3	Gatan Inc. (Pleasanton, USA)
DeNovix DS-11 spectrophotometer	Biozym Scientific GmbH (Hessisch Oldendorf, DEU)
FEMTO Plasma Cleaner	Diener electronic GmbH & Co. KG (Ebhausen, DEU)
GloQube	Quorum Technologies (Lewes, UK)
Hamilton Syringe (10 µl, 50µl, 100 µl, 1000µl)	Avanti Polar Lipids (Alabaster, USA)

High vacuum coating system MED020	BalTec (Pfäffikon, CH)
JEOL-JEM-1400 Transmission electron microscope	JEOL GmbH (Eching, DEU)
JEOL-JEM-3200 Transmission electron microscope	JEOL GmbH (Eching, DEU)
FEI Arctica	FEI, ThermoFisher Scientific GmbH (Hillsboro, USA)
FEI Spirit Transmission electron microscope	FEI, ThermoFisher Scientific GmbH (Hillsboro, USA)
FEI Titan KRIOS	FEI, ThermoFisher Scientific GmbH (Hillsboro, USA)
Leica EM ACE600	Leica Microsystems (Wetzlar, DEU)
LIQUIDATOR96 96 channel pipetting system	Steinbrenner Laborsysteme GmbH (Wiesenbach, DEU)
Magnetic stirrer	Bellco Biotechnology, Inc. (Vineland, USA)
Microfluidizer M-110-S	Microfluidics, (Newton, MA, USA)
Mini-Protean 3 SDS-PAGE gel electrophoresis system	BioRad Laboratories (Hercules, USA)
Peristaltic Pump Cyclo II	Carl Roth GmbH + Co. KG (Karlsruhe, DEU)
pH meter FiveEasy FE20 basic	Mettler Toledo (Columbus, USA)
Photometer 6131	Eppendorf (Hamburg, DEU)

Pipettes	Eppendorf (Hamburg, DEU)
Superose 6 Increase 5/150 GL	GE Healthcare Biosciences (Pittsburgh, USA)
Superdex 200 Increase 5/150 GL	GE Healthcare Biosciences (Pittsburgh, USA)
Superose 6 Increase 10/300 GL	GE Healthcare Biosciences (Pittsburgh, USA)
Superdex 200 Increase 10/300 GL	GE Healthcare Biosciences (Pittsburgh, USA)
Tabletop Centrifuge 5415 D	Eppendorf (Hamburg, DEU)
Tabletop Centrifuge 5424/ 5424 R	Eppendorf (Hamburg, DEU)
Tabletop Centrifuge 5427 R	Eppendorf (Hamburg, DEU)
Temperature Controlled Cell TCC-100	Shimadzu (Kyōto, JPN)
Thermomixer comfort	Eppendorf (Hamburg, DEU)
Thermomixer F1.5	Eppendorf (Hamburg, DEU)
Ultracentrifuge Optima XPN-80, Rotor: Type 45 Ti	Beckman Coulter (Palo Alto, US)
Ultrasonic Bath Sonorex Super RK 255/H	Bandelin (Berlin, DEU)
UV-Vis Recording Spectrophotometer, UV-2401 PC	Shimadzu (Kyōto, JPN)
Vacuum pump	Ilmvac/Fisher Scientific (Ilmenau, DEU)

LABORATORY VACUUM PUMPS	KNF Neuberger, Inc. (Trenton,USA)
Vitrobot	FEI ThermoFisher Scientific GmbH (Hillsboro, USA)
Vortex shaker Genie 2	Scientific Industries (New York, USA)
Water bath, TW2	Julabo GmbH (Seelbach, DEU)

2.1.2. Chemicals

Table 2: Chemical used in this work

Chemicals and Solutions	Manufacturer
Ampicillin	GERBU Biotechnik GmbH (Heidelberg, DEU)
Carbenicillin	AppliChem (Darmstadt, DEU)
Coomassie Brilliant Blue R250	Merck KGaA (Darmstadt, DEU)
DMSO	Sigma-Aldrich Chemie GmbH (München, DEU)
HEPES	GERBU Biotechnik GmbH (Gaiberg, DEU)
Hydrochloric acid	Applichem GmbH (Darmstadt, DEU)
Imidazole	Carl Roth GmbH + Co. KG (Karlsruhe, DEU)
IPTG	Sigma-Aldrich Chemie GmbH (München, DEU)

1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-dioleoyl-sn-glycero-3-phosphoserine (DOPS), 1,2-Dioleoyl-sn-glycero-3-phosphat (DOPA), Brain Polar Lipids (BPL)	Avanti Polar Lipids (Alabaster, USA)
Sodium chloride	Carl Roth GmbH + Co. KG (Karlsruhe, DEU)
Sodium dodecyl sulfate (SDS)	Boehringer
TCEP	Biosynth (Staad, CH)
Uranyl formate	Polysciences, Inc. (Warrington, USA)

2.1.3. Buffer solutions

Table 3: Media used in this work

Media		
LB media (1 l)	10 g	Bacterial Tryptone
	5g	Yeast Extract
	10 g	NaCl
SOC media (1 l)	20 g	Tryptone
	5g	Yeast Extract
	0.58 g	NaCl

	0.19 g	KCl
	2.03 g	$MgCl_2 \times 6 H_2O$
	2.46 g	$MgSO_4 \times 7 H_2O$
	40 ml	50 % glucose solution
M9 (1 l)	200 ml	5 x M9 Salze
Minimal media		
	1 ml	1M $MgSO_4$
	8 ml	50% Glucose
	1 ml	0.1 M $CaCl_2$

Table 4: Buffers used for protein purification in this work

Purification buffers		
Ni-IDA washing buffer	20 mM	Tris-HCl pH 8.5
	500 mM	NaCl
Imidazole elution buffer	500 mM	imidazole
	20 mM	Tris-HCl pH 8.5
	500 mM	NaCl
Dialysis buffer	50 mM	Tris-HCl pH 8.5
	150 mM	NaCl

	1 mM	EDTA
AEX no-salt buffer	20 mM	Tris-HCl pH 8.5
AEX salt buffer	20 mM	Tris-HCl pH 8.5
	1M	NaCl
SEC buffer	20 mM	Tris-HCl pH 8.5
	150 mM	NaCl
Nanodisc insertion, buffer A	20 mM	Tris-HCl pH 8.5
	150 mM	NaCl
	0.05%	0.05% Octyl β-D-glucopyranoside (w/v)
Nanodisc insertion, buffer B	50 mM	NaCl
	20 mM	Tris-HCl pH 7.5
Nanodisc insertion, buffer C		
	500 mM	NaCl
	50 mM	Tris-HCl pH 7.5
	50 mM	imidazole
Nanodisc insertion, buffer D	500 mM	NaCl
	50 mM	NaOAc, pH 5.0

Table 5. Other buffers used in this work

General buffers		
5 x M9 salts(1 l)	42,5 g	$Na_2HPO_4 \times 2 H_2O$
	15,0 g	KH_2PO_4
	2.5 g	NaCl
	5.0 g	NH_4Cl
Coomassie Destaining Solution	10 %	Acetic acid
Uranyl staining solution	375 mg	Uranyl formate
	6 µl	5 M NaOH

2.1.4. Commercial kits and disposable materials

Table 6: Consumables used in this work

Materials	Manufacturer
Amicon Filter Concentrators	Merck Chemicals GmbH (Darmstadt, DEU)
C-flat 2/1	Protochips, Inc. (Morrisville, USA)
Coppermesh G2400C grids	Plano GmbH (Wetzlar, DEU)
Filter Support 10 mm	Avanti Polar Lipids (Alabaster, USA)
Filterpaper Whatman No. 4	GE Healthcare Bioscience (Pittsburgh, USA)

Glass Slides	Thermo Fisher Scientific (Waltham, USA)
Mini-PROTEAN TGX Stain-Free Gels	Biorad Laboratories (Hercules, USA)
Parafilm	Pechiney Plastic Packaging (Bellevue, USA)
Quantifoil 2/1	Quantifoil Micro Tools GmbH (Großlöbichau, DEU)
Reaction tubes 0.5, 1.5, 2.0, 15, 50 ml	Sarstaedt (Nümbrecht, DEU)

2.1.5. Software

Table 7: Software used in this work

Software	Developer
Adobe Illustrator CS6 v16.0.0	Adobe Systems (München, DEU)
Adobe Photoshop CS6 v16.0.0	CCLRC Daresbury Laboratory (Daresbury, UK)
UCSF CHIMERA 1.13.1	University of California (San Francisco, USA)
crYOLO 1.3.0	MPI Dortmund (Dortmund, DEU)
EMAN2.1	Baylor College of Medicine (Houston, USA)
imodfit	IQFR- Rocasolano Physical Chemistry Institute (Madrid, ESP)

Phenix Software Package v1.13_2998	Phenix Industrial Consortium (Berkley, USA)
Microsoft Office Package 2011/2016	Microsoft (Redmond, USA)
MotionCor2	University of California (San Francisco, USA)
Sphire	MPI Dortmund (Dortmund, DEU)
TranSPHIRE	MPI Dortmund (Dortmund, DEU)

2.2. Methods

2.2.1. PA_7 and PA_7LF_3 purification

2.2.1.1. Transformation with heat shock

The sudden increasing of the temperature during the heat shock (Bergmans et al. 1981; Singh et al. 2010), leaded the formation of pores into the bacterial membrane allowing the permeation of the plasmid into the cell. The competent cells were prepared in house.

Around 100 ng of plasmid were added to 100 μl BL21(DE3) competent cells and incubated on ice for 30 min. Afterwards, the cells undergo to 90 sec of heat shock at 42°C, followed by 1 min incubation on ice before 800 μl of SOC-medium were added. After 60 min of incubation at 37 °C at 700 rpm, the cells were plated onto a LB-plate with 50 mM glucose added, plus 125 μg/ml of kanamycin and incubated to grow only transformed cells and incubated at 37 °C overnight. The cell colonies were used for plasmid isolation or culture expression. The double-stranded DNA concentration was determined using the DeNovix DS-11 spectophotometer (Biozym Scientific GmbH). The concentration is determined referring to the absorption properties of the π-system of the purine and pyrimidine structures of the nucleotides (the value of absorption equal to 1 corresponds to 50 μg/mL of dsDNA).

2.2.1.2. Gene expression in *E.coli*

Gene expression was performed using BL21(DE3) *E.coli* competent cells. The pET19b:His_{10}-PA plasmid containing the gene coding for PA_{83} (Fig. S1), confers ampicillin resistance. LB media was supplemented with 125 μg/ml of kanamycin and 75 μM of isopropyl-β-D-thiogalactopyranoside (IPTG) and inoculated with a full loop of cells (10 to 100 colonies per liter). The culture was incubated at 28 °C at 180 rpm for 24 h. Afterwards, the cells were harvested by centrifugation at 5.000 rpm for 15min and the pellet was stored at −20 °C eventually.

2.2.1.3. Cell disruption

Once the protein has been over expressed from the cells, the cell membrane has to be disrupted to release the soluble protein. The method adopted in this thesis is the cell mechanical disruption. The cell suspension (and later the lysate) goes for 3-4 times into the high pressure microfluidizer. Afterwards, the resulting lysate is centrifuged for 40 min at 38000 rpm at 4 °C to separate the soluble proteins from the cell debris. The supernatant containing the soluble protein is collected and immediately enters the protein purification pipeline.

2.2.1.4. Protein purification

The purification of PA_7 was carried out in several purification steps. First, PA_{87} was isolated from the majority of the other soluble proteins via its poly-His-tag, by Ni-IDA beads (Cube Biotech). After overnight dialysis, an anion-exchange MonoQ (AEX, GE Healthcare) further purified PA_{83}. To complete the PA_{83} purification, a size exclusion chromatography (SEC, GE Healthcare) was performed. Next, PA_{83} oligomerization was induced by addition of trypsin, generating PA_{63} that oligomerized forming PA_7. Another purification step by SEC (GE Healthcare) assured an almost pure PA_7. Finally, after having incubated LF with PA_7 in a 10:1 molar ratio for 2h, the last SEC (GE Healthcare) eluted the PA_7LF_3 complex.

2.2.1.5. Immobilized-metal affinity chromatography

Immobilized-metal affinity chromatography (IMAC) allows proteins with affinity tag, to non-covalently and reversibly bind specific molecules that are immobilized on a specific material. In the PA_{83} purification, the column matrix was composed by iminodiacetic acid (IDA) agarose with immobilized Ni^{2+} metal ions (Fig. S1). The non-specifically bound proteins were eluted from the beads by an initial washing step. The bound proteins can be collected via a specific buffer that includes a competitive ligand or by cleaving off (on column) the affinity tag from the protein of interest. In this work, the PA_{83} possessed an N-terminal His-tag, therefore a Ni-IDA was used to

separate it from the other soluble proteins. Imidazole has an analogous structure to the histidine side chain, therefore competing for Ni^{2+} binding side. The PA-beads non-covalent binding is replaced by competition with imidazole that is introduced with a gradient in the elution buffer, allowing the displacing of the tagged protein from the beads and consequent elution from the column. Despite the tag and the washing step, generally other proteins elute with the target ones, therefore other purification steps usually follow this one.

First, Ni-IDA beads (Cube Biotech) were equilibrated with the washing-buffer (Tab. 4). Afterwards, the beads were incubated with the supernatant for 2h at 4°C, on a shaker. Afterwards, beads and supernatant were transferred into an empty column and washed with 100 ml of washing-buffer per liter of main culture, at 4°C. The elution was carried out using 1 ml of imidazole elution-buffer (Tab. 4) per liter of main culture. The fractions were collected and analysed via SDS-page to verify the presence of PA_{83}. The fractions showing PA_{83} were dialyzed overnight with the dialysis-buffer (Tab. 4) into a semi-permeable membrane with cut off of 50 kDa.

2.2.1.6. Dialysis

Dialysis allows protein-buffer solution exchange. In this case the dialysis was settled in order to decrease the imidazole concentration, in preparation to the following AEX. A semi-permeable membrane having a cut-off appropriate to the protein molecular weight (MWCO) is filled with the protein solution. Afterwards, the filled membrane is placed into the SEC buffer (Tab. 4) and left with gentle stirring for O/N at 4 °C. The proteins at lower MW, compared to the MWCO, and the ions will diffuse through the membrane.

2.2.1.7. AEX (MonoQ)

In the AEX, the selective retention of proteins on the matrix surface is based on their isoelectric point (pI). AEX is characterized by having on the column matrix, positively charged residues. The protein solution is loaded in total absence of salt or with a very

low salt buffer, into the MonoQ (GE Healthcare). The negatively charged residues, present on the protein surface, allow the binding to the matrix. The elution is performed with a pH or salt gradient (NaCl), where the Cl^- displaces the protein from the matrix, which elutes from the column.

Before stating the purification, the column was washed several times with a no-salt buffer (Tab. 4) and several times with gradient from 0% to 100% of the salt buffer (Tab. 4). Later, the equilibration was done with 2 CV of no-salt buffer. PA_{83} was first diluted 1 to 5 with buffer A (Tab. 4) to a final NaCl concentration of 30 mM, then loaded into the column. Initially, the column was washed with buffer A for 2 to 3 CV and after protein elution was performed with an anion strength gradient ranging from 0% to 40 % of salt buffer, over 80 CV. The target protein was visualized by SDS-page and dialyzed O/N with SEC-buffer (Tab. 4).

2.2.1.8. PA_{83} SEC

SEC columns, separate protein mixtures based on their hydrodynamic radius (molecular weight and shape). SEC columns are pre-packed with polymer beads that create pores in which the protein particles, under a controlled flow and pressure, pass through. Compared to low MW proteins, higher MW ones run out from the column relatively fast. Aggregates and proteins with relatively high MW generally elute in the void volume, whereas smaller proteins insinuate in the tiny pores and elute at later elution volumes. The kind of SEC together with the column volume, strongly influence the elution profile.

After dialysis, the protein was usually concentrated to around 500 μl using concentrator tubes (Amiconn Ultra, 50 kDa MWCO). Around 500 μl PA_{83} was then loaded in the loop of the Akta system and injected into the superdex 200 increase. SEC column was previously equilibrated with 1.5 CV of SEC-buffer. The protein was eluted with 1 CV of SEC-buffer, respecting the maximum flow rate and pressure indicated by the manufacture.

2.2.1.9. Trypsin digestion and inhibition

The trypsin digestion was performed using 1 µg of trypsin for each mg of PA_{83} for 30 min at room temperature and in the dark. The enzyme was neutralized by addition of double molar excess of trypsin inhibitor (Sigma Aldrich). Trypsin cuts off 20 kDa from the N-terminus PA_{83}, leaving a PA_{63} and a 20 kDa fragment. PA_{63} spontaneously oligomerizes in solution forming PA_7.

2.2.1.10. SEC of PA_7

To separate PA7 from trypsin, trypsin inhibitor and the 20 kDa fragment, the protein mixture was loaded into a SEC. Around 500 µl of protein mixture was loaded in the loop of the Akta system and injected into the superdex 200i. SEC column was previously equilibrated with 1.5 CV of SEC-buffer (Tab. 4). The protein was eluted with 1 CV of SEC-buffer, respecting the maximum flow rate and pressure indicated by the manufacture.

2.2.1.11. Lethal complex formation

According to the manufacturer instructions, lyophilized LF (List Biological Lab. Inc., Lot#1692A1B) was resuspended in MilliQ H_2O to a final concentration of 2mg/ml and added with a molar ratio of 10:1 to the PA_7 solution. After incubation with LF in ice for 2h, the complex was purified by SEC using a Superdex 200i column, using SEC-buffer (Tab. 4), and then further analysed in down-stream applications.

2.2.1.12. Determination of protein concentration

The protein concentration was determined using the DeNovix UV/Vis Spectrometer DS-11+, that measures the native UV-absorption of the aromatic amino acids tyrosine, phenylalanine and tryptophan at 280 nm. For more accurate concentration estimation, the absorbance was determined considering also the specific extinction coefficient and MW calculated with the program Protoparam (ExPASy, GILL & VONHIPPEL 1989)

and applied to the Lambert-Beers law ($A = \varepsilon \times d \times c$), where ε is the molecular extinction coefficient, d the optical path length and c the protein concentration.

A protein volume of 2 µl was placed on the DeNovix UV/Vis Spectrometer DS-11+ tip and after having inserted the extinction coefficient and the molecular weight, the experimental concentration of the protein measured at 280nm was given.

2.2.1.13. SDS-PAGE

Electrophoresis techniques follow the principle that molecules presenting a charge, such as DNA, RNA or proteins migrate through a polymer- or gel-matrix when an electric field is applied (Raymond 1964; Shapiro et al. 1967). The protein is first denaturated by adding SDS solution and heating up to 95°C for 5 min. The high temperature unfolds the protein structure while the anionic detergent, sodium dodecyl sulfate (SDS solution) confers a global negative charge. The SDS hydrophobic carbon chains bind to the hydrophobic regions of the protein and thereby destabilizes and destroys the protein tertiary structure. Therefore, the protein migration depends only on its molecular weight. Therefore, by this method it is possible to initially estimate the MW of the proteins by comparing the position of the band of interest to the one of the marker, loaded in the same gel. The SDS-PAGE polyacrylamide polymer-matrix is produced by copolymerization of acrylamide and N,N'-methylene bisacrylamide. After electrophoretic separation, the protein bands were visualized by Coomassie Blue dye staining (Wilson et al. 2019).

2.2.2. PA_7 functional analysis

In order to fulfill the toxin action, some toxins undergo to conformational rearrangement, exposing trans-membrane domains that insert into the lipid bilayer of the cell membranes i.e. punching holes, destructing cell integrity or injecting toxic components (Schubert *et al.* 2018). Nanodiscs and liposomes are widely-used alternatives to cellular membranes for functional and structural analysis. Nanodiscs are composed by a circular membrane scaffold protein (MSP) that surrounds a mass

of lipids, mimicking the cell membrane. Instead, liposomes are vesicles of lipids that can vary in size, from nm to μm. Both methods, in principle, are successfully used in cryo-electron microscopy and tomography.

2.2.2.1. Protein insertion in nanodiscs

Ni-IDA beads were washed with 3 CV of ddH_2O, for 5 times, followed by equilibration with 3 CV of buffer A (Tab. 4). Afterwards, 500 µL of 0.2 µM PA_7 was added to the beads and incubated 20 min at room temperature. The beads were washed with buffer A to remove unbound PA_7, followed by 5 min incubation step in presence of 1M Urea at 37°C, and washed with buffer A. MSP1D1:POPC:sodium cholate ratio and preparation was done according to Akkaladevi *et al*, 2013. Afterwards the beads and the nanodiscs solution were carefully transferred in a semi-permeable membrane with a 12-14 kDa MWCO with buffer B (Tab. 4), and left it under stirring at 4 °C for 1, 2 or 3 days. To elute PA_7 pores inserted into nanodiscs, column material was incubated for 10 min on ice in buffer C. The eluted sample was concentrated and analyzed by negative staining EM.

2.2.2.2. Protein insertion into Liposomes

In order to constitute liposomes, POPC was initially solubilized in 5% Octyl β-D-glucopyranoside. Afterwards, the solubilized lipids were dialyzed (MWCO: 12–14 k) for 8–12 h at 4° C against buffer A (Tab. 4) and subsequently PA_7 into the pre-pore state was added to the lipids, following a 1: 10 molar ratio. To convert the PA_7 pre-pore to the pore state, the sample was dialyzed for 24–72 h (MWCO:12–14 k) against buffer D and later the samples were used for negative staining EM.

2.2.3.Structural biology methods

The typical biological sample for transmission electron microscopy is prepared in a liquid environment. The protein solution is fixed on a specimen holder that enters the

high vacuum chamber of the EM. The vacuum is essential to avoid the electron-air interfering. In the next chapters, two methods for sample preparation are described: negative staining and cryo vitrification.

An EM copper grid (diameter 3.5 mm) is covered with a thin carbon layer, and it represents the basic support for the biological liquid solution in analysis. The carbon layer can be either a continuous film (negative staining) or can have a holey carbon structure (for cryo analysis). The grid is organized squares (typically 200, 300 or 400 squares/par grid). A continuous carbon layer can be also manually added to the bare grid whenever is needed. The typical EM grid vary among the mesh size, hole size and hole spacing. The cryo EM grid type used for data collection was a Qf-gold R 2/1, which stands for a hole size of 2 µm and a hole spacing of 1 µm. In this work, gold worked best.

2.2.3.1. Negative staining EM: principle

Negative staining EM uses heavy metals solutions such as uranyl acetate or formate to stain the biological sample on the carbon coated cupper grid. The heavy metal solution surrounds the sample and by absorbing or scattering some of the electrons, generate phase contrast. The electron beam is well deflected by the metal ions and less by the sample, obtaining a good contrast in the image that successively is inverted to generate the reprojection of the sample on the grid. The structural information revealed from this technique are limited by the preparation itself. The integrity of the sample is affected by the drying procedure while preparing the negative stain grid. In addition, the projection of the particles depends on the shape of the particle itself and the homogeneity of the staining coverage. Furthermore, the resolution is limited to 20-25 Å, depending on the size of the negative stain particles which influence the final resolution. Negative stain electron microscopy is an easy and useful method largely used for checking the quality of the sample and generating an initial 3D volume used as initial model for further structural investigation via cryo-EM. The preparation of the negative stain grids requires only few minutes (see chapter 3.2.1). In addition, the technique is highly reproducible serving as quality control and

gives immediate information about the sample quality (overall heterogeneity, aggregation and impurities). A data set can be recorded and in the same day is possible to analyse the 2D-classification classes that help to visualize heterogeneity, like flexibility within the protein or different conformations. Further processing steps bring to a first low-resolution model serving as reference for a higher resolution refinement of a cryo-EM data set. In this thesis, negative staining EM served to visualize the PA_7 integrity, functionality (nanodiscs and liposomes essays) in addition to the quality of the PA_7LF_3 complex before proceeding to cryo-EM grid preparation.

2.2.3.2. Negative staining EM: grid preparation

The raw negative staining grid has to be covered with a homogenous carbon layer before incubating the protein solution on it. The negative staining grids can be purchased directly covered with a carbon layer, otherwise they are covered with a thin polymer film onto carbon can be coated after. First, 2 or 3 drops of collodion solution (Fluka) are dropped into a glass beaker (around 30 cm in diameter) filled with water, kept as still as possible. Afterwards, the 400-mesh copper grids (GS400C, Plano GmbH) are manually placed on the water-colloid surface, on top of the polymer layer (Booth et al., 2011; Ohi et al., 2004). When a water surface of around 4x4 cm^2 is covered, a piece of blotting paper is gently applied on the grids. The grids adhere on the paper surface that has sucked and blotted the solution around. The paper with the grids is transferred in a Petri dish and left upside down to allow a faster drying. Once the 2-3 days of drying are over, the grids are transferred into a Leica EM ACE600 (Leica Microsystems), initially a high vacuum (around 10^{-4} Pa). Afterwards, the carbon is evaporated or sputtered on the grid. The final carbon layer is in the range of 5-15 nm thickness that are normally associated to an optimal preparation and image contrast. Before sample application, the surface of each grid was cleaned making a cleaning cycle into a plasma cleaner (GloQube, Quorum Technologies) to generate a hydrophilic carbon layer that allows protein absorption (Booth *et al.* 2011).

