The University Of Sheffield.

Design and engineering of thermostable recombinant Protective Antigen – The next-generation Anthrax vaccine for disease control and prevention

By:

José Luis Avalos Díaz

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department/School of Chemical and Biological Engineering

February 2020

DECLARATION

I, Jose Luis Avalos Diaz, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university. Parts of this work are intended to be presented in form of scientific publications in the future.

Table of Contents

List of figures

List of tables

List of abbreviations

APR – Aggregation prone region
APS – Ammonium persulfate
AVA – Anthrax Vaccine Adsorbed
AVP – Anthrax Vaccine Precipitated
B. anthracis – *Bacillus anthracis*
B. subtilis – *Bacillus subtilis*
$CaCl_2$ – Calcium chloride
CMG2 – Capillary morphogenesis protein-2
CO_2 – Carbon dioxide
CV – Coefficient of variation
CVS – Column volumes
DDI – Distilled de-ionized
DEAE-column – Diethyl-aminoethyl column
DNA – Deoxyribonucleic acid
DNI - Dominant negative inhibitors
dNTP – deoxynucleotide triphosphate
DSF – Differential scanning fluorometry
E. coli – *Escherichia coli*
ED_{50} – effective dilution at 50% inhibition
EDTA – Ethylenediaminetetraacetic acid
EF – Oedema Factor
ELISA – Enzyme-linked immunosorbent assay
Eq. – Equation
EtOH – Ethanol
GK – Gate keepers
H_2O – Water
HIV – Human immunodeficiency virus
IPTG- Isopropyl β-D-1-thiogalactopyranoside
K_2HPO_4 – Potassium phosphate dibasic
KH_2PO_4 – Potassium phosphate monobasic
LF – Lethal Factor
MIDAS – Metal ion-dependent adhesion site
MP – Master-plate
NaOH – Sodium hydroxide
PA – Protective Antigen
PA(+++)– protease- aggregation-resistant rPA
PA63 – PA activated after cleavage
PAGE – Polyacrylamide gel electrophoresis
PEP – Post-exposure prophylaxis
pelB – Pectate lyase B
pI – Isoelectric point
Q-column – Quaternary amine column
rPA – Recombinant Protective Antigen
rPA – Recombinant Protective Antigen
SARS-CoV-2 – Severe acute respiratory syndrome coronavirus 2
SEM – Standard error of the mean
SDM – Site directed mutagenesis
SD – Standard deviation
SDS – Sodium dodecyl sulphate
TEM8 – Tumour endothelium marker-8
TEMED – Tetramethylethylenediamine
TNA – Toxin-neutralising antibody
VWA/1 – von Willebrand factor type A
WT – Wild-type

Acknowledgements

I would like to thank The University of Sheffield and my sponsor CONACyT for supporting this project.

This work would have not been possible without the guidance and supervision of Dr. Tuck S. Wong, who always encouraged me to become a better scientist. For his advice I will be forever thankful. I would also like to express my gratitude to Dr. Kang Lan for her invaluable support and advice in the lab. To my dear colleagues Bárbara, Miriam, Abdul, Inas, Powel, Yomi, Zaki (who never dropped a tennis set against me!), Melvin, Rob and Valeriane, I have nothing but gratitude and admiration for you all. Thanks for being there to share ideas, woes and joys.

The seasons did not pass quietly thanks to all the friends I made on the road. Jaimito, Miklos (best-house-mate-ever), Ryo, Josie, Liam, Ali, Tom, Emilio, ¡los quiero!

Finally, this thesis is dedicated to my family. I cannot express how much have I missed them all during these years. To my mom and dad for their unconditional love and support. To my brother Alan (whose laptop typed the whole of this thesis) and my sister Fabiola. And specially to my love, Magda.

Abstract

Anthrax is a deadly infectious disease whose aetiological agent is *Bacillus anthracis* (*B. anthracis*). Considered a long-known bioterrorist threat, Anthrax poses a Public Health challenge for effective and swift response in case of an outbreak. The current Anthrax vaccines are inactivated cell-free formulations of alum-precipitated antigens produced by non-pathogenic strains of *B. anthracis*. Efforts to produce a well-defined next-generation Anthrax vaccine have been focused on the production of recombinant Protective Antigen (rPA), one of the protein components of the Anthrax toxin. Although some rPA-based vaccines are under development, they have shown poor stability and short shelf-life. Efforts to improve their stability have been mainly focused on the development of novel adjuvants and delivery methods. This thesis proposes the generation of a thermostable rPA variant using protein engineering approach as a candidate for the development of a well-defined thermostable next-generation Anthrax vaccine. In this work, a bioprocess for soluble overexpression of rPA in *E. coli*, its recovery and purification was developed. Then, an rPA triple mutant was generated (PA(+++)) to obtain a protease- aggregation-resistant variant. PA(+++) was utilised as parental template for the protein engineering campaign. A semi-rational approach library was generated with mutations designed to improve PA(+++) hydrophobic core-packing. Also, a small rational approach library was generated following stability predictions using computational methods. Then, a semi-high throughput method was developed for screening thermostable variants based on detection of thermally induced aggregates with a polarity-sensitive fluorescent dye. Four variants resulted positive and were characterised using biophysical techniques. Intrinsic fluorescence spectroscopy and limited proteolysis assays demonstrated that thermostable variants maintained native structural conformation and biological function. Further biological function and immunisation studies are needed to determine if the new variants can elicit production of neutralising antibodies. This work contributed to the development of rationally designed vaccines and thermostable antigens. In conjunction with novel adjuvant technology, this approach can be an alternative to produce recombinant-based vaccines with longer shelf-life and achieve a long-desired unrefrigerated storage and transport.

Chapter 1. Introduction: Development of a next-generation Anthrax vaccine.

1.0 Abstract

This study aims to improve the thermostability of a next-generation Anthrax vaccine using recombinant DNA technology and semi-rational protein engineering approach. This introductory chapter covers the general aspects of Anthrax disease, including its aetiology and virulence factors. The chapter is mainly focused on reviewing Protective Antigen (PA), which is the target for the development of a well-defined recombinant Anthrax vaccine. The manufacturing of the current commercially available vaccines and the current efforts to produce a next-generation vaccine based on recombinant Protective Antigen (rPA) are as well reviewed.

1.1 Anthrax

Anthrax is a zoonotic disease caused by the Gram-positive spore-forming bacterium *Bacillus anthracis* (*B. anthracis*), a saprophytic soil organism whose spores germinate in the rhizosphere to establish populations of vegetative bacilli, thus grazing herbivores are its natural host (Saile and Koehler, 2006). Humans whose job involve working with animals or raw animal products are vulnerable to get in contact with Anthrax spores and get infected by inhalation, ingestion or dermal exposure (Williamson and Dyson, 2015).

Anthrax has accompanied humanity throughout history, dating back to Biblical times where the 5th plague described in the Book of Exodus is attributed to the disease. Other examples where Anthrax was purported during Ancient times events include the suggestion of inhalational Anthrax as the cause of the Plague of Athens, in Ancient Greece, and a

thorough description of its symptoms, by the renowned poet Virgil, during an epizootic in Ancient Rome. Symptoms associated with Anthrax are also described in records of epidemics and epizootics in Europe during pre-modern times (Sternbach, 2003). Genomic studies have shown that *B. anthracis* genetic diversity is higher in places where ancient and pre-modern Euro-Asian commerce routes were established, suggesting that migration and trade influenced global spread of *B. anthracis* (Van Ert *et al.*, 2007; Simonson *et al.*, 2009). In the 1870s, Koch established that *B. anthracis* was the causative agent of Anthrax, demonstrating for the first time the microbial aetiology of an infectious disease. Shortly after, Pasteur discovered that *B. anthracis* virulence could be abolished by heat treatment and that subsequent inoculation with the weakened pathogen could protect sheep against disease, creating the first attenuated vaccine (Carter, 1988). During XIX and early XX centuries, inhalational Anthrax became a feared disease in the textile industry of Yorkshire, England, where it was discovered that unprocessed wool contained Anthrax spores, leading to the establishment of norms to disinfect raw wool using formaldehyde, causing significant reductions on infections (Laforce, 1978).

Nowadays, the incidence of human Anthrax in developed countries is low. The disease has appeared in western countries involving imports of contaminated animal hides (Anaraki *et al.*, 2008; Marston *et al.*, 2011; Pullan *et al.*, 2015) and as injectional Anthrax, a new form of the disease contracted by heroin users in the United Kingdom and Europe (Berger, Kassirer and Aran, 2014). However, Anthrax is a prevalent disease in tropical developing countries with inadequate veterinarian surveillance and vaccination management (Chakraborty *et al.*, 2012; Emanuele Campese, 2015; Okello, Welburn and Smith, 2015). Recently, an Anthrax outbreak caused social disruption in an Artic region of Russia after a

heat-wave thawed the permafrost. The thawed surface left dormant Anthrax spores exposed on the soil, infecting thousands of reindeers that then spread the disease to humans, setting a precedent for latent climate-associated risks of outbreak (Timofeev *et al.*, 2018). Apart from naturally acquired infections, Anthrax has been utilised as a biological weapon using manipulated spores for its intentional release. The extent of social disruption that a biological attack can accomplish was demonstrated after the intentional distribution of Anthrax spores via the postal system in the United States, after the September 2001 terrorist attacks, causing 22 cases of Anthrax, including 5 deaths (Hughes, 2002). Thus, active surveillance and preparedness are major concerns for public health (Goel, 2015).

1.2 Anthrax virulence factors

The most lethal form of the disease is inhalational Anthrax, presenting a mortality of 90% if untreated, and 45% if timely diagnosed and treated with antibiotics (Holty *et al.*, 2006). Post inhalation, spores invade the respiratory tract where alveolar macrophages and lung dendritic cells engulf and transport them to the thoracic lymph nodes. Spores germinate on route, and vegetative bacilli propagate in the lymph nodes to commence virulence (Cleret *et al.*, 2014).

B. anthracis hijacks the host immune system by synthesizing an antiphagocytic capsule made of poly-γ-D-glutamic acid (PGA) and secreting toxins that kill phagocytes (Williamson and Dyson, 2015). These virulence factors are encoded in two plasmids: pXO1 (182 kb) and pXO2 (96 kb) (**Figure 1.1**). The pXO1 contains the genes *pagA*, *lef* and *cya* coding for the Anthrax toxin components: Protective Antigen (PA), Lethal Factor (LF), and Oedema Factor (EF), respectively. The plasmid pXO2 carries the *cap* genes that are involved in the synthesis of the PGA antiphagocytic capsule (Bourgogne *et al.*, 2003).

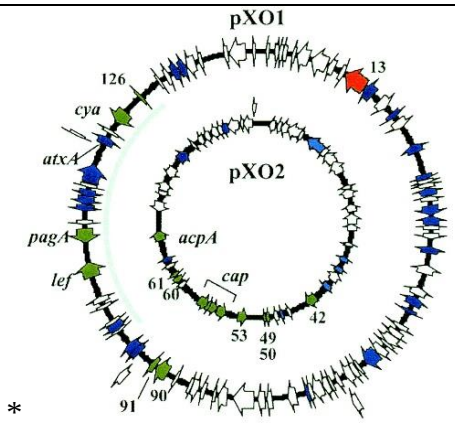


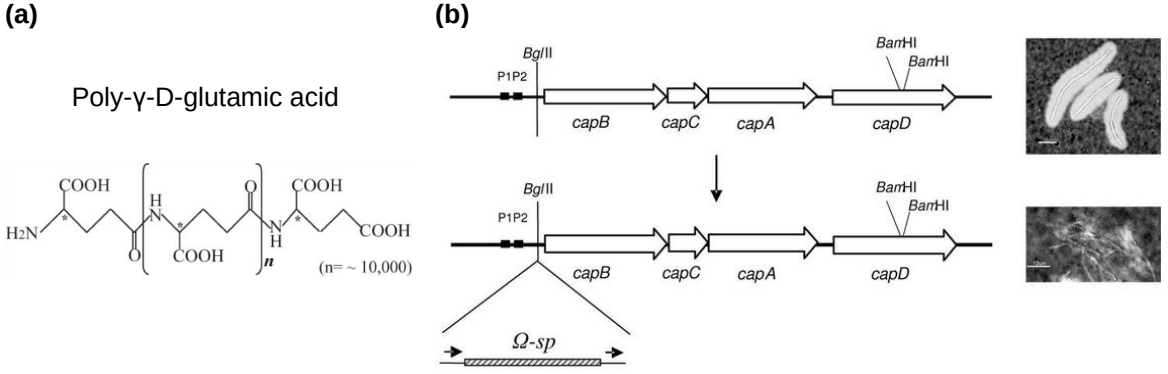
Figure 1.1. Virulent plasmids of Bacillus anthracis. pXO1 (182 kb) and pXO2 (96 kb) encoding the Anthrax virulence factors: Anthrax toxins and the poly-γ-D-glutamic acid, respectively.

*Modified from Bourgogne et al., (2003)

Figure 1.1 Virulent plasmids of Bacillus anthracis

1.2.1 Poly-γ-D-glutamic acid capsule

PGA is an anionic biopolymer, comprised of glutamic acid residues polymerised via formation of peptide bonds between the γ-carboxyl and α-amino groups (**Figure 1.2a**). Its biocompatibility properties have shown promising uses in the pharmaceutical and food industry (Hsueh *et al.*, 2017).



*Modified from Pomerantsev *et al.*, (2006)

Figure 1.2 *B. anthracis* capsule production. (a) Poly-γ-D-glutamic acid chemical structure. **(b)** Figure modified from Pomerantsev *et al.*, (2006) depicting operon *capBCADE* regulated by promoters P1 and P2. Truncation of transcription downstream P1-P2 abolished *B. anthracis* capsulation.

Figure 1.2 B. anthracis capsule production

B. anthracis produces poly-γ-D-glutamic acid to form an antiphagocytic capsule, key for pathogenesis. Genes for synthesis and attachment of the polymer onto the cell surface are encoded in plasmid pOX2, in the operon *capBCADE* (depicted in **Figure 1.1** as *cap* region, and in **Figure 1.2b**). Putative enzymes CapB, CapC, CapA, and 47-amino-acid peptide CapE, are involved in the polymerisation of PGA (Candela, Mock and Fouet, 2005), whereas CapD has proven to depolymerise PGA on the cell surface, releasing shorter D-glutamic acid peptide fragments that then attaches covalently to the peptidoglycan (Makino *et al.*, 2002; Candela and Fouet, 2005). The operon has two in-tandem promoters upstream of *cap*, P1 and P2, which are positively trans-regulated by the products of *acpA*, and in some extent by *atxA*, in presence of CO_2 (Uchida *et al.*, 1997). Gene *atxA* is harboured in pOX1 plasmid, exhibiting crosstalk between both virulence plasmids, yet there is evidence of an AtxA/CO_2-independent mechanism, enabling positive trans-acting regulation of *cap* without pOX1 (Drysdale, Bourgogne and Koehler, 2005).

The detailed molecular mechanisms for capsule formation, including crosstalk between virulent plasmids, CO_2-dependant regulation and genetic background remain unclear. Also, recent works have proposed variation in gene regulation among different *B. anthracis* strains due to differences in genetic background and plasmid mutations (Liang *et al.*, 2016, 2017). A deeper understanding of the implications of CO_2 in the genetic regulation of Anthrax pathogenicity could be exploited for the development of new therapies as well as for the development of novel genetic engineering tools for cell-line development and promoter engineering.

1.2.2 Anthrax toxin

The Anthrax toxin components are encoded in plasmid pOX1 (**Figure 1.1**). The product of genes *pagA*, *cya* and *lef* are Protective Antigen (PA), Oedema Factor (EF) and Lethal Factor (LF), respectively (**Figure 1.3**). Transcription of these genes is positively regulated by AtxA in presence of CO_2 (Dai *et al.*, 1995). It is now evident that AtxA is considered a global regulator of *B. anthracis* virulence, orchestrating the regulation of all virulence factors (Bourgogne *et al.*, 2003; Missiakas and Schneewind, 2017). However, the molecular mechanisms by which AtxA regulates expression of the toxin components remain unclear. Recent studies on *pagA* transcription mediated by AtxA found that a rich-TA sequence upstream *pagA* promoter acts as a silencer and suggested that AtxA competes with a repressor protein to release suppression of *pagA* transcription (Toyomane *et al.*, 2019). However, this study was performed using *E. coli* as heterologous model for transcription studies, thus further evidence using a *B. anthracis* strain as model is needed to validate the reported mechanisms of AtxA on gene regulation of the Anthrax toxin.

The Anthrax toxin is a classic AB toxin: an effector protein (A) assembled with a delivery protein (B) that mediates toxin translocation into the cell. The delivery protein is PA, that in combination with the effector proteins can form Lethal toxin (PA+LF) or Oedema toxin (PA+EF).

LF is a zinc-dependent metalloproteinase, whose molecular targets are the mitogen-activated protein kinase kinases (MEKs). Its action results in disruption of pathways involving mitogen-activated protein kinase (MAPK), impairing production of inflammatory cytokines involved in proliferation of $CD4^+$ T-cells and macrophages (Fang *et al.*, 2005; Liu, Moayeri and Leppla, 2014).

EF is a calmodulin-dependent adenylyl cyclase that catalyses synthesis of cAMP in a calmodulin-dependent way, leading to a dramatic elevation of cAMP levels that in consequence disrupts signalling involving protein kinase A (PKA) pathway (Liu, Moayeri and Leppla, 2014).

Toxaemia consequences in mice are lymphocytolysis, necrotic lesions in several tissues including adrenal glands, ileum, heart and kidney, ultimately leading to hepatic (by EF) or cardiovascular failure (by LF) (Firoved *et al.*, 2005; Liu *et al.*, 2013; Goldman, 2014).

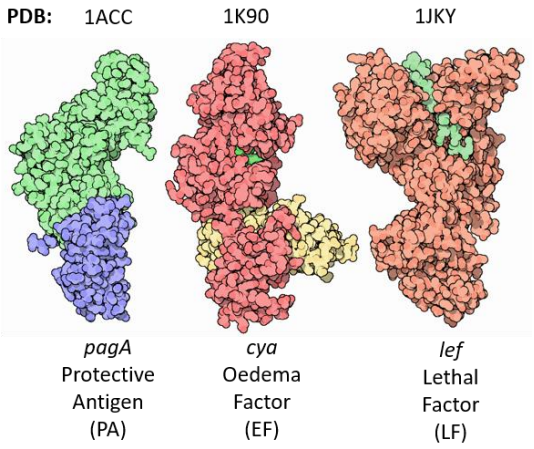


Figure 1.3 The Anthrax toxin components. The Anthrax toxin is a classic AB toxin whereby PA function is to deliver the effector proteins, LF and EF, into the host's cells

Figure 1.3 The Anthrax toxin components

1.3 The central role of Protective Antigen

Individually, the Anthrax toxin components are harmless. In order to achieve virulence, *B. anthracis* depends on PA for (1) toxin assembly and (2) toxin translocation.

(1) (**Section 1.3.1**) Toxin assembly starts with proteolytic activation of PA after binding its target receptors. PA cleavage triggers the formation of ring-shaped heptamers (or octamers) that serve as binding sites for LF/EF, forming a complex called the pre-pore.

(2) (**Section 1.3.2**) Toxin translocation starts with the endocytosis of the pre-pore by receptor-mediated endocytosis. Then, LF/EF translocation from late endosome into the cytosol is mediated by a conformational rearrangement of the pre-pore into a membrane-spanning pore.

PA accomplish these major conformational changes through the activity of its four domains (Petosa *et al.*, 1997) (**Figure 1.4**): Domain 1 binds two calcium ions and contains the cleavage site critical for protein activation by membrane proteases; Domain 2 is involved in the formation of a trans-membrane insertion pore that mediates toxin translocation; Domain 3 mediates oligomerisation during toxin assembly; Domain 4 binds to cell receptors.

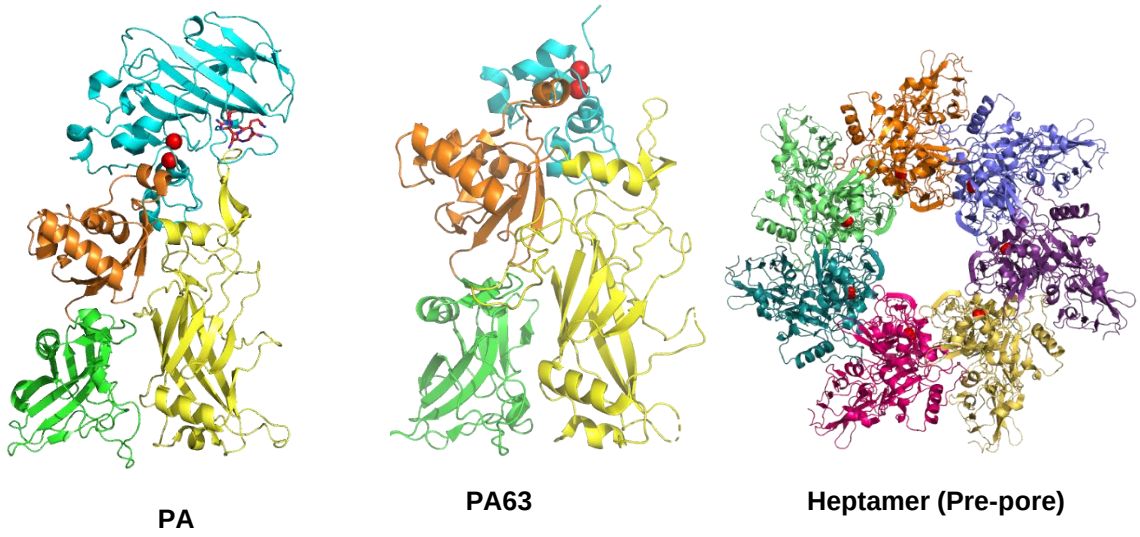


Figure 1.4. PA conformational rearrangements. PA (PDB: 3TEW) showing coloured domains. Domain 1: cyan; domain 2: yellow; domain 3: orange; domain 4: green. Domain 1 holds two Ca+2 ions (red) essential for biological function. Upon receptor binding, PA is cleaved in a furin-like proteolytic site (red sticks in domain 1) yielding PA63 (PDB: 1TZO, Chain A). PA63 triggers the formation of a heptamer (PDB: 1TZO) that serves as docking site for the effector proteins EF/LF, forming a structure known as the pre-pore.