2.2.3.3. Negative staining: solutions

Several staining solutions can be used for negative staining of proteins (uranyl acetate and formate). They mainly difference is represented by the quality of the protein contrast, the grain size and stability. Uranyl acetate and formate stain have an acid pH (pH 4 - 4.5) and this could influence the behavior and the protein structure when they are particularly sensitive to acidic pH values. The contrast obtained with these staining solutions is definitely better compared to other staining solutions having higher pH as sodium phosphotungstate (pH range from 5 to 8) and ammonium molybdate (pH range from 5 to 7).

The protein solution is applied on the surface of a carbon coated copper grid and embedded into a heavy metal solution (stain) that produces a negative image of the protein surface via the EM (Booth *et al.* 2011). The electron-dense staining solution covers the entire grid surrounding the protein and resulting in white molecules on dark background, in the so-called electron micrograph (Fig. 2.1).

The difference between the highly electron dense staining solution and the organic material, results in a high contrast. Limiting factor of this technique is represented by the protein deformation due to its dehydration. Additional limiting aspect of this technique is that the molecules are represented only on their surface because the stain cannot penetrate the molecules. As consequence, the maximal resolution estimation is around 20 Å (Ohio *et al.* 2004). Uranyl acetate and formate are both highly reactive and toxic and should be carefully handled. The contaminated materials (filter papers, tips) have to be properly discarded in the hazardous waste.

In this work, a standard negative staining solution of 0.75% of uranyl formate was prepared. For the staining preparation, 5 ml of water was boiled and 37.5 mg of uranyl formate powder (SPI-Chem) was immediately added. The yellow solution is left stirring in the dark for at least 5 min. Afterwards, the pH is adjusted by adding 6 µl of 5 M NaOH and eventually further corrected to a final pH of 4. The solution is left stirred again for additional 5 minutes in the dark. Finally, the solution is filtered through a 0.22 µm filter syringe into a plastic tube, covered with aluminium foil. The staining solution prepared as described is stable for up to a week. It has to be stored in the dark, at constant temperature and if it starts precipitating it has to be filtered again or freshly prepared.

2.2.3.4. Negative staining of proteins

Before applying the sample on the grid surface, the grid was made hydrophilic by glow-discharging it into a plasma cleaner (FEMTO, Diener electronics GmbH & Co. KG) in which the ionized gas charges the grid surface. The optimal protein distribution on the grid depends on the protein MW and shape, and generally it is optimized after every protein purification.

After the grid has been glow-discharged, it is fixed onto a reverse forceps. Two drops (10 µl each) of negative staining solution are placed on a clean piece of Parafilm. A small 4 µl protein solution drop is gently placed on the grid and incubated (from 30 s to 75s). Afterwards, the protein solution is gently blotted with a piece of filter paper (Whatman 4) placing the edge of the grid on to filter edge.

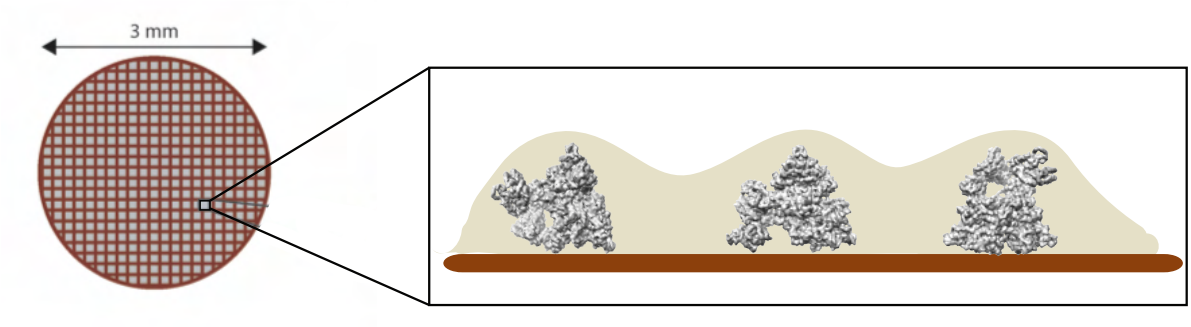


Figure 2.1 : Negative stain grid representation. (A) negative stain grid representation, highlighting the different squares. (B) Image of a close up view into a single square region. The particles, which adopt different orientation on the grid, are covered by a thin layer of negatively stained solution (beige).

The first negative staining drop is collected with the grid, incubated for 45 s and blotted away. This procedure is repeated for the second drop of staining solution. Finally, the grid is left for around 5 min at room temperature to air-dry (Fig. 2.1).

2.2.3.5. Negative stain: data collection and processing

The negative staining dataset of PA_7LF_3 heterocomplex was collected semi-automatically using the 2 x 2 spotscan function in the EMMENU software (TVIPS) interface with a defocus range that varied from –1.2 to -2.6 µm. More than 2.000 micrographs were collected at magnification of 120 k corresponding to a pixel size of 1.84 Å/pixel. The data set was manually cleaned (e.g. micrographs with drift or near to the grid edge, were excluded) and the particles were semi-automatically selected using e2boxer (part of the EMAN2 software package) (Tand *et al.* 2007). The picked particles were extracted using SPHIRE software package (Moriya *et al.* 2017). Afterwards, the extracted particles were classified into groups of similar views by the iterative Stable Alignment and Clustering approach (ISAC) implemented in SPHIRE (Z. Yang *et al.* 2012; Moriya *et al.* 2017).

2.2.3.6. Single particle cryo-EM

In single particle cryo-EM, the 3D-reconstruction at atomic resolution can be obtained averaging thousands of 2D projection of different particles orientations, in their native conformation. Recently, cryo-EM computational and technical advancements made this technique a good alternative among other high resolutions techniques. In the past, high resolution information of mostly symmetric proteins with MW above 100 kDa was possible to be obtained using cryo-EM (Bertesaghi *et al.* 2014; Campbell *et al.* 2015). Recently, the introduction of direct electron detectors (Binshtein *et al.* 2015) coined a new era for cryo-EM, defining the so-called “resolution revolution” (Kühlbrandt 2014b; Kühlbrandt 2014a). According to this method, the biological sample is fixed and embedded in an extra thin ice layer by fast sample freezing in liquid ethane at high freezing rate of ~108 K/s (Studer *et al.* 2008). The key aspect of the flash freezing is the consequent formation of amorphous but non-crystalline ice (vitrification) with a low density (Dubochet *et al.* 1988). Following this method, the protein structure is conserved without any particles deformation or damage due to the crystals ice formation.

Generally, when the electron beam interacts with the sample, the electrons are either elastically or inelastically scattered. In the elastic scattering event, the direction of

the electron changes but the velocity, or energy, does not. Differently, the inelastically scattered electrons damage the sample due to energy transferring, thus introducing noise. Comparing these two events, the probability of having inelastic scattering events (Henderson 1995) is higher than elastic scattering ones (Henderson 1995), that is why in cryo-EM a very small electron dose is applied to avoid addition radiation damage. The elastically scattered electrons produce two types of contrasts, thus defining the amplitude and phase contrast. The amplitude contrast is generated by the directional change of the electrons (strong directional change brings the electrons out from the detector). The relative areas in the micrograph, appear darker generating a contrast. The contrast is not strong since the biological material scatters the electrons with a very small bending angle that can be easily filtered out with an aperture. In addition, the elastically scattered electrons contribute to the phase contrast due to their interference with the unscattered electrons from the electron beam. Due to weak contrast in cryo-EM, the standard cryo-EM was addressed to protein with minimum a molecular weight of 90-100 kDa. This limitation changed with the introduction of the Volta Phase Plate (VPP) (Danev *et al.* 2014). The presence of the VPP brings to an additional phase shift to the unscattered electron beam that is translated to a contrast increase within the low dose cryo-EM images, without losing high-resolution information. The presence of the VPP brings higher contrast and therefore proteins with a MW lower than 50 kDa are visible when vitrified in a thin ice layer, placing the cryo-EM limit to its theoretical size limit of 38 kDa (Henderson *et al.* 1995). The new software development totally integrated the VPP option in the routinely data set collection, and many labs already implemented it, resolving structures below 100 kDa at high-resolution (e.g. hemoglobin) (Khoshouei *et al.* 2017), or streptavidin (Fan *et al.* 2019). All these developments and advancements made cryo-EM a competitive structural biological method that can stand bear to X-ray crystallography for the determination of atomic protein structures.

2.2.3.7. Cryo-EM grid preparation

Protein vitrification is conducted using cryo-plungers. The technique allows to (i) deposit the sample on the grid, (ii) blot the protein volume excess and (iii) immediately

proceed to a very fast grid freezing in liquid ethane, kept cooled at liquid nitrogen temperatures (-170 °C). The liquid nitrogen cannot be used for direct protein vitrification due to the Leidenfrost effect. Indeed, nitrogen forms a vapor layer on the grid surface that slows down the cooling process thereby resulting in the formation of ice crystals (Fig. 2.2).

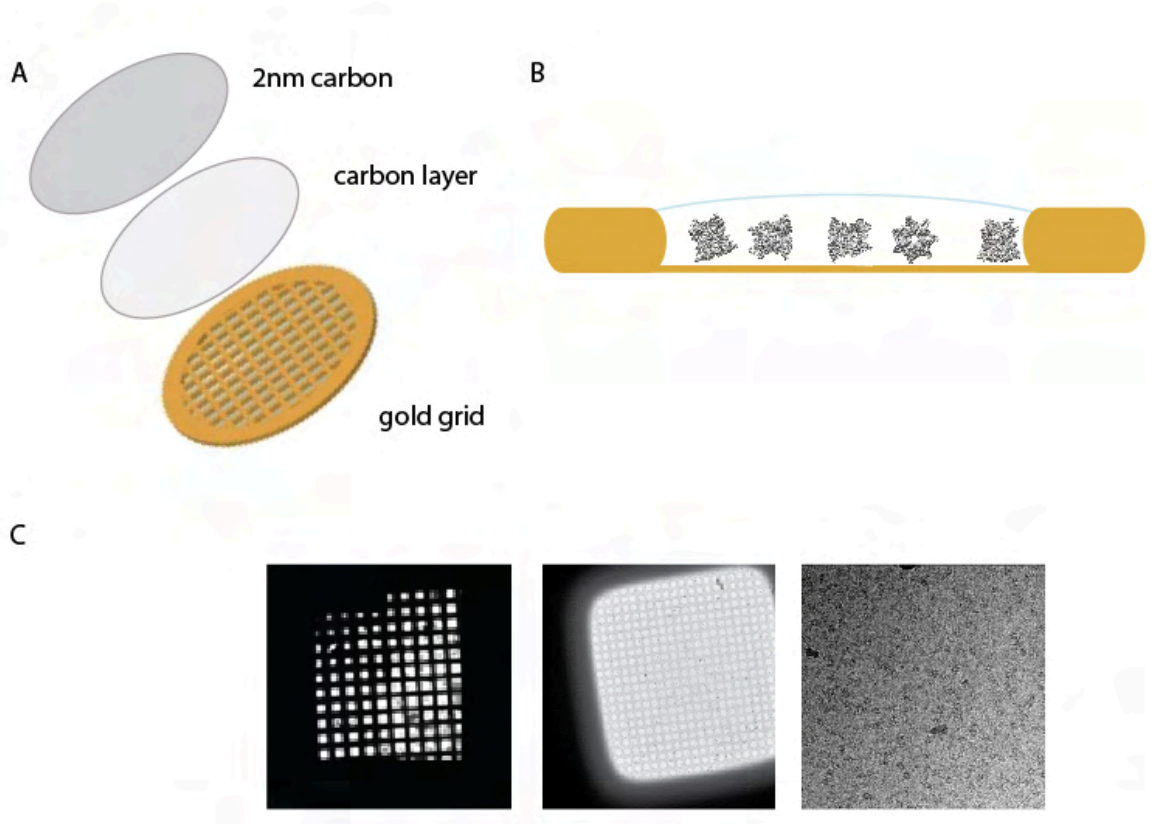


Figure 2.2 : Addional carbon-layer cryo-grid structure. (A) Schematic overview of a cryo grid with 3mm diameter used as support for cryo-EM analysis. The gold grid is covered with a first layer of carbon, and additional 2nm of carbon on top. (B) The proteins lye on the 2nm carbon layer and are coated with a thin layer of vitrified ice. (C) Micrographs of cryo grid atlas, square and hole. Images collected at Titan Krios microscope, 300 kV.

The ice thickness is one of the most important and crucial parameters that directly affects the final resolution. The ice layer should be thinnest possible and not have any contaminations. Anyway, the ice thickness have to contain the protein particles, while being thin enough to prevent excessive beam interference that would reduce image contrast (Orlova & Saibil 2011). Indeed, the ice thickness is correlated to the contrast between the protein and the water molecules and to maximize it, the ice layer should

have an optimal thickness (Hye-Jin Cho *et al.* 2013). However, a really thin ice layer can promote specific interaction between the protein and the air-water interface bringing to preferential orientation or even denaturation of the particles or displacing the particles to areas with thicker ice layer (visible for holey-carbon grids only). Using grids coated with carbon, graphene oxide or graphene, the film can improve the particles random orientation. Just before the sample application, the grids were cleaned and glow-discharged (Aebi & Pollard 1987) using a so-called plasma-cleaner (GloQube, Quorum Technologies). The cryo grids used during cryo screening were produced from two manufacturers: Qualtifoil Micro Tools (Qualtifoil grids) and Protochips (C-flats). The grids vary for the grid material (nickel & molybdenum, copper, gold), mesh number (200, 400) and hole size (from 1 to 2 µm). The glow discharged grid is fixed into a tweezer (Dumont) and loaded into the humidity-controlled chamber of the CryoPlunge 3 (Gatan), at RT. Subsequentely, 4 µl of PA_7LF_3 at concentration of ~0.06 mg/ml was loaded onto the gold Quantifoil R 2/1 holey carbon grid with a 2nm additional carbon support, and incubated into the chamber for 45s (95% humidity) before the excess of sample was automatically blotted (3s blotting time) and plunged into liquid ethane using a CryoPlunge 3 (Gatan). The vitrified protein is kept at liquid nitrogen temperature to avoid crystal ice formation and subsequent disruption of the sample. Usually, the grids are stored in specials containers filled with liquid nitrogen, in which the temperature is kept constant.
The cryo screening for protein concentration and ice optimization, was performed either manually at the Tecnai Spirit (120 kV) or semi-automatically on an Arctica microscope (FEI) (Fig 2.3).

2.2.3.8. Image processing and 3-D reconstruction

The cryo-EM data set of PA_7LF_3 was collected on C_s-corrected TITAN KRIOS transmission electron microscope (FEI), equipped with a high-brightened field-emission gun (XFEG, FEI) operated an acceleration voltage of 300 kV. The images were acquired on a K2 summit direct electron detector (Gatan) operated in continuing mode with calibrated pixel size of 1.07 Å/pixel, at 130.000 x magnification. Initially, the software acquires a grid atlas to obtain an overall image representing the whole grid (Fig 2.2

C). After square and relative holes selection, the images were collected automatically with EPU (FEI) software. The data set with a total of 5238 images was collected with 40 frames and an exposure of 15 s with a total dose of ~74.4 $e^{-}/Å^{-2}$. To generate an additional contrast, the data-set was collected with an applied defocus, that varied between -1.2 and -2.4 µm. All images were collected automatically using EPU (FEI), as automated data collection software. Motion correction was performed using the MotionCor2 program (Moriya et al., 2017).

2.2.3.9. Low-dose data acquisition

Radiation damage of biological sample remains a bottleneck of structural investigation performed via crystallographic or microscopy techniques that use radiation sources. The use of high X-ray and electron dose leads sample damage, altering and finally destroying the biological sample and the related structural information (Meshi *et al.* 2012). The definition of the electron dose while setting up the data acquisition is a crucial step that influences also the signal-to-noise ratio. The modern workflow provides automated low dose data collection, achieved by three basics steps: tracking, autofocus and exposure. Regular grids with fixed space between the foil holes and with fixed diameter are used, allowing the predefinition of settings for the automated acquisition. The tracking is made in the hole center, adjusting the foil hole diameter. The autofocus takes place on the carbon of the support foil, via beam tilt approach. The exposure depends on case to case and can be performed in the center or on the edge of the foil hole.

2.2.3.10. Direct detector cameras

The modern cryo-electron microscopes are equipped with direct electron detector cameras, that enable the direct detection of electrons on the chip, increasing the signal to noise ratio in comparison to the CCDs cameras which instead detect the electron indirectly (McMullan *et al.* 2014). The direct electron detector cameras save the individual series of exposure and stack of frames. These settings allow

computational correction of image blurring which interferes with the visualization of the image features. The instability of the sample stage and the motion induced by the illumination of the electron beam on the sample, generates image blurring. The motion correction is done by aligning several frames together. The most used cameras systems are the K2 summit camera (Gatan) and the Falcon direct electron camera (FEI).

2.2.3.11. Contrast transfer function (CTF) correction

CTF describes how the lens system and the imaging parameters of the EM affect the outcoming image. CTF is a Fourier based description of this phenomena and it is determined by several not fixed parameters such as the defocus, the astigmatism and the acceleration voltage. In addition, also characteristic proprieties of the EM can influence the CTF, such as lens aberration and beam coherence (Williams 2009). The typical CTF function oscillates around a zero contrast with a variable amplitude, with reversed phase, and a characteristic decreasing of the signal transfer while increasing of the spatial frequency (Zhu 1997). In order to obtain a high resolution cryo-EM map, the collected images influenced by the CTF need to be corrected for. In order to do so, the CTF parameters have to be accurately estimated.

2.2.3.12. Data pre-processing

Data pre-processing pipeline was performed with the lab internally established TRANSPHIRE software and it run on-the-fly during data collection (Stabrin *et al* 2020). As first step, the movies are renamed and sequentially numbered. The original images coming from the microscope are converted into several file formats (tif, jpeg, mrc) and transferred to the in-house cluster-system used for the main processing steps. A backup is also created on the in-house server, *billy*.

A first visual inspection is fundamental to discard images with high drift or bad ice quality. The micrograph frames were aligned and summed using MotionCor2 (Moriya et al., 2017), generating dose weighted and un-weighted full-dose images. All image

processing described in this thesis were carried out with the SPHIRE/EMAN2 software package (Moriya, T. *et al.* 2017). The un-weighted full-dose images were used for defocus and astigmatism estimation by CTER while Motion-corrected sums without dose weighting were used to determine the defocus and astigmatism in CTER. With the 'drift assessment' tool in SPHIRE, images with high defocus or as well with too high drift, were sorted out.

After having selected the best micrographs, 10 of the most diverse images (different defocus, contrast etc.) were manually picked to create a first reference for the automated picking done by crYOLO software (Wagner et al., 2020), implemented in SPHIRE (Stabrin et al., 2020). A total of 390 k particles were picked from the dose-weighted micrographs, with a final window size of 336x336 pixels. Particle alignment and two-dimensional classification was done using the *iterative and stable alignment and clustering* (ISAC), implemented in SPHIRE. During the 2D-classification, every particle is assigned into a class accordingly to its 3D-orientation. The particles assigned into a class get aligned by x- and y-shifts and by in-plane rotation until they superimpose. ISAC 2D-classes show structural features, as result of the averaging of the particle belonging to the same class, which boosts the signal-to noise ratio. 2D-classification is a useful tool that gives clear information about the data-set quality and allows an initial digital purification. The so-called Beautifier tool implemented in SPHIRE, generated sharpener 2D-classes, at the original pixel size. Being at the original pixel size and not at the scaled one, the beautified classes exhibited more high-resolution features, compared to the regular 2D-classes from ISAC. The best class averages were selected and all low resolution 2D-clesses were excluded. The selection led to final 213 k dose weighted and drift corrected clean particles. The 3D-reference used during data analysis was a chimera generated combining different published atomic structures (chapter 3.2.7). During the 3D-refinement no symmetry constrain was imposed (sxmeridian in SPHIRE, adopting C1) and the post-processing brought to a first 3.5 Å electron density map of PA_7LF_3 complex. Nevertheless, several 3D-classification and class rotation were required to properly identify and separate the particles belonging to three complexes PA_7LF_{2+1A}, PA_7LF_{2+1B} and $PA_7LF_{(2+1A)'}$. The 3D-refinement implement in SPHIRE is an automated refinement program, called MERIDIAN. The program generates projection parameters of each particle in the x- and

y-axis as well as the Eulerian angles (ϕ, θ, ψ) (Moriya et al. 2017). The output of every 3D-refinement were two unfiltered half volumes, which were used in the final post-processing step to estimate the overall resolution and calculate the final 3D-reconstruction. Important to specify that the electron densities coming from 3D-classificaton or from previous 3D-refinement, served as reference for the subsequence 3D-refinement runs.

2.2.3.13. Post-processing and resolution estimation

The two unfiltered volumes coming as MERIDIEN output were used in the post-processing step to calculate the Fourier Shell Correlation (FSC). Initially, MERIDIEN splits the data set in two equal subsets that are independently refined, thus bringing to two half volumes. The FSC represent the correlation of the two half set volumes in the Fourier shell as a function of the spatial frequency. The final resolutions of the 3D-reconstruction were estimated according to the 0.143 criterion, reported also as the gold standard FSC criterion (Scheres & Chen 2012). Post-processing of the unfiltered half volumes generated from MERIDIEN were run using the *PostRefiner* tool implemented in SPHIRE (Stabrin et al., 2020). The *PostRefiner* tool essentially combines the two half maps and subsequently low-pass filters the combined volumes according to the estimated resolution. In order to boost the structural features and to exclude the surrounding noise, a tight adaptive mask and a negative B factor of -170 A^2 were applied. In addition, the *Local Resolution* tool present in SPHIRE allowed to calculate and analyse the local resolution of the PA_7LF_3 complexes. Using UCFS CHIMERA (Pettersen et al. 2004) it was possible to differently colour the maps according to their local resolution. The final densities were locally filtered according to the estimated local resolution, or the local de-noising filter LAFTER was applied to recover features enhancing the signals.

2.2.3.14. Model Building, refinement and validation

As first step to build the heptamer ring, a single monomer from the PA_8 crystal structure [PDB: 3HVD] was independently fitted into every monomer of the $PA_7LF_{3\text{-}masked}$ map, using rigid body fitting tool in Chimera (Petterson et al 2004). Similarly,

the unbound LF crystal structure (PDB: 1J7N) was fitted into the two LFs bound to the PA_7 ring, LF_1 and LF_2. The initial model was flexibly fitted using iMODFIT (Lopez-Blaco JR et al, 2013) and the followed refinement runs were a combination of manual model building in COOT (Emsley P *et al*, 2004) and real-space refinement in PHENIX (Liebschener B *et al*, 2019). Finally, the unresolved loop regions were deleted and where the electron density was poor, the side chain information were removed for less well-resolved regions. The $PA_7LF_{3\text{-masked}}$ model served as starting model for PA_7LF_{2+1B} and it was initially fitted into the map using rigid-body fitting. In the PA_7LF_{2+1B} , LF_2 was fitted into the density corresponding to LF_3. Similarly to $PA_7LF_{3\text{-masked}}$, a restrained refinement in PHENIX was run on the entire model, followed by additional manual refinement with COOT and real-space refinement in PHENIX. The PA_7LF_{2+1B} served as starting model for the PA_7LF_{2+1A} density, and the procedures is similar to the one described for the PA_7LF_{2+1B} , with the initial fitting based on $PA_7LF_{3\text{-masked}}$.

For both maps, unresolved loop regions were deleted and side chain removed when the density was poorly resolved. Geometrics of the final refined models were obtained either from PHENIX or calculated using the Molprobity server (http://molprobity.biochem.duke.edu).

2.2.3.15. Bioinformatical analysis and visualization tools

The figures presented in this work were obtained using Adobe illustrator and Adobe Photoshop CS6. The program UCSF Chimera was used of visualization of protein structures and movies realization (Petterson et al. 2004). The H++ webserver was used to protonate the different subunits complexes at pH 8.0 (Anandakrishnan et al 2012). The webserver based on the pH value, calculated the pKa values of charged, acidic or basic residues and eventually added hydrogen atoms to the pdb model depending on the protonation state. The protonated model was used for surface colouring of the different structures according to their electrostatic Coulomb potential (-12 to +12 kcal/mol) using the propper tool in UFSC CHIMERA named *Coulombic surface colouring*. Additionally, the hydrophobicity of the structures and surfaces was visualized in UCSF CHIMERA, following the Hessa and Heijne (Hessa et al 2005), using

the so called *Define attribute tool*. The different complexes were superimposed and compared using the *MatchMaker* tool implemented in UCFS CHIMERA. Instead, with *Morph Conformations* tool the conformational changes between 2 or more structures were easily visualized.

3. Results

3.1. Gene expression and protein isolation

The PA_{83} construct was transformed into *E.coli* BL21(DE3) cells, following the protocol described in chapter 2.1.1. The pre-culture was scaled up to 10 L and the cells were broken using the French press. Afterwards, the separation of the soluble protein from membranes and membrane-bound proteins was performed by a high-speed centrifugation (chapter 2.2).

PA_7LF_3 was purified via a series of purification steps, using Ni-IDA beads, AEX and SEC.

In order to optimize the purification strategy, the PA_{83} construct was modified at the N-terminus by an additional His_{10}-tag (Fig. S1). Based on the high affinity of the poly-His_{10}-tag to the Ni-IDA beads, the target protein was retained on the beads, while the unbound proteins were washed away. After an extensive wash with the washing buffer (20 mM Tris-HCl pH 8.5, 500 mM NaCl), the target protein was eluted using an imidazole elution buffer (500 mM imidazole, 20 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM EDTA) (chapter 2.3.3.1). However, other soluble proteins with unspecific binding to the Ni-IDA beads were co-eluted together with the target protein, resulting in the impure elution fractions. The fractions with the target protein were collected, dialyzed overnight and concentrated to a final volume of 1 ml (fractions from 1 to 6, Fig. 3.1-A).

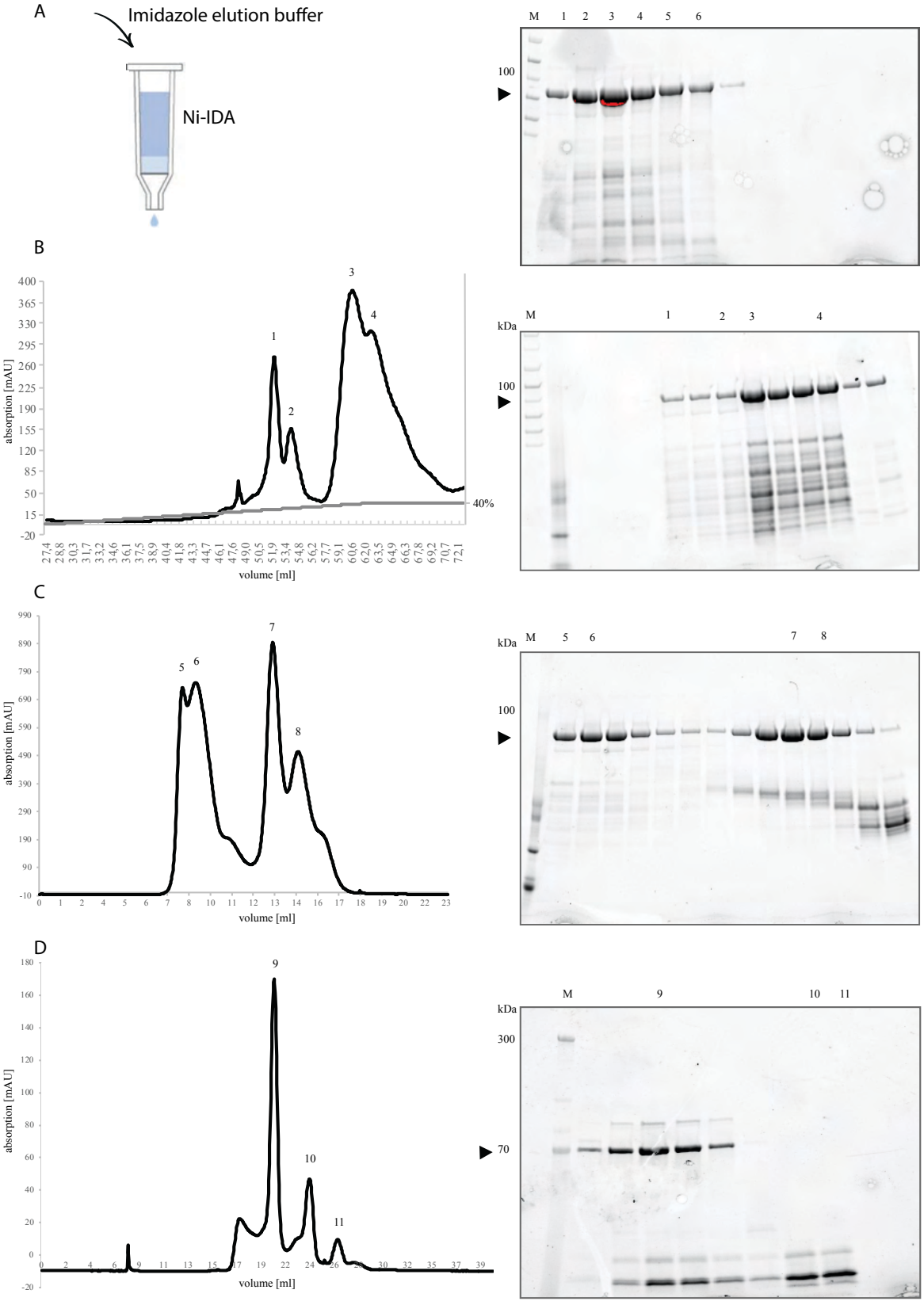

A
Imidazole elution buffer
Ni-IDA
B
absorption [mAU]
volume [ml]
40%
C
absorption [mAU]
volume [ml]
D
absorption [mAU]
volume [ml]
M
kDa
100
300
70

Figure 3.1 : Chromatograms and SDS-PAGE gels corresponding to PA_7 purification steps. (A) Representation of Ni-IDA gravity flow system and corresponding SDS-PAGE. The gel shows the elution of PA_{83} in all fractions reported on the gel. (B) Anion exchange chromatogram (AEX) and corresponding SDS-page gel. The elution was performed with a salt gradient from 0 to 40% of 500 mM NaCl, in 80 min (C) SEC performed via Superdex 200i of PA_{83}. Peak 7 and 8 corresponds to the PA_{83} elution. (D) SEC performed with Superdex 200i of the PA_7 obtained after cleavage with trypsin. The small first shoulder corresponds to the void volume while peak 9 to the PA_7 elution. The later peaks represent the elution of the 20 kDa protein fragment and the residual trypsin and trypsin inhibitor. In every gel, the target protein is highlighted by a black arrowhead.

The protein solution was diluted with no-salt buffer (20 mM Tris-HCl pH 8.5) to reduce the salt concentration and immediately injected into the AEX column (Fig. 3.1B). As reported from the AEX gel, the target protein was present in almost every fraction and the contaminations could not be improved (Fig. 3.1B-right). The fractions corresponding to peak 1 to 4 were collected and dialyzed overnight.

Afterwards, the protein solution was concentrated and injected into the Superdex 200i eluting an almost pure PA_{83} (Fig. 3.1C). The SDS-gel also reveals the presence of a small amount of other proteins with lower MW (Fig. 3.1C).

The pre-protein PA_{83} was incubated with trypsin (Sigma Aldrich), and the reaction quenched with trypsin inhibitor (Sigma Aldrich, see chapter 2.3.3.5). This step cut the 20 kDa N-terminus from the PA_{83}, leaving PA_{63} that spontaneously assembled in complexes composed by seven monomers PA (PA_7).

To purify PA_7 from the 20 kDa N-terminus fragment, trypsin, trypsin inhibitor and other impurities, the protein mixture undergo through another round of SEC (Fig.3.1D). Pure PA_7 eluted in peak 9 (Fig.3.1D). Peak 9 and the fractions before and after were collected and concentrated to a final volume of 500 µL.

3.2. PA_7 structural analysis

The four purification steps resulted in pure PA_7 that was further analyzed by negative staining EM. This technique was used to judge on the sample quality and define the particle distribution of PA_7 on the grid.

3.2.1. PA_7 Negative staining EM

During negative staining EM, the particle distribution on the grid was optimized. To reach an optimal particle concentration (Fig. 3.2), 0.04 mg/ml of PA_7 were applied on the grid and stained using uranyl acetate as negative staining solution. The unique disc shape disposition allowed a direct count of the protomers forming the complex. In addition, the central hole is directly visible also in a negatively stained micrograph, contributing to design a unique complex shape. The background looked clean since only a small amount of lower MW proteins were still present in the sample.

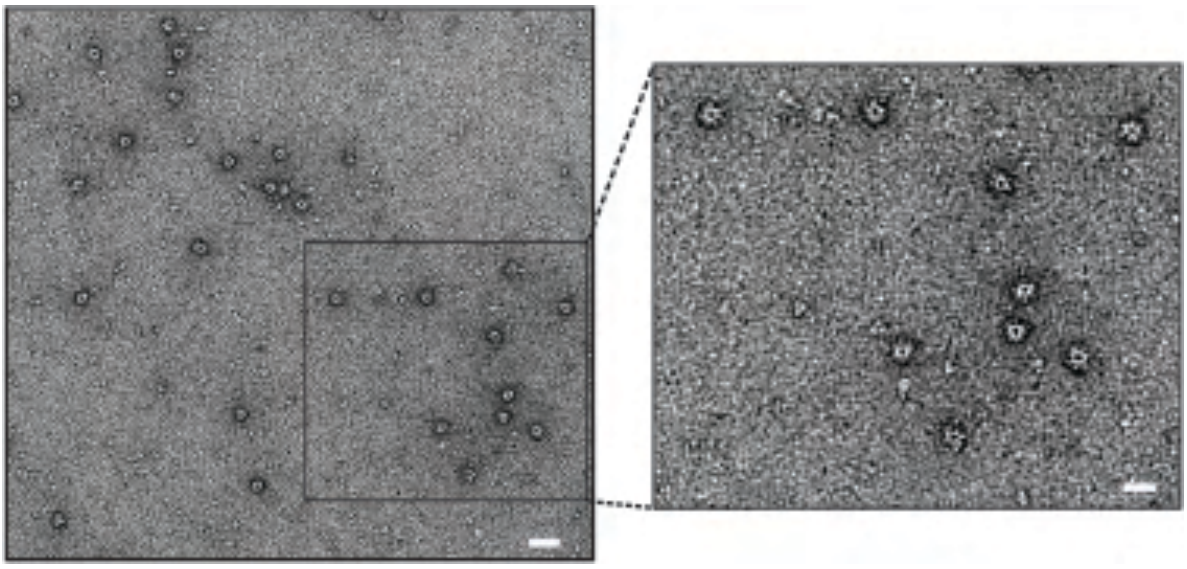


Figure 3.2 : Negative staining micrograph of PA_7 after SEC. Original micrograph showing PA_7 distribution on the negative stain grid (scale bar, 40 nm) and zoom in of an area highlighting the PA_7 ring shape (scale bar, 20 nm).

3.2.2. Reconstitution of PA_7 in membrane-mimicking systems.

It is known from the literature (Jiang et al., 2015) that acidic pH leads to a massive conformational rearrangement of the PA_7 complex. In this study, I verified the correct assembly of the pre-pore PA_7, its integrity and functionality, changing the environmental pH and therefore inducing a conformational rearrangement (Fig. 3.3, 3.4 and 3.5) (Jiang *et al* 2015). The massive structural rearrangement that leads PA to the pore conformation, terminates with a structure comparable to a flower (Fig 3.3 B). From the literature it is also known (Fabre *et al.* 2016), that the tip of PA_7 pore-state tends to easily aggregate in solution. Indeed, the tip is highly hydrophobic and in absence of membranes or membrane-mimicking systems, they result in a completely disordered assembly of proteins in solution (Fabre *et al.* 2016).

In order to obtain an image of the pore state, the hydrophobic tips were shielded using artificial membranes: liposomes or nanodiscs (Efremov et al., 2017; Gatsogiannis et al., 2016).

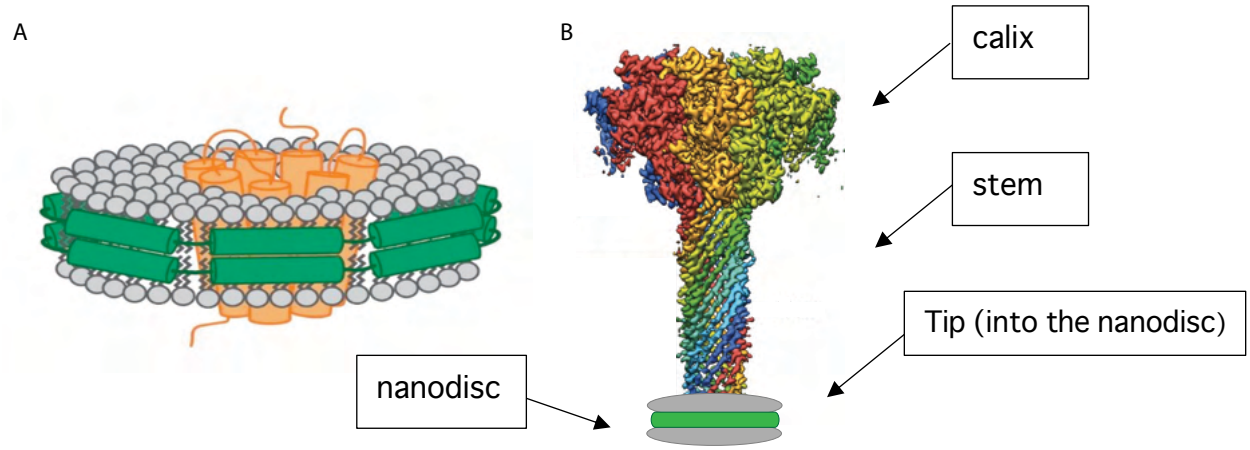


Figure 3.3 : Scheme of membrane bilayer (nanodisc) and representation of PA_7-pore state insertion into a nanodisc. (A) Cartoon representation of a nanodisc, composed of MSR in green and lipids in grey as well as the hydrophobic protein in orange as example of protein insertion. (https://cube-biotech.com/s-products/nanodisc-products) (B) PA_7 pore state inserted in a lipid nanodisc. The PA_7 hydrophobic tip inserts into the double lipid layer of a nanodisc. Figure adapted from https://www.technologynetworks.com/proteomics/articles/nanodiscs-the-mobile-home-for-membrane-proteins-32691.

The preformed nanodiscs used (Cube Biotech) were composed of a lipid bi-layer disc out of different lipids, like 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine phosphatidylcholine (POPC), which are surrounded and kept in shape by a membrane scaffold protein disc (MSP) (Fig. 3.3) (Gatsogiannis et al., 2016). Nanodiscs have a homogenous size and shape and therefore are suitable for single particle cryo-EM analysis, contributing to the determination of high-resolution structures (Efremov *et al.,* 2017).

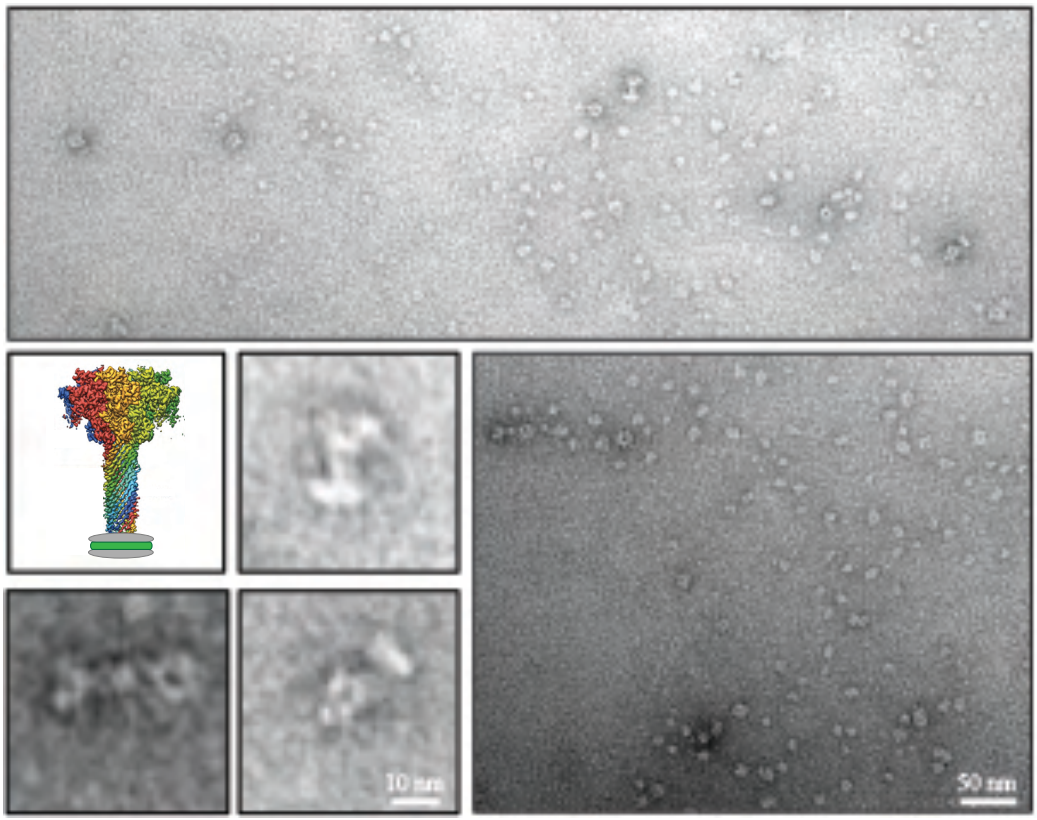


Figure 3.4 : Reconstitution of PA_7 into nanodiscs after pore transition. Representative negative staining electron micrograph of PA_7 reconstituted in lipid nanodiscs (MSP1D1) with individual single particles highlighted in single boxes. 3D-model of PA_7 complex pore state inserted into lipid nanodiscs.

Despite the quite low yield (only few particles inserted into the nanodiscs per micrograph), the PA_7-pore state inserted into the nanodisc was visible by negative staining EM (Fig. 3.4). The typical funnel like shape of PA_7 pore-state inserted into a nanodisc is simplified in a cartoon-like structure (Fig. 3.4). The protocol of the protein reconstitution using nanodiscs was optimized. Different incubation times between the fixed-protein on the Ni-IDA column material, the nanodisc sizes, the buffer pH, the nanodisc types (preformed or not) and the washing buffer were tested (chapter 2.3.6.1).

Liposomes were also used as other membrane-mimicking system (Fig. 3.5). Typically, nanodiscs have all the same structure and dimension (30 nm to several μm) (Akbarzadeh et al., 2013) while liposomes can vary in both. Indeed, liposomes consist in at least one lipid bilayer which is composed of phospholipids. Due to their larger and inhomogeneous size, they are not suitable for single particle cryo-EM analysis, but they are often used in cryo-electron tomography (cryo-TEM). The self-assembled liposomes for PA_7 pore-state reconstitution were prepared with POPC (Fig. 3.5), as reported in the material and methods part (chapter 2.3.6.2).

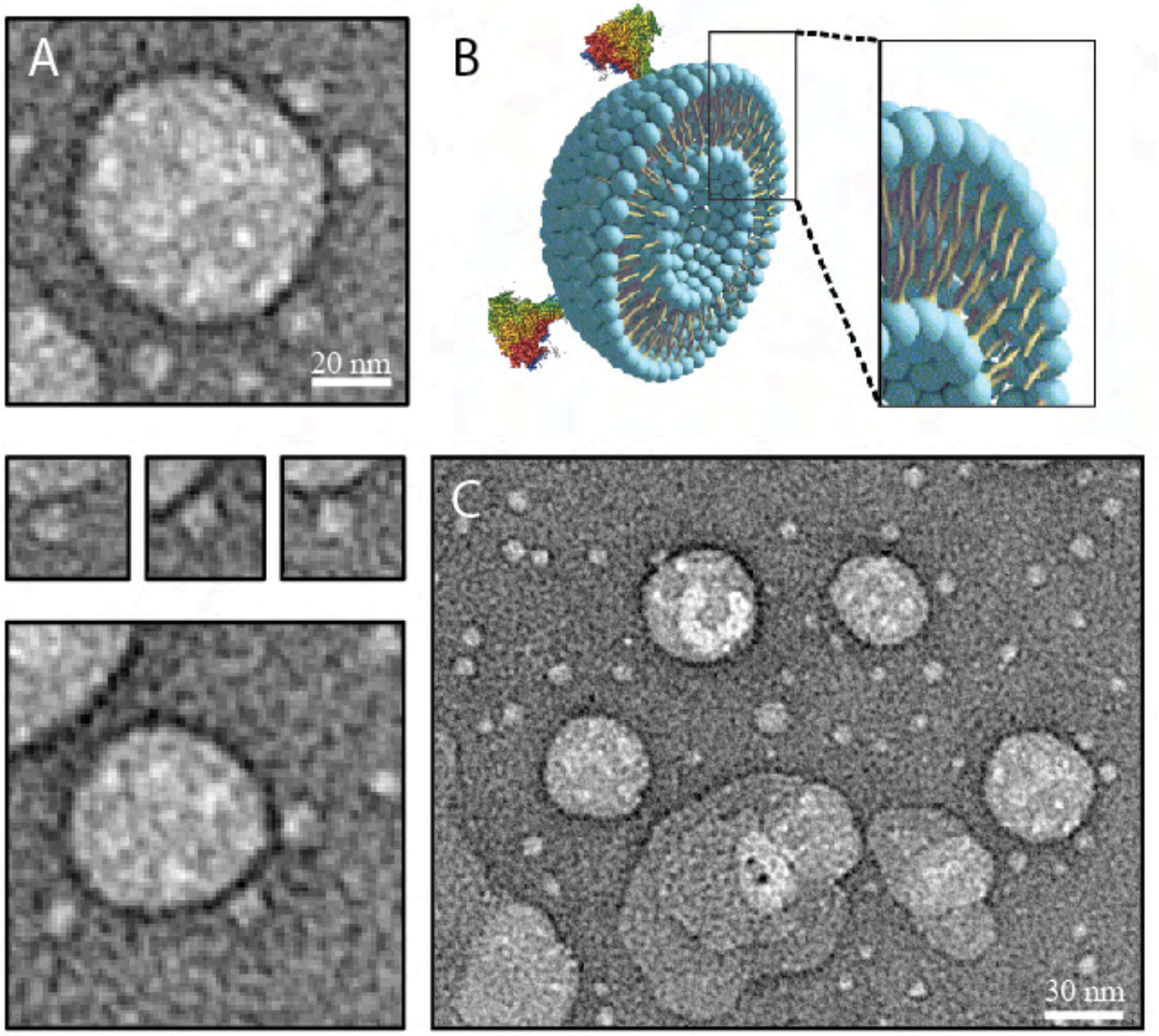


Figure 3.5 : Reconstitution of PA_7 into liposomes after pore transition. (A) Representative negative staining micrographs of PA_7 pore-state inserted into liposomes vesicles. (B) Scheme of liposome structure and simulation of PA_7 pore-state insertion. Highlight of the liposome lipid bilayer structure in the box. (C) Representative micrograph showing several liposomes at different sizes decorated with PA_7 pore-state particles.

Anyhow, the analysis was qualitative and the observed PA_7 pore-state particles inserted into the nanodiscs and into liposomes ensured that the complexes were indeed functional.

Accordingly, to the negative stained micrographs, more PA_7-pore particles were inserted into the liposomes, compared to the PA_7-pore state inserted into nanodiscs (Fig. 3.5). The different yield could be explained because of the different surface tension between liposomes and nanodiscs as well as a different procedure adopted

for the pH shift and incubation, as explained in the relative material and method section (chapter 2.3.6.1 and 2.3.6.2).

In order to increase the yield of PA pore-state inserted into the nanodiscs, a different immobilizing method should be adopted. The high affinity between EF and LF with PA could be used to functionalize the Ni-beads (Narahari et al., 2015), obtaining a more stable anchoring system.