1.3.1 Toxin assembly

1.3.1.1 PA receptor binding

The PA receptors are tumour endothelium marker-8 (TEM8) and capillary morphogenesis protein-2 (CMG2). Studies on the physiological function of these receptors implicate them in the regulation of extracellular matrix homeostasis and angiogenesis, suggesting an important role in extracellular matrix remodelling (St. Croix *et al.*, 2000; Reeves, Charles-Horvath and Kitajewski, 2013; Besschetnova *et al.*, 2015).

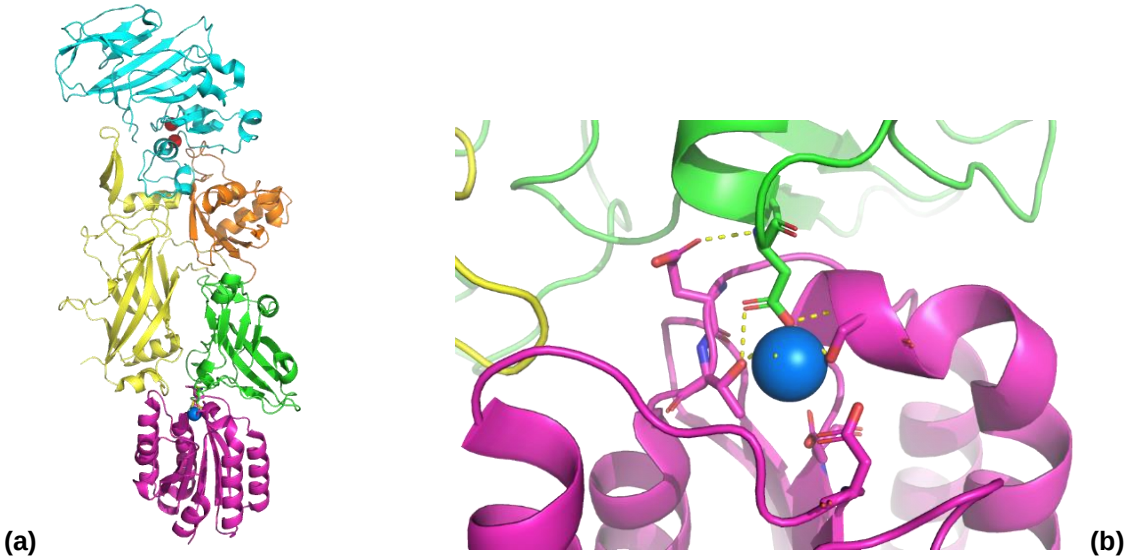


Figure 1.5. PA bound to VWA/1 domain of receptor CMG2. PA binds the domain VWA/1 of receptors CMG2 and TEM8 via domain 4 (green) and domain 2 (yellow). **(a)** (PDB: 1T6B) VWA/1 domain is in magenta. The blue sphere is a Mn^{+2} ion. **(b)** Residue D683 (green sticks) participates in the coordination of the divalent cation in VWA/1 during receptor binding.

Figure 1.5 PA bound to VWA/1 domain of receptor CMG2

PA binds to TEM8/CMG2 in their extracellular von Willebrand factor type A (VWA/1) domain (**Figure 1.5**). VWA/1 contains a metal ion-dependent adhesion site (MIDAS) that mediates PA binding in presence of a divalent cation (Ca^{2+}, Mg^{2+} or Mn^{2+}).

Binding is conducted via coordination of the divalent cation by residues within MIDAS, PA residue D683 and water (Bradley *et al.*, 2003). Although VWA/1 domain is highly conserved in both receptors (60% identical), *in vitro* studies have shown that PA presents ~ 1000-fold stronger binding affinity towards CMG2 related to TEM8 (Wigelsworth *et al.*, 2004; Scobie *et al.*, 2005).

CMG2 is the relevant receptor for Anthrax toxicity. CMG2-null mice are resistant to Anthrax toxin effects and can survive infection after spore challenge with non-encapsulated toxigenic *B. anthracis*. Thus concluding that toxaemia lethality in late stage infection is mediated by CMG2 (Liu *et al.*, 2009). To investigate the role of CMG2 during early stages of infection, the same research group developed a mice model with myeloid cells lacking CMG2, demonstrating that PA is essential to establish infection (Liu *et al.*, 2010). The full mechanism involving PA on infection's early stage is unclear and its elucidation could fully explain PA role as "protective antigen" against disease in the current Anthrax vaccines.

TEM8 and CMG2 are endothelial receptors that have been found overexpressed in some tumours, thus there have been efforts to exploit PA specificity towards these receptors to develop PA-based tumour markers for diagnosis and as an anti-cancer therapy (Samples, Willis and Klauber-DeMore, 2013; Bachran and Leppla, 2016; Chen *et al.*, 2016). Also, anti-toxin approaches can be directed towards inhibition of receptor binding. Recent studies suggest that PA exhibits proteolytic activity that is necessary for receptor binding, being a potential target for development of competitive inhibitors (Storm *et al.*, 2018).

1.3.1.2 PA oligomerization and the pre-pore

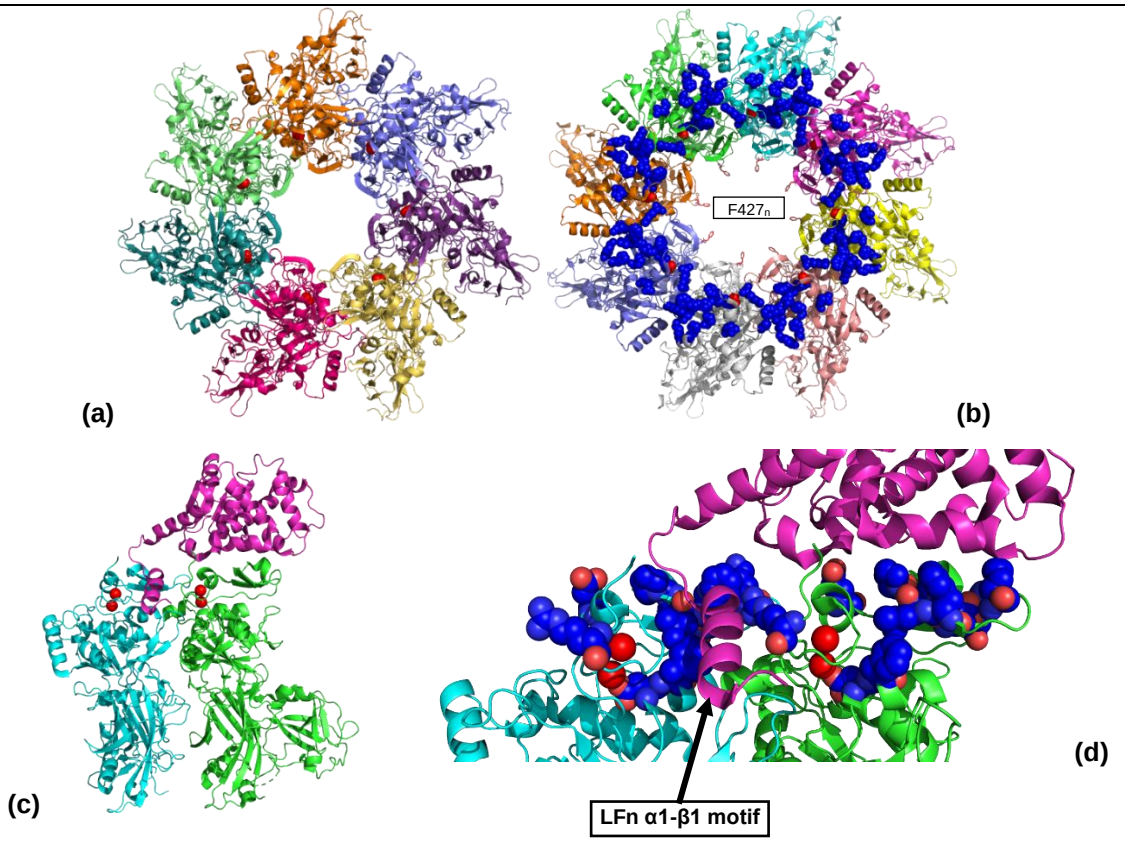


Figure 1.6. PA63 oligomers form a ring-shaped structure known as the pre-pore. (a) Heptamer (PDB: 1TZO). **(b)** Octamer (PDB: 3HVD). Blue spheres are the binding sites for lethal factor. Sticks facing the pre-pore lumen represent the φ-clamp formed by F427 from each protomer ($F427_n$). **(c)** LF N-terminal (LFn) bound to two PA63 protomers (cyan and green). Red spheres are PA63 Ca^{2+} ions (PDB: 3KWV) **(d)** LFn α1-β1 motif bound to PA63 α-clamp. This site promotes the entry of LF into the pre-pore upon pore formation whereby LF is translocated under a proton gradient.

Figure 1.6 PA63 oligomers form a ring-shaped structure known as the pre-pore

After binding its target receptor, PA (83 kDa) is proteolytically activated by a furin-like protease, realising a 20 kDa N-terminal fragment whereas its C-terminal 63 kDa (PA63) moiety remains bound to the cell receptor. This event triggers PA oligomerisation, comprising 7 or 8 protomers that form a ring-shaped structure called the pre-pore, while creating binding sites for EF and LF in PA63 Domain 1 (Fabre *et al.*, 2016), allowing three

(heptameric PA) or four (octameric PA) LF/EF proteins to bind onto the pre-pore (Kintzer *et al.*, 2009) (**Figure 1.6**), producing an AB-like toxin. PA oligomerisation can also be triggered by serum proteases thus toxins can assemble in lymph and plasma (Moayeri, Wiggins and Leppla, 2007).

The pre-pore is an important structural precursor that mediates virulence. When the toxin is endocytosed by a host cell it is confined in endosomal vesicles. Endosome acidification due to low pH conditions in the cytosol causes structural changes in oligomeric PA that triggers its conversion to a membrane-spanning pore that translocate LF/EF from endosome to cytosol (Bann, 2012). Thus, the molecular function of the pre-pore is to stabilise PA and prevent unspecific pore formation out of the cytosol.

It is suggested that EF and LF binding motifs and ligand interactions with PA are similar (Cunningham *et al.*, 2002). LF has a N-terminal α1-β1 motif (residues 29–50) that interacts with PA63 at the top of the ring-shaped pre-pore, in a site formed between two adjacent domain 1 called the α-clamp (**Figure 1.6c**). Hydrophobic and aromatic residues surrounding the α-clamp destabilise LF/EF structure, functioning as well as an unfolding machine during translocation (Brown, Thoren and Krantz, 2015). The pre-pore structure is held by interactions between domain 3 and domain 1. Domain 2 faces the pre-pore lumen whereby residues F427 from each protomer form an internal ring (**Figure 1.6a**) which plays a major role during protein translocation and as pH sensor for pore formation, this structure is known as the φ-clamp (Jiang *et al.*, 2015). Domain 4 presents a hinge-like behaviour that allows rearrangement of PA63 conformation while bound to receptor. (Feld *et al.*, 2010, 2012).

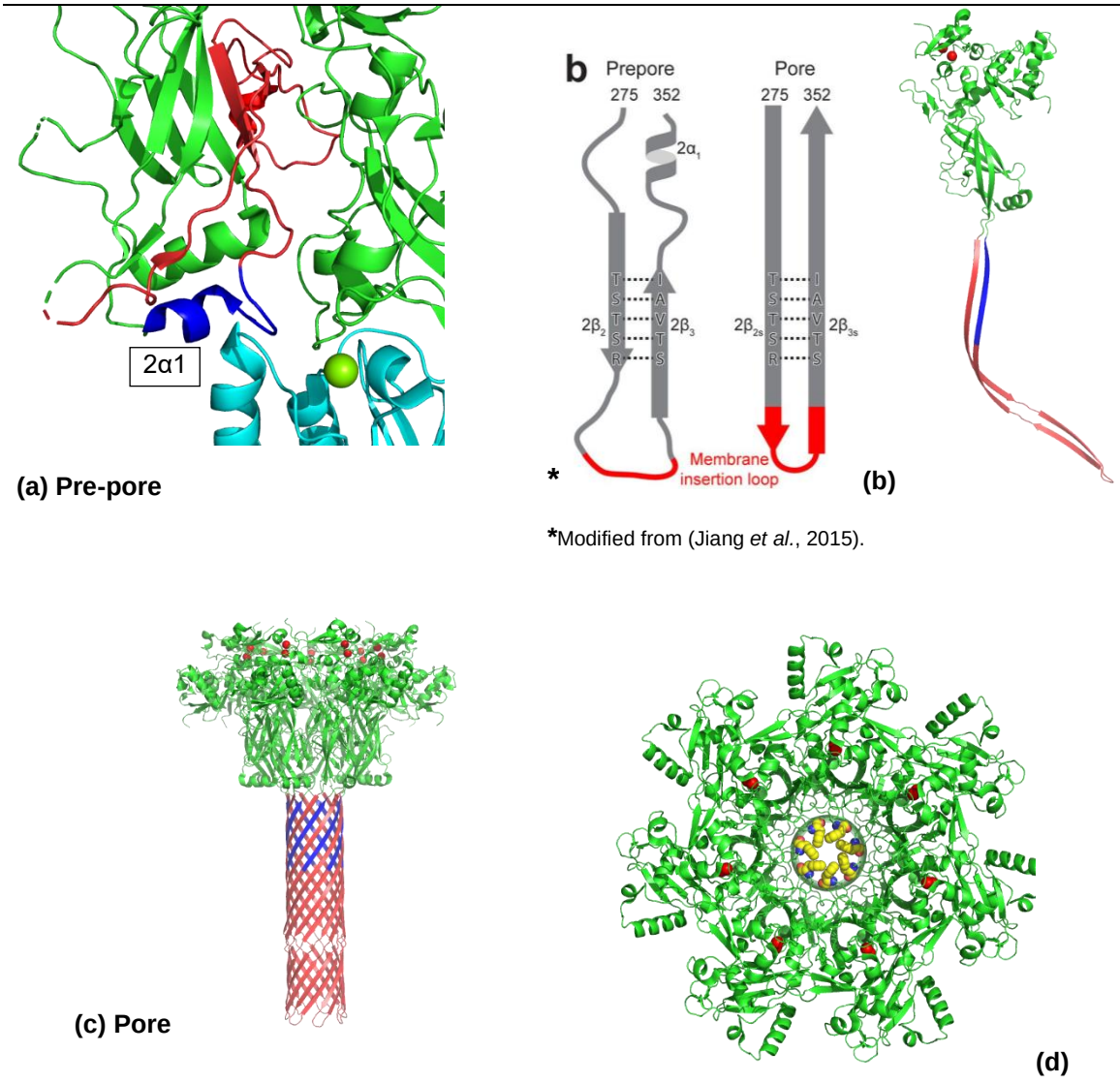


Figure 1.7 Pre-pore to pore transition. (a) Pre-pore (green) bound to CMG2 receptor (cyan) (PDB: 1TZN). The β-barrel and transmembrane insertion loop are in red. The motif 2α1 (blue) is bound to CMG2. **(b)** The motif 2α1 dissociates from receptor driven by pH decline in the endosomes, leading the transition from pre-pore to pore (PDB: 3J9C). **(c)** The pore has a flower-on-a-stem structure (PDB: 3J9C). **(d)** View of the pore lumen form the top. Yellow spheres are residues F427 forming the φ-clamp.

1.3.2 Toxin translocation: transmembrane pore

The AB toxin, comprising effector proteins LF/EF bound to the PA63 pre-pore (**Figure 1.6**), is internalised via receptor-mediated endocytosis (Abrami, Leppla and Gisou Van Der Goot, 2006; Abrami *et al.*, 2010). PA remains bound to its receptor in the endosome, functioning as a pH sensor to trigger PA conformational rearrangements that lead to the formation of a transmembrane pore that serves as translocase for EF/LF. (Pilpa *et al.*, 2011; Chadegani *et al.*, 2014). This conversion is mediated by the dissociation of a loop in PA domain 2 (residues 340 – 351 forming a loop and the 2α1 turn, as shown in **Figure 1.7a,b**) bound to the surface of receptor CMG2 when pH reaches 5.0 – 6.0 (Chadegani *et al.*, 2014). Domain 4 flexibility allows PA to remain bound to the receptor, presenting a hinge-like motion (Feld *et al.*, 2012). The rearrangement of domain 2 reconfigures the pre-pore conformation and triggers the formation of a transmembrane β-barrel (**Figure 1.7b**) (Feld *et al.*, 2012). The β-barrel is formed by domain 2 β2- β3 strands, one form each PA63 protomer conforms the pore (**Figure 1.7c**) (Jiang *et al.*, 2015). Jiang *et al.*, (2015) resolved the pore crystal structure that is described as a flower-on-a-stem, comprising: (1) the corolla, which is the top of the ring that contains the binding sites for LF/EF as well as the α-clamp that facilitates LF/EF entry into the pore, (2) the calyx, which contains the φ-clamp, composed of F427 residues facing the pore lumen that are thought to catalyse LF/EF translocation (Janowiak, Fischer and Collier, 2010) (**Figure 1.7d**), (3) and the stem, that comprises the transmembrane β-barrel. This type of structure is found in other bacterial toxins, like α-haemolysin secreted by *Staphylococcus aureus* (De and Olson, 2011).

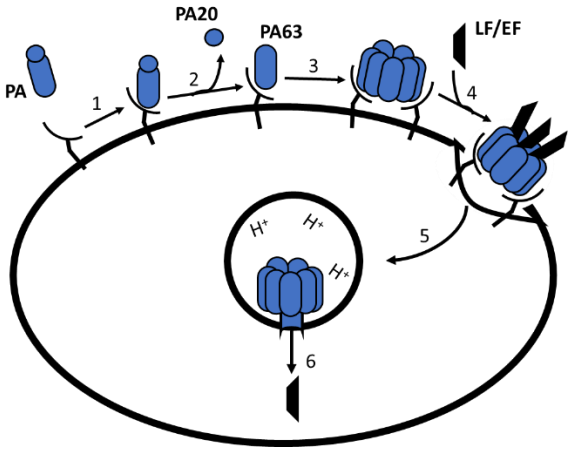


Figure 1.8 Toxin assembly and translocation is mediated by PA. (1) PA binds cell receptors CMG2/TEM8. (2) PA is cleaved by membrane proteases into PA63 upon binding its receptor. (3) PA63 forms oligomers (heptamers or octamers) producing a ring-shaped structure known as pre-pore. (4) LF/EF bind the pre-pore to assembly the Anthrax toxin. (5) The pre-pore is internalised via receptor-mediated endocytosis. (6) A pH-driven transition from pre-pore into a transmembrane pore occurs in the endosome and LF/EF are translocated via proton gradient and the translocase activity of the pore.

Figure 1.8 Toxin assembly and translocation is mediated by PA

EF/LF translocation is then driven by a proton gradient and catalysed by the pore translocase activity (Fabre *et al.*, 2016). Typically, experiments to study activity of the PA channel are performed on planar lipid bilayers (Blaustein *et al.*, 1989; Schiffmiller, Anderson and Finkelstein, 2015; Fabre *et al.*, 2016). A pH gradient is created between the cis and the trans sides of the membrane and PA channel activity is detected by changes in conductance. PA channel activity as well as docking and translocation of LF are voltage dependent and therefore can be modulated by controlling the potential across the membrane (Blaustein *et al.*, 1989). The feature of being able to effectively control PA channel activity by potential difference and to perfuse the inactivated/deficient channels or protein excess makes planar lipid bilayer the method of choice to study the PA pore and to study potential inhibitors (Omersa, Podobnik and Anderluh, 2019). The conclusions of these experiments are that PA

channels are strongly cation selective (Blaustein *et al.*, 1989; Schiffmiller, Anderson and Finkelstein, 2015); channels can be tightly controlled by modulating voltage and that both pore formation and channel activity are pH dependent (i.e. the pore is open at pH < 6.8) (Fabre *et al.*, 2016).

1.4 Protective Antigen and Anthrax vaccines

Vaccines are biological formulations that confer protection against a disease by mimicking the intrusion of an infectious agent. Vaccines are one of the most effective ways to prevent infectious diseases and their continuous development is essential to broad the protection against old-known infectious agents as well as new (Delany, Rappuoli and De Gregorio, 2014).

The development of the Anthrax vaccine has more than a century of history since Koch and Pasteur started their aetiology studies. Veterinary vaccines have long been used for livestock vaccination using live attenuated strains of *B. anthracis* (Turnbull, 1991). The use of non-pathogenic *B. anthracis* strains for the production of inactivated cell-free vaccines for humans was established since the early 1950s (Belton and Strange, 1954). This technology has been applied for the development of the current Anthrax vaccines whose approved licenses date from the 1970s. Anthrax Vaccine Precipitated (AVP), the vaccine manufactured in the UK by Porton Biopharma Ltd, contains high titres of PA produced in non-capsulated *B. anthracis* 34F2 Sterne strain grown statically in Thomson bottles containing 500 mL of medium supplemented with yeast hydrolysates and activated charcoal to induce high production of PA (Charlton *et al.*, 2007). Anthrax Vaccine Adsorbed (AVA or BioThrax), the US version of the vaccine manufactured by Emergent BioDefense Operations Lansing LLC (Emergent Biosolutions), is produced by synthesising PA in the non-capsulated *B.*

anthracis V770-NP1-R strain grown in stirred-vessel aerobic fermentation using minimal medium supplemented with bicarbonate (Puziss *et al.*, 1963). After fermentation, culture supernatants are recovered and filter sterilised. The final formulation of AVP is a mixture of alum-precipitated antigens (Charlton *et al.*, 2007), and in the case of AVA antigens are alum-adsorbed (Farchaus *et al.*, 1998). These formulations rely on PA to elicit protection against disease, however LF is found in significant amounts in AVP, as well as traces of EF and other undefined antigens (Whiting *et al.*, 2004; Charlton *et al.*, 2007; Whiting, Wheeler and Rijpkema, 2014). PA degradation, likely due to protease secretion by *B. anthracis* during fermentation, has also been reported (Charlton *et al.*, 2007; Whiting, Wheeler and Rijpkema, 2014). Although AVP is an effective vaccine, the presence of undefined components is a safety concern. Moreover, the complex mixture hinders thorough characterisation and potential lot-to-lot variation could affect vaccine quality.

1.5 Next-generation Anthrax vaccines: production of recombinant Protective Antigen

It is unclear how PA ultimately protects against Anthrax disease. Some reports show that immunisation with PA can prevent the installation of a systemic infection in mice and rabbits rather than responding during later stages of toxaemia (Merkel *et al.*, 2013). However, it is well accepted that neutralising antibodies against PA are the main correlates of protection (Quinn *et al.*, 2004; Laws *et al.*, 2016). Consequently, the main efforts to produce a well-defined next-generation Anthrax vaccine are focused on the production of recombinant PA (rPA)-based vaccines (Kaur, Singh and Bhatnagar, 2013).