3.2.3. PA_7LF_3 complex formation

After PA_7 assembly and verification of its integrity both in the pre-pore (Fig 3.2) and pore state (Fig. 3.4 and 3.5), LF was added to PA_7 solution to form the PA_7LF_3 complex. Indeed, PA_7 was incubated with LF (List Biological Lab) at a 1:10 molar ratio, and left on ice for at least 2 h. After incubation, PA_7LF_3 complex was isolated and purified from the excess of LF and other residual proteins via SEC (Fig. 3.1 D). The elution profile showed two relevant peak s (Fig. 3.6). Peak 13 represented the elution of the PA_7LF_3 complex mixed to aggregates (Fig. 3.6-A and B), whereas peak 14 represented the pure PA_7LF_3 complex elution (Fig. 3.6-A and C). The earlier peak 12 corresponds to the void volume (Fig. 3.6-A).

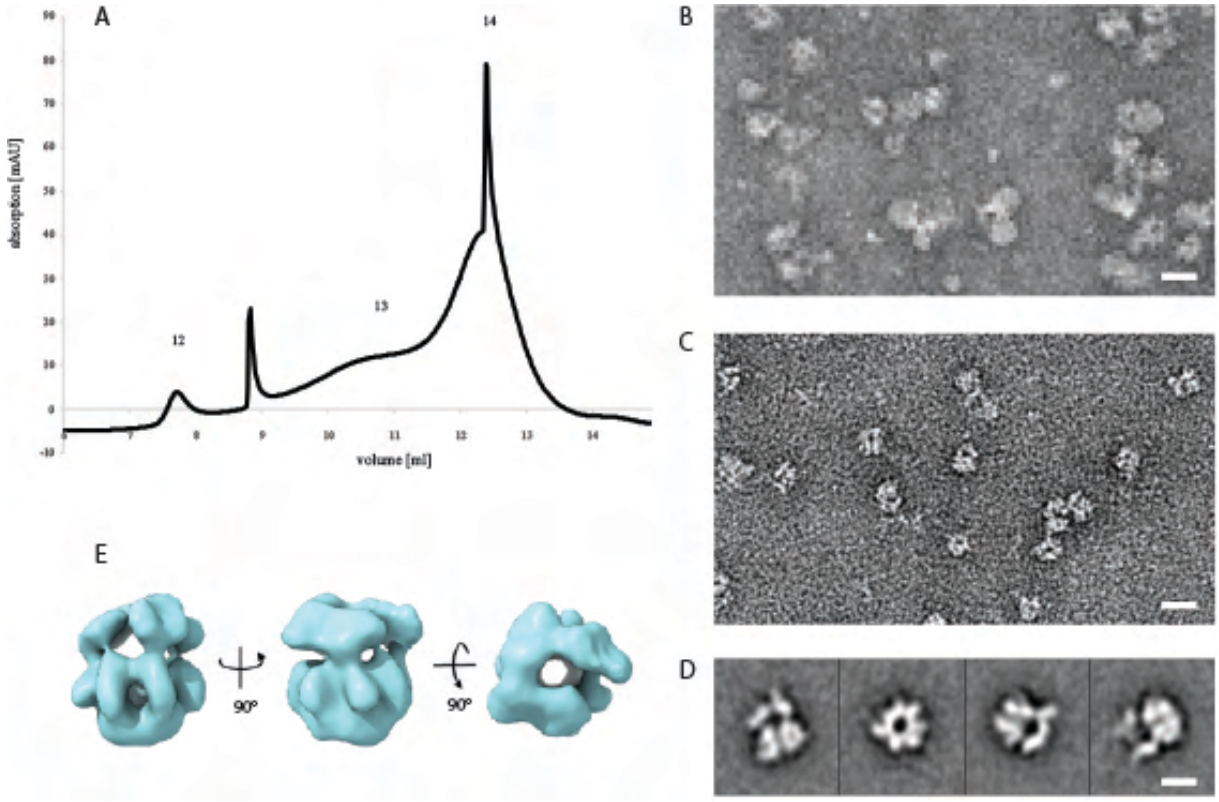


Figure 3.6 : PA_7LF_3 SEC chromatogram, negative staining EM analysis and first 3D-reconstruction. (A) SEC chromatogram of the PA_7-LFs complex. Peak 12 corresponds to the void volume. (B) Peak 13 negative stained micrograph reports the PA_7-LFs mixed with aggregates. Scale bar 25 nm. (C) Peak 14 negative stained micrograph shows well defined particles of PA_7-LFs complex. Scale bar 25 nm. (D) Representative 2D classes of PA_7LF_3 as side and top views, obtained with ISAC. In addition, the LFs are particularly visible in the side view classes. (E) First 3D-negatively stained reconstruction of the PA_7LF_3 complex. The bulky PA_7 ring sits at the bottom of the structure whereas 3 LFs sit on top of the PA_7 ring.

An initial negatively stained data set of the fraction 14 was collected and processed in SPHIRE. The 2D-classes (Fig. 3.6-D) confirmed the high purity of the protein purification. Indeed, the picked particles had expected dimensions and the LFs were well visible on top of the PA_7 (Fig. 3.6-C). The particles sit on the carbon layer of the negatively stained grid in random orientations (Fig. 3.6-C and D. In the magnified boxes (Fig.3.6-D) some particles orientations are highlighted. In the typical bottom view, PA_7LF_3 appear undecorated, showing the central channel free. In the side views, two additional densities protrude from the PA_7 ring, corresponding to the LFs which sits on top of the PA_7. The 3D model gives an initial indication about the LFs disposition on PA_7. As already mentioned in the previous chapters, the negative stain analysis provided fundamental indication about the quality of the protein preparation, but also

about the stability of the complex and its stoichiometry. According to this reconstruction, the complex had a jellyfish shape with on-top-densities relative to the 3 LFs molecules. This led to the assumption that the C7-symmetric PA_7 is broken when PA_7LF_3 is formed, generating a C1 hetero-complex. The 3D reconstruction had a final resolution of around 19 Å based on the 0.5 criterion.

3.2.4. Cryo-EM on PA_7LF_3

While negative stained EM reveals first general aspects of protein structure, cryo-EM analysis enables more detailed (on a near atomic level) view of the target protein. Therefore, after having performed negative stained EM analysis, the PA_7-LF_s complex was induced (Fig.3.6) and therefore analysed via cryo-EM.

A droplet of protein solution was applied on additional carbon-coated cryo-grids and then vitrified in a thin layer of amorphous ice (chapter 2.4.1.6). Collection of several thousands of images followed by data processing revealed that what was supposed to be a single complex state, was instead a multiple one (Fig. 3.10 and 3.11).

3.2.5. Vitrification of PA_7LF_3

With the final goal of obtaining well-distributed particles embedded in amorphous ice, the vitrification step was optimized. Noteworthy, the protein concentration and the blotting conditions have a great impact on the ice homogeneity and distribution.

Initially, the PA_7-LFs complex was loaded on holey carbon grids. The cryo-screening confirmed what was already published by Fabre et al. 2016. The particles have a strong predominance to adapt the same orientation on ice.

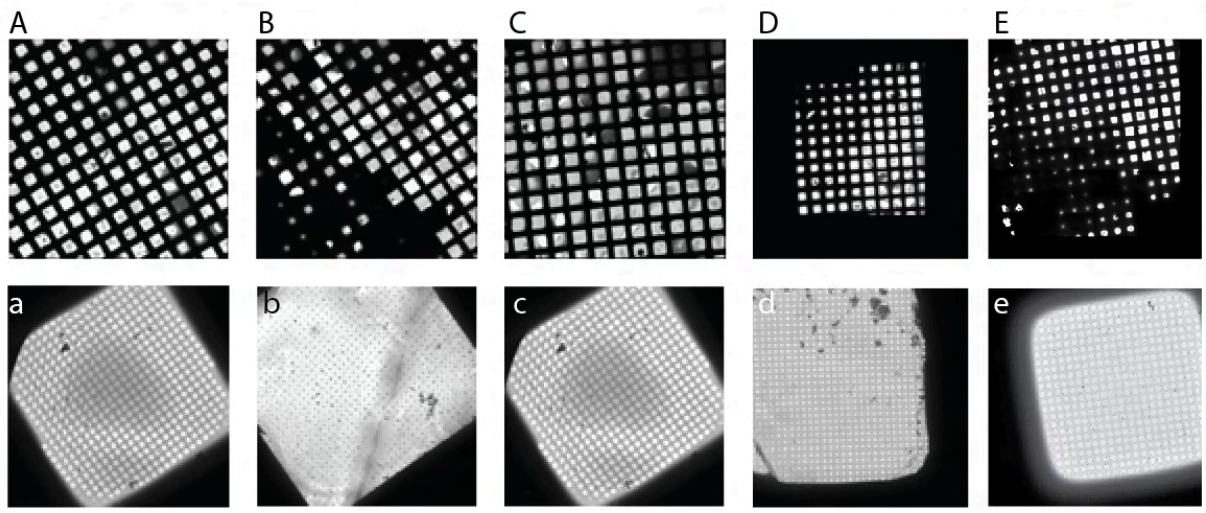


Figure 3.7 : Atlas and some example of a cryo-grid square imaged during cryo-EM screening at Arctica EM. (A-a) CF (R2/1) with additional 2 nm carbon layer: atlas and relative square as an example. The atlas shows darker spots of extremely thick ice next to totally dried squares. The image of the the square shows the high difference of ice thickness between its center and the edges. (B-b) CF 1/2 with additional 1 nm of carbon layer (made in-house): atlas and one square as an example. The 1 nm additional carbon grids were generally more fragile compared with the 2 nm ones. As the atlas shows, the grid did not have a progressive ice thickness gradient. Many squares resulted empty even with extremely short blotting times (< 1s). (C-c) Qf 2/1 with additional 2 nm additional carbon layer. The squares had a central thicker ice block, missing a smooth ice gradient, all over the grid. (D-d) Qf 1/2 with 1 nm additional carbon layer. The carbon layer was too thin and broke up after few seconds of imaging at the EM. A thicker carbon layer showed the same problem as in (A). (E-e) gold Qf 2/1 with additional 2 nm of carbon. The atlas shows a well distributed ice gradient all over the grid. The bigger squares have a thinner ice layer while the smaller squares a thicker one.

Therefore, the choice turned toward grids with additional support that could prevent the particles to orient on preferential sites. Generally, once the grid type was chosen, two main aspects are extensively tuned: (i) particles concentration and (ii) ice thickness/quality. The (i) final particle concentration and distribution on a cryo-grid was also dependent on the characteristics of the cryo-grid itself. Normally, the additional carbon layer of the cryo-grid is rapidly saturated by the particles. Typically, the particles concentration is at almost same order of magnitude as in negative stain. This advantage was fundamental, because when concentrating the sample (> 0.3 mg/ml), it tended to aggregate. Despite this great positive aspect, the 2 nm of carbon layer reduces the overall image contrast. Notably, the ice thickness and the protein

concentration are in an interdependent relation that is why those two parameters were extensively optimized.

The general problem faced in this study was the thickness and distribution of the vitrified ice (Fig.3.7). The ice distribution was not homogeneous. The cryo-micrographs showed areas with either dried or too thin ice that rapidly melted during imaging. Often, the areas with thin layer of ice were immediately next to areas with extremely thick ice (case observed also in the same grid hole) (Fig. 3.7 a-c). No features were visible underneath the ice block, with an ice-film thickness significantly greater than the size of the particles, thus resulting in loss of useful signal in the image (Glaser et al. 2016). This situation persisted changing the blotting time, the chamber humidity of the plunger and was independent on the grid type (Fig. 3.7). Only few holes in between the thick and ice-dried parts were suitable for data collection. Nevertheless, even if some holes has an ice thickness suitable for structural studies, the micrograph quality was not good enough. Indeed, the ice was often contaminated (ice crystals).

Using the self-coated holey carbon grids, with different carbon layer thickness (1 nm, 2 nm, 3 nm, 4 nm), the ice homogeneity slightly improved (Fig. 3.7-B-b). This results in more holes potentially suitable for data collection. But still, squares with generally patchy ice were observed and were not good enough for data collection.

In order to make a more uniform ice layers, the grid (CF or Qf) was pre-treated with tween-20. Before applying the sample, a drop of complex buffer with 0.05% tween-20 was applied on the grid, incubated 45 s and blotted away as usual (chapter 2.4.1.6). After 10 min, the protein complex was applied on the grid surface and the grid was prepared in the conventional way (chapter 2.4.1.6). The grid was plunged and afterwards imaged at the EM, but the general Atlas and relative micrographs did not show any improvement at the ice level.

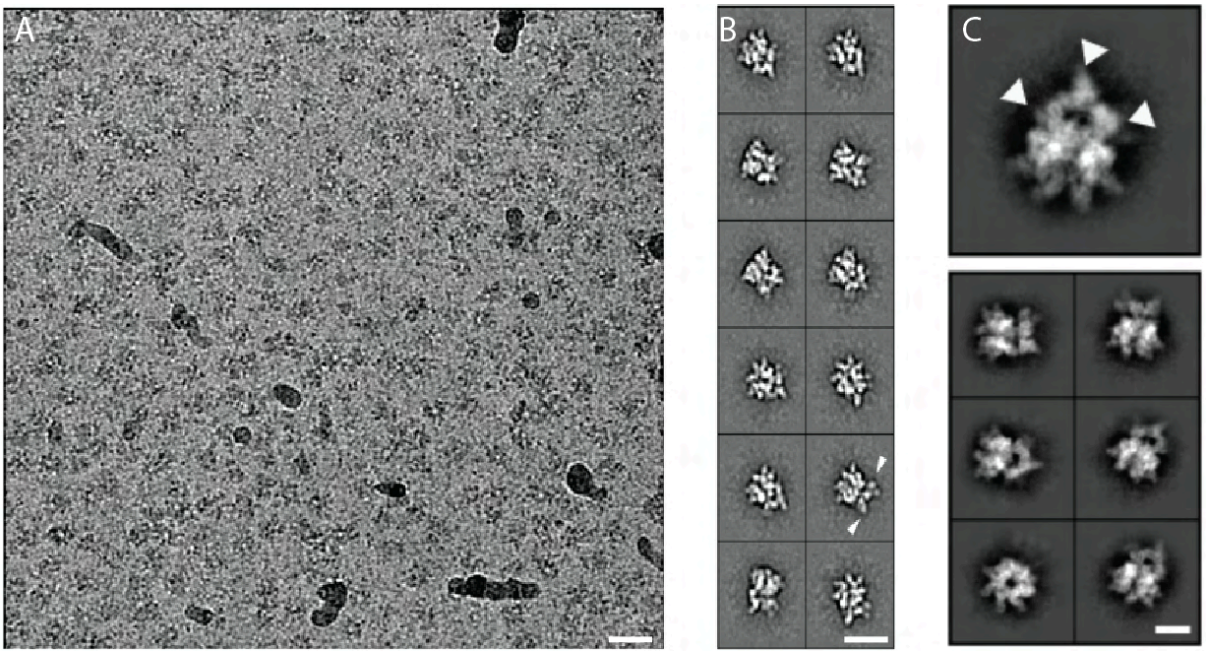


Figure 3.8 : PA_7LFs cryo micrograph and 2D-classes. (A) Representative cryo-EM micrograph of the PA_7LFs data set collected at Krios 300 kV. The particles sit predominantly on the side, with few examples on the top and bottom view. Scale bar 40 nm. (B) Representative 2D-classes showing different views. The white arrows indicate the LFs sitting on top of PA_7 ring. Scale bar 20 nm. (C) Representative Beautified 2D-classes showing different views. The white arrows indicate the three LFs sitting on top of PA_7 ring. Scale bar 15 nm.

UltrAuFoil R2/1 2nm carbon, usually used in cryo-EM, present 2 layers of gold covered with additional 2nm of carbon. Instead, the QuantiFoil R2/1 with 2nm carbon have a single gold layer, covered with a standard carbon layer and an additional 2nm carbon on top (chapter 2.4.1.6). These gold grids presented a better ice distribution compared to the classic additional carbon grids tested. Despite that, the best result was reached using Qf-gold R 2/1 with 2nm of additional carbon layer (Fig.3.7-E-e). Those grids were usually used as support for cell growth and subsequent preparation for tomography analysis. The ice distribution had gradients on the grid and into the squares, and tuning with the usual parameters the ice thickness obtained was suitable for cryo-EM data-collection. Once the ice thickness was optimized, the particles were clearly visible (Fig. 3.8-A). The particle concentration was adjusted in order to get a uniform protein distribution which adheres on the carbon layer (Fig.3.8-A). As

mentioned before, although the grids with additional carbon reduce the overall image contrast in general, the particles here appeared crispy, easily recognizable without need of any filter-tool (Fig. 3.8 A).

3.2.6. Processing and 3D structure of PA_7LF_3

Following the cryo-EM grid optimization and preparation, the grids were imaged at the TITAN Krios 300 kV microscope (details in the material and methods section 2.4.1.6, Tab. S1).

During data acquisition, TRANSPHIRE, a fully-automated on-the-fly processing pipeline that simplifies and accelerates the data-collection (Stabrin *et al.* 2020) was used. It allowed an easy handling of big data-sets as well as a direct monitoring of key parameters (CTF estimation, particle picking, initial 2D clustering and 3D refinement in parallel to image recording). At the end of micrograph collection and transfer, a manual selection discarded low-quality micrographs as well as the ones with high drift, contamination or with broken holes. This finally results 5238 clean micrographs from the three independently collected data sets (Fig. 3.9).

A total of 10 micrographs was picked manually using boxmanager.py and the selected particles were used as a template for the automated picking with crYOLO (Wagner et al., 2018, Wagner et al., 2020). Initially, the three datasets with a total of 390 k particles were independently analyzed giving three separated sets of ISAC 2D-classes (Fig. 3.9). The independent 2D classes showed a comparable quality among the datasets and therefore, the relative particles were merged after 2D-class selection. This led to a total of 217 k final particles. The selected 2D classes were later used in the beautify run (Fig. 3.8-C). In SPHIRE, the beautify provides the 2D-classes at their original pixel size, increasing the contrast and highlighting high-resolution features (Fig. 3.8-C).

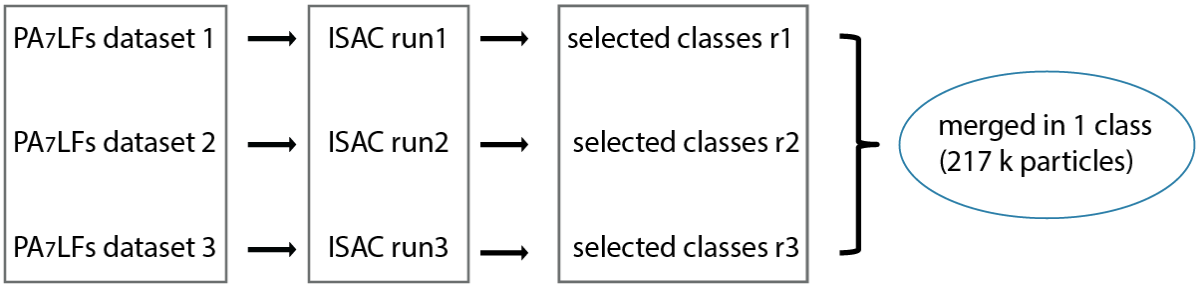


Figure 3.9 : Processing pipeline of the 3 PA_7LF_3 complex datasets. Initially, the three datasets were processed separately. The 2D-classes resulted from ISAC runs, were sorted and the best looking 2D-classes were selected and merged in one single pdb-file.

All the single particle analysis steps were performed in SPHIRE (Moriya et al., 2017), supported by Chimera for visual presentation (Pettersen et al. 2004). After 2D classification, the selected particles were included in the 3D-refinement (Meridien) (Fig. 3.10). Usually the Meridien run is preceded by 3D-model generation with VIPER. Unfortunately, this usual approach did not result in a reasonable model. Most probably, the destroyed symmetry of a bulky PA_7 from the 3 LFs, caused the failure of the model generation.

Hence, the 3D-reference used during data analysis was a Chimera generated one, obtained combining different published atomic structures. Indeed, PA_7 crystal structure (PDB: 1TZO) was decorated with 3 full-length LFs (PDB: 1J7N), which N-terminal domains were docked to PA_7 based on the PA_8LF_4 structure (PDB: 3KWV). This chimeric structure was then converted into an electron density (sp_pdb2em functionality in SPHIRE). The resulting map was filtered to 30 Å and used as reference during 3D-refinement in SPHIRE.

Using the hybrid reference created in Chimera, the 217 k particles went through a 3D-refinement and resulted in a reconstruction (Fig. 3.10-A), at an overall resolution of 3.5 Å. However, the density map corresponding to the LFs area was partially unassignable. This could be explained as (i) a result of a symmetry mismatch after the 3 LFs bind on the PA_7 upper rim, or due to (ii) a different conformation adopted by the LFs bound to PA_7.

3D-classification on the refined volume A helped clarifying this hypothesis by generating 3 classes: class B, C and D (Fig. 3.10 B, C and D). The 3D-class C and D differed mainly on the overall resolution of the LFs, in which C had a slightly better defined LFs densities. Instead, the 3D-class B showed a different LFs resolution gradient and LF disposition compared to C and D. Indeed, LF_1 had the lowest resolution instead of the LF_3, as observed in map C and D. With these settings, the 3D-class B could not overlap on C and D. After multiple assays, 3D-class B superimposed to C and D after an anti-clockwise rotation of 103°. With this rotation, it became clear that, what was initially supposed to be LF_1, was in reality LF_3. Therefore, LF_3 from 3D-class B resulted in a new binding position on PA_7, which could not be observed on C and D.

The 3 particle stacks from rotated-3D-class B and 3D-classes C and D were merged, respecting the new orientation. Before analysing all the LFs binding positions, LF_3 was masked out to improve the resolution on the rest of the complex, using the resulting structure as reference for further processing steps. The 3D-refinement resulted in a final 3.5 Å map, named $PA_7LF_{3masked}$ (Fig. 3.10, map E).

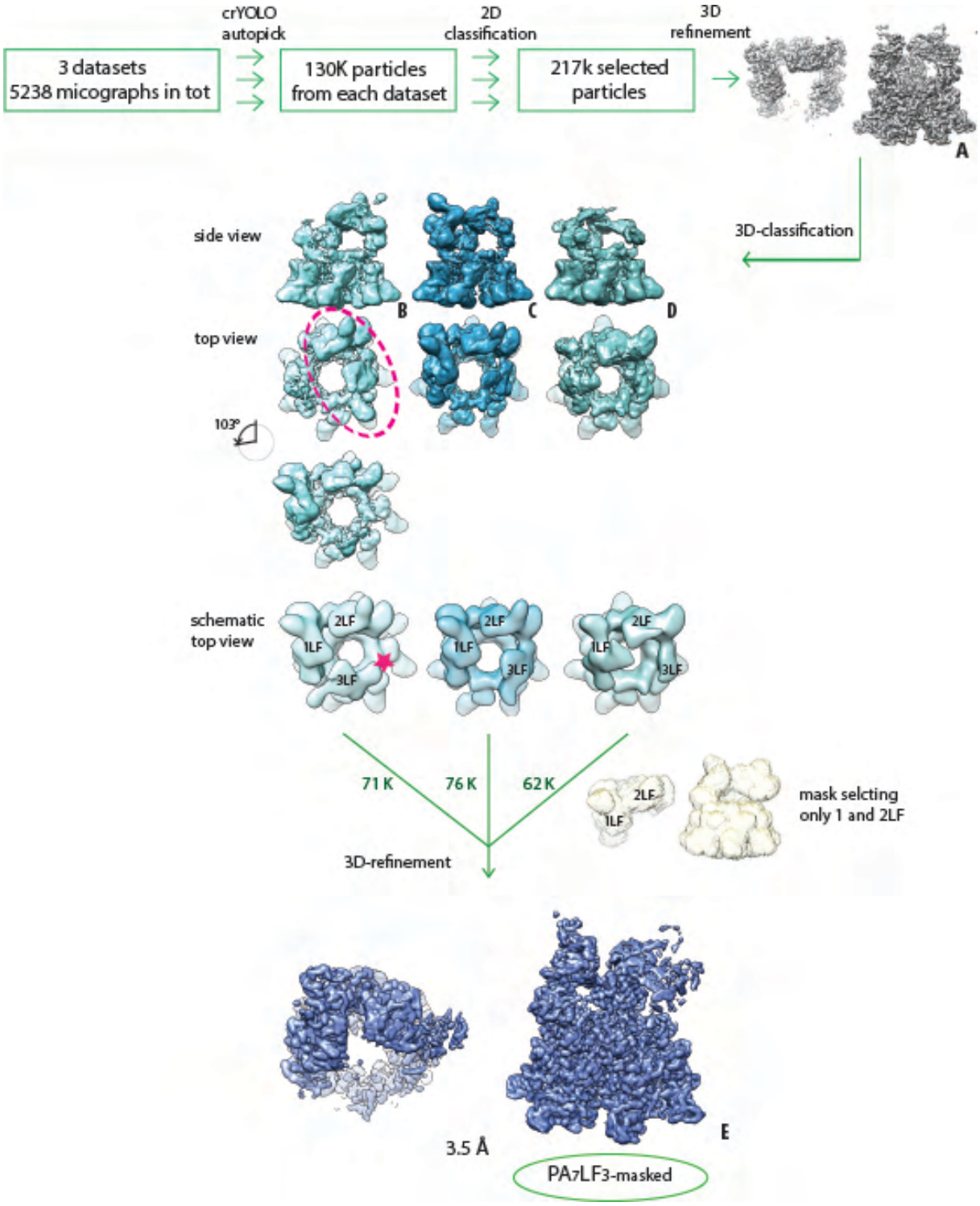


Figure 3.10 : Flowchart of the processing pipeline generating $PA_7LF_{3\text{-masked}}$ EM electron density (first part, second part follows in Fig. 3.11). The single particle processing workflow shows an initial 3D-classification and the subsequent rotation by 103° of class B to match 3D-classes C and D. The subsequent 3D-refinement included a mask selecting only LF_1 and LF_2, generating $PA_7LF_{3\text{-masked}}$. The pink circle highlights the "rotated" disposition of LF_1 and LF_2, whereas the pink star indicates the LF_3 binding position in 3D-class B.