Emergent BioSolutions has a rPA-based vaccine in the pipeline under the commercial name PreviThrax. The vaccine comprises highly pure rPA produced using a non-

encapsulated non-toxigenic *B. anthracis* strain (Farchaus *et al.*, 1998). Originally, the rPA gene sequence, *pagA* (GenBank: AF268967), was derived from plasmid pOX1 and cloned into *E. coli* using plasmid pBR32 (Vodkin and Leppla, 1983) and subsequently subcloned into *B. subtilis* using a compatible plasmid to obtain the clone rPA102 (Ivins and Welkos, 1986), which is employed for recombinant expression in *B. anthracis*. The attenuated *B. anthracis* strain used for production is ΔSterne-1 (pPA102) CR4: a non-toxigenic ($pOX1^-$), non-encapsulated ($pOX2^-$), asporogenic ($spoOA^-$) strain carrying the rPA102 construct (Worsham and Sowers, 1999). Using the native secretion system for rPA expression has the advantage of ensuring native folding of mature rPA, as well as allowing accumulation in the culture medium facilitating protein recovery. *B. anthracis* was preferred over *B. subtilis* because the protease secretion profile of ΔSterne-1 (pPA102) CR4 is lower related to *B. subtilis* and yields could be further improved by the addition of EDTA as protease inhibitor (Wu *et al.*, 1991). Novel approaches for production of rPA with *B. anthracis* include the use of genetically engineered strains that secretes less proteases (Pomerantsev *et al.*, 2011) and the use of promoter engineering for overexpression of rPA via trans-regulatory elements (Pomerantsev *et al.*, 2017; Sharma *et al.*, 2018). These and other approaches have been reported for production of well-defined rPA-based vaccines (Williamson *et al.*, 2005; Watkinson *et al.*, 2013), however regulatory approval has been halted due to issues regarding rPA stability (Kaur, Singh and Bhatnagar, 2013).

As described in **Section 1.3**, PA is a complex protein whose functionality depends on dramatic conformational rearrangements, displaying domains with high flexibility. Indeed, reports have shown that rPA can be denatured under relatively low temperature (37 °C) (Radha *et al.*, 1996), an indication of poor thermostability derived from its inherent

flexibility. Thus, manufacturers and researchers have allocated resources towards stabilisation of rPA-based vaccines, mainly through the development of novel adjuvants, carriers and lyophilised formulations (Wang *et al.*, 2012; Feinen *et al.*, 2014; Hassett *et al.*, 2015; Savransky *et al.*, 2019; Weir *et al.*, 2019). There is an empirical connection between pharmaceutical stability and conformational stability that dictates that the degradation pathways of therapeutic proteins can be inhibited if the conditions that promote their native folded state are provided, thus reducing protein motions via lyophilised formulations is very effective to achieve vaccine stability.

The advent of rationally designed vaccines opened the avenue to produce novel recombinant-based vaccines with improved features, including better stability (Giuliani *et al.*, 2006; Rappuoli *et al.*, 2016; Wan *et al.*, 2019). PA structure and function have been thoroughly studied during the years. Moreover, some of the PA protective epitopes have been mapped (McComb and Martchenko, 2016). Consequently, in this thesis we propose the application of protein engineering approach to design structure-based mutations to improve the hydrophobic core-packing of rPA, avoiding residues forming protective epitopes or that are essential for rPA functionality. Proteins with improved packing demonstrate reduced flexibility and higher thermostability (Sammond *et al.*, 2016), thus our aim is to obtain a thermostable rPA variant as a candidate for the development of a well-defined thermostable next-generation Anthrax vaccine.

1.6 Thesis content

The content of this thesis unfolds as follows. **Chapter 2** contains detailed description of the materials and methods used in this work and should be used as a reference to consult detailed descriptions on experiments. The next three chapters are experimental chapters which are

structured to present an introduction, aims, results and discussion, and conclusions. Methods are generally described in the results section and **Chapter 2** should be consulted for detailed experimental conditions and methods. **Chapter 3** comprises the designing of a bioprocess for soluble overexpression, recovery and purification of rPA and the generation of a protease-aggregation-resistant triple mutant (PA(+++)). **Chapter 4** presents the protein engineering approach campaign, comprising the generation of a semi-rational approach library and a rational approach library using structure-based design and computational methods aiming to improve PA(+++) hydrophobic core-packing, as well as the development of a semi-high throughput method for screening of thermostable variants. **Chapter 5** comprises the biophysical characterisation of the thermostable variants obtained after protein engineering approach. Finally, **Chapter 6** includes a general discussion of our findings, concluding remarks and prospects on rationally designed and thermostable vaccines.

Chapter 2. Materials and methods

2.1 Materials and equipment

2.1.1 List of chemicals

Potassium dihydrogen phosphate (KH_2PO_4), Kanamycin, Tris-base molecular biology grade, Sodium hydroxide (NaOH), β-meracaptoethanol, Acrylamide - Solution (30%), Tetramethylethylenediamine (TEMED) were obtained from **AppliChem GmbH**. EDTA disodium was obtained from **FORMEDIUM**. Glycerol, acetic acid and Sodium dodecyl sulphate (SDS) were obtained from **VWR**. PageRuler Broad Range Unstained was obtained from **Thermo Fisher Scientific**. Protein Broad Range Standard was obtained from **New England Biolabs Ltd**. Protein staining RAPIDstain was obtained from **G-Biosciences.** Ammonium sulphate, Monopotassium dihydrogen phosphate, Disodium hydrogen phosphate, Isopropyl β-D-1-thiogalactopyranoside (IPTG), Protease inhibitor: Complete Mini protease inhibitor cocktail tablets EDTA-free, sucrose, MOPS, HEPES, ethanol (HPLC-grade), HCl, Nile red, Thioflavin T, and all other reagents were obtained from **Sigma-Aldrich**.

2.1.2 List of enzymes and kits

Lysozyme was obtained from **Thermo Fisher Scientific**. 10x Trypsin solution from porcine pancreas, trypsin inhibitor from Glycine max (soybean), Deoxyribonuclease I from bovine pancreas (DNAse), Ribonuclease A from bovine pancreas (RNAse) were obtained from **Sigma-Aldrich**.

Q5 High-Fidelity DNA Polymerase (2 U/µL), T4 DNA Ligase (400,000 U/mL), DpnI (20,000 U/mL), T4 Polynucleotide Kinase (10,000 U/mL), T4 DNA ligase 10x Buffer with

10 mM ATP, 5x Q5 reaction buffer, dNTPs (10 mM) were obtained from **New England Biolabs**. QIAprep Spin Miniprep and QIAquick PCR Purification Kit were obtained from **QIAGEN**. NativePAGE™ Bis-Tris 4 – 16% gel and NativeMark™ Unstained protein and NativePAGE™ Bris-Tris buffers were obtained from **Invitrogen**.

2.1.3 Bacterial strains and plasmids

E. coli DH5α, *E. coli* BL21 (DE3) and plasmid pET24a(+) were obtained from **Novagen**.

2.1.4 Culture media (per 1 L)

Culture media reagents were obtained from **FORMEDIUM**. *2xTY media*: 16 g Tryptone, 10 g Yeast extract, 5 g NaCl; *TYE agar***:** 10 g Tryptone, 5 g Yeast extract, 8 g NaCl, 15 g Agar granulated. Overnight cultures were incubated in 5 mL 2xTY contained in 50-mL Falcon tubes using an ES-20 shaker-incubator (**Grant Instruments**). Inoculated agar plates were incubated in an INCU-Line incubator.

2.1.5 List of materials and equipment

Gels images were taken with GenoSmart2; sterile syringe filter 0.2 µm Cellulose Acetate, Nylon Membrane Filters 0.45 µm (**VWR**). 96-well round-bottom microplate, Greiner clear-bottom 96-well microtiter plate, black-bottom black 96-well microtiter plate (**Sigma-Aldrich**). Thermocycler Mastercycler personal; Microcentrifuge 5415R; BioPhotometer Plus UV/Vis photometer; Eppendorf UVette, thermoblock (**Eppendorf**). Centrifuge with temperature control: Multifuge 3 S-R; electrophoresis chamber for native-PAGE (**Thermo Fisher Scientific**). Electrophoresis chamber for SDS-PAGE (**Bio-Rad Laboratories**). JENWAY 3510 pH Meter (**JENWAY**). SpectraMax M2e microplate reader (**Molecular Devices**). AKTA FPLC system, HiTrap™ Q HP 5 mL column, HiTrap™ DEAE FF 5 mL,

HiLoadTM 26/600 Superdex^{TM} 200 pg, Whatman^{TM} 0.45 µm pore-size nylon membrane filters, 0.45 µm syringe filters and 0.2 µm filter unit are from **General Electric**.

2.2 Protein expression

2.2.1 Genetic construct pET24a-rPA

The rPA construct was synthesised and codon-optimised by GenScript using the gene *pagA*. The original PA signal peptide was replaced by the pelB signal peptide and restriction sites NdeI and EcoRI were added for cloning (see Appendix 1 for full sequence). The construct was previously cloned into pET24a by standard molecular biology techniques. The plasmid was transformed into *E. coli* DH5α and kept stored in a 25% glycerol stock at – 80 °C.

2.2.2 Bacterial transformation by $CaCl_2$ heat shock

Plasmids were isolated from *E. coli* DH5α or *E. coli* BL21 (DE3) from 5 mL overnight cultures of 2xTY with 50 mg/mL kanamycin, grown at 37 °C, shaking, using QIAprep Spin Miniprep Kit. For transformation, an overnight culture of *E. coli* DH5α or *E. coli* BL21 (DE3) was prepared in 5 mL of 2xTY, growing at 37 °C, shaking. Next day, 5 mL of fresh 2xTY medium was inoculated with 50 µL of overnight culture, grown at 37 °C, shaking, until OD_{600} reached 0.6 – 0.7. Then, 1 mL of culture was transferred to sterile 1.5 mL microcentrifuge tube and cells were centrifuged at 2,800 rpm for 2 min. Cells were resuspended with 500 µL of pre-chilled 50 mM $CaCl_2$ and centrifuged at 13,500 rpm for 1 min. Supernatant was discarded, cells were resuspended in 500 µL pre-chilled 50 mM $CaCl_2$ and incubated for 10 min, in ice. Then, 1 µL of plasmid DNA (~ 100 ng/µL) was added. After 10 min incubation, the suspension was incubated 1 min at 42 °C in thermoblock and immediately cool-down in ice. Then, 800 µL of pre-warmed 2xTY was added and incubated at 37 °C for 1 hr, shaking.

Finally, 150 µL of cell suspension was plated on TYE agar plates prepared with 50 mg/mL kanamycin. Plates were incubated overnight at 37 °C.

2.2.3 Protein expression

Protein expression was performed with *E. coli* BL21 (DE3). A single colony of *E. coli* BL21 (DE3) transformants was used to inoculate 5 mL of 2xTY media with 50 mg/mL kanamycin, cultured overnight at 37 °C, shaking. Next day, 250 mL conical flask containing 50 mL of 2xTY with 50 mg/mL kanamycin was inoculated with 250 µL of overnight culture. Cells were grown at 37 °C, 250 rpm until OD_{600} reached 0.4 - 0.6. Protein expression was induced by the addition of 0.1 mM IPTG (final concentration) and expression was conducted at 20 °C, 220 rpm for 24 hrs. Cells were harvested by centrifugation at 6,000 xg, 6 °C for 10 min. Cell pellets were stored at – 20 °C.

2.3 Analysis of protein expression

2.3.1 SDS-PAGE

10% SDS-PAGE gels were cast using Bio-Rad manual casting equipment as follows. For one gel: the resolving gel solution was prepared with 2.05 mL DDI water, 1.65 mL 30% acrylamide/Bis, 1.25 mL 1.5 M Tris-HCl (pH 6.8), 0.5 mL of 10% (*w*/*v*) SDS, 2.5 µL TEMED, and 25 µL of 10% APS; the stacking gel solution was prepared with 2.05 mL DDI water, 1.65 mL 30% acrylamide/Bis, 1.25 mL 0.5 M Tris-HCl (pH 6.8), 0.5 mL of 10% (*w*/*v*) SDS, 5 µL TEMED, and 25 µL of 10% APS. For sample preparation, protein solutions were mixed 1:1 with 2xSDS reducing buffer (65 mM Tris-HCl, pH 6.8; 25% (*v*/*v*) Glycerol, 2% (*w*/*v*) SDS, 0.01% (*w*/*v*) Bromophenol blue) supplemented with β-mercaptoethanol (5% (*v*/*v*)), and then incubated 5 min at 94 °C, in thermoblock. Then, denatured proteins were

centrifuged at 13,500 rpm for 1 min. Samples were loaded onto gel wells and electrophoresis was run in Bio-Rad electrophoresis chamber at constant voltage of 200 V for 50 min. Gels were stained with RAPIDstain solution and distained with DDI water.

2.3.2 Analysis of soluble and insoluble proteins

During protein expression, aliquots of 2 mL were taken after 2, 6, and 24 hrs post induction. Cells were harvested by centrifuge at 2800 rpm for 5 min. The pellets were stored at – 20 °C for later protein expression analysis. Pellets were resuspended in 200 µL of 50 mM phosphate buffer (pH 7.5) supplemented with 0.1 mg/mL of lysozyme, followed by 1 hr incubation in ice. After lysis, 800 µL of 50 mM phosphate buffer were added to the suspension and mixed thoroughly with pipette, then soluble and insoluble fractions were separated by centrifuge at 13,500 rpm for 5 min. Soluble fraction (supernatant) was mixed 1:1 with 2xSDS (see **Section 2.3.1**) and incubated 5 min at 94 °C, in thermoblock. Insoluble fraction (pellet) was resuspended in 200 µL of 1xSDS reducing buffer (see **Section 2.3.1**), supplemented with 5% (*v*/*v*) β-mercaptoethanol, and then incubated 5 min at 94 °C. Samples were then centrifuged at 13,500 rpm for 1 min. Soluble and insoluble fractions were analysed by SDS-PAGE.

2.4 Protein recovery and purification

2.4.1 Protein recovery by osmotic shock

Harvested cells from 50 mL culture were recovered from – 20 °C and pellets were thawed in ice. Cells were resuspended in 40 mL of an ice-cold hypertonic solution (0.1 M Tris- BASE buffer with 0.5 mM EDTA and 20 % (*w*/*v*) sucrose; pH 8.0 adjusted with acetic acid) supplemented with one protease inhibitors tablet, 1 µg/mL DNases and 1 µg/mL RNases, incubated in ice for 15 min and centrifuged at 6,000 xg for 10 min at 6 °C. The supernatant

was collected and kept in ice; the pellet was resuspended in 10 mL of ice-cold water, incubated in ice for 15 min and centrifuged at 6,000 xg for 10 min at 6 °C. The supernatant was combined with the supernatant that was previously kept in ice, and the mixture was centrifuged at 6,000 xg for 40 min at 6 °C. The supernatant was recovered and filtered with a 0.45 µm pore size membrane filter.

2.4.2 Protein purification

rPA and rPA variants were purified using AKTA FPLC system. All buffers and solutions were filtered using a 0.45 µm pore size membrane filter and kept in the fridge at 4 °C prior use. Q-column was equilibrated with 5 cvs of Buffer A (100 mM Tris- BASE, pH 8.0) and the filtered protein extract from osmotic shock was loaded onto the column to capture rPA (or rPA variant). The column was washed with 5 cvs of Buffer A. Proteins were eluted with a 0 – 1 M NaCl gradient with Buffer B (100 mM Tris- BASE, 1 M NaCl; pH 8.0). Fractions containing rPA (or rPA variant) were mixed and diluted 10x with Buffer A. Then, a DEAE-column was equilibrated with 5 cvs of Buffer A and the diluted proteins were loaded. The column was washed with 5 cvs of Buffer A and then the proteins were eluted with a 0 – 1 M NaCl gradient with Buffer B. Fractions containing rPA (or rPA variant) were mixed and loaded onto a HiLoad Superdex 200-pg gel-filtration column equilibrated with 1.5 cvs of Gel-filtration Buffer (10 mM HEPES, 50 mM NaCl, 20% glycerol; pH 7.0). Isocratic elution was performed with 1.5 cvs of Gel-filtration Buffer. Fractions containing rPA (or rPA variant) were mixed, then separated in 1 mL aliquots, frozen with liquid nitrogen and stored at – 80 °C. Collected fractions were analysed with SDS-PAGE.

2.5 Site-directed mutagenesis on rPA

2.5.1 Primers list

Mutagenic partially overlapping primers T576E and S559L were designed using OneClick program (Warburton *et al.*, 2015) and mutagenic primer FFD_{315}-AAA_{315} was designed using NEBaseChanger (New England Biolabs). Primers were synthesised by Eurofins.

Table 2.1 Mutagenic primers to obtain PA(+++)

Mutation	Primers	
T576E	Fwd	AAC ATC TAC GAA GTT CTG GAT AAG ATC AAG CTG AAC GCT AAG ATG A (46)
	Rev	ATC CAG AAC TTC GTA GAT GTT GGT CGC GTT CAG TTC AGC CAG TTG A (46)
S559L	Fwd	CAG CAA ACG CTG CAG AAC ATC AAA AAT CAA CTG GCT GAA CTG AAC G (46)
	Rev	GAT GTT CTG CAG CGT TTG CTG ATC GAA GTT GAA GTC AAA TTC GGT G (46)
FFD_{315}-AAA_{315}	Fwd	GGC GAT TGG CGG TTC CGT GTC A (22)
	Rev	GCC GCT GAT GCA TGC ACT TCG GC (23)

Table 2.1 Mutagenic primers to obtain PA(+++)

2.5.2 Generation of PA(+++): an aggregation-resistant protease-resistant variant

PA(+++) was obtained after three stages of site-directed mutagenesis on rPA. Mutagenic primers are listed in **Table 2.1**. Mutations T576E and L566I were reported by Ganesan *et. al* (2016) to improve rPA solubility. Mutation FFD_{315}-AAA_{315}, was designed to remove a chymotrypsin proteolytic site.

Stage 1: Plasmid pET24a-rPA was used as parental DNA template. Point mutation T576E was introduced by SDM using partially overlapping primers (**Table 2.1**) for whole plasmid amplification. The reaction was prepared as described in **Table 2.2** and performed as described in **Table 2.3**. The PCR product was treated with 1 µL DpnI for 2 hrs, at 37 °C. The product pET24a-rPA-T576E (pET24a-PA(+)) was transformed into *E. coli* DH5α.

Stage 2: plasmid pET24a-PA(+) was isolated from *E. coli* DH5α and used as parental DNA template. Point mutation S559L was introduced by SDM using partially overlapping primers (**Table 2.1**) for whole plasmid amplification. PCR and DpnI treatment were performed as described in Stage 1. The product pET24a-rPA-T576E-S559L (pET24a-PA(++)) was transformed into *E. coli* DH5α.

Stage 3: plasmid pET24a-PA(++) was isolated from *E. coli* DH5α and used as parental DNA template. Mutation FFD_{315}-AAA_{315} was introduced by SDM using mutagenic primers designed by NEBaseChanger (**Table 2.1**). PCR reaction was prepared as described in **Table 2.2** and reaction was conducted as described in **Table 2.3**. The PCR product was treated with 1 µL DpnI for 2 hrs, at 37 °C. The product was purified using QIAquick PCR Purification Kit. Then, phosphorylation and ligation steps were conducted as follows. A reaction tube containing 2x final concentration of T4 DNA Ligase buffer (from a stock of T4 DNA ligase 10x Buffer with 10 mM ATP) was used to incubate ~ 50 ng of purified PCR product with 0.5 µL of T4 Polynucleotide Kinase (from a 10,000 U/mL stock), topped with DDI water to a final volume of 20 µL, and incubated at 37 °C for 30 min. Then, in the same reaction tube, 2 µL of T4 DNA Ligase (from a 400,000 U/mL stock) was added and the reaction was performed at 16 °C, overnight. The result was the construct pET24a-rPA-T576E-S559L-(FFD_{315}-AAA_{315}), hereafter called PA(+++). The PA(+++) construct was transformed into *E. coli* DH5α and kept in a 25% glycerol stock. Mutations were confirmed by sequencing.

Table 2.2 PCR reaction mixture for SDM on rPA		
Component	Stock Concentration	Amount per Reaction
DDI Water	–	36 µL
Q5 buffer	5×	10 µL
dNTPs	10 mM each	1 µL
DNA template	100 ng/µL	0.5 µL
Fwd Primer	20 µM	1 µL
Rev Primer	20 µM	1 µL
Q5 HF DNA Polymerase	2 U/µL	0.5 µL

Table 2.2 PCR reaction mixture for SDM on rPA

Table 2.3 PCR reaction programme for SDM on rPA						
	T576E		S559L		FFD_{315}-AAA_{315}	
Step	T˚(˚C)	Time	T˚(˚C)	Time	T˚(˚C)	Time
1	98	30 s	98	30 s	98	10 s
2	98	8 s	98	8 s	98	10 s
3	55	20 s	55	20 s	67	30 s
4	72	3 min 8 s	72	3 min 8 s	72	3 min 8 s
5	Go to Step 2, repeat 19 times				Go to Step 2, repeat 24 times	
6	72	2 min	72	2 min	72	2 min
7	8	Hold	8	Hold	8	Hold

Table 2.3 PCR reaction programme for SDM on rPA

2.6 Semi-rational approach protein engineering

A semi-rational approach mutant library (**Table 2.6**) was designed in collaboration with Dr. Antonina Andreeva (MRC Laboratory of Molecular Biology, Cambridge) using structure-based analysis of rPA (using the high-resolution crystal structure PDB: 3TEW) aiming to improve hydrophobic core-packing and domain interface contacts. The variant PA(+++) was used as parental template and starting point for thermostability improvement.