In order to separate the two possible binding positions of LF_3, particles from $PA_7LF_{3masked}$ undergo through a new 3D-classification, using a focusing mask on the 3 LFs. This generated two new 3D-classes, named F and G (Fig. 3.11). These two 3D-classes and relative 3D-refined density maps (H and I) exactly pointed out the two possible LF binding positions. The 3D-class G presented the LF_3 in the same position as observed in B, naming the 3D-refined density map I, PA_7LF_{2+1B} (3.8 Å, Fig. 3.11). Instead, the other 3D-refined density H, obtained from 3D-class F, was not clearly resolved. Therefore, the particles of the 3D-refined density H undergo through an additional 3D-classification run. The obtained two additional 3D-classes (L and M) confirmed the hypothesis and highlighted the two possible conformations adopted by LF_3. Indeed, LF_3 in L and M adopted the same binding position as previously showed in C and D. This time, the difference was at the C-terminal domain. In the 3D-class M, the C-terminus LF_1 and LF_2 pointed inward the PA lumen whereas in the 3D-class L, the C-terminal domains did not. The 3D-refined electron density relative to the 3D-class L was named $PA_7LF_{(2+1A)}$ (4.2 Å), whereas from the 3D-class M, $PA_7LF_{(2+1A')}$ (4.3 Å, Fig. 3.11).

Finally, PA_7LF_3 complex revealed to be a mixture of 3 conformations $PA_7LF_{(2+1B)}$, $PA_7LF_{(2+1A)}$, and $PA_7LF_{(2+1A')}$, resolved at final resolutions of 3.8 Å, 4.2 Å and 4.3 Å, respectively (Fig. S2). The resolved electron densities differ not only in the position of LF_3 bound to PA_7 ($PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$) but also in LF C-terminal conformation ($PA_7LF_{(2+1A')}$).

$PA_7LF_{3\text{-masked}}$ was the best resolved density map. The fitting of the atomic model and the subsequent refinement of the fitted atomic structures (details in the material and methods chapter 2.4.2.6) resulted in an overall resolution of 3.5 Å. The relative atomic model served as the first template for the fitting in map $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$. Individual atomic structures fitting and relative refinements were independently calculated (chapter 2.4.2.6) and the results are listed in Tab. S1.

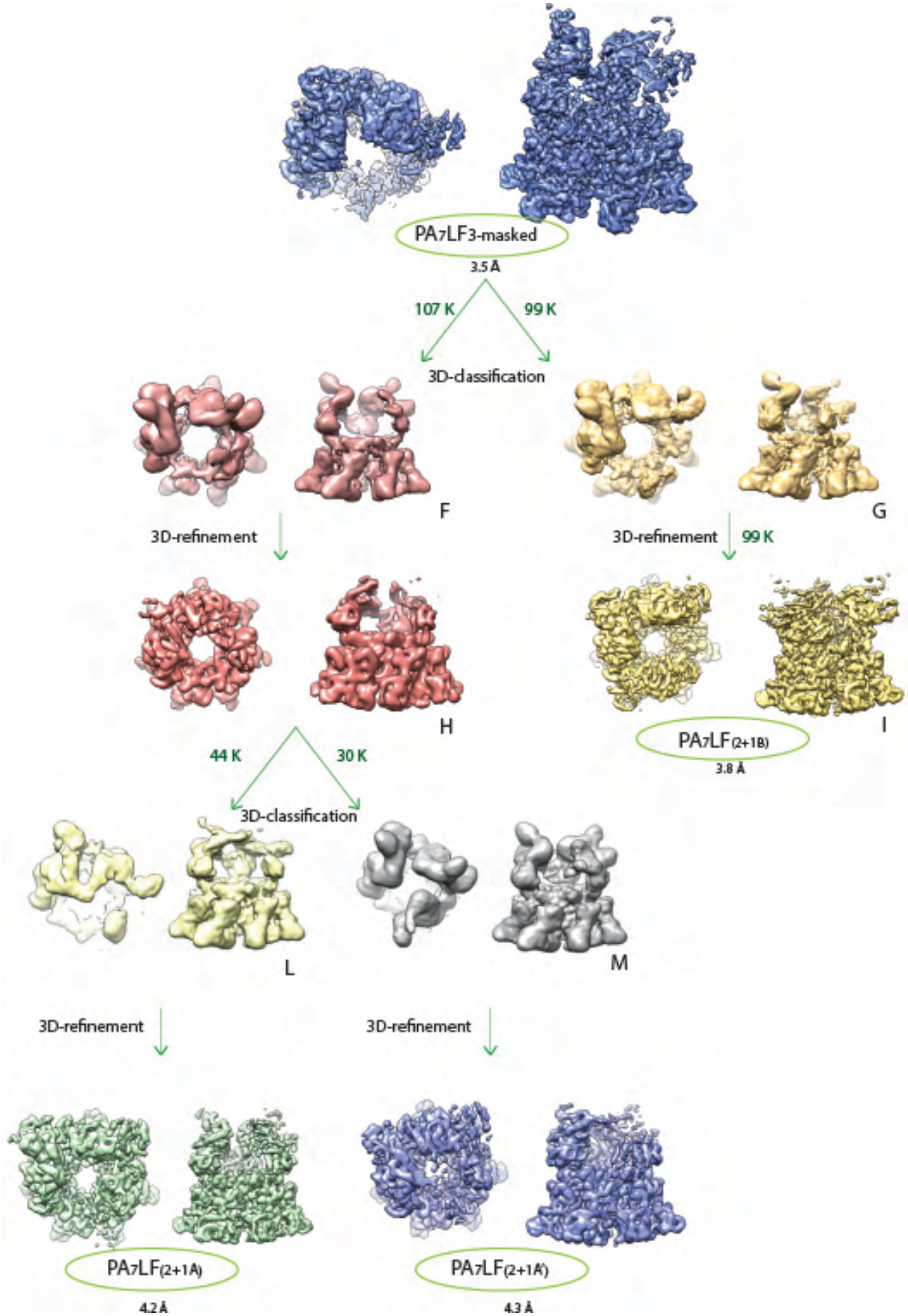


Figure 3.11 : Flowchart of the processing pipeline starting from $PA_7LF_{3\text{-masked}}$ that generates $PA_7LF_{(2+1B)}$ $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1A')}$ electron densities (second part). Several 3D-classification cycles allowed to

highlight the intricated series of hidden PA_7LF_3 structures. A 3D-classification starting from the particles of the $PA_7LF_{3\text{-masked}}$ structure, revealed the presence of additional two classes, named F and G, which generated the relative 3D-refined volumes H and I ($PA_7LF_{(2+1B)}$). After 3D-classification, the 3D-refined volume H showed the presence of other two possible complex assemblies (L and M), which 3D-refined volumes generated $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1A')}$.

3.2.7. Model building, refinement and validation

The $PA_7LF_{3\text{-masked}}$ model building started from PA_7. A single protomer from PA_8 crystal structure (PDB: 3HVD) was initially used as model and fitted into the PA_7 density, by rigid body fitting in Chimera (Pettersen et al. 2004). Afterwards, the LF crystal structure (PDB: 1J7N) was fitted into the two LFs electron densities, also via rigid body fitting in Chimera. Initially, the LF crystal structure sequence was divided in fragments based on the local resolution values (Fig. 3.12). LF_1 and LF_2 sequences were manually refined from amino acid 52 to 253 (N-terminal) and from amino acid 550 to 600 (C-terminal). LF_1 and LF_2 N-terminal regions were better resolved, compared to the relative C-terminal parts. The binding of the LF to the PA_7 made the LF N-terminal area more stable, impacting also on the final resolution since more flexible regions generally show a lower resolution. Contrary to the other LFs, the N-terminal domain of LF_2 was manually refined from amino acid 32 (instead then amino acid 52). LF_2 N-terminal electron density from amino acid 32 to 52 protruded into the PA_7 lumen ring (Fig. 3.19), as also described from Feld et al. 2010.

Instead, the remaining LFs electron density fragments were too poorly resolved to allow a correct fitting and are not discussed in this thesis. Indeed, the central region showed patchy electron densities and moving towards the C-terminal part, the electron density became less continuous and a correct fitting was difficult. The LF_3 was the poorest resolved reconstruction, because of its high flexibility at its un-bound C-terminal part.

The resulting $PA_7LF_{3\text{-masked}}$ first model was flexibly fitted using iMODFIT (Lopéz-Blanco & Chacón 2013) and further refined, optimized with a combination of manual model building in COOT (Emsley et al. 2010) and real-space refinement in the PHENIX software package (Terwilliger et al. 2008). Unresolved loop regions were deleted and side chain information removed when the EM density was less well-resolved. Geometries of the final refined models were obtained either from PHENIX or calculated with the Molprobity server (Williams et al. 2018).

3.2.8. Local Resolution

Resolution is the most important parameter that informs about the quality of an electron density map obtained via cryo-EM analysis (Vilas et al 2018). The FSC (chapter 2.4.2.5) is used for estimating the global resolution of an EM map but it does not give information about the variance in the different areas of the map.

In the four maps presented, the resolution is higher at the PA_7 core structure and lower at the complex periphery, due to an increased exposition to the solvent. The resolution variation becomes clearer observing the LFs. Indeed, the LF N-terminus binds the PA_7 thus providing higher stability to the LF and PA structure. Whereas, the bulky LF C-terminal part has a minimal interaction with the adjacent LF therefore showing higher flexibility. Among the LFs, LF_1 is the best resolved while LF_3 presents the weakest density and therefore the worst resolution (Fig. 3.19).

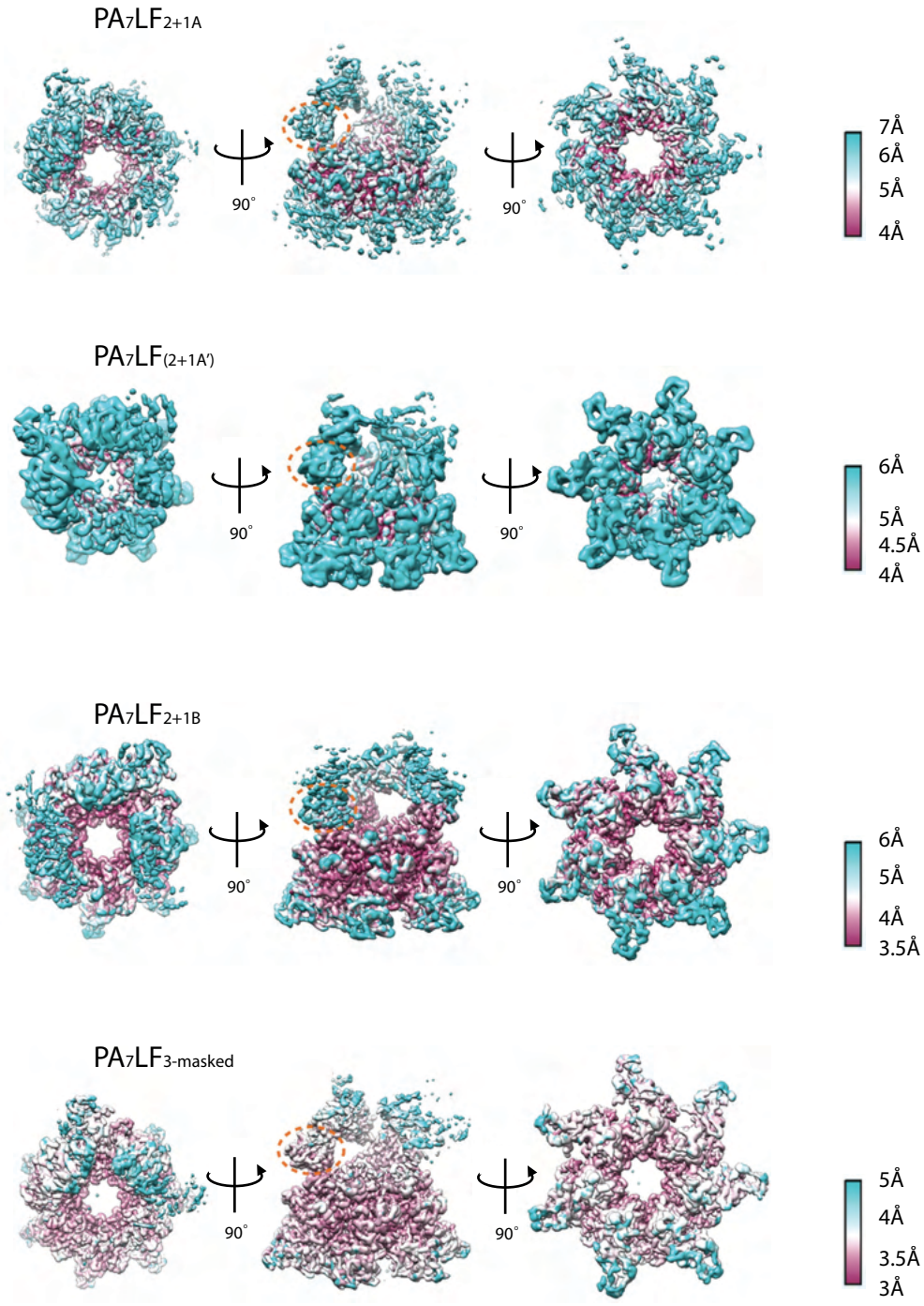


Figure 3.12 : PA_7LF_{2+1A}, $PA_7LF_{(2+1A)'}$, PA_7LF_{2+1B}, and $PA_7LF_{3masked}$ local resolution. The structures are oriented from top to bottom, side, and bottom to top view, all coloured by local resolution. The reference of the colours was provided at the scale bar legend on the bottom right corner of every complex structure.

3.2.9. Structural analysis and comparison between $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$

Initial structural studies began comparing PA_{63} protomers from PA_7 of $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$ complex. The seven PA_{63} protomers belonging to the same complex structure, were superimposed (Fig. 3.13 A-B-C). The superimposition showed that in all three complexes ($PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$), the protomers belonging to the same complex almost perfectly overlapped on each other, with only little chain movements (Fig. 3.13 A-B-C). Independently from the complex analyzed, the PA_{63} protomers did not show major rearrangements (Fig. 3.13 A-B-C). Furthermore, a single PA_{63} protomer from $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$ was chosen and superimposed to the PA_{63} protomer from the PA_7 crystal structure (PDB: 1TZO) (Fig. 3.13-D). This superimposition showed an almost perfect alignment of all structures with not relevant movements (Fig 3.13 D). Noteworthy, the loop region 2β2-2β3 (residues 300–323), which is implicated in pore formation was only resolved in the crystal structure but not in $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$. This indicates a high flexibility of the loop (Fig 3.13 D, highlighted with an arrowhead), which is in line with previous MD simulations (Alisaraie et al 2016). The high flexibility of this loop could represent the preamble of the pore state formation. During model building, the loop was indeed deleted from the three structures because of the poor resolution of the relative EM map. The superimposition showed that binding of the LFs did not induce to any major structural rearrangement of the PA_7 monomers in the $PA_7LF_{3\text{-masked}}$ PA_7LF_{2+1A} and PA_7LF_{2+1B} complexes (Fig. 3.13). In contrast, Ren et al. 2004 suggested that LF binding results in a distortion of the symmetric PA_7 ring, thereby facilitating the passage of cargo through the enlarged lumen. This suggestion is not confirmed by $PA_7LF_{3\text{-masked}}$ PA_7LF_{2+1A} and PA_7LF_{2+1B} structures (Fig. 3.13 A-B-C).

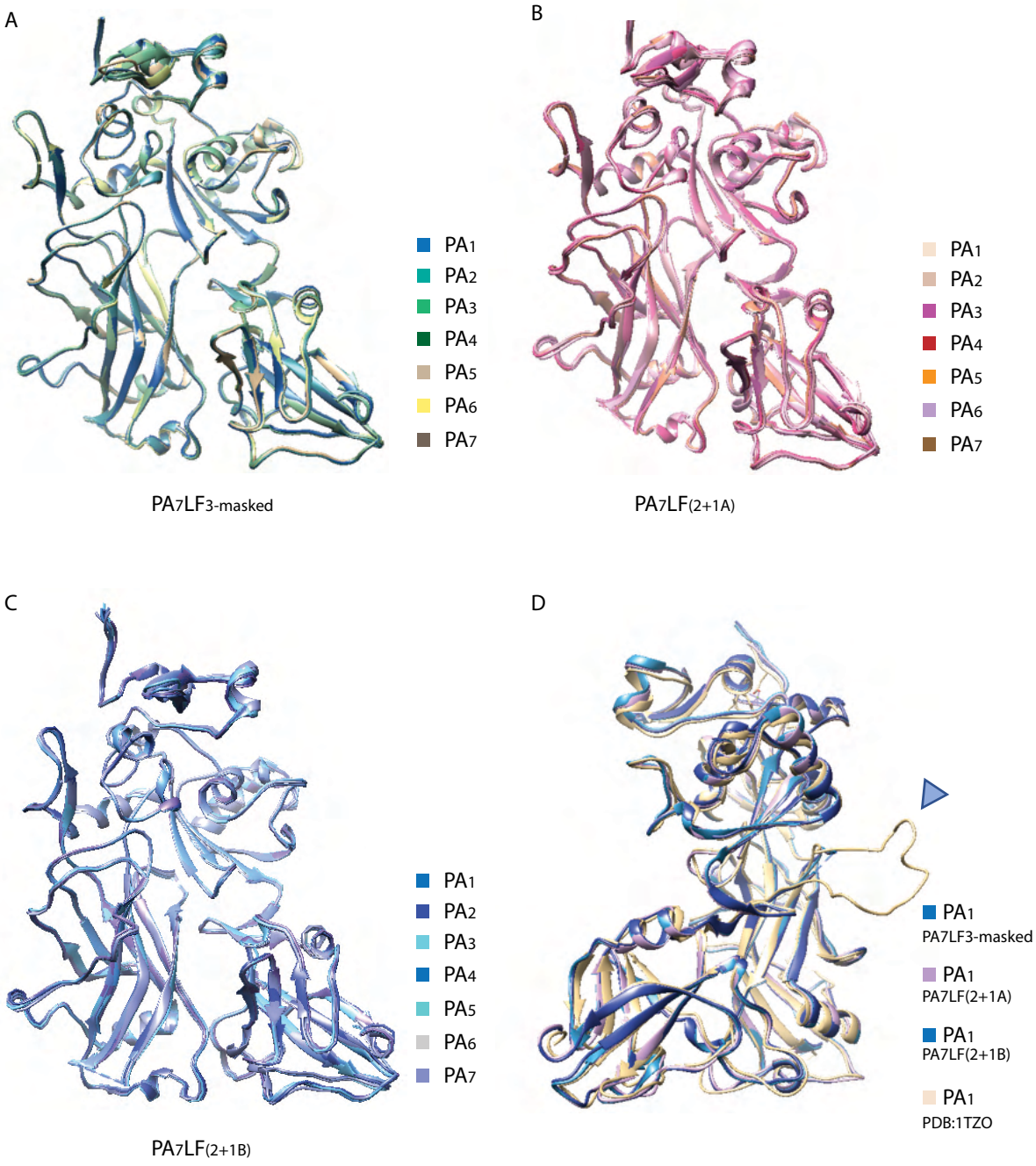


Figure 3.13 : Structural comparison between the PA_{63} protomers of $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2\text{-}1A)}$ and $PA_7LF_{(2\text{-}1B)}$ and between a single monomer of PA_{63} from each complex and a PA_{63} protomer from PA_7 crystal structure (PBD: 1TZO). (A) Superposition of the seven PA_{63} protomers of $PA_7LF_{3\text{-masked}}$, which are colored in different green hues. (B) Superposition of the seven PA_{63} protomers of $PA_7LF_{(2+1A)}$, which are colored in different pink hues. (C) Superposition of the seven PA_{63} protomers of $PA_7LF_{(2+1B)}$, which are colored in different blue hues. (D) Superposition of the PA_{63} protomer from $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$ and from the PA_7 crystal structure (PDB: 1TZO). The superimposition is rotated by 90 ° on the y axes to show the loop region 2β2-2β3 (residues 300–323) resolved only in the PA_{63} protomer from the crystal structure (PDB: 1TZO), highlighted with an arrowhead.

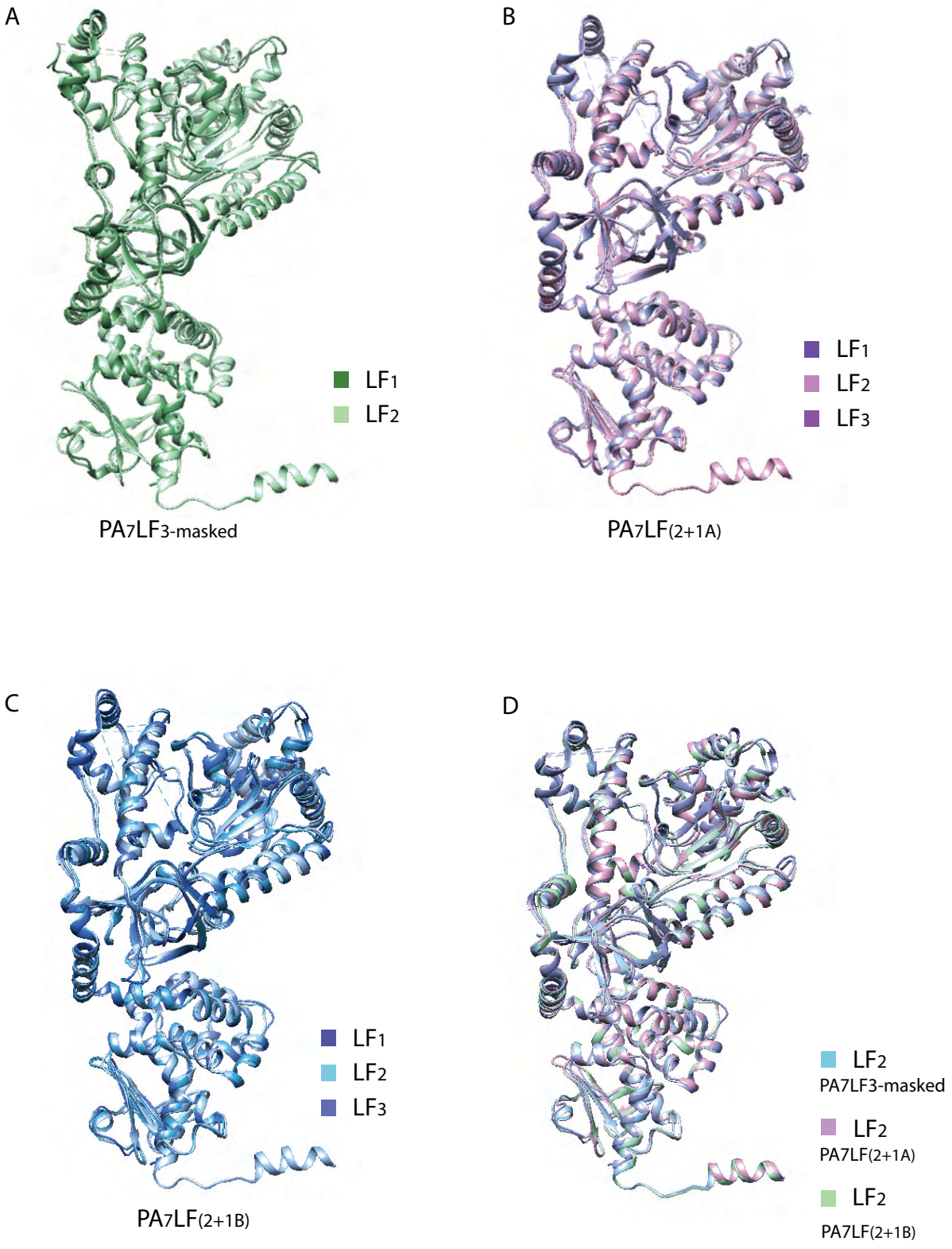


Figure 3.14 : Structural comparison between the LFs from $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2\text{-}1A)}$ and $PA_7LF_{(2\text{-}1B)}$ structures. (A) Superimposition of LF_1, LF_2 and LF_3 from $PA_7LF_{3\text{-masked}}$, in different green hues. (B) Superimposition of LF_1, LF_2 and LF_3 from $PA_7LF_{(2+1A)}$, in different pink hues. (C) Superimposition of LF_1, LF_2 and LF_3 from $PA_7LF_{(2+1B)}$, in different blue hues. (D) Superimposition of LF_2 from $PA_7LF_{3\text{-masked}}$ (light blue), $PA_7LF_{(2+1A)}$ (pink) and $PA_7LF_{(2+1B)}$ (light green).

Despite the different LF binding positions determine the generation of different PA_7LF_3 complexes ($PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$), superimposition of the LF_1, LF_2 and LF_3 from the same complex showed that LF structure organization does not vary. Indeed, both the C-terminal and the N-terminal LF domains from LFs overlapped with minimal movements of the chains (Fig. 3.14 A, B, C). Furthermore, the three LF_2 belonging to $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$ were superimposed to verify if LFs from different complexes assemblies have differences in the structure. The three LF_2 atomic models perfectly overlapped with only minor movements (Fig. 3.14 D), showing that LF_2 adopts the same conformation in all three complexes. Furthermore, LF_2 from $PA_7LF_{3\text{-masked}}$ was used as reference and overlapped with the unbound LF crystal structure (PDB: 1J7N), (Fig. 3.15 A).

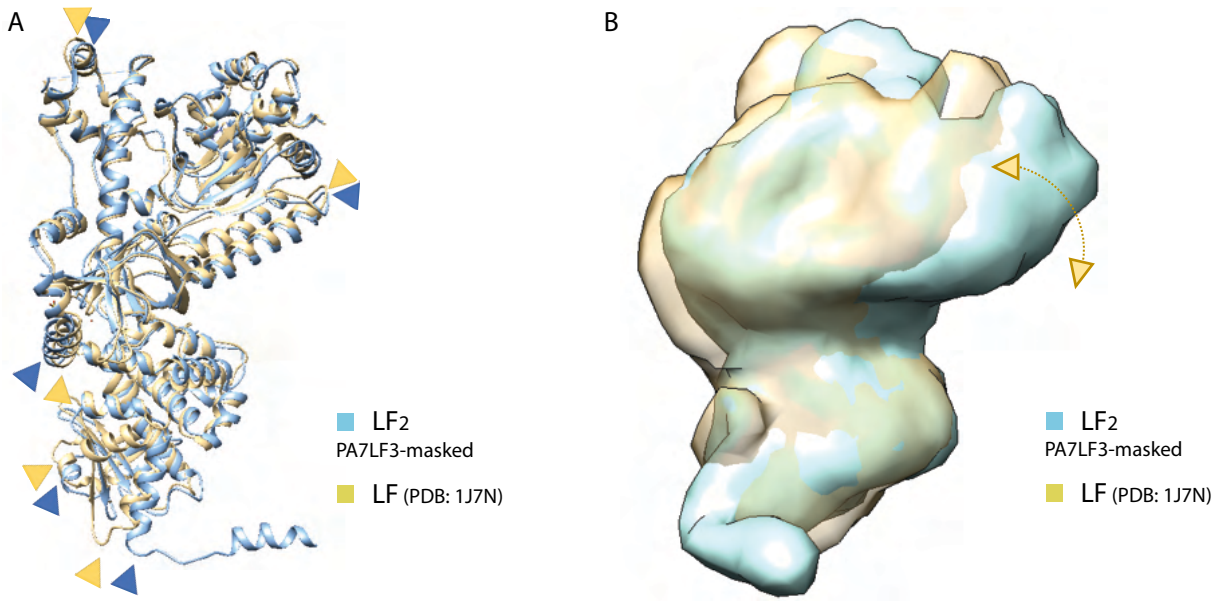


Figure 3.15 : Superimposition of LF_2 from $PA_7LF_{3\text{-masked}}$ and the unbound-LF crystal structure (PBD: 1J7N)
(A) Superposition of LF_2 (PA_7LF_3-masked, light blue) and unbound-LF (PDB: 1J7N, yellow) structure. The yellow and blue arrowheads indicate similar positions on the two structures, highlighting the structural movement. (B) Superposition of LF_2 ($PA_7LF_{3\text{-masked}}$, light blue) and unbound LF (PDB: 1J7N, yellow) filtered EM volumes, aligned from their N-terminal domains. The yellow arrows mimic the LF structural movement after PA_7LF_3 complex formation.

Differently than the superimposition between LFs from the same or different complexes, this superimposition clearly shows differences between the bound (light

blue, LF_2 from $PA_7LF_{3\text{-masked}}$ complex) and un-bound LFs (gold, PDB: 1J7N, yellow) (Fig. 3.15 A-B). The schematic representation (Fig. 3.15 B) illustrates a rotation of the C-terminal domain that occurs between unbound (gold) and bound (light blue) LF. Basically, the C-terminal domain rotates with respect to the N-terminal domain (Fig. 3.15 A-B) resulting in a more compacted structure. Indeed, the LFs interact via their N-terminal domain with the PA_7 but also forms a continuous chain of head-to-tail interaction with each other via their C-terminal domain, in which the three LFs come closer to each other (Fig. 3.16). The binding of LF to the PA_{63} protomer involves indirectly also its bulky C-terminal domain, such as the adjacent PA_{63} protomer is not accessible for another binding. Because of this singular binding scheme, a single LF basically binds to one PA_{63} protomer, but occupies also the adjacent ones thus finally blocking two PA_{63}. In this chain of LFs, the bulky C-terminus of one LF binds to the N-terminus of the next one, defining a sort of directionality in the complex (Fig. 3.16). Following this binding scheme, when two LFs are bound on the PA_7, three free PA binding sites are left but only two can be occupied because of steric clashes. Based on the binding position of LF_3 on PA_7, a different PA_7LF_3 complex can be generated (PA_7LF_{2+1A} or PA_7LF_{2+1B} complex) (Fig. 3.17). A saturated PA_7 can bind up to three LFs on its surface, thus leaving an empty binding site which generates a symmetry mismatch in the complex.

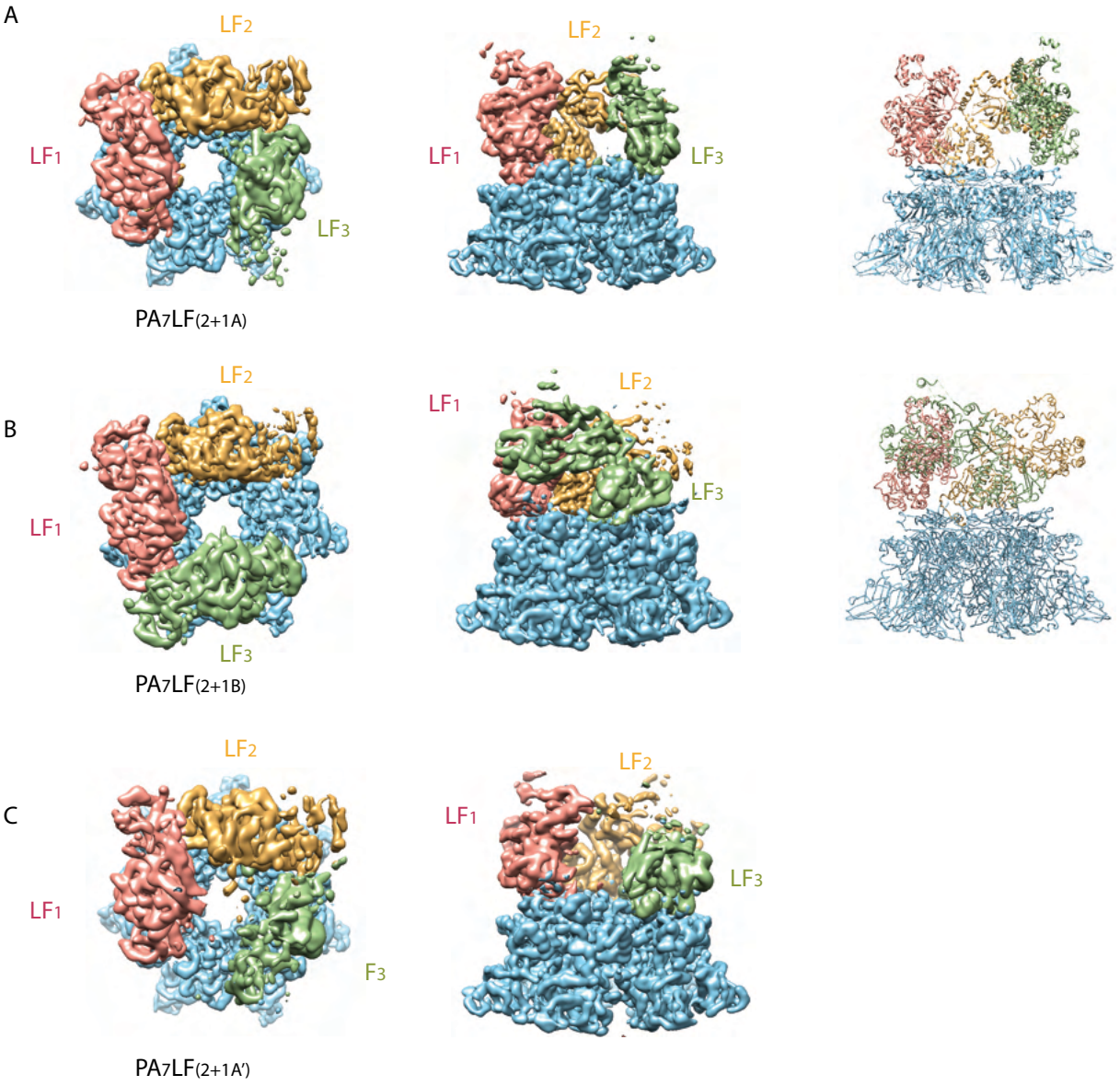


Figure 3.16 : Cryo-EM electron densities of the PA_7LF_3 complexes. For all complexes, LF_1 in dark pink, LF_2 in gold, LF_3 light green and PA_7 in light blue. (A) Top view and side view of the color-coded segmented cryo-EM density map of $PA_7LF_{2+1A.}$ The naming of the three LFs is based on their respective local resolution with LF_1 being the best-resolved, while LF_3 showed the weakest density in all reconstructions. Three LF bind to the PA_7 and form a continuous chain of head-to-tail interactions. Atomic model is shown on the left. (B) Same as in (A) for the PA_7LF_{2+1B} complex. (C) Same as in (A) for the $PA_7LF_{(2+1A)'}$ complex. The densities corresponding to the LFs in structure (C) were not resolved well enough to allow the fitting of an atomic model.

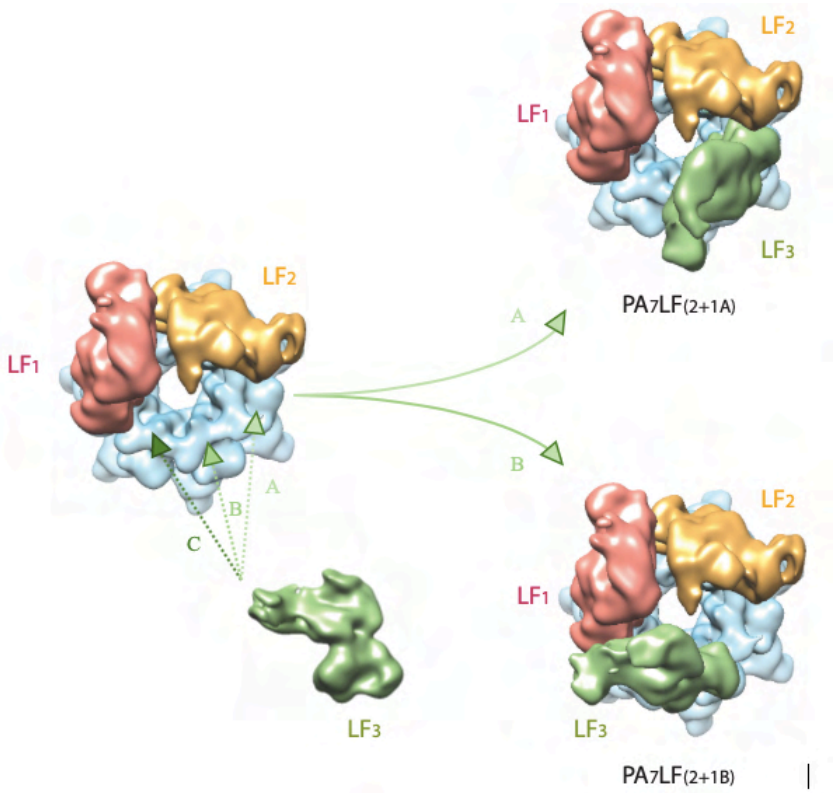


Figure 3.17 : Schematic representation of PA_7LF_3 toxin assembly. After the first two LFs bind PA_7, LF_3 can potentially bind to the tree binding sites left, named as A, B and C. Due to steric clashes, only A and B can be really occupied by the LF_3, generating $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1B)}$.

3.2.10. N-terminal LFs structural analysis

The major difference between any of the LFs from $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$ complexes and the unbound-LF (PBD: 1J7N) is the rotation of the LF-bound C-terminal domain toward its N-terminal domain (Fig 3.15).

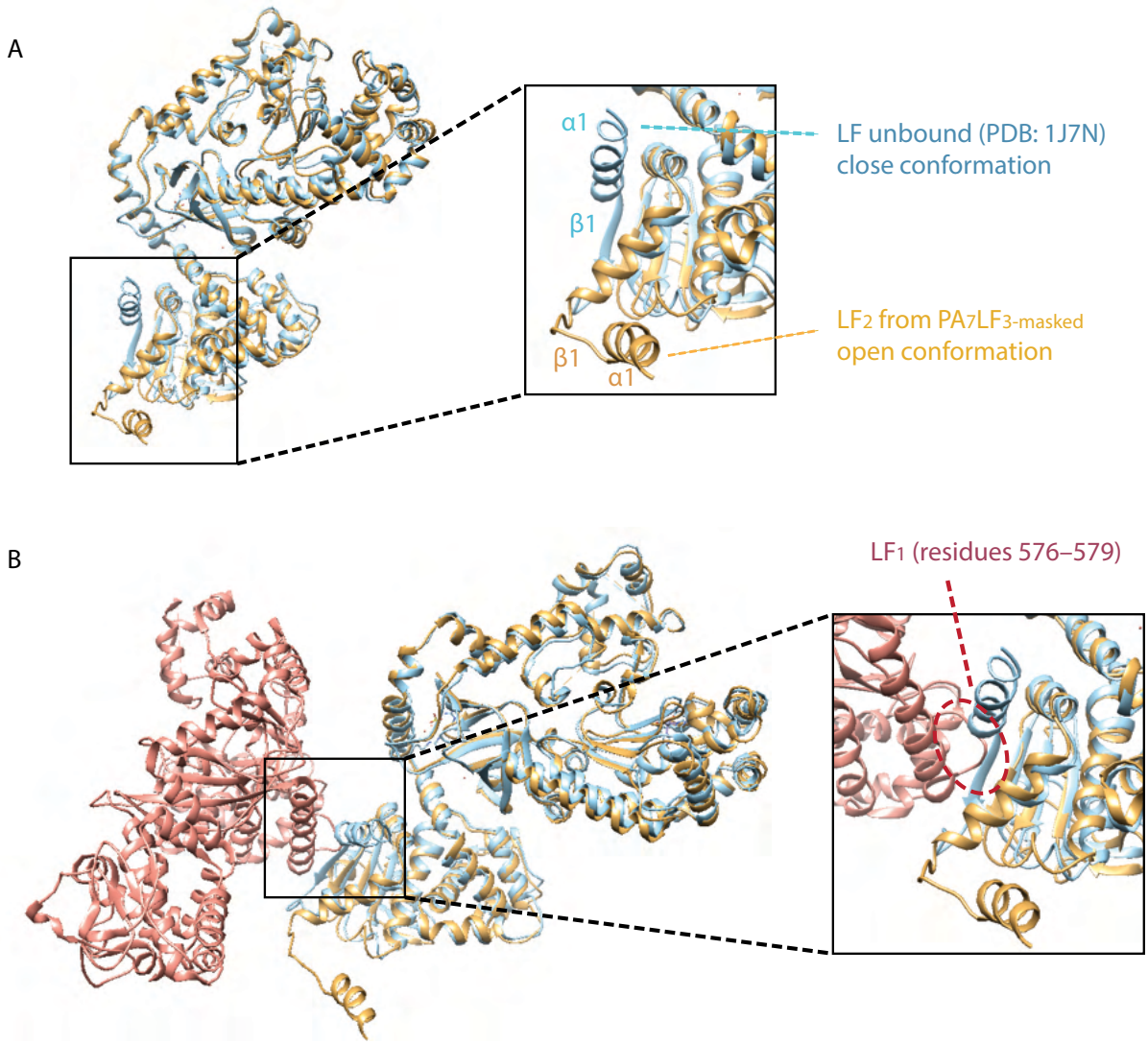


Figure 3.18 : Interface area between two adjacent LFs in the $PA_7LF_{3\text{-masked}}$ complex. (A) Superimposition of unbound-LF (PDB: 1J7N, light blue) and LF_2 from $PA_7LF_{3\text{-masked}}$ (gold). Highlight on the α1-β1 in the open conformation (pointing downward) of LF_2 from $PA_7LF_{3\text{-masked}}$ (gold) and close (pointing upward) conformation in the unbound-LF. (B) Simulation of the interface between LF_1 with LF_2 in the open (gold) and closed state (light blue), docked in the $PA_7LF_{3\text{-masked}}$ electron density. Steric clashes between LF_1 (pink) and the adjacent LF_2 in the closed state (light blue). Highlight of the clashes between the residues 576-579 from LF_1 and the α1-β1of LF_2 (closed state).

In addition to this, the LF_2 N-terminus region from $PA_7LF_{3\text{-masked}}$, $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$ adopts a different conformational state compared to the other LFs. Indeed, LF_2 α1-β1 is the only one clearly adopting the open conformation (as described in Feld et al. 2010). In the open conformation it can interact with the α-clamp region located between two adjacent PA_{63} protomers (Fig. 3.18 B and Fig. 3.19).

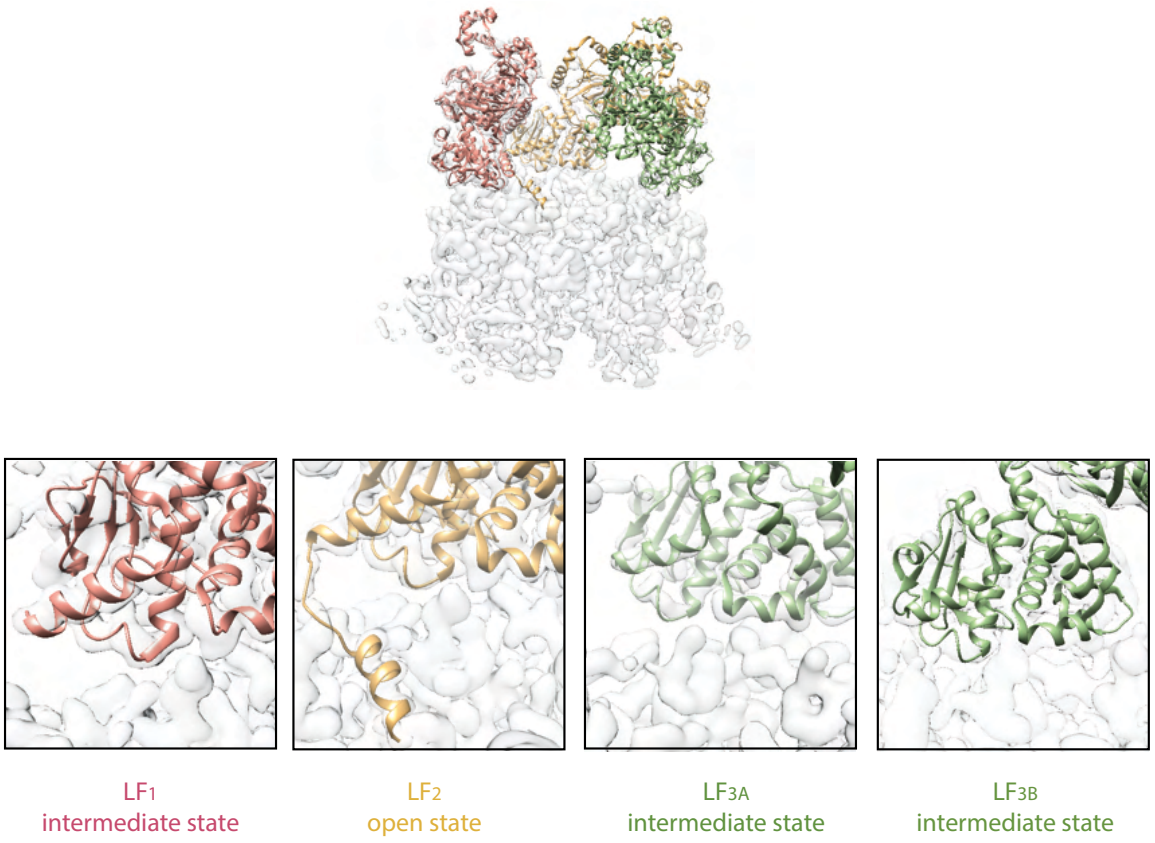


Figure 3.19 : LF-PA interfaces. Overall shape of the $PA_7LF_{(2+1A)}$ in transparent. LF_1 in pink, LF_2 in gold, LF_{3A} and LF_{3B} in green. Highlight of the LF-PA interfaces in the boxes. In the second box, LF_2 α1-β1 inserts into the PA lumen, interacting with the PA α-clamp. The α1-β1 EM density is not visible for LF_1, LF_{3A} and LF_{3B}.

Different situation is for LF_1, LF_{3A} and LF_{3B} α1-β1 electron density, which was not visible (Fig 3.19) either in the open state nor in the closed state. Moreover, the α1-β1 cannot adopt the closed state because of the presence of steric clashes between α1-β1 and the loop region (residues 576-579) of the previous LF (Fig. 3.18 A). This evidence brought to define a new state, called "intermediate" in which the α1-β1 adopts a highly flexible conformation which takes place in a time frame between the toxin assembly and its subsequent translocation state.

3.2.11. Interaction sites between LF and PA in the PA_7LF_3 complex

The LF-PA interaction is mediated by a large planar interface that involves the domain I of LF (Fig. 1.6) and the LF binding site of two adjacent PAs. This interaction is characterized by a large hydrophobic core, in presence of electrostatic interactions (Fig. 3.21).