2.6.1 Creation of a semi-rational approach mutant library

Plasmid pET24a-PA(+++) was used as parental DNA template for SDM. The mutant library was generated by SDM as follows. For each position, two parallel PCR reactions were set, each reaction tube containing only one of the mutagenic partially overlapping primers (forward primer or reverse primer) **(Table 2.4)** for whole plasmid amplification, as described elsewhere (Edelheit, Hanukoglu and Hanukoglu, 2009) and as follows. After 10 cycles (the reaction conditions are described in **Table 2.5**), the reaction was put on hold and the tubes were mixed into a single reaction tube and then redistributed into two tubes, equal volumes. Then, the reaction was carried out for additional 20 cycles. The reaction products were combined and treated with 1 µL DpnI for 2 hrs at 37 °C. Then, the reaction product was transformed directly into the expression strain *E. coli* BL21 (DE3) using $CaCl_2$ heat shock method (see **Section 2.2.2**). The library was arrayed in a 96-well round-bottom plate, each well inoculated with a single colony. The sampling diversity (*S*) (i.e. number of picked colonies) for each position was considered according to its corresponding primer degeneracy (**Table 2.6**), adjusted to reduce bias using a sampling factor (*N*), as calculated by Equation 2.1 (**Eq. 2.1**):

Eq. 2.1. $$S = \frac{N}{\frac{less\ likely\ codon\ \%}{100}}$$

Where *S* is sampling (number of colonies to pick); *N* is a sampling factor to reduce bias, which is 2 (*N* = 2); *less likely codon %*, is the percentage of the less likely codon for desired substitutions (see **Table 2.6**).

Table 2.4. Mutagenic partially overlapping primers for semi-rational approach protein engineering SDM.

For position 63:	**For position 644:**
63:F 5'–TTCCAATCTRTCATTTGGAGTGGCTTTATCAAAGTGAAAAAGTCGG–3'	**644:F** 5'–GGTTATATTCTGGAAATCGAAGATACCGAAGGCCTGAAAGAAGTCA–3'
63:R 5'–ACTCCAAATGAYAGATTGGAAATACTGGTTTTCACTCGGGATATTT–3'	**644:R** 5'–TTCGATTTCCAGAATATAACCACTCAGGATCTTGCGAATATCTTTG–3'
For position 81:	**For position 674:**
81:F 5'–ACCTTCGCGVTCAGCGCCGACAATCATGTGACCATGTGGGTTGATG–3'	**674:F** 5'–GACGGCAAGGTGTTCATCGATTTCAAGAAGTACAACGACAAACTGC–3'
81:R 5'–GTCGGCGCTGABCGCGAAGGTGTATTCATCCGACTTTTTCACTTTG–3'	**674:R** 5'–ATCGATGAACACCTTGCCGTCCTGGCGCAGACTAGAGATGTTCAGC–3'
For position 89:	**For position 698:**
89:F 5'–CATGTGACCCTGTGGGTTGATGACCAGGAAGTTATCAACAAAGCCT–3'	**698:F** 5'–AAAGTTAACATCTACGCGGTGACCAAGGAAAACACGATTATCAATC–3'
89:R 5'–ATCAACCCACAGGGTCACATGATTGTCGGCGCTCGTCGCGAAGGTG–3'	**698:R** 5'–CACCGCGTAGATGTTAACTTTATAATTCGGGTTGGAGATGTACAGC–3'
For position 91:	**For position 283:**
91:F 5'–ACCATGTGGMTCGATGACCAGGAAGTTATCAACAAAGCCTCCAACT–3'	**283:F** 5'–CAAAATACGRGCTCTGAAACCCGTACGATTAGTAAAAACACCTCTA–3'
91:R 5'–CTGGTCATCGAKCCACATGGTCACATGATTGTCGGCGCTCGTCGCG–3'	**283:R** 5'–GGTTTCAGAGCYCGTATTTTGGGTGCTCTGATCTTCGTTCTTCGAC–3'
For position 96:	**For position 340:**
96:F 5'–GACCAGGAAATSATCAACAAAGCCTCCAACTCAAACAAGATCCGTC–3'	**340:F** 5'–TCTCTGAGTRMCGCCGGTGAACGTACCTGGGCAGAAACGATGGGCC–3'
96:R 5'–TTTGTTGATSATTTCCTGGTCATCAACCCACATGGTCACATGATTG–3'	**340:R** 5'–TTCACCGGCGKYACTCAGAGAGTGGTCGATTGCAACGGTACTAGAA–3'
For position 262:	**For position 566:**
262:F 5'–TATCCGATCMTCCACGTTGATATGGAAAACATCATCCTGTCGAAGA–3'	**566:F** 5'–AAAAATCAARTCGCTGAACTGAACGCGACCAACATCTACGAAGTTC–3'
262:R 5'–ATCAACGTGGAKGATCGGATATGCTGCAACCAGCGGATGACGTGCT–3'	**566:R** 5'–CAGTTCAGCGAYTTGATTTTTGATGTTCTGCAGCGTTTGCTGGTCG–3'
For position 264:	**For position 434:**
264:F 5'–ATCGTGCACATCGATATGGAAAACATCATCCTGTCGAAGAACGAAG–3'	**434:F** 5'–CCGATCACGMTCAACTATAATCAGTTTCTGGAACTGGAAAAAACCA–3'
264:R 5'–TTCCATATCGATGTGCACGATCGGATATGCTGCAACCAGCGGATGA–3'	**434:R** 5'–ATTATAGTTGAKCGTGATCGGGGTTGAGGAGAAGTCATCCTGTGCG–3'
For position 266:	**For position 489:**
266:F 5'–CACGTTGATCTGGAAAACATCATCCTGTCGAAGAACGAAGATCAGA–3'	**489:F** 5'–GAAACCACGGTGCGCATTATCTTTAATGGCAAAGATCTGAACCTGG–3'
266:R 5'–GATGTTTTCCAGATCAACGTGCACGATCGGATATGCTGCAACCAGC–3'	**489:R** 5'–GATAATGCGCACCGTGGTTTCTTGAATCTGCGGCAGAACTTCGCTC–3'
For position 350:	**For position 491:**
350:F 5'–GCAGAAACGCTGGGCCTGAACACCGCAGATACGGCTCGTCTGAACG–3'	**491:F** 5'–ACGGCACGCCTGATCTTTAATGGCAAAGATCTGAACCTGGTGGAAC–3'
350:R 5'–GTTCAGGCCCAGCGTTTCTGCCCAGGTACGTTCACCGGCCAGACTC–3'	**491:R** 5'–ATTAAAGATCAGGCGTGCCGTGGTTTCTTGAATCTGCGGCAGAACT–3'
For position 364:	**For position 526:**
364:F 5'–AACGCTAATCTGCGCTATGTGAATACCGGTACGGCGCCGATCTACA–3'	**526:F** 5'–CTGAAAGAAACGCTGAAGATTGCATTTGGCTTCAATGAACCGAACG–3'
364:R 5'–CACATAGCGCAGATTAGCGTTCAGACGAGCCGTATCTGCGGTGTTC–3'	**526:R** 5'–AATCTTCAGCGTTTCTTTCAGGGTCATGTCCGGTTTCGTGGTTTCC–3'
For position 381:	**For position 589:**
381:F 5'–CTGCCGACCMTGAGTCTGGTCCTGGGCAAAAATCAGACCCTGGCTA–3'	**589:F** 5'–AAGATGAACCTGCTGATCCGTGACAAGCGCTTCCATTACGATCGTA–3'
381:R 5'–GACCAGACTCAKGGTCGGCAGAACGTTGTAGATCGGCGCCGTACCG–3'	**589:R** 5'–ACGGATCAGCAGGTTCATCTTAGCGTTCAGCTTGATCTTATCCAGA–3'
For position 432:	
432:F 5'–TCAACCCCGCTGACGATGAACTATAATCAGTTTCTGGAACTGGAAA–3'	
432:R 5'–GTTCATCGTCAGCGGGGTTGAGGAGAAGTCATCCTGTGCGTTCAGG–3'	

Table 2.4 Mutagenic partially overlapping primers for semi-rational approach protein engineering SDM

According to sampling (*S*) for each position, colonies were picked to inoculate 150 µL of 2xTY media with 50 mg/mL kanamycin in a 96-well microtiter plate. Wells B2, E6 and G11 were inoculated with parental PA(+++) as internal control. Wells A2, H2, A11 and

H11 were inoculated with parental PA(+++) as internal control for mock expression (null-induction) (**Figure 2.0**). Also, a 96-well roud-bottom plate was fully inoculated with only PA(+++), this is named PA(+++)-MasterPlate and it was used for culturing and screening method optimisation (**Section 2.6.2.1** and **2.6.2.2**). Plates were incubated for 12 hrs at 30 °C, 1050 rpm. Then, 50% glycerol was added to the cultures, for a final 12.5 % glycerol concentration. The plates were stored at – 80 °C.

Table 2.5 PCR reactions and programme conditions for whole plasmid SDM with partially overlapping primers.

PCR Reaction tubes				PCR programme		
	Stock [ ]	TUBE F	TUBE R	STEP	T (ºC)	Time
DDI H_2O		18.5 µL	18.5 µL	1	98	30 s
Q5 BUFFER	5x	5 µL	5 µL	2	98	8 s
dNTPs	10 Mm each	0.5 µL	0.5 µL	3	55	20 s
DNA template	100 ng/µL	0.25 µL	0.25 µL	4	72	3 min 8 sec
Primer F	20 µM	0.5 µL	-	5	Go to step 2, repeat 9 times	
Primer R	20 µM	-	0-5 µL	6	PAUSE. Mix tubes F and R. Redistribute.	
Q5 HF Polymerase	2 U/µL	0.25 µL	0.25 µL	7	98	8 s
TOTAL		25 µL	25 µL	8	55	20 s
				9	72	3 min 8 sec
				10	Go to step 7, repeat 19 times	
				11	72	2 min
				12	8	HOLD

Table 2.5 PCR reactions and programme conditions for whole plasmid SDM with partially overlapping primers.

Table 2.6 PA(+++) semi-rational approach mutant library: Sampling diversity					
Position	Primer degeneracy	Degeneracy % (of desired substitutions)	Desired %	Undesired %	Sampling (S)
A63	2	I (50%), V (50%)	2 (100%)	0%	4
T81	3	L (33.33%), I (33.33%), V (33.33%)	3 (100%)	0%	6
M89	1	L (100%)	1 (100%)	0%	2
V91	2	L (50%), I (50%)	2 (100%)	0%	4
V96	2	I (50%), M (50%)	2 (100%)	0%	4
V262	2	L (50%), I (50%)	2 (100%)	0%	4
V264	1	I (100%)	1 (100%)	0%	2
M266	1	L (100%)	1 (100%)	0%	2
M350	1	L (100%)	1 (100%)	0%	2
I364	1	L (100%)	1 (100%)	0%	2
T381	2	L (50%), M (50%)	2 (100%)	0%	4
I432	1	L (100%)	1 (100%)	0%	2
M434	2	L (50%), I (50%)	2 (100%)	0%	4
A489	1	V (100%)	1 (100%)	0%	2
I491	1	L (100%)	1 (100%)	0%	2
A526	1	T (100%)	1 (100%)	0%	2
I589	1	L (100%)	1 (100%)	0%	2
V644	1	L (100%)	1 (100%)	0%	2
T674	1	V (100%)	1 (100%)	0%	2
V698	1	I (100%)	1 (100%)	0%	2
D283	2	G(50%), S(50%)	2 (100%)	0%	4
L340	4	T(25%), N(25%), D(25%)	3 (75%)	1 (25%)	8
Y436	4	A(25%), V(25%), L(25%)	3 (75%)	1 (25%)	8
L566	2	V(50%), I(50%)	2 (100%)	0%	4

Table 2.6 PA(+++) semi-rational approach mutant library: Sampling diversity

Figure 2.1 Semi-rational approach library array. Wells were inoculated with a single colony from transformants containing PA(+++) mutants, except for: Red: wells inoculated with PA(+++) (WT) for internal control (B2, E6, G11). Green: wells inoculated with PA(+++) (WT) for mock expression (null-IPTG) (A2, H2, A11, H11).

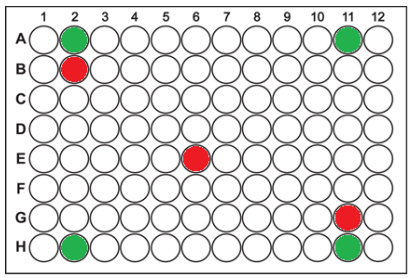


Figure 2.1 Semi-rational approach library array

2.6.2 Semi-rational approach library screening method

2.6.2.1 Development of a semi-high throughput screening method for thermostable PA(+++) variants

A semi-high throughput screening method for thermostability based on detection of heat-induced aggregates with extrinsic fluorescence using Nile red was developed. Nile red is a polarity-sensitive fluorescent dye whose fluorescence increases when bound to hydrophobic regions of proteins (Sackett and Wolff, 1987). This feature can be exploited to detect the solvent-exposed hydrophobic regions of protein aggregates and partially unfolded proteins. Our screening involves the protein expression of the mutant library in 96-well plate format. PA(+++) variants are extracted by osmotic shock (adapted to 96-well plate format from **Section 2.4.1**), then protein aggregation is induced by heat in presence of Nile red. A variant is considered positive (for thermostability) if Nile red fluorescence intensity is weaker related to that of the original PA(+++). Positive variants are isolated and expressed in 50 mL flask culture for validation.

Before screening our library, the aggregation temperature of PA(+++) was established using protein extracts from 50 mL flask expression. Also, method conditions for 96-well

cultivation, expression, and recovery, as well as optimal Nile red concentration, were established in optimisation assays using the PA(+++)-MasterPlate. PA(+++)-MasterPlate is described in **Section 2.6.1**.

2.6.2.1.1 Defining the aggregation temperature of PA(+++)

PA(+++) aggregation temperature was defined using periplasmic extracts derived from osmotic shock of harvested cells obtained from a 50 mL expression culture. Extracts were subjected to a range of temperatures (20, 37, 55, 65 and 70 °C) and aggregates were detected using Nile red, using three independent replicates for each condition, as described below.

PA(+++) was expressed in 50 mL flask culture as described in **Section 2.2.3**. Aliquots of 2 mL were taken after 24 hrs expression and cells were harvested at 2,800 rpm, 6 °C for 5 min using a microcentrifuge. Pellets were stored at – 20 °C for later use. Protein extracts were obtained by osmotic shock as described in **Section 2.4.1**, with the following modifications. Cells were thawed on ice and resuspended in 1.6 mL ice-cold hypertonic solution (see **Section 2.4.1**), incubated for 15 min ice. Then, the suspension was centrifuged at 2,800 rpm, 6 °C for 10 min. Supernatants were transferred into new 2 mL centrifuge tubes and stored in the fridge at 4 °C for later. Pellets were resuspended in 0.4 mL ice-cold DDI water, incubated for 15 min. Then, the suspension was centrifuged at 2,800 rpm, 6 °C for 10 min. Supernatants were combined with the extracts previously stored at 4 °C.

The temperatures tested were room temperature (25 °C), 37, 55, 65 and 70 °C, using three independent replicates for each condition (i.e. each condition uses a protein extract extracted from 3 different 2 mL harvested cells tubes), as follows. 100 µL of PA(+++) protein extract was transferred into a Greiner clear-bottom 96-well microtiter plate containing 100 µL of 100 mM MOPS buffer (100 mM MOPS, 200 mM NaCl; pH 7.0). The plate was mixed

briefly using a thermoblock. Then, 20 µL of Nile red were added to a final concentration of 5 µM from a working solution prepared in 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0) containing 50% (*v*/*v*) ethanol. The plate was briefly mixed with a thermoblock. The plate was covered with a lid and sealed with tape, then protein aggregation was induced after incubation for 10 min with corresponding treatment (25, 37, 55, 65 or 70 °C) using a thermoblock. The plate was immediately cool down at 4 °C. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation and 625 nm emission, read from bottom. Corresponding blank background was subtracted (200 µL of 50 mM MOPS buffer with 5 µM Nile red).

2.6.2.1.2 Protein expression and recovery in 96-well microtiter plates

PA(+++)-MasterPlate was retrieved from – 80 °C and thawed at room temperature. Then, it was replicated into a 96-well round-bottom plate prepared with 150 µL of 2xTY media with 50 mg/mL kanamycin. Cells were cultured at 30° C for 14 hrs, 1050 rpm. After incubation, the replica was replicated into a plate prepared with 100 µL fresh 2xTY media with 50 mg/mL kanamycin. Cells were cultured at 37 °C, 1050 rpm. OD_{600} was monitored with SpectraMax M2e microplate reader until 0.5 – 0.6 average within the plate was reached. Then, protein expression was induced with a solution of 0.3 mM IPTG prepared in 2xTY media supplemented with 50 mg/mL kanamycin, for a final concentration of 0.1 mM IPTG. Protein expression was conducted at 20 °C, 900 rpm for 24 hrs in a thermoblock.

Cells were harvested by centrifuge at 4,000 rpm, 6 °C for 10 min. Harvested cells were resuspended with 50 µL of ice-cold hypertonic solution (100 mM Tris- BASE, 20 % (*w*/*v*) sucrose, 0.5 mM EDTA, pH 8.0 adjusted with acetic acid, containing protease inhibitors) mixed for 5 min using the thermoblock and incubated at 4 °C for 15 min. Then,

200 µL of ice-cold DDI water was added, the plate was mixed for 1 min using the thermoblock and then incubated for 15 min at 4 °C. The plate was centrifuged at 4,000 rpm, 6 °C for 25 min, to obtain the protein extracts.

2.6.2.1.3 Defining Nile red optimal concentration

The protein extracts recovered from PA(+++)-MasterPlate culture were used to define the optimal concentration of Nile red. Aggregates were stained with 5, 10, 15, 20, 25 µM Nile red, using 7 independent replicates for each condition (e.g. for condition 1, A1:G1 wells containing PA(+++)-MasterPlate protein extracts were transferred to A1:G1 wells of the assay plate and stained with 5 µM Nile red), as described below.

50 µL of protein extract from wells A1:G6 were transferred to a black clear-bottom Greiner 96-well microtiter plate containing 100 µL of 100 mM MOPS buffer (100 mM MOPS, 200 mM NaCl; pH 7.0) and 50 µL DDI H_2O. Wells H1:H6 were prepared with only 200 µL of 50 mM MOPS, used as background. The plate was mixed briefly using a thermoblock. Then, 20 µL of Nile red were added to a final concentration of 5, 10, 15, 20 and 25 µM from a working solution prepared in 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0) containing 50% (*v*/*v*) ethanol. The plate was mixed briefly with a thermoblock. The plate was covered with a lid and sealed with tape, then protein aggregation was induced after incubation for 10 min at 65 °C using a thermoblock. The plate was immediately cool down at 4 °C. Then, 50 µL of protein extracts from wells A7:G11 were transferred to the assay plate containing 100 µL of 100 mM MOPS buffer and 50 µL DDI H_2O. Then, the fresh extracts were stained with 5, 10, 15, 20 and 25 µM Nile red, as described above. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation and 625 nm emission, read from bottom. Corresponding blank backgrounds were

subtracted (X µM Nile red in MOPS buffer). A saturation curve was fitted using one-site specific binding equation by GraphPad Prism 7.0.

2.6.2.1.4 Defining the fluorescence/protein concentration relationship

PA(+++)-MasterPlate was cultured, proteins expressed and recovered as described in **Section 2.6.2.1.2**. Different volumes of protein extract were transferred to a black clear-bottom Greiner 96-well microtiter plate containing 100 µL of 100 mM MOPS buffer and topped to 200 µL with DDI water. The plate was mixed briefly using a thermoblock and OD_{280} absorbance was recorded with SpectraMax M2e microplate reader. Then, 20 µL of Nile red were added to a final concentration of 25 µM from a working solution prepared in 50 mM MOPS buffer containing 50% (*v*/*v*) ethanol. The plate was mixed briefly with a thermoblock. The plate was covered with a lid and sealed with tape, then protein aggregation was induced after incubation at 65 °C for 10 min using a thermoblock. The plate was immediately cool down at 4 °C. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation and 625 nm emission, read from bottom. The relationship was fitted to a linear regression and correlation coefficient were calculated with GraphPad Prism 7.0.

2.6.2.2 Semi-rational approach library screening: detection of heat-induced aggregates with Nile red

The 96-well round-bottom plate containing the semi-rational approach library was retrieved from – 80 °C and thawed at room temperature. Culture, protein expression and recovery of protein extracts were performed as described in the optimisation experiments (**Section 2.6.2.1.2**), with the following modification. Protein expression was induced with a solution of 0.3 mM IPTG prepared in 2xTY media supplemented with 50 mg/mL kanamycin, for a final induction concentration of 0.1 mM IPTG, except for wells A2, H2, A11 and H11

whereby only 50 µL of 2xTY with kanamycin (for mock expression, IPTG free) were added. Proteins were expressed at 20 °C, 900 rpm for 24 hrs in a thermoblock.

100 µL of protein extracts were transferred to a black clear-bottom Greiner 96-well microtiter plate containing 100 µL of 100 mM MOPS buffer (100 mM MOPS, 200 mM NaCl; pH 7.0). The plate was mixed briefly using a thermoblock and OD_{280} absorbance was recorded with SpectraMax M2e microplate reader. Then, 20 µL of Nile red were added to a final concentration of 25 µM from a working solution prepared in 50 mM MOPS buffer containing 50% (*v*/*v*) ethanol. The plate was mixed briefly with a thermoblock. The plate was covered with a lid and sealed with tape, then protein aggregation was induced after incubation at 65 °C for 10 min using a thermoblock. The plate was immediately cool down at 4 °C. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation and 625 nm emission, read from bottom. Fluorescence spectra were recorded at 590 – 680 nm emission, 10 nm step, after excitation with 540 nm. Relative stability calculations were performed using instrument readings (without subtracting blanks).

Positive candidates (hits) were cultured in 50 mL for validation. Protein expression and recovery was performed as described in **Section 2.2.3** and **Section 2.4.1**, respectively. The protein extracts were diluted 1:1 with 100 mM MOPS buffer. Aliquots of 100 µL were transferred from the diluted extracts to a black clear-bottom Greiner 96-well microtiter plate containing 100 µL of 100 mM MOPS buffer (100 mM MOPS, 200 mM NaCl; pH 7.0) and the screening assay was performed as described above using 6 experimental replicates. Validated candidates were isolated. Their plasmids were extracted and transformed into *E. coli* DH5α for storage in 25% glycerol stock. Mutations were confirmed by sequencing.

2.7 Rational approach protein engineering

2.7.1 Mutational stability predictions using FoldX

The empirical force-field algorithm FoldX was used to predict the mutational stability of the variants designed in **Section 2.6.2**. FoldX is fast enough for a desktop computer and can be used as a plugin in the graphical user interface for visualisation of molecules, YASARA. High-resolution crystal structure 3TEW (PDB: 3TEW) and default parameters were used as input. First, structure energy was minimised with the FoldX RepairPDB function. Then, the BuildModel function was used for energy calculations to predict the free energy change upon mutation (ΔΔG). Variants that yielded ΔΔG < – 0.5 kcal/mol were selected to generate our rational approach library.

2.7.2 SDM for rationally designed variants

Generation of a small rational approach library was performed using SDM on PA(+++) as described in **Section 2.6.1**, using the following primers:

Table 2.7 Rational approach library primers

Mutation	Primers	
V264I	Fwd	5'–ATCGTGCACATCGATATGGAAAACATCATCCTGTCGAAGAACGAAG–3'
	Rev	5'–TTCCATATCGATGTGCACGATCGGATATGCTGCAACCAGCGGATGA–3'
A489V	Fwd	5'–GAAACCACGGTGCGCATTATCTTTAATGGCAAAGATCTGAACCTGG–3'
	Rev	5'–GATAATGCGCACCGTGGTTTCTTGAATCTGCGGCAGAACTTCGCTC–3'
T674V	Fwd	5'–GACGGCAAGGTGTTCATCGATTTCAAGAAGTACAACGACAAACTGC–3'
	Rev	5'–ATCGATGAACACCTTGCCGTCCTGGCGCAGACTAGAGATGTTCAGC–3'

Table 2.7 Rational approach library primers

2.7.3 Screening of rationally designed variants

Plasmids harbouring rPA, PA(+++) and the rational approach variants (**Table 2.7**) were isolated from *E. coli* DH5α glycerol stocks and then transformed into *E. coli* BL21 (DE3) for

protein expression. Expression and recovery of protein variants was carried out as described in **Section 2.2.3** and **Section 2.4.1**, respectively. Protein extracts concentration was adjusted with 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0) to an OD_{280} of 0.9, measured with cuvette using a BioPhotometer Plus UV/Vis photometer. Aliquots of 500 µL were then transferred into 1.5 mL centrifuge tubes, using three experimental replicates for each condition. Then, protein extracts were heated at 40, 45, 50, 55 and 60 °C for 60 min. The tubes were then centrifuged at 13,500 rpm for 1 min and then aggregates were resuspended and mixed thoroughly with pipette. Then, 150 µL of native protein and protein aggregates were transferred into a black clear-bottom Greiner 96-well microtiter plate. Aggregates were then stained with 1 µL of Nile red for a final concentration of 5 µM, from a working solution of 750 µM Nile red prepared in pure ethanol, and briefly mixed with a thermoblock. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation, 625 nm emission, read from bottom.

2.8 Characterisation

After screening of semi-rational and rational approach libraries, the following variants were expressed in 50 mL and purified for characterisation: PA(+++)-T81I, PA(+++)-T674V, PA(+++)-V264I, PA(+++)-A489V; as well as variant PA(+++) and the original rPA.

For characterisation assays, pure protein concentration was determined by OD_{280} absorbance using the Beer-Lambert law (**Eq. 2.2**).

Eq. 2.2 $$A = \varepsilon c l$$

Where A is absorbance at 280 nm, c is molar concentration, ε is molar absorption coefficient and l is the cell pathlength. Protein concentration was calculated using the mass

extinction coefficient ($\varepsilon_{0.1\%}$) of rPA calculated by ProtParam tool (by ExPASy), using **Equation 2.3** (**Eq. 2.3**), as derived from **Eq. 2.2**:

Eq. 2.3 $$C = \frac{A}{l\ \varepsilon_{0.1\%}}$$

Where C is concentration in mg/mL, A is absorbance at 280 nm, l is 1 cm, $\varepsilon_{0.1\%}$ is 0.935 $(mg/mL)^{-1}(cm)^{-1}$. Absorbance was measured using a BioPhotometer Plus UV/Vis photometer (Eppendorf, Stevenage, UK) using 1 cm pathlength UV cuvette (Eppendorf UVette).

2.8.1 Thermostability characterisation assay: detecting thermally induced aggregates with Nile red.

2.8.1.1 Optimisation: optimal Nile red concentration and aggregation temperature

Thermal aggregation assay was optimised using PA(+++). All experiments were performed using three independent replicates.

Pure PA(+++) was retrieved from – 80 °C and thawed in ice. Then, PA(+++) concentration was adjusted to 120 µg/mL with 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0). Aliquots of 350 µL were transferred into 1.5 mL centrifuge tubes and then incubated at 94 °C for 5 min using a thermoblock and immediately plunged in ice. Tubes were then centrifuged at 13,500 rpm for 1 min. Protein aggregates were thoroughly mixed with pipette. Then, 100 µL of protein aggregates, as well as 100 µL of native PA(+++) (i.e. without heat treatment) were transferred into a black clear-bottom Greiner 96-well microtiter plate. A stock solution of 110 mM Nile red was prepared in pure ethanol. The following Nile red working solutions were prepared from stock: 110 µM Nile red in pure ethanol; 110 µM

Nile red in 50 mM MOPS buffer containing 50% ethanol. Then, aggregates and native PA(+++) were stained with Nile red working solutions to final concentrations: 10 µM Nile red, 10% ethanol; 10 µM Nile red, 5% ethanol; 10 µM Nile red, 1% ethanol; 1 µM Nile red, 1% ethanol. The final volume was 110 µL, adjusted with 50 mM MOPS buffer. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation, 625 nm emission, read from bottom. Emission spectra were recorded after 540 nm excitation at 580 – 690 nm, step 10 nm, using a 570 nm filter. All measurements were corrected for background fluorescence (i.e. 50 mM MOPS buffer with corresponding Nile red concentration.

To investigate the effect of ethanol on aggregates structure, PA(+++) aggregates were stained with Thioflavin T. PA(+++) aggregates were prepared as described above and 100 µL were transferred into a black clear-bottom Greiner 96-well microtiter plate. Then, 10 µL of 50 mM MOPS buffer were added to the free-ethanol group and 10 µL of ethanol were added to the 10% ethanol group. Thioflavin T solution was prepared in 50 mM potassium buffer (50 mM potassium buffer, 150 nm NaCl; pH 7.0). PA(+++) aggregates were stained with 5 µM Thioflavin T, final concentration. Fluorescence was measured with SpectraMax M2e microplate reader with 410 nm excitation and 480 nm emission, read from bottom. Unpaired Student's *t* test was performed using GraphPad Prism 7.0.

To determine the aggregation temperature, a 120 µg/mL PA(+++) solution was prepared as described above and aliquots of 350 µL were transferred into 1.5 mL centrifuge tubes. Then, the protein solutions were incubated at 55, 60 °C for 60 min, and 94 °C, for 5 min, using a thermoblock, and immediately cool-down in ice. One tube was left in ice for native protein staining (i.e. PA(+++) without heat treatment). Tubes were centrifuged at 13,500 rpm for 1 min, and then aggregates were thoroughly mixed with pipette. Then, 100

µL of aggregates from each treatment group, as well as 100 µL of native PA(+++), were transferred into a black clear-bottom Greiner 96-well microtiter plate, containing 10 µL of 50 mM MOPS buffer. Native PA(+++) and aggregates were stained with 1 µM Nile red, final concentration, using a 110 µM Nile red working solution prepared in pure ethanol. Fluorescence was measured with SpectraMax M2e microplate reader with 540 nm excitation, 625 nm emission, read from bottom. Emission spectra were recorded at 580 – 690 nm, using a 570 nm filter, step 5 nm, after 540 nm excitation. All measurements were corrected for background fluorescence (i.e. 50 mM MOPS buffer with 1 µM Nile red).

2.8.1.2 Thermostability characterisation of rPA variants

Thermal aggregation assay was performed on variants PA(+++)-T81I, PA(+++)-T674V, PA(+++)-V264I, PA(+++)-A489V; as well as on variant PA(+++) and the original rPA. The assay was performed using three independent replicates.

Pure protein solutions were retrieved from – 80 °C and thawed in ice. Then, protein concentration was adjusted to 60 µg/mL with 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0). The assay was performed using the optimal conditions defined in **Section 2.8.1.1** and as follows.

Aliquots of 350 µL were transferred into 1.5 mL centrifuge tubes. Protein solutions were incubated at 55 °C for 15, 30 and 60 min, immediately cool-down in ice. Tubes were centrifuged at 13,500 rpm for 1 min. Aggregates were thoroughly mixed with pipette. Then, 100 µL of denatured protein, as well as 100 µL of native protein, were transferred into a black clear-bottom Greiner 96-well microtiter plate, containing 10 µL of 50 mM MOPS buffer. Native and denatured proteins were stained with 1 µM Nile red, final concentration, using a 110 µM Nile red working solution prepared in pure ethanol. Fluorescence was

measured with SpectraMax M2e microplate reader with 540 nm excitation, 625 nm emission, read from bottom. Emission spectra were recorded at 580 – 690 nm, step 5 nm, using a 570 nm filter, after 540 nm excitation. All measurements were corrected for background fluorescence. Barycentric mean was calculated for each spectrum using **Eq. 2.4**:

$$\textbf{Eq. 2.4} \qquad \lambda_{bcm} = \frac{\sum_{\lambda=m}^{n} F_{\lambda}\lambda}{\sum_{\lambda=m}^{n} F_{\lambda}}$$

Where λ_{bcm} is barycentric mean; λ is wavelength; F_{λ} is fluorescence at wavelength λ; *m* is 580 nm; *n* is 690 nm.

2.8.2 Urea denaturation and tryptophan fluorescence assay

Urea denaturation assay was performed on variants PA(+++)-T81I, PA(+++)-T674V, PA(+++)-V264I, PA(+++)-A489V; as well as on variant PA(+++) and the original rPA. Experiments were performed in two independent replicates.

A tube of pure protein solution was retrieved from – 80 °C and thawed in ice. A solution of 8 M urea was prepared in 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0). A range of urea concentrations was achieved by adding protein aliquots to 1.5 mL-tubes containing 8 M urea-MOPS buffer, yielding 150 µL of protein-urea solutions with final urea concentrations of 1, 2, 4 and 6 M. Tubes containing the protein-urea solutions were incubated overnight (~ 18 hrs) at room temperature. Next day, denatured protein solutions were diluted and brought to equal concentration using the corresponding urea buffer. A fresh protein sample was as well prepared in 50 mM MOPS buffer (urea-free). Native and denatured protein solutions were equilibrated for 1 hr at room temperature. Then, 150 µL were transferred into a black opaque-bottom 96-well plate and intrinsic fluorescence emission spectra from 300 – 410 nm, 5 nm step, were recorded after 280 nm excitation using

a SpectraMax M2e microplate reader. All measurements were corrected for background fluorescence. Barycentric mean was calculated using **Eq. 2.4**, with *m* = 305 nm and *n* = 410 nm.

2.8.3 Limited proteolysis and pre-pore formation

Oligomerisation was triggered by limited proteolysis with trypsin on variants PA(+++)-T81I, PA(+++)-T674V, PA(+++)-V264I, PA(+++)-A489V; as well as on variant PA(+++) and the original rPA.

A 1x solution of trypsin was prepared form a 10x stock trypsin solution from porcine pancreas (10x solution from Sigma-Aldrich is 25 mg/mL) using HEPES buffer (10 mM HEPES, 50 mM NaCl; pH 7.0), also a solution of 1 mg/mL soybean trypsin inhibitor was prepared in HEPES buffer. Pure protein solutions were retrieved form – 80 °C and thawed in ice. Protein concentration was adjusted to 150 µg/mL with HEPES buffer. Then, 1x trypsin solution was added to a final trypsin concentration of 1.25 µg/mL. Proteolysis was carried out at room temperature for 15 min and the reaction was stopped by adding soybean trypsin inhibitor solution from the 1 mg/mL stock to a 10 µg/mL final concentration. Digested proteins were analysed by SDS-PAGE.

2.8.3.1 Native-PAGE

Proteins were cleaved with trypsin as described in **Section 2.8.3** and pre-pore were assembled onto a Q-column as follows. Ion-exchange chromatography with Q-column was performed using FPLC AKTA. Trypsin digested protein was diluted 20x with Buffer A (100 mM Tris-BASE; pH 8.0), and loaded onto a Q-column pre-equilibrated with Buffer A. Elution was

performed with 0 – 1 M NaCl gradient with Buffer B (100 mM Tris-BASE, 1 M NaCl; pH 8.0). Fractions containing PA63, analysed by SDS-PAGE, were pooled and stored at 4 °C.

Oligomers were detected by Native-PAGE using a pre-casted NativePAGE™ Bis-Tris 4 – 16% gel (Invitrogen). Purified pre-pore, as well as intact native proteins (used as control), were stored overnight at 4 °C. Next day, samples were mixed with 1x sample buffer, as described by supplier, and kept on ice. NativeMark™ Unstained protein standard was used as reference. Samples and standard were loaded onto gel and electrophoresis was run with NativePAGE™ anode and light-blue cathode buffer for 120 min at constant voltage of 150 V, at room temperature. The gel was stained with RAPIDstain solution and destained with DDI water.

2.9 Proteomic analysis

Fractions containing rPA from gel filtration were pooled and the proteolytic fragments were resolved by SDS-PAGE. Bands were revealed using silver stain and fragments 40 kDa and 50 kDa were excised form the gel manually. The bands were de-stained and digested in-gel with trypsin for mass spectrometry analysis coupled to liquid chromatography (LC-MS/MS). Mass spectrometry data was analysed using Mascot software search engine to identify sequences from the *E. coli* B/BL21-DE3 reference database (UniProt Proteome ID: UP000002032), including the rPA sequence as an additional entry to the reference proteome. Both fragments matched rPA, with a sequence coverage of ~ 25%.

Chapter 3. Bioprocess design for expression, recovery and purification of recombinant Protective Antigen

3.0 Abstract

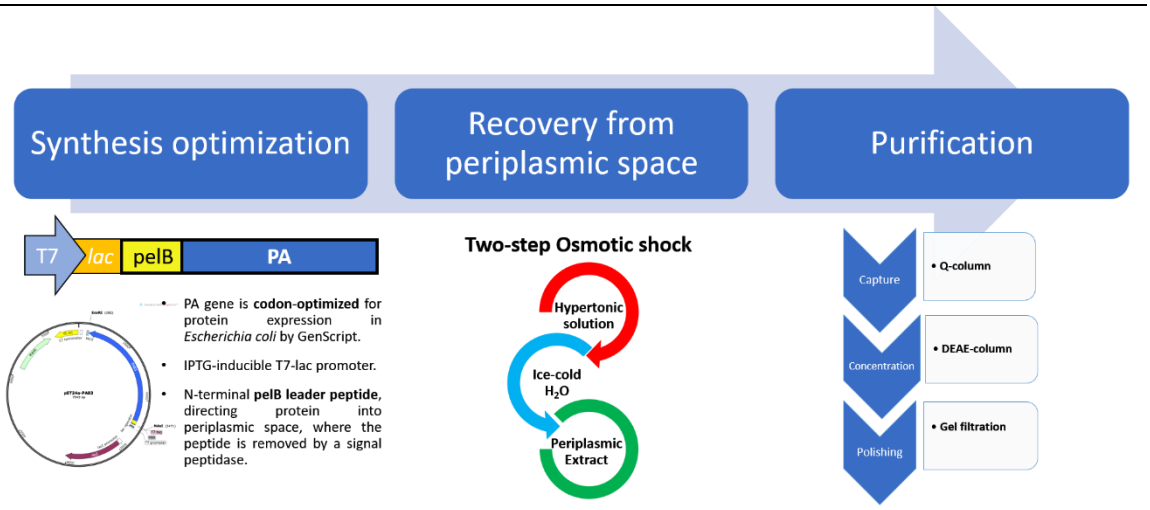


Figure 3.0 Graphical abstract. Bioprocess design for soluble expression, recovery and purification of recombinant Protective Antigen.

The current Anthrax vaccines are manufactured using non-pathogenic *Bacillus anthracis* strains, comprising a sterile-filtrate formulation of alum precipitated (or adsorbed) antigens. Although effective, this type of vaccines yields a complex mixture that hinders full characterisation and presents lot-to-lot variation. The immunogenic component that confer protection against disease in these vaccines is Protective Antigen (PA), a component of the Anthrax toxin, thus recombinant PA (rPA) is a candidate for the development of a well-defined next-generation Anthrax vaccine. Indeed, the proposal of this thesis is the synthesis of a thermostable version of rPA, designed by protein engineering approach, as a candidate to develop a well-defined thermostable next-generation Anthrax vaccine. In this chapter, the

designing of a bioprocess for soluble overexpression, recovery and purification of rPA is described. The soluble overexpression of rPA in *E. coli* BL21 (DE3) could be achieved using a codon-optimised construct and periplasmic expression. rPA was recovered in soluble form using osmotic shock. Pure rPA was obtained after two subsequent anion-exchange chromatography steps and a final polishing step with gel filtration. We found that rPA was being cleaved by proteases in a chymotrypsin-sensitive site, thus a protease- aggregation-resistant mutant, PA(+++), was generated by site-directed mutagenesis. PA(+++) was successfully expressed and purified using the established bioprocess for rPA, demonstrating the reproducibility and robustness of the bioprocess. PA(+++) will be used in the next chapters as the initial template for the protein engineering approach campaign.

3.1 Introduction

Anthrax is a deadly infectious disease caused by the spore-forming bacterium *Bacillus anthracis* (*B. anthracis*). It is primarily a disease of livestock, yet humans can get infected when encountering spores in products derived from infected animals (e.g. hides, wools, hair) or by direct contact with contaminated animal carcases (Saile and Koehler, 2006; Williamson and Dyson, 2015). Anthrax is a prevalent disease in tropical developing countries with inadequate vaccination of livestock. Cases are rare in developed countries but people at risk of infection are veterinarians, laboratory workers handling *B. anthracis*, workers handling imported animal products and members of the military (Head, Rubinstein and Meyers, 2016). The severity of the disease depends on type of exposure: cutaneous Anthrax lethality is 20%, 25 – 75% for gastrointestinal, and higher than 80% for inhalation Anthrax. Opportune diagnostic and antibiotic treatment significantly reduces lethality (Williamson and Dyson, 2015). Anthrax is recognized as a latent bioterrorist threat due to the potential of *B. anthracis*

spores to be released as a weaponised agent, thus it is a major concern for Public Health preparedness, to respond swiftly to a potential outbreak (Goel, 2015).

B. anthracis virulence factors include an anti-phagocytic capsule and two binary toxins. The toxins consist of three proteins: Protective Antigen (PA), Lethal Factor (LF) and Oedema Factor (EF), that combine to form Lethal Toxin (PA+LF) or Oedema Toxin (PA+EF) (Liu *et al.*, 2015). PA plays a central role during disease by binding cell receptors and translocating the toxin into the cells. It has four domains to perform its distinct functions: domain 4 binds cell receptors; domain 1 is cleaved by membrane proteases to activate binding of effector proteins LF/EF; domain 3 is involved in the formation of a ring-shaped heptamer; domain 2 forms a channel with translocase activity (Petosa *et al.*, 1997). It is well-known that immunisation with PA elicits production of neutralising antibodies that confer protection against Anthrax disease (Ivins *et al.*, 1998; Weiss *et al.*, 2006). The manufacture of the currently licenced Anthrax vaccines, Anthrax Vaccine Adsorbed (AVA) and Anthrax Vaccine Precipitated (AVP), involve the production of high titres of PA using non-capsulated toxigenic *B. anthracis* strains, yielding a cell-free filter-sterilised formulation of alum hydroxide adsorbed (AVA) or alum precipitated (AVP) antigens (Kaur, Singh and Bhatnagar, 2013). Although effective, vaccines of this form comprise a complex mixture that hinders full characterisation, present lot-to-lot variation and traces of the other toxin components which are believed to contribute to vaccine reactogenicity (Turnbull, 2000). PA is also extensively studied to understand Anthrax pathogenesis due to the central role it plays during virulence. Consequently, there is an interest to safely produce high-quality pure PA for comprehensive characterisation studies and development of a well-defined next-generation vaccine.

The proposal of this thesis is the production of a thermostable version of recombinant PA (rPA) as a candidate to develop a well-defined thermostable next-generation Anthrax vaccine. This chapter will describe the designing of a bioprocess for overexpression of soluble rPA in *E. coli*, its recovery and purification. *E. coli* is extensively used for heterologous protein expression due to its well-known genetics and capacity to rapidly grow high density cultures on inexpensive media. The continuous development of mutant host strains and novel cloning vectors makes it the most versatile option for recombinant expression of prokaryotic proteins and in some cases for eukaryotic proteins without post-translational modifications (Zarschler *et al.*, 2013; Wurm *et al.*, 2016; Zieliński *et al.*, 2019). Bioprocess designing for production of a recombinant protein is often a trial-and-error endeavour whereby decisions are made according to the ultimate use of the protein or its characterisation requirements. Our goal is the development of thermostable rPA using protein engineering approach (**Chapter 4**) as candidate for a next-generation Anthrax vaccine. Thus, our bioprocess must have the robustness to purify different variants of rPA and yield pure natively folded rPA at laboratory scale for characterisation (**Chapter 5**) and at the same time its designing should aim to be suitable for manufacturing scalability. In order to achieve these goals, a literature review of recombinant expression and purification of rPA was compiled and critically analysed to devise a strategy.

3.2 Aims

- To achieve overexpression of soluble rPA in *E. coli* using recombinant DNA technology tools.
- To design a robust down-stream process for recovery and purification of rPA and rPA variants.

3.3 Results and Discussion

Table 3.1 enlists different strategies reported for rPA expression in *E. coli*, as well as their corresponding methods for recovery and purification.

3.3.1 Upstream processing: Overexpression of soluble rPA in *E. coli* BL21 (DE3) was achieved by periplasmic expression and using a codon-optimised construct.

rPA is an aggregation prone protein (Radha *et al.*, 1996) and its expression in *E. coli* often results in formation of inclusion bodies (See **Table 3.1**). The literature review summary in **Table 3.1** shows that rPA can be expressed in soluble form when protein expression is directed towards the periplasmic space or by using solubility tags. Also codon-optimised constructs with periplasmic expression at < 30 °C showed improved solubility.