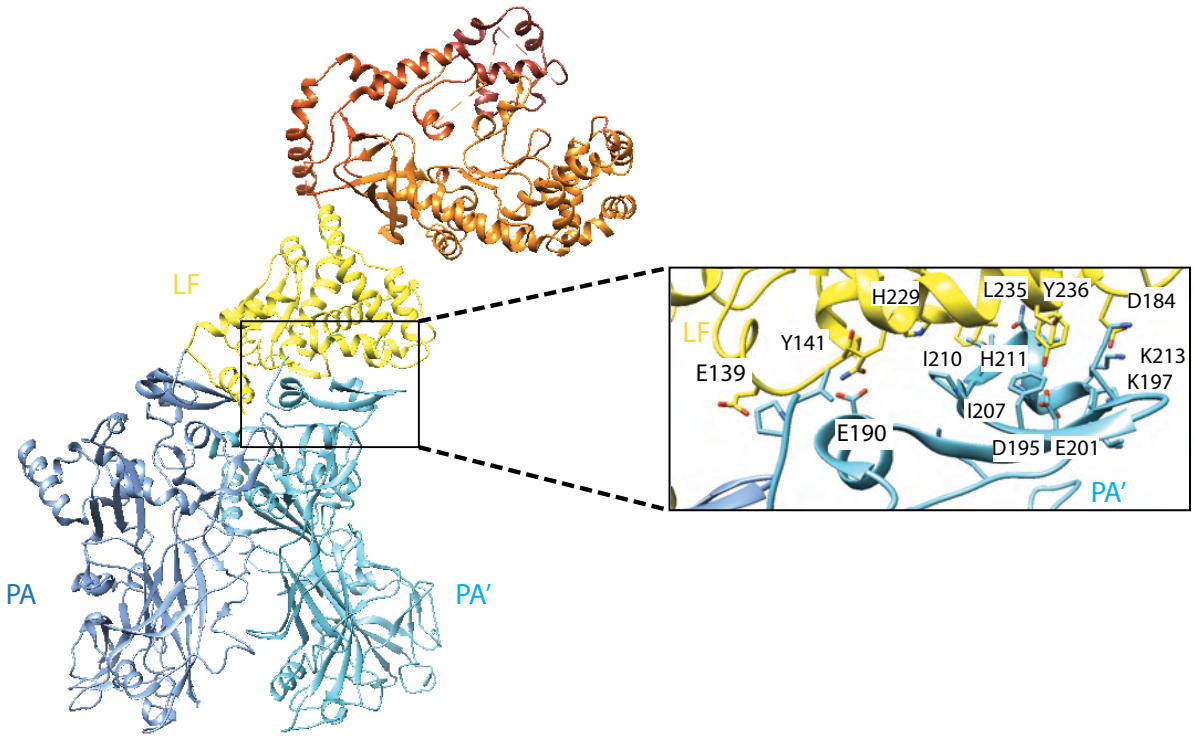


Figure 3.20 : interfaces between LF and PA. (A) The N-terminal domain of LF binds PA forming a large planar interface with two adjacent PA molecules (LF_2 in yellow, PA and PA' in light blues). A black square indicates the interaction interface between all three molecules. The inset shows a close-up of the interaction regions (with contributing residues labelled) that generate a central hydrophobic core surrounded by electrostatic interactions.

All LFs present the same interaction with the relative PA, and this site is well resolved and conserved in all complexes described in this work, with little differences due to variation in the resolution. The large interaction site between LF and PA is very similar to the one reported for PA_8LF_4 (Feld et al. 2010).

The hydrophobic interface involves PA_C I207 and I210, and LF L235, H229 and Y236. The PA_C I210 phenol hydroxyl forms a hydrogen-bonding network with PA_C H211 and D195 near the center of the hydrophobic interface. Additional electrostatic interactions surround this hydrophobic core. The carboxyl side chain of PA_C Glu190 forms a pair of hydrogen bonds with the LF T141. PA_C K197 forms a salt bridge with D184.

The PA α-clamp subsite interacts with the N-terminus LF adopting the open state (Fig. 3.22). In this interface, the N-terminal LF β1 (Ile43 and Lys45) forms β-sheet hydrogen bonds parallel to L203 of PA. PA P205, which is positioned at the end of PA β13, is involved in parallel-sheet interactions with the N-terminal LF β1. In general, the N-terminal LF side chains are not intensively-packed with side chains of the PA α-clamp cleft, in contrast to the hydrophobic core of the PA-LF interface (Fig. 3.21). In addition, PA contacts the side chains of LF M40 and H35 through backbone interactions.

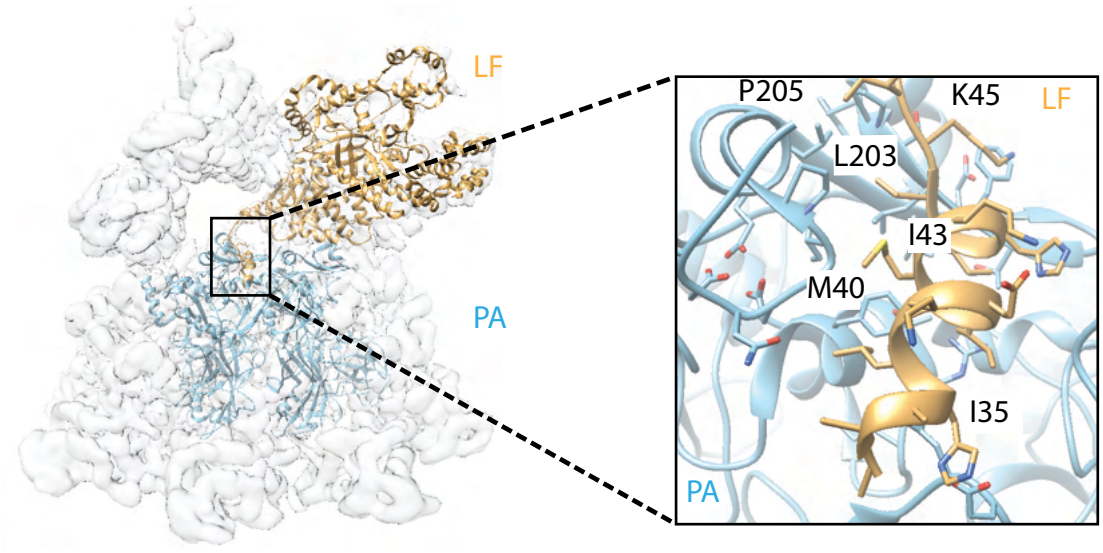


Figure 3.21 : PA-LF_2 N-terminal interface. The N-terminal α-helix LF interacts with the α-clamp region, located between two adjacent PA molecules. The black square highlights the interaction interface between the two LFs, and the contributing residues are labelled.

The other important interaction site to mention is the one between the C-terminus and the N-terminus of adjacent LFs. It is a relatively small interaction interface. Anyway, it contributes to increase the stability of the highly flexible C-terminal region. The interface consists of the helix-loop region (residues 572-579) relative to the C-terminus LF which is involved in a small interaction surface with the helix-helix-ß-sheet motif (residue 52-84) of the adjacent LF (Fig. 3.23). More in detail, the LF N-terminal residues L63, L71 and I81 form a central hydrophobic cavity that interacts with LF C-terminal residue Y579. Moreover, in the LF C-terminal ß-sheet region, there is a potential backbone-backbone hydrogen bond between K578 and I81 relative to the adjacent N-terminal LF.

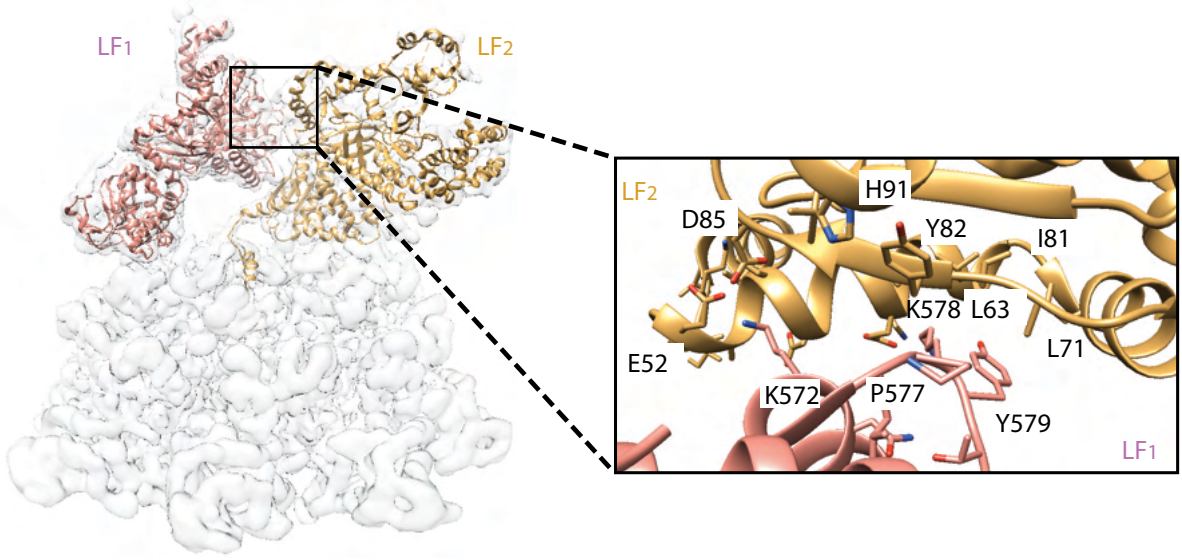


Figure 3.22 : Interface between two adjacent LFs. A relatively small interaction interface between the C-terminal domain of LF_1 and the N-terminal domain of the adjacent LF_2. The positions of LF_1 (pink) and LF_2 (gold) are shown relative to the overall shape of the PA_7LF_3 complex, which is represented as a transparent EM volume. A black square highlights the interaction interface between the two LFs, and the contributing residues are labelled.

Another hydrophobic interaction is the one between of the C-terminal LF P577 and N-terminal LF Y82, which is further stabilized by H91. Another potential salt bridge could be the one bewteen LF C-terminal K572 and E52 or D85 of the adjacent N-terminal LF. All those interactions listed above mediate the binding between two adjacent LFs, LF_1-LF_2 and LF_2-LF_3.

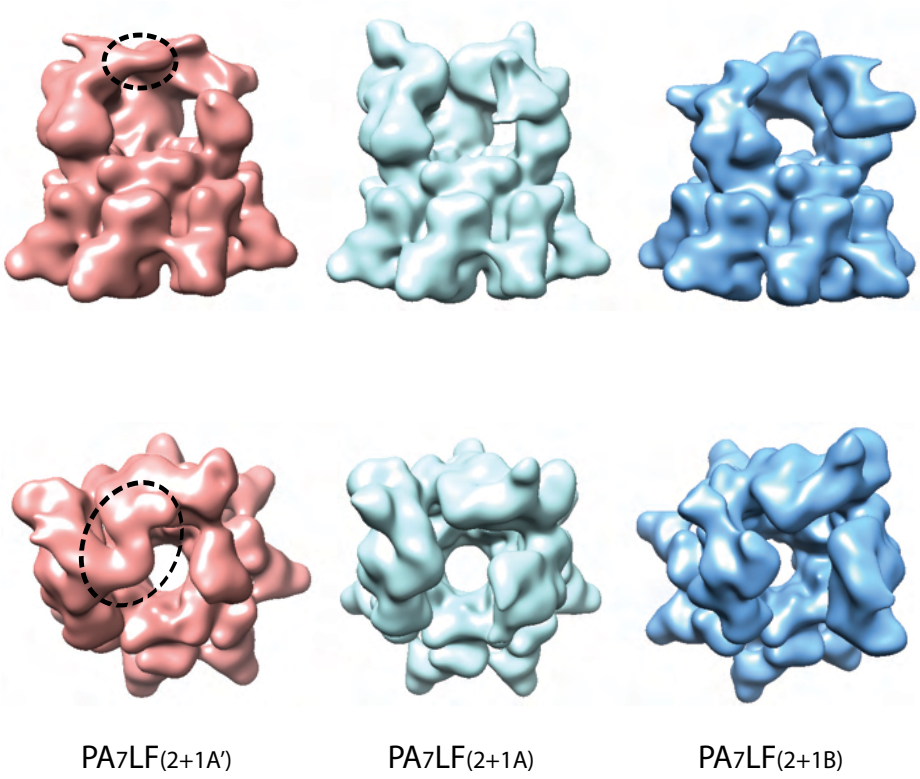


Figure 3.23 : LF-LF interface between their relative C-terminal domains. Top and side views of the low-passed filtered maps of the three PA_7LF_3 complexes: $PA_7LF_{(2+1A')}$ in pink, $PA_7LF_{(2+1A)}$ in light blue and $PA_7LF_{(2+1B)}$ in blue. Volumes are shown at the same threshold. All the PA_7LF_3 complexes share the same LFs scheme, which creating a head-to-tail interaction sequence via their N- and C-terminal domains. An additional interface was identified in the $PA_7LF_{(2+1A)'}$ reconstruction. Here, LF_1 and LF_2 C-terminal domains interact with each other close to the central axis of the complex. This region is highlighted by dashed black circles in both side and top view.

Despite a quite poor local resolution at LF_2-LF_{3A} and LF_{3B}-LF_1 interfaces that does not allow an accurate fitting of the side chains, the structures from LF_1 and LF_2 have been flexibly fitted in and since all the structures are almost identical at the backbone level (RMSD of 0.84 Å and 0.96 Å), they should have a highly similar interaction pattern. Nevertheless, all the interfaces described for PA-LF or between LF-LF have a limit of freedom of movement especially at the N-terminal region.

In addition to those interaction sites, in the $PA_7LF_{(2+1A)'}$ two LFs most likely interact with each other via the upper C-terminal region in proximity of the central axis of the molecule (Fig. 3.24). Also Fabre et al. 2016 proposed an additional interface at the C-terminal region, but it was localized far off from the central axis.

The evidences from this work, show a different behaviour than those suppositions.

4. DISCUSSION

The anthrax outbreak in 1979 in USSR (Guillemin J. et al 2002) or the more recent use of the spores of *B. anthracis* as bioweapon (2001 and 2010, Jernigan, J. A. et al. 2001) drawn the attention on how the related scientific progresses are still of great importance.

During the years, many studies (medical, biochemical, biological etc.) reveal fundamental aspects that allowed the introduction of the vaccine and therapies (Mayo clinic family health book, 5th edition, Kummerfeldt, C. E. et al., 2014). Even if important questions have been answered over the years, there are still some points that need to be precisely delineated.

The lethal complex, is the most dangerous toxin combination secreted from *B. anthracis.* So far, individual structures of the single molecules are available, but a high-resolution structure of the formed lethal complex was still missing.

In this thesis I presented the structure of the three lethal complex combinations, named as $PA_7LF_{(2+1B)}$, $PA_7LF_{(2+1A)}$ and $PA_7LF_{(2+1A')}$ for which I was able to propose a new translocation model. In this model, PA translocates the LFs through the cellular membrane into the cytosol (Antoni C. and Quentin D. et al 2020).

This study contributes giving important information to uncover the anthrax translocation machinery.

Possible mechanism adopted by PA to translocate the LF

My structures showed that α1-β1 of two LFs present in a complex, adopt a so-called "intermediate" state in which the N-terminus does not point upwards (close state) and neither interacts with the α-clamp of PA (open state). Instead and due to its high flexibility, it fluctuates into the lumen. This intermediate state has not been described previously. For instance, in the PA_8LF_4 crystal structure the N-terminus domain of every LFs was in the open state.

Considering the results presented here, PA_8LF_4 conformation is doubtful. PA_8LF_4 could have been obtained by averaging the asymmetric unit of PA_8LF_4 crystal, which groups two PAs and the relative N-terminal LF, therefore cutting out the conformation adopted by the near LFs of the complex. Another possible explication is related to the difference in dimension between the PA_7 and the PA_8 ring. Indeed, PA_8 could provide more space for the LF N-termini to adopt the open state and then interact with the α-clamp of the PA. But, if all the N-termini LF are in the open state, which LF would translocate first through the PA channel? The channel is quite narrow, allowing only one unfolded chain to pass through at one time. In this case, the process would be stochastic.

Instead, the structures described in this thesis suggest another explanation of the process. As also suggested from Fabre et al. 2016, the order of the translocation can not be stochastic. They suggested that the first LF to be translocated is the one which N-terminal domain does not interact with an adjacent LF. Afterwards, the next LF would be the one freed from the inhibitory binding of the first LF and so on.

Based on the results of this work, I can hypothesize two possible translocation events (Fig. 4.1): (i) the LF_2 having the α1-β1 in the open position interacts with the α-clamp and is translocated first. In a second moment, LF_1 or LF_3 (intermediate state) would pass to an open state and translocate through the PA_7 channel and so on. The hypothesis would suggest that when α1-β1 is not bound to the α-clamp, due to its high flexibility it probably fluctuates into the cavity of the PA lumen. This state has been introduced as "intermediate" state and is located in a time frame after the toxin assembly and before the toxin translocation. It can be considered as a waiting state for the channel to be freed from the previous α1-β1 that interacts with the α-clamp, then occupying it (Fig. 4.1). Otherwise, (ii) LF_1 or LF_3 are translocated before LF_2 (Fig. 4.1). Due to the binding disposition, these two options would be independent on an adjacent factor for LF translocation. Both scenarios are possible, but according to this thesis results, the first hypothesis is more likely to be correct. The α-clamp is known to have the characteristic of unfolding polypeptides chains in a sequence-depending manner. The region between two adjacent PAs stabilizes unfolding intermediates, thus establishing a mechanical strain. Therefore, it is more likely that the LF_2 α1-β1 are

unfolded at the beginning of the translocation processes and that the LF_2 will be progressively unfolded while passing through the PA channel. Interestingly, despite the binding between C-terminal LF_2 and N-terminal LF_3, the C-terminal LF_2 has higher flexibility compared to the LF_1, suggesting that the interaction of the N-terminal domain with the PA α-clamp is reflected in a destabilization of the entire molecule. Being the C-terminal LF_2 indirectly destabilized by the α-clamp, the relative energy barrier for the unfolding of the molecule is lower, further supporting that LF_2 is translocated first. After LF_2 is translocated through the PA-pore channel, either LF_1 or LF_3 can potentially follow. Both LF_1 and LF_3 are in the intermediate state. For that it is not possible to predict which one is second to be translocated (Fig. 4.1).

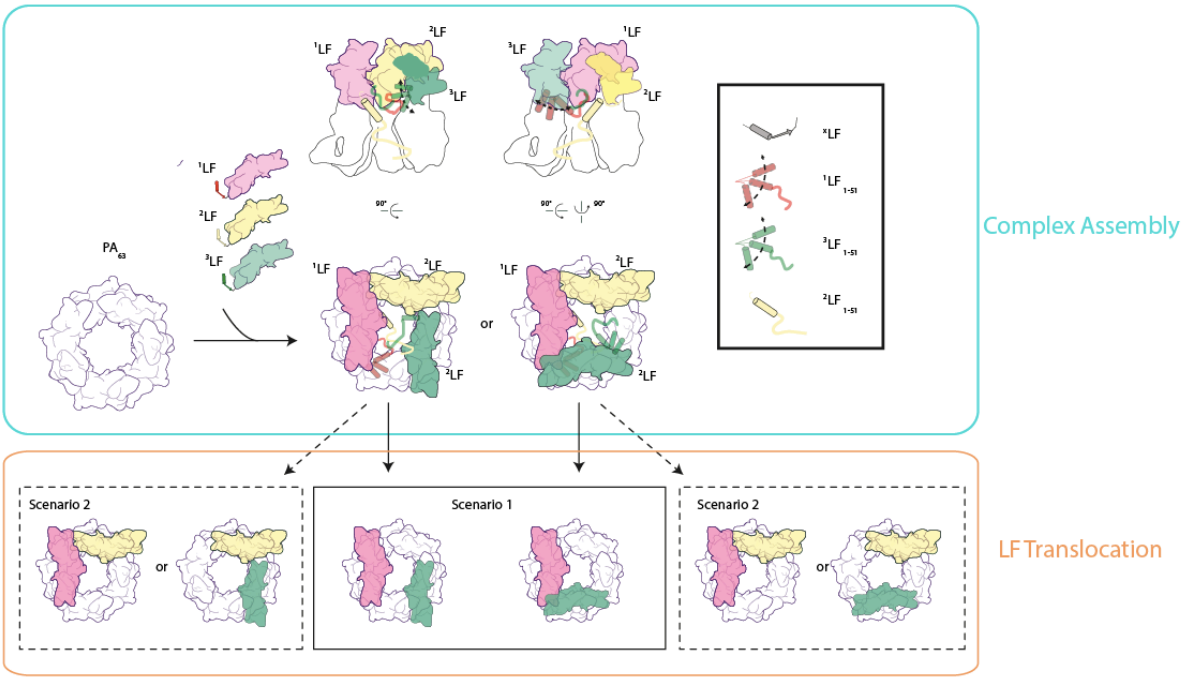


Figure 4.1 : Representation of PA_7LF_3 complex assembly and subsequent LFs translocation. Representation of the receptor-mediated PA_7 assemblies on the host cellular surface, followed by three LFs which crown PA_7, thus forming the fully loaded PA_7LF_3 (anthrax lethal complex). Due to the loss of symmetry after LFs binding on PA_7, the last LF has the possibility to bind PA_7 on two different sites, thus generating two distinct PA_7LF_3 complexes. Interestingly, N-terminal LF_2 undergoes to a conformational rearrangement passing from "closed" to an "open" state. The open state presents an additional interaction site between the LF N-terminal α1-β1 and the PA α-clamp. In contrast, LF_1 and LF_3 N-terminals are in a new state, defined "intermediate", which is a different state from the open and closed ones. With this new structural information, two possible scenarios can be envisaged to describe the LF translocation.

Scenario I propose LF_2 being translocated first, followed by either LF_1 or LF_3. Instead, scenario II indicates LF_1 or LF_3 as the first to be translocated. Figure adapted from Antoni and Quentin et al. 2020.

PA_7LF_3 toxin assembly

Accordingly to the previously published studies, two possible ways of toxin assembly are envisaged: (i) the toxin assembly is mediated by the anthrax-receptor (ATR) and it takes place on the cellular surface, while the other scenario (ii) hypothesizes an independent assembly in solution (ATR-independent), like in the bloodstream. These two possible ways are not mutually exclusive. In the first hypothesis (i) the ATRs present on the cellular surface bind full-length PA which later is cleaved in its active form and then assembles to form PA_7 and PA_8. Next, the LF/EF are recruited on the PA surface.

According to Kintzer et al. 2010 *in vitro* mass spectrometry data, PA oligomerization happens via even-numbered states (PA_2, PA_4 and so on) being consistent with the proposed dimerization of the ATR.
During this study, the PA_7 has been preassembled in the absence of ATR or LF/EF, that rather corresponds to hypothesis (i) (Fig. 4.1).
Furthermore, LF_2 is most probably the first LF to be translocated (Fig. 4.1) but also the first one that binds to PA_7 during toxin assembly. When the LF binds to the PA_7 ring, its conformation passes from “closed” to “open”, and since the LF_3 can bind to two different PAs, this leads to two different complexes. Following this hypothesis, the assembled toxin has three LFs bound on the PA_7 ring, of which two are in the “intermediate” state (LF_1 and LF_3) and one in the open state (LF_2).

Additional interface between C-termini of LFs.

As described above, the main and common interaction sites in the $PA_7LF_{(2+1A)}$, $PA_7LF_{(2+1B)}$ and $PA_7LF_{(2+1A')}$ complexes are represented by: (i) the LF N-terminal domain with the PA protomer (Fig. 3.21) and the (ii) the LF C-terminal domain with the N-terminal domain of the adjacent LF (Fig. 3.23). In the $PA_7LF_{(2+1A')}$ structure there is an additional interface at the C-terminal region of LFs already indicated by Fabre et al. 2016. But differently from what they proposed, this interface is localized far off from

the central axis. Indeed, analysing the low pass filtered map, two LFs most likely interact with each other via their C-terminal domains, near to the central axis of the complex, thus opening a new interface to investigate in the future.

References

- Abrami, L.; Leppla, S.H.; van der Goot, F.G. Receptor palmitoylation and ubiquitination regulate anthrax toxin endocytosis. *J. Cell Biol.* 2006, *172*, 309–320
- Abrami, L.; Kunz, B.; van Der Goot, F.G. Anthrax toxin triggers the activation of src-like kinases to mediate its own uptake. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 1420–1424
- Abrami, L.; Bischofberger, M.; Kunz, B.; Groux, R.; van der Goot, F.G. Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors. *PLoS Pathog.* 2010, *6*, e1000792.
- Abrami, L.; Brandi, L.; Moayeri, M.; Brown, M.J.; Krantz, B.A.; Leppla, S.H.; van der Goot, F.G. Hijacking multivesicular bodies enables long-term and exosome-mediated long-distance action of anthrax toxin. Cell Rep. 2013, 5, 986–996.
- Akkaladevi N, Hinton-Chollet L, Katayama H, Mitchell J, Szerszen L, Mukherjee S, Gogol EP, Pentelute BL, Collier RJ, Fisher MT. Assembly of anthrax toxin pore: lethal-factor complexes into lipid nanodiscs. Protein Sci. 2013 May;22(5):586-94.
- Baron, S. & Turnbull, P. C. B. Bacillus. (1996).
- Bartesaghi, A. et al., 2014. Structure of beta-galactosidase at 3.2-angstrom resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), pp.11709–11714.
- Belyy A. Merino F., Mechold U. & Raunser S. Mechanism of actin-dependent activation of nucleotidyl cyclase toxins from bacterial human pathogens. Nat. Commun. 2021.
- Bergmans, H.E., Van Die, I.M. & Hoekstra, W.P., 1981. Transformation in Escherichia coli: stages in the process. *Journal of bacteriology*, 146(2), pp.564–570.
- Binshtein, E. & Ohi, M.D., 2015. Cryo-electron microscopy and the amazing race to atomic resolution. *Biochemistry*, 54(20), pp.3133–3141.