Overexpression of soluble rPA was achieved using a synthetic construct codon-optimised for *E. coli* expression by GenScript (see the synthetic sequence in Appendix I) using the sequence coding for mature PA (i.e. without signal peptide) as found in the *pagA* gene (GenBank: AF268967). The original PA signal peptide was replaced by the 22-amino-acid signal peptide from pectate lyase B (pelB), as expressed by *Erwinia carotovora* EC (Lei *et al.*, 1987). The rPA synthetic construct was cloned into plasmid pET24a (**Figure 3.1**). The pET system is a powerful tool for expression of high yields of protein under the control of the IPTG inducible T7/lac promoter (William Studier *et al.*, 1990). The bacterial host must provide the T7 RNA polymerase, thus pET is compatible with *E. coli* strains lysogenized with the gene DE3 coding for T7 RNA polymerase (bacteriophage T7 gene 1) under control of lac promoter (Dubendorf and Studier, 1991).

Table 3.1 PA expression, recovery and purification systems found in the literature.

Ref	Genetic construct	Protein expression and solubility	Recovery and Purification
[1]	In-house engineered plasmid pMS1: - **IPTG-inducible** T7/lac promoter. Native *pagA* sequence **OmpA signal peptide**.	*E. coli* BL21(DE3) Induction: 1 mM IPTG at 37 ºC, 2 hrs. PA was found **soluble**.	- Periplasmic extract was recovered by **osmotic shock**. Proteins were **precipitated with ammonium sulphate**. - PA was purified using **three chromatography steps**: Dialyzed sample was loaded on a **hydroxyapatite column**; fractions containing PA were dialyzed and passed through a **DEAE-Sepharose CL-4B column**. Pooled fractions containing PA were precipitated with **ammonium sulphate**, then resuspended and dialyzed. Proteins were loaded onto a **Sephadex G-100 column**. Purified PA protein was dialyzed against formulation buffer.
[2]	Plasmid pQE30: - **IPTG-inducible** T5/lac promoter. - **6xHis Tag**. Native *pagA* sequence.	*E. coli* SG13009 (pREP4) Induction: 1 mM IPTG at 37 °C, 5 hrs. PA was found to form **inclusion bodies**.	- Inclusion bodies were **solubilized** in 6 M Guanidine-HCl. - Protein was purified under denaturing conditions by **affinity chromatography** using nickel nitrilotriacetic acid **(Ni-NTA)** column. Protein was refolded whilst bound to the column by gradual removal of the denaturant agent. PA was further purified using **anion-exchange chromatography** with **Mono-Q column** as final step.
[3]	Same genetic construct from **[1]** - **IPTG-inducible** T7/lac promoter. Native *pagA* sequence **OmpA signal peptide**.	*E. coli BL21 (DE3)* Induction: *0.5 mM IPTG at 30 °C, 3 hrs.* PA was found **soluble**.	- Periplasmic extract was recovered by **osmotic shock.** - Periplasmic extract was loaded onto a **Resource Q column** and PA was captured by **anion-exchange chromatography**. Pooled fractions containing PA were saturated to 70% with **ammonium sulphate**. Finally, PA was purified by **hydrophobic interaction chromatography** using a **phenyl sepharose column**.
[4]	In-house engineered plasmid based on ColE1 vector: - **IPTG/*lacI*^{q}** promoter. *pagA* sequence **codon-optimized**. **OmpA signal sequence**.	*E. coli* K-12 Induction: 1 mM IPTG at 37 °C, 3 hrs. PA was found **soluble** and recovered from periplasm.	- Periplasmic extract was recovered by **osmotic shock**. - Clarified supernatant was loaded to a **Q-Sepharose-HP column**, pooled fractions containing PA were diluted and loaded onto a **Bio-Gel hydroxyapatite HT column.** Finally, **ion-exchange chromatography** in **Q-Sepharose-HP column** as final separation step. Resulting soluble PA fractions were dialyzed versus final formulation buffer.
[5]	Expression plasmid pET-30a: - **IPTG-inducible** T7/lac promoter. Native *pagA* sequence.	*E. coli Rosetta™ 2 (DE3)* Induction: *0.5 mM IPTG at 30 °C, 4 hrs.* *PA was found to form **inclusion bodies**.*	- Inclusion bodies were recovered by **sonication**, then solubilized with high concentrations of urea and refolded using dialysis against urea-free buffer. Dialyzed solution was saturated to 60% with **Ammonium sulphate**. - The suspension was clarified and then applied to a **Phenyl-Sepharose High Performance XK16/20** for **hydrophobic-interaction chromatography.** Pooled fractions containing PA were incubated with **ammonium sulphate** to 60% saturation. PA was reconstituted in 2XPBS and dialyzed against 2xPBS.
[6]	Expression plasmid pGEX-6p-1: - **IPTG-inducible** tac promoter - **GST-Tag** Native *pagA* sequence.	*E. coli* BL21-CodonPlus (DE3)-RIL0 Induction: 0.2 mM IPTG at 15°C, 5 hrs. PA was obtained in **soluble** form.	- Harvested cells were resuspended in lysis buffer and disrupted using **French press**. - Cell lysate was centrifuged and supernatant was incubated with GST-Binding Resin for purification by **Glutathione S-transferase(GST)-agarose affinity chromatography.**
[7]	Expression plasmid pGEX-6p-1: - **IPTG-inducible** tac promoter - **GST-Tag** Native *pagA* sequence.	*E. coli* BL21 (DE3) Induction: 1 mM IPTG at 37 °C, 6 hrs. PA was expressed in **soluble** form.	- Harvested cells were disrupted by **sonication**. - Cell lysate was clarified by centrifugation and the supernatant was incubated with **GST-Binding Resin** for purification by **Glutathione S-transferase (GST)-agarose affinity chromatography.**
[8]	Expression plasmid pET32c: - **IPTG-inducible** T7/lac promoter. - **His-Tag**; **Trx-Tag** Native *pagA* sequence.	*E. coli* (DE3-pLysS) Induction: 1 mM IPTG at 22 °C, 5 hrs PA was found in **soluble** form.	- Harvested cells were disrupted **sonication.** - Purification was performed by **immobilized metal ion affinity chromatography** (IMAC) using nickel trinitriloacetic acid (Ni-NTA) resin followed by **anion exchange chromatography** using a DEAE-column.
[9]	Expression plasmid pET22b: - **IPTG-inducible** T7 promoter. - pelB **signal peptide**. Native *pagA* sequence	*E. coli* BL21 (DE3) Induction: 1 mM IPTG at 25 °C, 18 hrs PA was found **soluble** and recovered from periplasm.	- Periplasmic extract was recovered by **osmotic shock** and using **lysozyme**. - Periplasmic extract was subjected to **anion-exchange chromatography** using **Q-column**. Fractions containing PA were then loaded into a **hydroxyapatite column** (**CHT Ceramic Hydroxyapetite**).

[1] (Sharma *et al.*, 1996), **[2]** (Gupta, Waheed and Bhatnagar, 1999), **[3]** (Ahuja, Kumar and Bhatnagar, 2001), **[4]** (Laird *et al.*, 2004), **[5]** (Lu *et al.*, 2009), **[6]** (Wu *et al.*, 2010), **[7]** (Uchida *et al.*, 2012), **[8]** (Suryanarayana *et al.*, 2016), **[9]** (Ganesan *et al.*, 2016)

Table 3.1 PA expression, recovery and purification systems found in the literature

Periplasmic expression was preferred because the use of fusion-tags can modify the conformational stability of the protein, which is essential to ensure vaccine effectiveness (Scheiblhofer *et al.*, 2017), and its removal would involve an extra purification step. The use of pelB for periplasmic expression of recombinant proteins in *E. coli* have yielded natively folded proteins, with no signs of alteration of their conformational stability (Lei *et al.*, 1987; Singh *et al.*, 2013; Zhang *et al.*, 2018), and it has been used for expression of rPA in another study (Ganesan *et al.*, 2016).

SDS-PAGE analysis of soluble and insoluble protein fractions after 2, 6 and 24 hrs of expression (**Figure 3.2**) showed that rPA was successfully overexpressed in *E. coli* BL21 (DE3) in soluble form after induction with 0.1 mM IPTG (final concentration) at 20 °C, grown for 24 hrs in rich 2xTY medium.

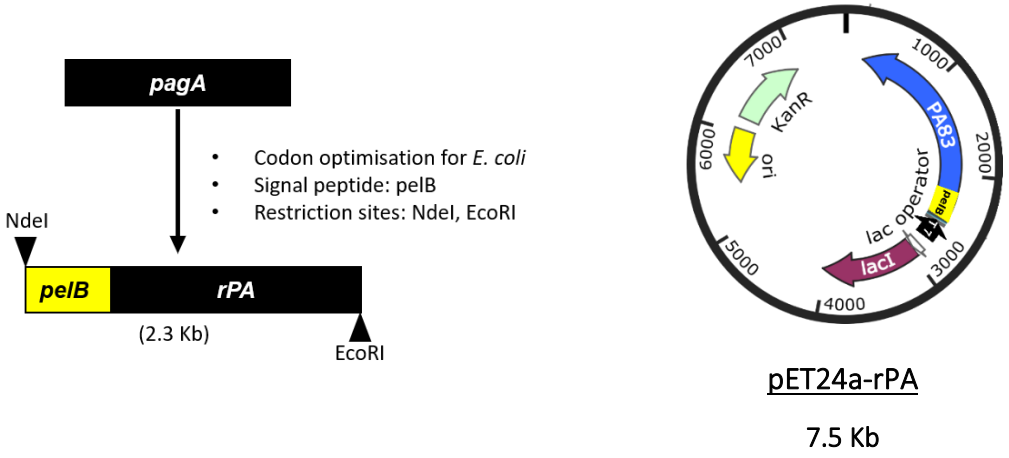


Figure 3.1. rPA synthetic construct. See the full synthetic sequence in Appendix I. The PA gene *pagA* was codon-optimized for expression in *E. coli*. The original N-terminal signal peptide from PA was replaced by the signal peptide pelB. Restriction sites were added for cloning into pET24a(+), to obtain construct pET24a-rPA.

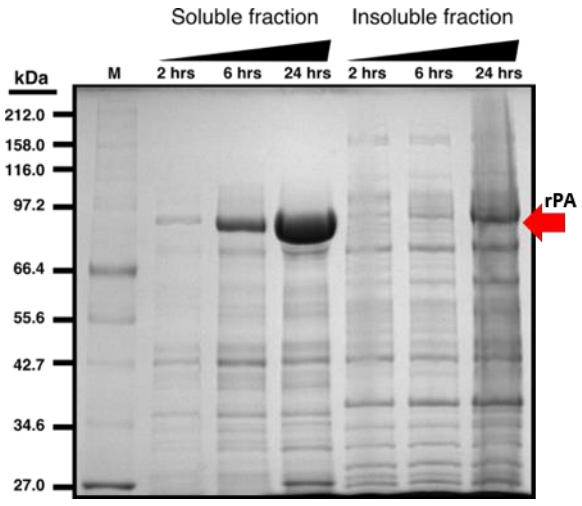


Figure 3.2 Soluble protein analysis. SDS-PAGE of soluble and insoluble fractions after lysozyme lysis of *E. coli* BL21 (DE3) expressing rPA (Red arrow, 83 kDa). Protein expression was induced with 0.1 mM IPTG (final concentration), cultured at 20 °C, 220 rpm. Samples were taken 2 hrs, 6 hrs and 24 hrs post induction.

Figure 3.2 Soluble protein analysis

The role of codon usage in heterologous protein expression is not fully understood, however rare codons have been found to significantly affect gene transcriptional and translational efficiency, leading to translational errors and low protein yields (Wu *et al.*, 2004; Boël *et al.*, 2016). Thus, codon usage optimisation has relevant biotechnological applications and many computational algorithms have been developed to design DNA sequences with optimal codon bias for *E. coli* expression (Tuan-Anh *et al.*, 2017). Codon optimisation has been applied for the development of the rPA-based vaccine candidate SparVax-L (Altimmune, Inc.), currently undergoing early clinical trials (Altimmune, 2018). SparVax-L is a lyophilized vaccine candidate comprising pure rPA produced in *E. coli* cytoplasm, purified from inclusion bodies (Williamson *et al.*, 2005). Our results and those reported in the literature suggest that soluble overexpression of rPA in *E. coli* can only be achieved by periplasmic expression and that codon bias has an influence only in yields related to efficient transcription and translation. These results can be expected due to the high propensity of PA to form aggregates (Radha *et al.*, 1996; Ganesan *et al.*, 2016). Indeed, SDS-PAGE analysis

of soluble and insoluble protein fractions show some rPA in the insoluble fraction 6 hrs post-induction (**Figure 3.2**), although it is mostly present in soluble form.

3.3.2 Downstream processing: soluble rPA can be recovered from periplasmic space using osmotic shock.

Periplasmic expression is advantageous for recovery of soluble rPA by osmotic shock, resulting in a protein extract with less impurities compared to a full cell-lysate method. As shown in **Figure 3.3b**, rPA was successfully recovered from the periplasmic space after a two-step osmotic shock. The osmotic shock method (**Figure 3.3a**) consists in subsequent cell washes with hypertonic solution and ice-cold water, in presence of EDTA, whereby sudden osmotic transition prompts the release of periplasmic proteins through a permeabilised cell wall, creating the effect of a molecular sieve of ~ 100 kDa cut-off (Vázquez-Laslop *et al.*, 2001).

The reason behind successful periplasmic expression may be explained by the way PA is naturally secreted. PA secretion in *B. anthracis* is mediated by a N-terminal signal peptide and under the Sec-dependent pathway, whereby PA is directed towards the membrane in an unfolded conformation, then the signal peptide is cleaved during the translocation process and finally PA folds extracellularly with the aid of extracellular chaperones (Williams *et al.*, 2003; Sarvas *et al.*, 2004). In this work, the original PA signal peptide was replaced by the pelB sequence. PelB enables protein post-translational translocation mediated by the Sec-dependant pathway (Yoon, Kim and Kim, 2009; Zhang *et al.*, 2018), thus mimicking the manner how PA is naturally secreted. Efforts to produce rPA using *Bacillus subtilis*, a gram-positive bacterium closely related to *B. anthracis* with a similar secretory machinery, are halted due to rPA degradation by extracellular proteases

(Baillie, Johnson and Manchee, 1994). Indeed, rPA degradation has been found in the vaccine AVP (Charlton *et al.*, 2007) and a similar outcome in AVA is expected as protease secretion is a natural characteristic of *Bacillus* sp. Recent studies have genetically engineered a *B. anthracis* strain that lacks extracellular proteases for overexpression of rPA (Pomerantsev *et al.*, 2017; Sharma *et al.*, 2018), producing a promising system not only for rPA expression but also as a novel alternative for secretion of recombinant proteins. However, the use of non-pathogenic *B. anthracis* strains still raise safety concerns and regulatory agencies require their handling as class III pathogen (Leppla, 1988; Plaut *et al.*, 2018).

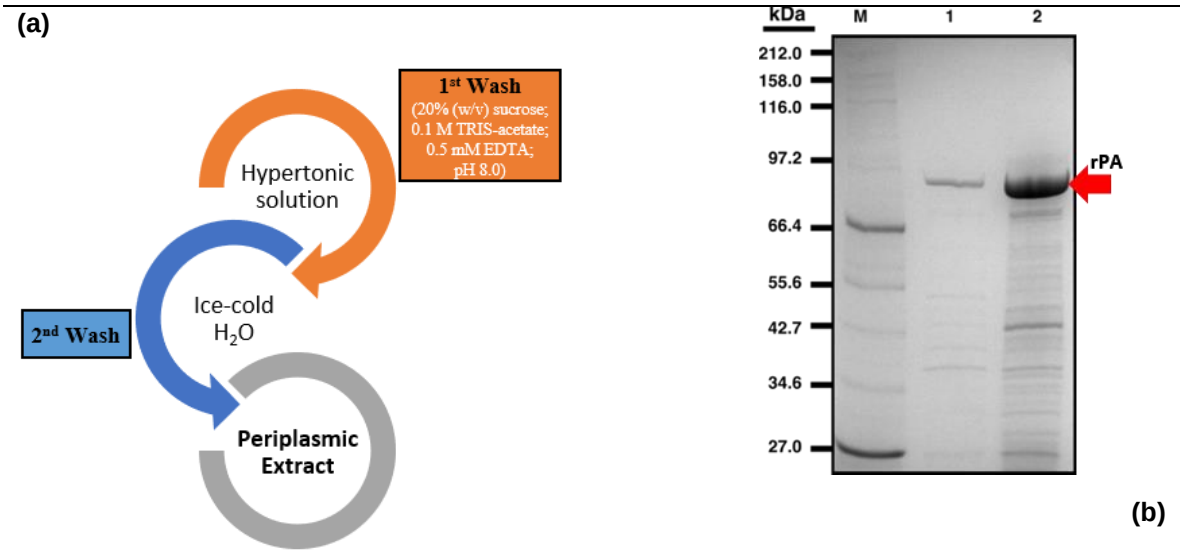


Figure 3.3 rPA recovery by osmotic shock. (a) Schematic describing the osmotic shock method for recovery of rPA. **(b)** SDS-PAGE of rPA recovered from the periplasmic space by osmotic shock. M: protein ladder. Lane 1 represents the protein crude obtained after washing cells with hypertonic solution (20% (*w/v*) sucrose solution); the protein crude obtained from the following washing with ice-cold water is represented in lane 2.

3.3.3 Downstream processing: rPA was purified by anion-exchange chromatography and gel filtration.

The development of an effective purification process is based on deep knowledge of the biophysical characteristics of the target protein and the process is usually established via experimental trial and error. rPA purification has been achieved using several chromatography methods, including ion-exchange chromatography (**Table 3.1**). Ion-exchange chromatography is used for capturing proteins depending on their surface net-charge using ion exchanger matrices. The net-charge of a protein can be determined by its isoelectric point (pI): at buffer pH above its pI, the protein is negatively charged; and at buffer pH below its pI, the protein is positively charged. rPA pI is 5.6 (as calculated by ProtParam (Gasteiger *et al.*, 2005)), hence it presents a global negative charge at physiological conditions, thus able to bind to positively charged matrices, such as quaternary amines (Q) and diethylaminoethyl groups (DEAE). Bound proteins are then separated by differences on their net-charge using a linear salt gradient, typically from 0 – 1 M, to achieve stepwise elution.

The purification process comprises two subsequent steps of anion-exchange chromatography using chromatography columns with matrices of different ionic strength: (1) rPA capture from periplasmic extract using Q-column, a strong anion-exchanger; (2) rPA concentration using DEAE-column, a weak anion-exchanger. The logic behind is that the weaker ionic strength of the DEAE-column would retain rPA and elute remaining protein impurities. Then, a third step comprising gel filtration as final polishing step separates proteins by size, yielding pure rPA (**Figure 3.4**). rPA eluted from Q-column and DEAE-column using a 0 – 1 M NaCl gradient at ~ 0.35 and ~ 0.25 M NaCl, respectively. Pure rPA was eluted by isocratic gel filtration.

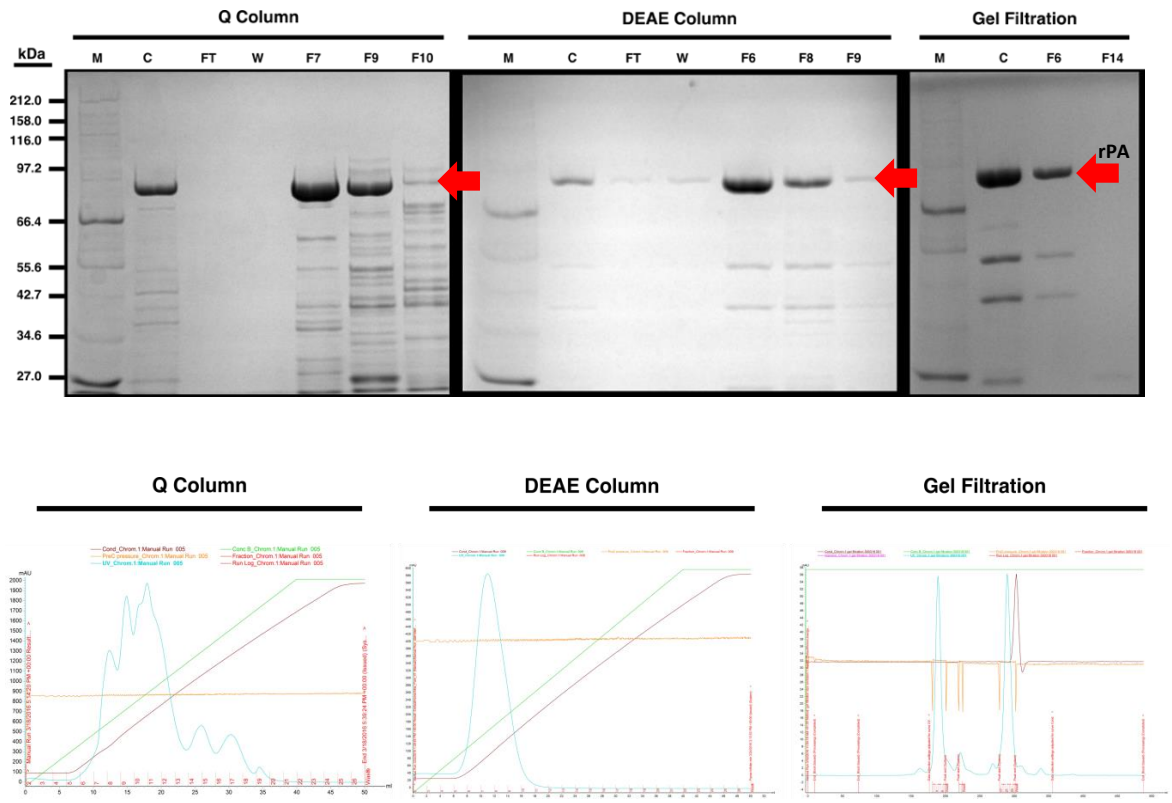
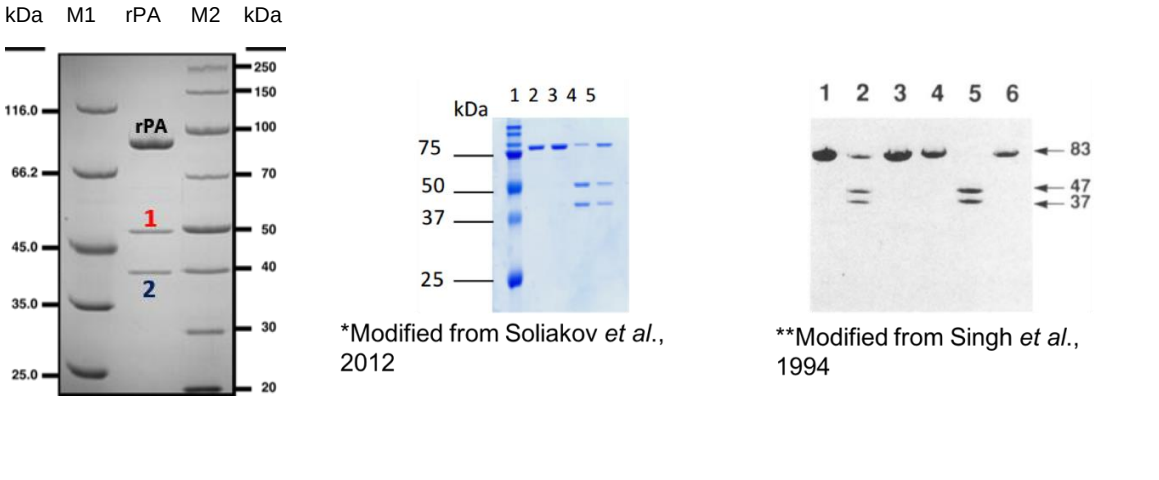


Figure 3.4 rPA purification process. SDS-PAGE analysis of eluted protein fractions (F#) containing rPA (red arrow, 83 kDa) during each purification step. For all gels: lane M is the protein ladder; lane C is the protein crude; lane FT represents the proteins in the flow-through (i.e. proteins that did not bind the column); lane W represents proteins detached (washed) from the column after 5 cvs with loading buffer (Buffer A). Chromatograms are presented as displayed by UNICORN Control Software for each purification step: blue lines are U.V signals; green lines are salt concentration (%); and red lines are conductivity

3.3.4 rPA is cleaved by proteases in a chymotrypsin-sensitive site

(a)



(b)

1

Protein sequence coverage: 28%

Matched peptides shown in ***bold red***.

```
  1 EVKQENRLLN ESESSSQGLL GYYFSDLNFQ APMVVTSSTT GDLSIPSSEL
 51 ENIPSENQYF QSAIWSGFIK VKKSDEYTFA TSADNHVTMW VDDQEVINKA
101 SNSNKIRLEK GRLYQIKIQY QRENPTEKGL DFKLYWTDSQ NKKEVISSDN
151 LQLPELKQKS SNSRKKRSTS AGPTVPDRDN DGIPDSLEVE GYTVDVKNKR
201 TFLSPWISNI HEKKGLTKYK SSPEKWSTAS DPYSDFEKVT GRIDKNVSPE
251 ARHPLVAAYP IVHVDMENII LSKNEDQSTQ NTDSETRTIS KNTSTSRTHT
301 SEVHGNAEVH ASFFDIGGSV SAGFSNSNSS TVAIDHSLSL AGERTWAETM
351 GLNTADTARL NANIRYVNTG TAPIYNVLPT TSLVLGKNQT LATIKAKENQ
401 LSQILAPNNY YPSKNLAPIA LNAQDDFSST PITMNYNQFL ELEKTKQLRL
451 DTDQVYGNIA TYNFENGRVR VDTGSNWSEV LPQIQETTAR IIFNGKDLNL
501 VERRIAAVNP SDPLETTKPD MTLKEALKIA FGFNEPNGNL QYQGKDITEF
551 DFNFDQQTSQ NIKNQLAELN ATNIYTVLDK IKLNAKMNIL IRDKRFHYDR
601 NNIAVGADES VVKEAHREVI NSSTEGLLLN IDKDIRKILS GYIVEIEDTE
651 GLKEVINDRY DMLNISSLRQ DGKTFIDFKK YNDKLPLYIS NPNYKVNVYA
701 VTKENTIINP SENGDTSTNG IKKILIFSKK GYEIG
```

2

Protein sequence coverage: 20%

Matched peptides shown in ***bold red***.

```
  1 EVKQENRLLN ESESSSQGLL GYYFSDLNFQ APMVVTSSTT GDLSIPSSEL
 51 ENIPSENQYF QSAIWSGFIK VKKSDEYTFA TSADNHVTMW VDDQEVINKA
101 SNSNKIRLEK GRLYQIKIQY QRENPTERGL DFKLYWTDSQ NKKEVISSDN
151 LQLPELKQKS SNSRKKRSTS AGPTVPDRDN DGIPDSLEVE GYTVDVKNKR
201 TFLSPWISNI HEKKGLTKYK SSPEKWSTAS DPYSDFEKVT GRIDKNVSPE
251 ARHPLVAAYP IVHVDMENII LSKNEDQSTQ NTDSETRTIS KNTSTSRTHT
301 SEVHGNAEVH ASFFDIGGSV SAGFSNSNSS TVAIDHSLSL AGERTWAETM
351 GLNTADTARL NANIRYVNTG TAPIYNVLPT TSLVLGKNQT LATIKAKENQ
401 LSQILAPNNY YPSKNLAPIA LNAQDDFSST PITMNYNQFL ELEKTKQLRL
451 DTDQVYGNIA TYNFENGRVR VDTGSNWSEV LPQIQETTAR IIFNGKDLNL
501 VERRIAAVNP SDPLETTKPD MTLKEALKIA FGFNEPNGNL QYQGKDITEF
551 DFNFDQQTSQ NIKNQLAELN ATNIYTVLDK IKLNAKMNIL IRDKRFHYDR
601 NNIAVGADES VVKEAHREVI NSSTEGLLLN IDKDIRKILS GYIVEIEDTE
651 GLKEVINDRY DMLNISSLRQ DGKTFIDFKK YNDKLPLYIS NPNYKVNVYA
701 VTKENTIINP SENGDTSTNG IKKILIFSKK GYEIG
```

Figure 3.5 rPA proteolysis by chymotrypsin. **(a)** SDS-PAGE analysis of rPA after gel filtration showing rPA (83 kDa) and two shorter peptides: peptide 1 (~ 50 kDa) and peptide 2 (~ 40 kDa). M1: Unstained Protein Molecular Weight Marker; M2: PageRuler Unstained Broad Range Protein Ladder. Gels on the right correspond to limited proteolysis assays with chymotrypsin on recombinant PA found in the literature (Singh *et al.*, 1994; Soliakov *et al.*, 2012). **(b)** Sequence coverage of peptides 1 and 2 as revealed by mass spectrometry. Mascot database searching against the *E. coli* (strain B/BL21-DE3) reference proteome, plus rPA sequence. Matched peptides 1 and 2 are shown in red against the full-length amino acid sequence of rPA. Both peptides 1 and 2 correspond to rPA.

SDS-PAGE analysis of purified rPA after gel filtration (**Figure 3.4,** Gel Filtration F6) showed three distinct bands, the one on top corresponding to rPA (83 kDA) and two unknown smaller peptides (~ 50 kDa and ~ 40 kDa). When compared with other studies, these bands resemble the pattern left after cleaving PA with chymotrypsin (**Figure 3.5a**) (Singh *et al.*, 1994; Soliakov *et al.*, 2012). In order to determine if the smaller peptides correspond to cleaved rPA or to *E. coli* proteins, the fragments were excised from an SDS-PAGE stained with silver-staining to perform mass spectrometry (MS). MS data was acquired by LC-MS/MS and analysis determined that the fragments match rPA C-terminal and N-terminal peptides, respectively, consistent with digestion patterns of rPA treated with chymotrypsin protease (**Figure 3.5b**).

3.3.5 Generation of PA(+++): a protease-resistant aggregation-resistant variant

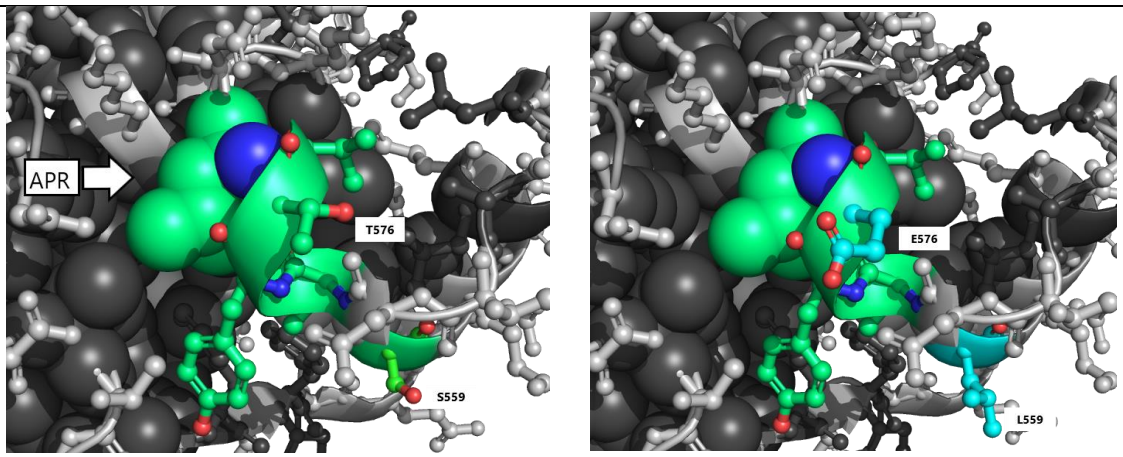


Figure 3.6 PA(+++): a protease- aggregation-resistant variant. The variant PA(+++), produced by SDM on rPA, contains three modifications. The mutation FFD_{315}-AAA_{315} removes a chymotrypsin-site found in a loop in domain 2 unsolved in the crystal structure. Mutations T576E and S559L (right, in cyan) were designed by Ganesan *et al* (2016), T576E was reported to silence the solvent-exposed aggregation-prone region (APR) $NIYTVL_{578}$ (in green) on domain 3 and S559L was reported as compensatory mutation. PDB: 3TEW. Sticks are solvent-exposed residues. Spheres are buried residues.

As shown in **Figure 3.5**, the bioprocess for expression and purification of rPA yielded a cleaved protein, whose fragments correspond to the chymotrypsin-sensitive site FFD_{315} that lies in domain 2 in a highly flexible loop, unsolved in the crystal structure (Petosa *et al.*, 1997). The chymotrypsin-sensitive site was removed by replacing FFD_{315} for AAA_{315} with site-directed mutagenesis (SDM). This mutant was reported to significantly reduce the translocation activity of PA in studies searching for dominant negative inhibitors (DNI) (Singh *et al.*, 1994)**.** DNIs of PA can assemble the Anthrax toxin and bind cell receptors but are unable to translocate the effector proteins into the cytosol (Mourez *et al.*, 2003). These mutations are relevant for development of rPA-based vaccine candidates as well as anti-toxin therapeutics. Also, two additional mutations were introduced: T576E and S559L (**Figure 3.6**), from an rPA variant reported to prevent aggregation (Ganesan *et al.*, 2016). Previous studies showed that a solvent-exposed hydrophobic patch in domain 3 (sequence $NYTVL_{578}$) served as nucleation site for aggregation of rPA (Ganesan, Watkinson and Moore, 2012). Mutation T576E was designed to silence this aggregation-prone region (APR). The logic behind this type of mutation is the substitution of a residue within an APR (T576) for a charged residue (576E), thereby charged residues repel and hinder intermolecular interactions between APRs of monomers. These mutations are called gatekeepers (GKs) and there is evidence to suggest their appearance as a natural selection mechanism that modulates protein abundance (De Baets *et al.*, 2014). T576E was predicted using SolubiS (Van Durme *et al.*, 2016), a webserver for protein design that scans for GKs to reduce aggregation propensity without disrupting the thermostability of the native protein structure (i.e. it ignores the hydrophobic regions in the protein's core). The analysis is based on the TANGO algorithm for prediction of APRs (Fernandez-Escamilla *et al.*, 2004) and FoldX for calculations of mutational free energy change (Schymkowitz *et al.*, 2005), thus it mutates all

residues within the predicted APR for D, E, R, K and P and estimates the mutation effect on the thermostability of the protein. Mutation S559L was reported as a compensatory mutation for T576E (Ganesan *et al.*, 2016), suggesting that mutation T576E is destabilising despite of its anti-aggregation character. Even though aggregation issues were not encountered during bioprocessing of rPA, these mutations will be useful for objectives described in the following chapters of this thesis, comprising the protein engineering campaign on rPA aiming to improve its thermostability.

3.3.6 Purification of two rPA variants demonstrates bioprocess reproducibility.

PA(+++), an aggregation- protease-resistant variant containing the substitutions T576E, S559L and FFD_{315}-AAA_{315} was produced by SDM. Also, an rPA variant containing only mutations T576E and S559L, PA(++), was as well produced to compare purification results. As shown in **Figure 3.7**, the absence of the 40 and 50 kDa fragments in SDS-PAGE after gel filtration demonstrate that mutation FFD_{315}-AAA_{315} in PA(+++) successfully removed the chymotrypsin-sensitive site. On the other hand, variant PA(++) (without FFD_{315}-AAA_{315}), still shows the protein fragments corresponding to cleavage by chymotrypsin. Moreover, the ability to purify two rPA variants demonstrate the reproducibility and robustness of the bioprocess.

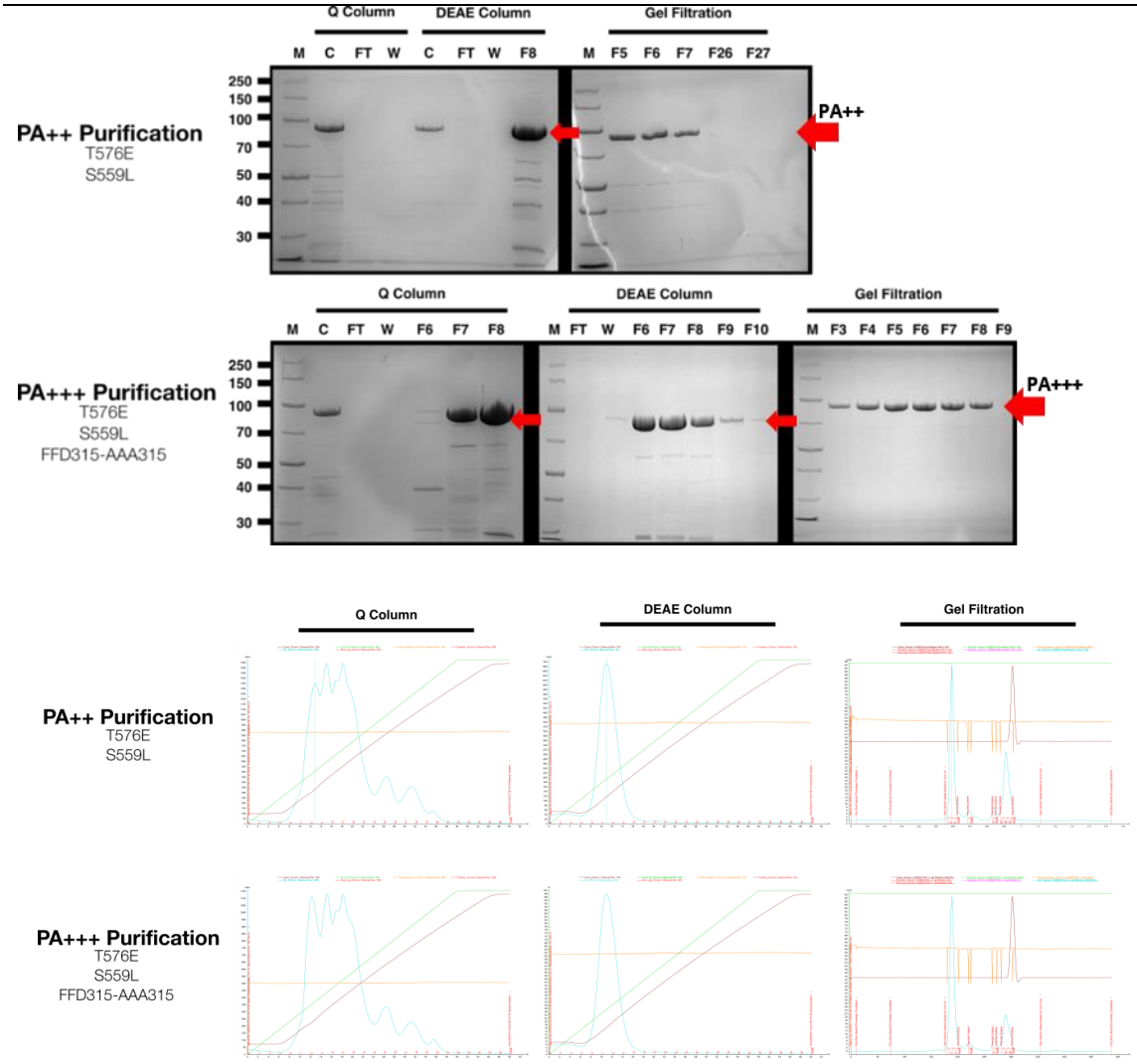


Figure 3.7 PA(+++) purification demonstrates bioprocess reproducibility. The variant PA(+++), containing mutations T576E, S559L and FFD_{315}-AAA_{315}, was purified and analysed by SDS-PAGE. The variant PA(++), containing only mutations T576E and S559L, designed by Ganesan *et al* (2016), was purified as well for comparison. As shown in SDS-PAGEs Gel Filtration, PA(+++) successfully removed the chymotrypsin-sensitive site, whereas variant PA(++) shows the 40 and 50 kDa fragments from chymotrypsin digestion. The ability to purify two distinct rPA variants using the same method for production of pure rPA demonstrate the reproducibility and robustness of our bioprocess. Chromatograms are presented as displayed by UNICORN Control Software for each purification step: blue lines are U.V signals; green lines are salt concentration (%); and red lines are conductivity

3.4 Conclusion

Overexpression of soluble rPA in *E. coli* BL21 (DE3) was achieved using a codon-optimised construct and periplasmic expression at 20 ºC with low inducer concentration (0.1 mM IPTG) in rich 2xTY medium. Soluble rPA can be recovered from the periplasmic space using osmotic shock method. rPA was captured from periplasmic extract by ion-exchange chromatography using Q-column, a strong anion exchanger. Then, rPA was further purified by ion-exchange chromatography using DEAE-column, a weak anion exchanger. A final polishing step with gel filtration yielded pure rPA. SDS-PAGE analysis on pure rPA showed that rPA is cleaved by proteases in a chymotrypsin-sensitive site. The proteolytic fragments were analysed by MS, revealing that the fragments were indeed rPA peptides consistent with chymotrypsin proteolysis. Then, rPA cleavage by proteases was confirmed after the generation of a protease-resistant variant using SDM on the chymotrypsin-sensitive site FFD_{315}. In consequence, rPA triple mutant, PA(+++), was generated using SDM to introduce the following mutations: (1) the chymotrypsin-sensitive site FFD_{315} was substituted with AAA_{315}; (2) T576E and (3) S559L. Mutations (2) and (3) are reported in the literature to reduce the aggregation propensity of rPA. Also, mutant PA(++), carrying only mutations (2) and (3), was produced only for comparison. Both variants were expressed and purified using the established bioprocess for rPA. It was concluded that mutation FFD_{315}-AAA_{315} successfully removed the chymotrypsin-sensitive site. The successful soluble expression and purification of rPA and two distinct rPA variants demonstrate the robustness of the bioprocess.

PA(+++) will be utilised in the following chapters of this thesis as initial template to improve its thermostability using protein engineering approach aiming to produce a candidate for a thermostable next-generation Anthrax vaccine.

Chapter 4. Semi-rational and rational protein engineering approach for the improvement of recombinant Protective Antigen thermostability

4.0 Abstract

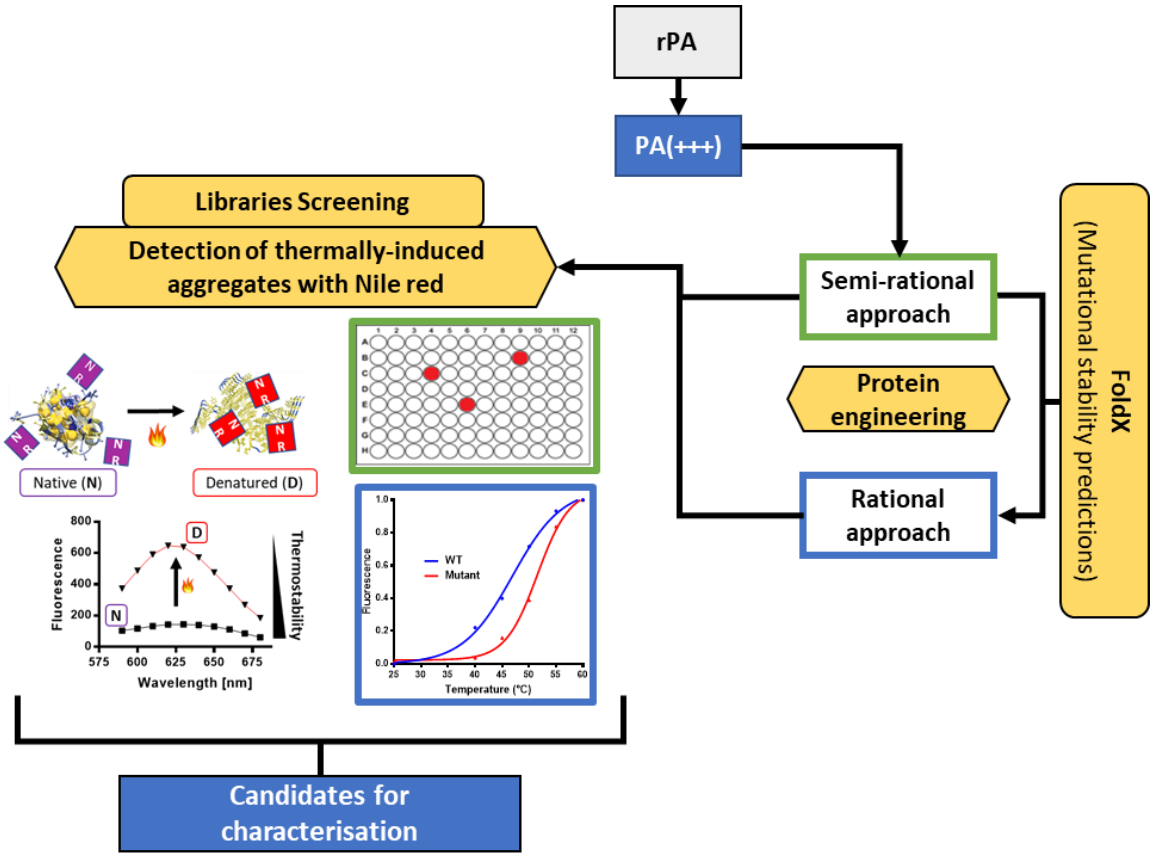


Figure 4.0 Graphical abstract. The protein engineering process of rPA.