- Bischofberger, M., Iacovache, I. & van der Goot, F.G., 2012. Pathogenic pore-forming proteins: function and host response. *Cell host & microbe*, 12(3), pp.266–275.
- Blewitt, M.G., Chung, L.A. & London, E., 1985. Effect of pH on the conformation of diphtheria toxin and its implications for membrane penetration. *Biochemistry*, 24(20), pp.5458–5464.
- Booth, D.S., Avila-Sakar, A. & Cheng, Y., 2011. Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. *Jove- Journal of Visualized Experiments*, (58), p.e3227.
- Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414,** 225–229 (2001).
- Brown, M. J., Thoren, K. L. & Krantz, B. A. Role of the α Clamp in the Protein Translocation Mechanism of Anthrax Toxin. *J. Mol. Biol.* **427,** 3340–3349 (2015).
- Campbell, M.G. et al., 2015. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *eLife*, 4, p.e06380.
- Chand, H. S. *et al.* Discriminating virulence mechanisms among Bacillus anthracis strains by using a murine subcutaneous infection model. *Infect. Immun.* **77,** 429–435 (2009).
- Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, 2020 (https://www.cdc.gov/anthrax/bioterrorism/threat.html)
- Cosentino, K., Ros, U. & García-Sáez, A.J., 2016. Assembling the puzzle: Oligomerization of α-pore forming proteins in membranes. *Biochimica et biophysica acta*, 1858(3), pp.457–466.
- Danev, R. et al., 2014. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44), pp.15635–15640.

- Deuquet J., laush E., Van der Goot F., The dark sides of capillary morphogenesis gene 2. The EMBO Journal 31(1):3-13, 2011.
- DUBOCHET, J. et al., 1988. Cryo-Electron Microscopy of Vitrified Specimens. *Quarterly Reviews of Biophysics*, 21(2), pp.129–228.
- Duesbery, N.S.; Webb, C.P.; Leppla, S.H.; Gordon, V.M.; Klimpel, K.R.; Copeland, T.D.; Ahn, N.G.; Oskarsson, M.K.; Fukasawa, K.; Paull, K.D.; et al. Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. Science 1998, 280, 734–737.
- Dunkelberger, J.R. & Song, W.-C., 2010. Complement and its role in innate and adaptive immune responses. *Cell research*, 20(1), pp.34–50.
- Efremov, R.G., Gatsogiannis, C., and Raunser, S. (2017). Lipid Nanodiscs as a Tool for High- Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. Meth. Enzymol. *594*, 1–30.
- Emsley, P. et al., 2010. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4), pp.486–501.
- Fabre, L. *et al.* Structure of anthrax lethal toxin prepore complex suggests a pathway for efficient cell entry. *J. Gen. Physiol.* **148,** 313–324 (2016).
- Feld, G. K., Kintzer, A. F. & Krantz, B. A. Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. (2010). doi:10.2210/pdb3kwv/pdb
- Fan, X. et al., 2019. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nature communications*, 10(1), pp.–11.
- Garcia-Saez, A.J., 2012. The secrets of the Bcl-2 family. *Cell Death and Differentiation*, 19(11), pp.1733–1740
- Gatsogiannis C, Lang AE, Meusch D, Pfaumann V, Hofnagel O, Benz R, Aktories K, Raunser S. A syringe-like injection mechanism in Photorhabdus luminescens toxins. Nature. 2013 Mar 28;495(7442):520-3.

- Gatsogiannis, C. *et al.* Membrane insertion of a Tc toxin in near-atomic detail. *Nat. Struct. Mol. Biol.* **23,** 884–890 (2016).
- Gatsogiannis, C. et al., 2018. Tc toxin activation requires unfolding and refolding of a β- propeller. *Nature*, 9, p.185.
- Gilfoyle, D. Anthrax in South Africa: economics, experiment and the mass vaccination of animals, c. 1910-1945. *Med Hist* **50,** 465–490 (2006).
- Goel, A. K. Anthrax: A disease of biowarfare and public health importance. World J Clin Cases 3, 20–33 (2015).
- Goldberg, A. B. & Turk, B. E. Inhibitors of the Metalloproteinase Anthrax Lethal Factor. *Curr Top Med Chem* **16,** 2350–2358 (2016).
- Gómez, I. et al., 2014. Bacillus thuringiensis Cry1A toxins are versatile proteins with multiple modes of action: two distinct pre-pores are involved in toxicity. *The Biochemical journal*, 459(2), pp.383–396.
- Guillemin J. The 1979 anthrax epidemic in the URSS: applied science and political controversy. Proc.Amer. phil. Soc. 2002.
- Hanks, S. *et al.* Mutations in the gene encoding capillary morphogenesis protein 2 cause juvenile hyaline fibromatosis and infantile systemic hyalinosis. *Am. J. Hum. Genet.* **73,** 791–800 (2003).
- HENDERSON, R., 1995. The Potential and Limitations of Neutrons, Electrons and X- Rays for Atomic-Resolution Microscopy of Unstained Biological Molecules. *Quarterly Reviews of Biophysics*, 28(2), pp.171–193.
- Jang, J., Pentelute, B. L., Collier, R. J. & Zhou, Z. H. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521,** 545–549 (2015).
- Jang j., Cho M., Chun JH., Cho MH., Park J., Oh HB., Yoo CK., Rhie G. The Poly-γ-d-Glutamic Acid Capsule of Bacillus anthracis Enhances Lethal Toxin Activity. Infect Immun. 2011; 79(9): 3846-3854
- Jiao, G.-S. *et al.* Small molecule inhibitors of anthrax edema factor. *Bioorg. Med. Chem. Lett.* **28,** 134–139 (2018).

- Katayama, H. *et al.* Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles. *Proceedings of the National Academy of Sciences* **107,** 3453–3457 (2010).
- Khoshouei, M. et al., 2017. Cryo-EM structure of haemoglobin at 3.2 angstrom determined with the Volta phase plate. *Nature communications*, 8(1), pp.–6.
- Kühlbrandt, W., 2014a. Biochemistry. The resolution revolution. *Science (New York, N.Y.)*, 343(6178), pp.1443–1444.
- Kühlbrandt, W., 2014b. Microscopy: Cryo-EM enters a new era. *eLife*, 3, p.e01963.
- Kummerfeldt, C. E. Raxibacumab: potential role in the treatment of inhalational anthrax. *Infect Drug Resist* **7,** 101–109 (2014).
- Hugh-Jones, M.E.; de Vos, V. Anthrax and wildlife. *Rev. Sci. Tech.* **2002**, *21*, 359–383.
- Iacovache, I., van der Goot, F.G. & Pernot, L., 2008. Pore formation: an ancient yet complex form of attack. *Biochimica et biophysica acta*, 1778(7-8), pp.1611–1623.
- Liu, S.; Crown, D.; Miller-Randolph, S.; Moayeri, M.; Wang, H.L.; Hu, H.J.; Morley, T.; Leppla, S.H. Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin *in vivo*. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12424–12429.
- Lopéz-Blanco, J.R. & Chacón, P., 2013. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *Journal of Structural Biology*, 184(2), pp.261–270.
- Mayo clinic family health book, 5th edition
- Makino, S.; Watarai, M.; Cheun, H.; Shirahata, T.; Uchida, I. Effect of the lower molecular capsule released from the cell surface of Bacillus anthracis on the pathogenesis of anthrax. J. Infect. Dis. 2002, 186, 227–233
- Meusch, D. et al., 2014. Mechanism of Tc toxin action revealed in molecular detail. *Nature*, 508(7494), pp.61–65.

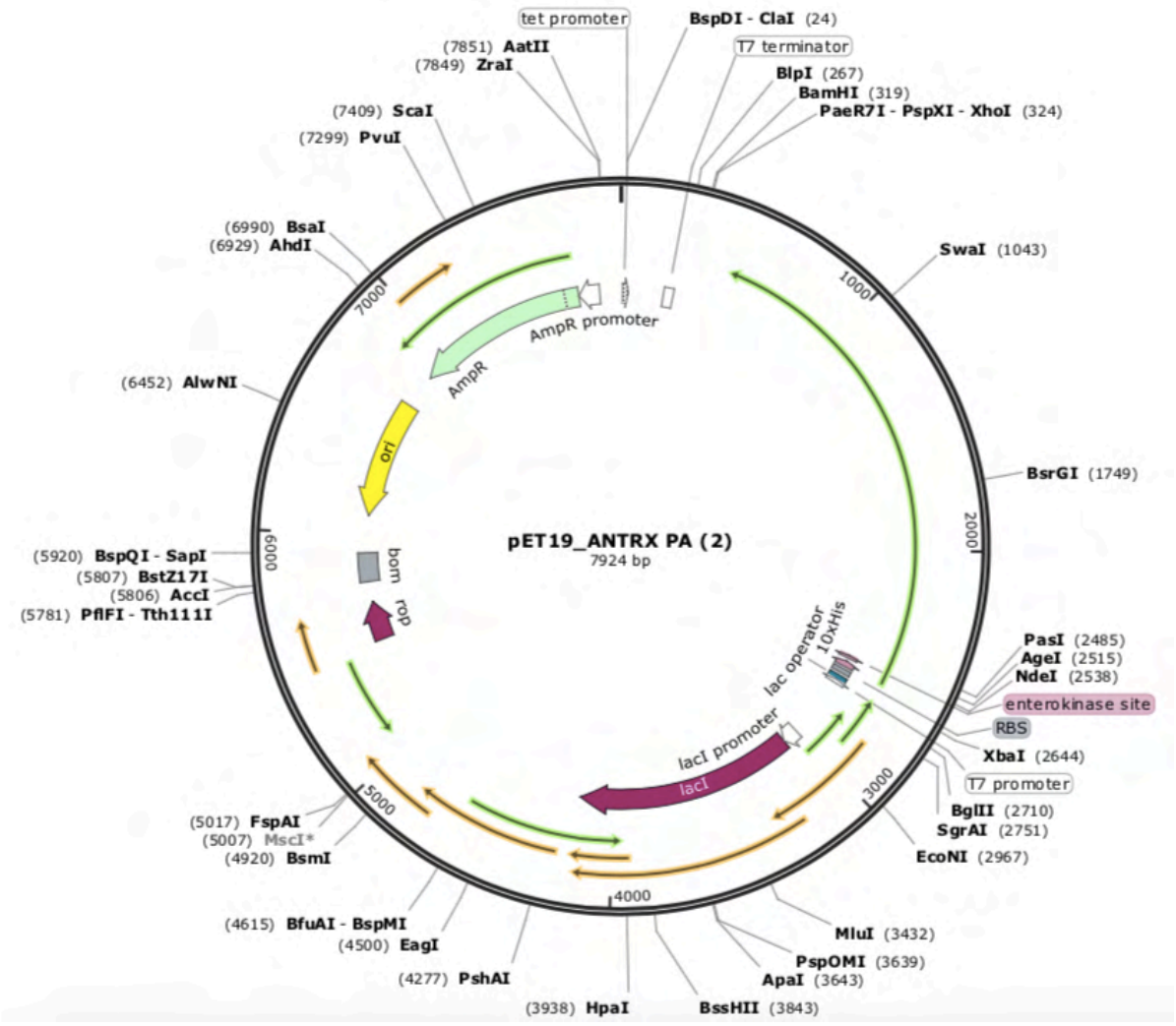
- Moore, C.A.; Milano, S.K.; Benovic, J.L. Regulation of receptor trafficking by GRKs and arrestins. *Annu. Rev. Physiol.* **2007**, *69*, 451–482.
- Moriya, T. et al., 2017. High-resolution Single Particle Analysis from Electron Cryo- microscopy Images Using SPHIRE. *Jove-Journal of Visualized Experiments*, (123).
- Ohi, M. et al., 2004. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biological procedures online*, 6, pp.23–34.
- Pannifer, A. D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414,** 229–233 (2001).
- Penczek PA, Fang J, Li X, Cheng Y, Loerke J, Spahn CMT. CTER—Rapid estimation of CTF parameters with error assessment. Ultramicroscopy. North-Holland; 2014;140: 9–19. pmid:24562077
- Parker, M.W. & Feil, S.C., 2005. Pore-forming protein toxins: from structure to function. *Progress in biophysics and molecular biology*, 88(1), pp.91–142.
- Peraro, M.D. & van der Goot, F.G., 2016. Pore-forming toxins: ancient, but never really out of fashion. *Nature reviews. Microbiology*, 14(2), pp.77–92.
- Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385,** 833–838 (1997).
- Pettersen, E.F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–1612.
- Rainey, G. J. A. *et al.* Receptor-specific requirements for anthrax toxin delivery into cells. *Proceedings of the National Academy of Sciences* **102,** 13278–13283 (2005).
- Raufman JP. Cholera. *Physiology in medicine*. 1998
- Raymond, S., 1964. Acrylamide gel electrophoresis. *Annals of the New York Academy of Sciences*, 121(2), pp.350–365.

- Roderer, D. & Raunser, S., 2019. Tc Toxin Complexes: Assembly, Membrane Permeation, and Protein Translocation. *Annual review of microbiology*, 73(1), pp.annurev–micro–102215–095531.
- Santelli, E., Bankston, L. A., Leppla, S. H. & Liddington, R. C. Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* **430,** 905–908 (2004).
- Pita R and Guranatna R Anthrax as a Biological Weapon: From World War I to the Amerithrax Investigation. 2009
- Shen Y., Zhukovskaya LN.. Guo Q., Florian J., Tang W. Calcium-independent calmodulin binding and two-metal–ion catalytic mechanism of anthrax edema factor. EMBO journal 2004
- Shapiro, A.L., Viñuela, E. & V Maizel, J., Jr., 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and biophysical research communications*, 28(5), pp.815–820.
- Shi, C.-S. & Kehrl, J.H., 2019. Bcl-2 regulates pyroptosis and necroptosis by targeting BH3-like domains in GSDMD and MLKL. *Cell death discovery*, 5(1), pp.151–13.
- Scheres, S.H.W. & Chen, S., 2012. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9(9), pp.853–854.
- Schubert E, Vetter IR, Prumbaum D, Penczek PA, Raunser S. Membrane insertion of α-xenorhabdolysin in near-atomic detail. Elife. 2018 Jul 16;7:e38017.
- Scobie, H.M.; Rainey, G.J.A.; Bradley, K.A.; Young, J.A.T. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 5170–5174
- Schubert, E. et al., 2018. Membrane insertion of α-xenorhabdolysin in near-atomic detail. *eLife*, 7, p.30.

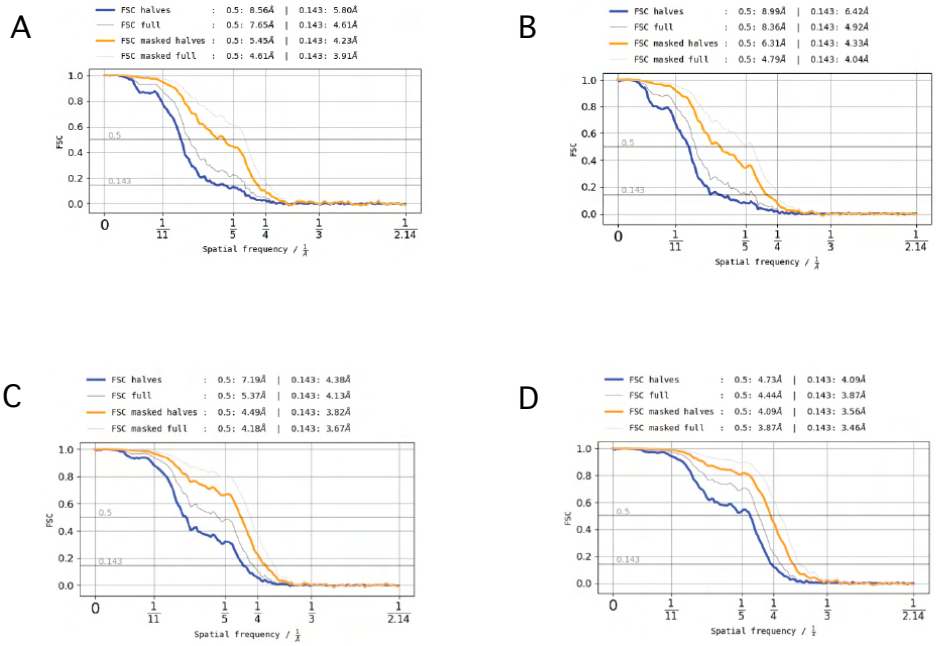
- Singh, M. et al., 2010. Plasmid DNA transformation in Escherichia Coli: effect of heat shock temperature, duration, and cold incubation of CaCl2 treated cells. *International Journal of*
- Stabrin M, Schoenfeld F, Wagner T, Pospich S, Gatsogiannis C, Raunser S. TranSPHIRE: automated and feedback-optimized on-the-fly processing for cryo-EM. Nat Commun. 2020 Nov 11;11(1):5716.
- Studer, D., Humbel, B.M. & Chiquet, M., 2008. Electron microscopy of high pressure frozen samples: bridging the gap between cellular ultrastructure and atomic resolution. *Histochemistry and cell biology*, 130(5), pp.877–889.
- Tanaka, K. et al., 2015. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. *Nature communications*, 6(1), p.6337.
- Tang, G. et al., 2007. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1), pp.38–46.
- Tang Wj., Guo Q. The Adenylyl Cyclase Activity of Anthrax Edema Factor. *Mol Aspects Med* 2009 Dec; 30 (6): 423-430
- Terwilliger, T.C. et al., 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography*, 64(Pt 1), pp.61–69.
- Teter, K. Toxin instability and its role in toxin translocation from the endoplasmic reticulum to the cytosol. *Biomolecules* **3,** 997–1029 (2013).
- Tonello, F., Naletto, L., Romanello, V., Dal Molin, F. & Montecucco, C. Tyrosine-728 and glutamic acid-735 are essential for the metalloproteolytic activity of the lethal factor of Bacillus anthracis. *Biochem. Biophys. Res. Commun.* **313,** 496–502 (2004).
- P. Turnbull. Medical Microbiology. 4th edition. 1996
- P. Turnbull and O. Cosivi. Anthrax in humans and animals. 2008 https://www.who.int/csr/resources/publications/anthrax_webs.pdf
- Ulmer T.Soelaiman S., Li S., Klee C. B., Tang W., Bax A. *Biol Chem* 2003 Aug 1;278(31):29261-6

- Vigneux F, Zumbihl R, Jubelin G, Ribeiro C, Poncet J, Baghdiguian S, Givaudan A, Brehélin M. The xaxAB genes encoding a new apoptotic toxin from the insect pathogen Xenorhabdus nematophila are present in plant and human pathogens J Biol Chem 2007 Mar 30;282(13):9571-9580.
- Xu, H. et al., 2014. Innate immune sensing of bacterial modifications of Rho GTPases by the Pyrin inflammasome. *Nature*, 513(7517), pp.237–241.
- Williams, C.J. et al., 2018. MolProbity: More and better reference data for improved all- atom structure validation. *Protein science : a publication of the Protein Society*, 27(1), pp.293–315.
- Wilson, J.S. et al., 2019. Identification and structural analysis of the tripartite α-pore forming toxin of Aeromonas hydrophila. *Nature communications*, 10(1), p.2900.
- Yamashita, D. et al., 2014. Molecular basis of transmembrane beta-barrel formation of staphylococcal pore-forming toxins. *Nature communications*, 5, p.4897.
- Yang, Z. et al., 2012. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure (London, England : 1993)*, 20(2), pp.237–247.
- Yuan, H., Johnson, S. L., Chen, L. H., Wei, J. & Pellecchia, M. A Novel Pharmacophore Model for the Design of Anthrax Lethal Factor Inhibitors. *Chemical Biology & Drug Design* **76,** 263–268 (2010).
- Young, J. A. T. & Collier, R. J. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **76,** 243–265 (2007).

Supplementary figures



Supplementary Figure 1 : PA_{83} plasmid. The plasmid shows the ampicillin (Amp) resistance, the enterokinase cutting site (enterokinase site) and the 10x His-tag.



Supplementary Figure 2 : FSC graphs. FSC curves between two independently refined half-maps of PA_7LF_{2+1A} (A), $PA_7LF_{(2+1A)'}$ (B), PA_7LF_{2+1B} (C) and $PA_7LF_{3\text{-masked}}$ (D), respectively.

Supplementary table

Table S1. Data collection, refinement and model building statistics.

	$PA_7LF_{3\text{-masked}}$	PA_7LF_{2+1B}	PA_7LF_{2+1A}
Microscopy and cryo-EM			
Microscope	Titan Krios		
Voltage [kV]	300		
Defocus range [μm]	-1.2 to -2.6		
Camera	K2 Summit		
Pixel size [Å]	1.07		
Total electron dose [$e^-/Å^2$]	74.4		
Exposure time [s]	15		
Frames per movie	40		
Number of images	5238		
Map resolution [Å]	3.5	3.8	4.2
Model statistics (phenix)[a]			
Molprobity score	3.00	2.42	
EMRinger	2.34	1.93	
Bond RMSD [Å]	0.008	0.004	
Angle RMSD [°]	0.723	0.753	
Ramachandran favored [%]	90.26	92.10	
Ramachandran outliers [%]	0.03	0.23	

Model statistics (iMODFIT)			
Molprobity score	3.01	2.43	2.63
EMRinger	0.61	0.46	1.18
Bond RMSD [Å]	0.011	0.005	0.009
Angle RMSD [°]	1.201	0.701	0.937
Ramachandran favored [%]	84.37	89.06	86.05
Ramachandran outliers [%]	0.96	0.09	0.02

[a]A detailed description of which parts of the models were refined with phenix or flexibly fitted using iMODFIT can be found in the Material and Methods chapter.

Publications and Conference Contribution

Publications

- Wagner T, Merino F, Stabrin M, Moriya T, **Antoni C**, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, Quentin D, Roderer D, Tacke S, Siebolds B, Schubert E, Shaikh TR, Lill P, Gatsogiannis C, Raunser S. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol. 2019 Jun 19;2:218. doi: 10.1038/s42003-019-0437-z. eCollection 2019.
- **Antoni C***, Quentin D*, Lang AE, Aktories K, Gatsogiannis C, Raunser S. Cryo-EM structure of the fully-loaded asymmetric anthrax lethal toxin in its heptameric pre-pore state. PLoS Pathog. 2020 Aug 18;16(8):e1008530. doi: 10.1371/journal.ppat.1008530. eCollection 2020.
- Fitzian K, Brückner A, Brohée L, Zech R, **Antoni C**, Kiontke S, Gasper R, Linard Matos AL, Beel S, Wilhelm S, Gerke V, Ungermann C, Nellist M, Raunser S, Demetriades C, Oeckinghaus A, Kümmel D. TSC1 binding to lysosomal PIPs is required for TSC complex translocation and mTORC1 regulation. Mol Cell. 2021 Jul 1;81(13):2705-2721.e8. doi: 10.1016/j.molcel.2021.04.019. Epub 2021.
- Klink B U*, Herrmann E*, **Antoni C**, Langemeyer L, Kiontke S, Gatsogiannis C, Ungermann C, Raunser S, Kümmel D. Structure of the Mon1-Ccz1 complex reveals molecular basis of membrane binding for Rab7 activation. PNAS, 2022.

Conference contribution

Antoni C. & Raunser, S. Unravelling structural secrets of anthrax toxin complexes. Oral presentation. IMPRS-CMB, Max-Planck Institute of Molecular Physiology, Dortmund, Germany, 2019.

Acknowledgements

The PhD student enters the working life on the tips of his feet. Insure and motivated, he builds his knowledge stone after stone, in both personal and professional meaning. He lives behind his comfort zone, made by family and friends, driving towards the dream of the knowledge.

During this journey, the personal and professional support make these intricated wheels properly turning. Therefore, I would like to thank every member of the dept.3 that at different level provided professional support, kindness, comprehension and friendship. I really enjoyed the company and atmosphere of my office in B2.24. The constant support of each other made possible to face every situation from different angles, constantly learning new ways.

My gratitude is to Oliver and Daniel for their extreme availability, professionality and deep interest in everyone's project. To Christos that initiated me in the cryo-EM field and guide me all long my time at MPI. To Christa and Lucia, precious shoulders from the beginning to the end of this journey. To Stefan being my mentor, giving me this great opportunity, interesting scientific questions and reinforcing my passion for the scientific research. To Prof. Dr. Roland Winter for having accepted to review my PhD thesis and being my relator.

To my family for their support and strength. To my grandparents for their unconditional love. To my little Chloé, my sweetest sunshine ever.