The potential of therapeutic proteins is often limited by issues related to poor stability. Efforts to develop a next-generation Anthrax vaccine have been hampered due to such stability constraints. It is well-known that antibodies directed against Protective Antigen (PA), a protein component of the Anthrax toxin, are the main correlates of protection against Anthrax disease. Thus, efforts to develop a well-defined next-generation Anthrax vaccine are focused

on the production of recombinant PA (rPA). However, rPA is a thermolabile protein found to aggregate at relatively low temperatures (37 °C) and that has shown poor long-term stability issues in current recombinant vaccine candidates. Our proposal is that the stability of rPA-based vaccines can be improved by generating a thermostable rPA variant using protein engineering approach. In this Chapter, we used semi-rational and rational approach protein engineering to generate thermostable variants by designing mutations that improve the protein's hydrophobic core-packing. Libraries were generated using PA(+++) (see **Chapter 3**) as parental template. A semi-rational approach library was constructed using structure-based analysis of an atomic-resolution X-ray crystal structure of PA for selection of hotspot positions and mutations designed to improve hydrophobic core-packing. The mutational effect of these substitutions was predicted using the empirical force field algorithm FoldX and a small-size rational approach library was constructed with the mutations that resulted beneficial. Then, libraries were screened for thermostability. A semi-high throughput method was developed for screening of the semi-rational approach library, consisting of library expression in 96-well plates following recovery of periplasmic extracts and subsequent detection of thermally induced aggregates using the polarity sensitive dye Nile red. Three hits were isolated to perform a validation assay, one resulting positive (variant PA(+++)-T81I). A method for screening small-size rational approach libraries was developed as well. Rational approach variants were expressed in 50 mL flask culture, recovered in periplasmic extracts by osmotic shock and thermostability was indirectly measured by detection with Nile red of aggregates produced at a temperature range. Variants PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V resulted positive. Positive variants form both semi-rational and rational approach were selected for characterisation in **Chapter 5**.

4.1 Introduction

A protein is considered stable if it can fold to its native conformation and perform its function under given conditions. Proteins fold according to intrinsic thermodynamic parameters that determine their equilibrium state, whereby their structural conformation is reversibly shifting towards the most favourable energetic state. Although comprehensive understanding of protein folding remains one of the greatest conundrums of science, it is well accepted that all the information a protein needs to fold is encoded in its sequence and that it is finely tuned by evolution towards optimal function under native conditions (Dobson, 2003). Such native conditions are often dramatically different from those of biotechnological interest, being the reason behind the poor stability issues (e.g. thermolability) that limit the manufacture of recombinant proteins in the biopharmaceutical industry (Cromwell, Hilario and Jacobson, 2006; Manning *et al.*, 2010).

Conformational stability is of special importance for the development of next-generation vaccines manufactured with recombinant proteins because preservation of the structural conformation of the antigen is essential to effectively display protective epitopes (Scheiblhofer *et al.*, 2017). The target to develop a recombinant protein-based vaccine for Anthrax is the toxin component Protective Antigen (PA). PA undergoes significant conformational rearrangements to accomplish its natural function, thus it is naturally inclined to experience high conformational flexibility (Feld *et al.*, 2012; Jiang *et al.*, 2015). When recombinantly produced PA (rPA) is combined with alum salts adjuvants in vaccine formulations its inherent flexibility contributes to stability loss by promoting the formation of unfavourable interactions within the protein and with the adjuvant, which in turn promotes the occurrence of partially folded moieties that are detrimental for vaccine efficacy due to

loss of protective epitopes and promotion of unspecific immune responses (Wagner *et al.*, 2012). Moreover, partially unfolded intermediates can trigger irreversible denaturation by aggregation and promote chemical alterations such as deamidation, shortening vaccine shelf life. Indeed, rPA is considered a thermolabile, aggregation-prone protein (Radha *et al.*, 1996; Chalton *et al.*, 2007) and also have shown propensity to suffer deamidation during short-term storage (Powell *et al.*, 2007). Although focused mutagenesis on PA has been widely used to understand its functionality, only few studies have explored mutational effects on PA stability, including solubility improvement (Ganesan *et al.*, 2016), reduction of its deamidation propensity (Verma and Burns, 2018) and improvement of stability via electrostatic contributions of ion pairs (Iyer *et al.*, 2013).

The hypothesis of this thesis is that rPA stability can be improved through protein engineering approach by designing mutations to improve its hydrophobic core-packing. This type of engineering approach is designed to reduce conformational flexibility generating protein variants that maintain their native functional state at higher temperatures (i.e. improved thermostability) (Radestock and Gohlke, 2011; Angarica and Sancho, 2012; Rathi *et al.*, 2016), thus the ultimate aim is to obtain a functional rPA variant with improved thermostability as a candidate to develop a next-generation thermostable Anthrax vaccine. **S**emi-rational protein engineering approach was used to select hot-spot positions by scrutiny of a high resolution (1.45 Å) X-ray crystal structure of PA, investigating core residues in strategic positions, such as those involved in cavity formation or with buried polar side chains with unsatisfied hydrogen bonds, and replacing them for bulkier hydrophobic amino acids, some positions involving more than one substitution. Also, the computational force field algorithm FoldX (Schymkowitz *et al.*, 2005) was used to predict beneficial mutations on the

previously selected hot-spots to create a small rational approach library. The protease-aggregation-resistant variant PA(+++) (as designed in **Chapter 3**) was used as parental template for mutagenesis and as starting point for the protein engineering campaign. Then, a semi-high throughput stability screening method was developed for selection of positive variants for subsequent characterisation in **Chapter 5**.

4.2 Aims

- To use semi-rational protein engineering approach to select hot-spot positions for improvement of hydrophobic core-packing using structure-based analysis.
- To create a semi-rational approach library using PA(+++) as parental template to generate variants with improved hydrophobic core-packing aiming to improve PA(+++) thermostability.
- To create a small rational approach library using the computational force field algorithm FoldX to predict beneficial mutations on previously selected hot-spots.
- To design a semi-high throughput method for screening of variants with improved thermostability.

4.3 Results and discussion

4.3.1 Protein engineering: generating genetic diversity

A protein engineering campaign comprises the generation of genetic diversity followed by screening a library for variants with a desired phenotype. Genetic diversity is typically achieved by introduction of random mutations as part of an iterative directed evolution approach (Tee and Wong, 2013). However, since the native folded state is only marginally stable (Pace *et al.*, 1996), it is expected that most random mutations will be deleterious,

detrimental for protein stability (Bloom *et al.*, 2006). Moreover, an antigen must display the same protective epitopes as during a natural infection in order to elicit immunity, thus a random approach is not ideal to develop an engineered recombinant vaccine because it may alter relevant protective epitopes. On the other hand, knowledge-based approaches use structural, sequence and functional information to select specific positions for focused mutagenesis following a rational design to obtain small high-quality libraries (Lutz, 2010). Thus, the criteria to select positions for focused mutagenesis included: (1) excluding residues within experimentally characterised PA protective epitopes (McComb and Martchenko, 2016) and (2) excluding experimentally characterised key functional residues reported to elicit production of toxin-neutralising antibodies (Rosovitz *et al.*, 2003; Zhou, Carney and Janda, 2008; Kelly-Cirino and Mantis, 2009; Smith *et al.*, 2012; Chi *et al.*, 2015).

Mutations were introduced by whole-plasmid site-directed mutagenesis (Edelheit, Hanukoglu and Hanukoglu, 2009), using the variant PA(+++) as parental template. PA(+++) is a protease- aggregation-resistant variant, designed and described in **Chapter 3**. PA(+++) contains mutations T576E and S559L, designed by Ganesan *et al.* (2016) to improve rPA solubility while conserving its native structure and biological activity. These mutations successfully supressed a kinetically-trapped intermediate involved in the promotion of intermolecular interactions that cause initial nucleation during early aggregation events of rPA at room temperature (Ganesan, Watkinson and Moore, 2012). Yet, improvement of folding cooperativity did not show an effect on temperature of unfolding, as midpoint thermal unfolding transition (Tm) of the mutant (47.8± 0.13 °C) was only marginally higher than that of rPA (46.4± 0.13 °C), as determined by differential scanning calorimetry (DSC) (Ganesan *et al.*, 2016). Therefore, the decision to use PA(+++) as starting point for thermostability

improvement was based on: (1) the T576E-S559L mutant retains rPA biological activity and structure, (2) presents a similar midpoint thermal unfolding transition related to rPA, (3) immunisation in mice generates antibodies that recognise rPA, (4) improved solubility is a desired feature to improve shelf-life of therapeutic proteins and (5) PA(+++) stability is more robust as result of the suppression of a kinetically-trapped intermediate in rPA, thus the sequence space available to explore thermostable mutations is broaden and can be studied in a semi-high throughput manner (e.g. thermal shift assays). Additionally, PA(+++) contains mutation FFD-AAA_{315} that removes a chymotrypsin-sensitive site found to cause rPA proteolysis during bioprocessing (see **Chapter 3**). This mutation is harboured in rPA membrane insertion loop (MIL) (residues 303-324), involved in toxin translocation (Petosa *et al.*, 1997). The MIL is not resolved in any available crystal structure and its removal in fact improves crystal diffraction (Feld *et al.*, 2012), thus it is expected for this highly dynamic loop to have a small contribution on folding and global conformational stability of the protein. Therefore, it was assumed that FFD-AAA_{315} is not deleterious. Moreover, mutations on this site produced a dominant negative inhibitor that prevents toxin translocation while maintaining rPA ability for toxin assembly and receptor binding (Singh *et al.*, 1994).

Semi-rational approach

Position	Buried	Mutation
A63	YES	I V
M89	YES	L
V91	YES	L I
V96	YES	I M
V262	YES	L I
V264	YES	I
M266	YES	L
M350	YES	L
I364	YES	L
I432	YES	L
M434	YES	L I
A489	YES	V
I491	YES	L
A526	YES	T
I589	YES	L
V644	YES	L
V698	YES	I
D283	NO	S G
L340	NO	T N D
L566	NO	I V
T81	YES	I V L
T381	YES	L M
T674	YES	V
Y436	NO	A L V

Rational approach

Position	Buried	Mutation
A489	YES	V
V264	YES	I
T674	YES	V

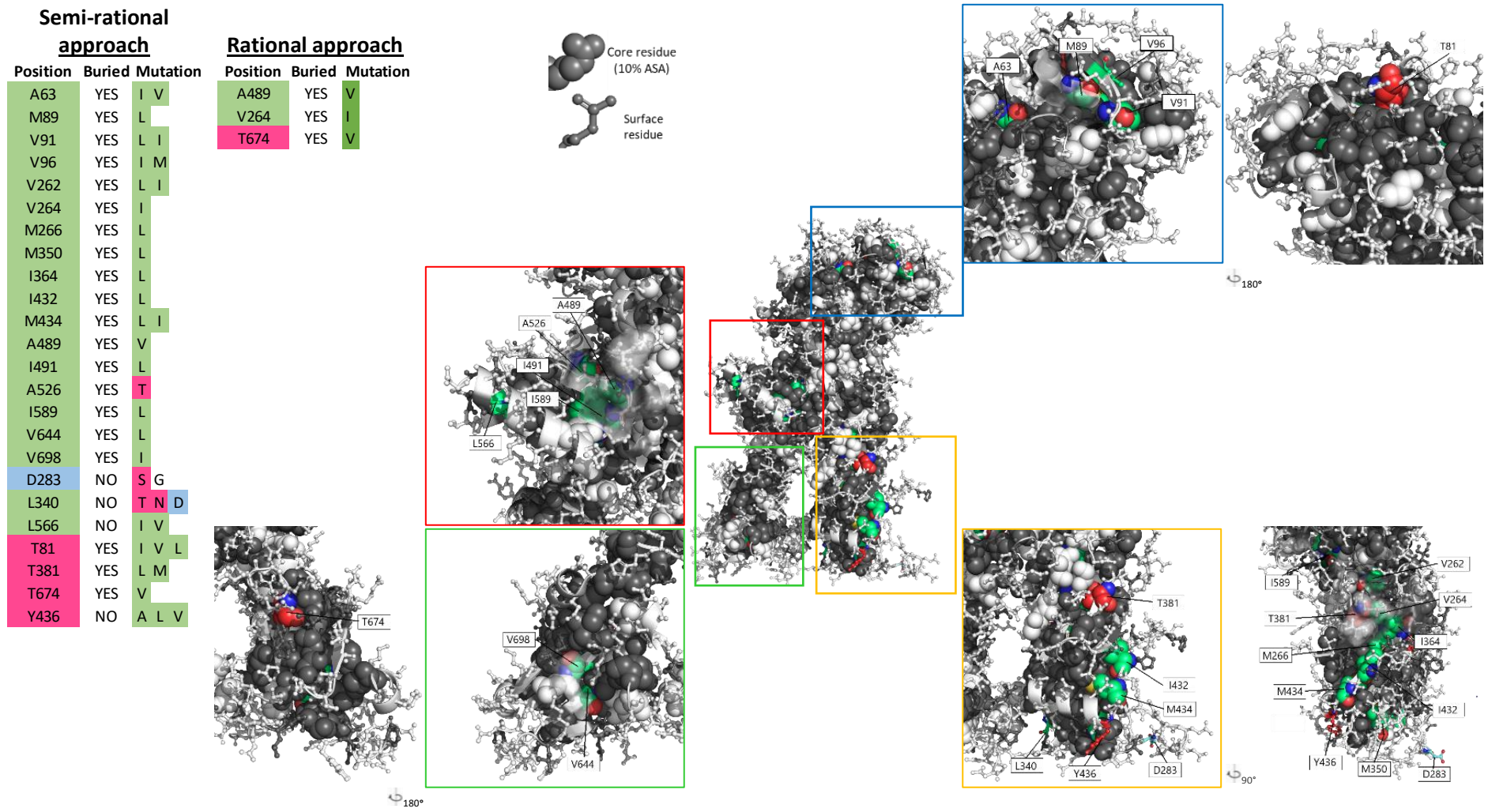


Figure 4.1. Semi-rational and rational approach libraries. A semi-rational approach library was created aiming to improve rPA hydrophobic core-packing. Positions were selected by structure-based design using the PA atomic-resolution X-ray crystal structure 3TEW (PDB ID). Rational approach was applied to generate a small library using the force field algorithm FoldX to predict beneficial mutations on the preselected residues. Substitutions that predicted $\Delta\Delta G < -0.5$ kcal/mol were selected. Focused mutagenesis was performed on variant PA(+++). Colour code of mutations by residue type: Green = apolar; Red = polar; Cyan = charged. Buried residues are $\leqslant$ 10% accessible surface area (ASA), depicted as spheres. Sticks are solvent-exposed residues. Black residues are apolar and white residues are polar or charged.

4.3.1.1 Semi-rational approach library: structure-based design.

Assuming a two-state equilibrium model between folded and unfolded state, the thermal stability of a protein can be defined as the balance of forces between the stabilising non-covalent interactions that conform the native folded state against the unfolded state. The use of structure-based design to improve protein thermostability assumes that a high-resolution crystal structure represents the native folded state of the protein and that can be used to make stability predictions based on the effect of stabilising non-covalent interactions (Pace *et al.*, 1996; Lamazares *et al.*, 2015; Malito, Carfi and Bottomley, 2015). Such interactions are temperature dependent, thus thermostable proteins attained improved amino acid interactions that shift equilibrium towards the native state at higher temperature related to meso stable proteins (Razvi and Scholtz, 2006; Ashenberg, Gong and Bloom, 2013). The underlying thermodynamic mechanism of this balance of forces is known as well as conformational stability, that is the difference between the free energy change (ΔG) of transition from folded to unfolded state.

It is considered for most proteins that the main driving force behind folding is the hydrophobic effect (Camilloni *et al.*, 2016), whereby a hydrophobic protein core is formed as consequence of the burial of clustered hydrophobic residues avoiding contact with water. Then, the polypeptide backbone and side chains rearrange to form electrostatic interactions that further stabilise the protein by forming secondary structures, in consequence the protein's core is shielded from water and the surface is mainly populated with polar residues that interact favourably with the solvent. These events can be described as the creation of interaction networks during folding, including hydrophobic contact networks (e.g. hydrophobic core-packing), hydrogen bond networks (e.g. secondary structures, surface

hydration) and other electrostatic interactions (e.g. salt bridges and aromatic interaction networks) (Pucci and Rooman, 2017). Structural studies on thermostable proteins attempt to elucidate the role of these networks and their contribution to thermal stability relative to meso stable homologs (Sammond *et al.*, 2016). Interaction networks are often interdependent, so it is difficult to elucidate their individual effects on global conformational stability, yet several structural studies have concluded that thermostable proteins have significantly enhanced hydrophobic core-packing and reduced flexibility (Vieille and Zeikus, 2001; Northey, Di Nardo and Davidson, 2002; Maugini *et al.*, 2009; Angarica and Sancho, 2012; Sammond *et al.*, 2016; Zhang *et al.*, 2016; Sumi and Koga, 2019). Indeed, thermostable variants have been obtained by protein engineering approach designing mutations aiming to improve hydrophobic core-packing and conformational rigidity (Joo *et al.*, 2010; Yu and Huang, 2014; Liu *et al.*, 2016; Roth *et al.*, 2017; Li *et al.*, 2018). Also, directed evolution experiments resulting in mutants with improved hydrophobic interactions have generated enzymes with improved thermostability (Wu and Arnold, 2013; Alfaro-Chávez *et al.*, 2019).

In this work, a semi-rational approach mutant library was designed in collaboration with Dr. Antonina Andreeva (MRC Laboratory of Molecular Biology, Cambridge). Mutations were designed by structure-based analysis of PA (PDB: 3TEW) aiming to improve hydrophobic core-packing and reduce protein's flexibility. The crystal structure 3TEW has the highest resolution among resolved PA crystal structures (1.45 Å) (Feld *et al.*, 2012). Residues with the potential to improve hydrophobic networks in the protein core or its periphery were chosen as hot-spot positions and substituted for bulkier hydrophobic residues (Rose *et al.*, 1985), except for positions D283, L340 and Y436 (**Figure 4.1**). Position D283 is part of a solvent-exposed loop in domain 2, so mutations were designed to improve surface

hydration (D283S) or reduce the loop's flexibility (D283G). Position L340 was picked because the crystallization results report significant overlaps with surrounding residues, suggesting high flexibility, thus saturation mutagenesis using residues with similar steric constraints but different chemical groups (L340N, L340D, L340T) was performed on this position. Position Y436 is found in a helix C-terminus in domain 2, in the vicinity of a buried hydrophobic cluster, so hydrophobic residues were introduced to improve helix-terminal and adjacent hydrophobic interactions (Y436A, Y436L, Y436V).

4.3.1.2 Rational approach library: using the computational algorithm FoldX for mutational stability predictions.

Rational approach protein engineering was used to generate a small mutant library. Generally, after selecting hot-spot residues for mutagenesis based on structural or evolutionary information, semi-rational approach generates genetic diversity by introducing a set of substitutions based on empirical information (e.g. increasing hydrophobicity and core-packing increases thermostability) or by saturation mutagenesis (Lutz, 2010; Zhang *et al.*, 2016); on the other hand, rational approach uses computational tools to design focused libraries based on computational prediction algorithms. A rational approach is particularly powerful for designing thermostable libraries because, according to distributions of mutational stability and evolutionary studies, most substitutions are likely to be neutral (Tokuriki *et al.*, 2007; Ashenberg, Gong and Bloom, 2013). Predicting mutational stability also yields significantly smaller libraries of typically ten to a few hundreds of variants, thus a selection screening can be developed without the caveats of miniaturised high-throughput assays or the variants can be directly characterised after purification using calorimetry assays.

Mutational stability is defined as the difference in ΔG of a mutant related to the ΔG of the WT, and it is expressed as ΔΔG. As the native ↔ unfolded transition has a negative ΔG (– ΔG) because adopting the native state is energetically more favourable than the unfolded state, mutational stability is also negative (– ΔΔG) if the mutant has improved conformational stability related to the WT. Thus, most computational algorithms predict mutational stability using an energy function, either by iterative methods that find a minimum energetic state or by using energy functions that assign scores and weights to a set of empirical parameters (Broom *et al.*, 2017). FoldX is an empirical force-field algorithm (Schymkowitz *et al.*, 2005). Calculations are based on empirical terms described in simple energetic contributions, as shown in equation 4.1 (**Eq. 4.1.**), that result in the ΔG (in kcal/mol) of a protein.

Eq. 4.1. $$\Delta G = a \cdot \Delta G_{vdw} + b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + f \cdot \Delta G_{el} + g \cdot \Delta G_{kon} + h \cdot T\Delta S_{mc} + k \cdot T\Delta S_{sc} + l \cdot \Delta G_{clash}$$

Using an atomic-resolution three-dimensional structure, FoldX can calculate the energy difference between the wild-type and a variant ($\Delta\Delta G = \Delta G_{wt} - \Delta G_{variant}$). The fixed backbone (from the 3D-structure) is accounted in **Eq. 4.1** as entropy penalties ($T\Delta S_{mc}$ and $T\Delta S_{sc}$, are entropy penalties from fixed backbone chain and side chains, respectively). The energetic terms in the function are weighted using constants ($a \ldots l$) obtained by training the algorithm with experimental ΔΔG values from a mutant database to represent the best fit (R = 0.8; SD = 0.46 kcal/mol) (Guerois, Nielsen and Serrano, 2002).

Briefly, the force field is calculated by simulating the folding process calculating the desolvation and van der Waals forces transition from unfolded to folded state. As such, it assumes a hypothetical unfolded state structure, whose energetic properties are implied in

the weight terms ($a \ldots l$), thus absolute protein ΔG predictions (folded ↔ unfolded) are expected to have a large error because of lack of knowledge of the unfolded state structure. Therefore, FoldX works with the above-mentioned high accuracy only when the ΔG reference is an atomic-resolution three-dimensional structure of the wild-type protein and calculating the ΔΔG against a protein variant ($\Delta\Delta G = \Delta G_{wt} - \Delta G_{variant}$). The desolvation contributions of each amino acid were derived from physicochemical experiments on amino acids free energy of transfer from solvents with different polarity (apolar ↔ polar) aiming to mimic the transition that an amino acid in a polypeptide suffers during unfolding (buried ↔ solvent-exposed). Depending on residue type the contributions are separated in polar (ΔG_{solvP}) and hydrophobic (ΔG_{solvH}), and balanced by considering the burial of each residue calculating its solvent accessibility using the atomic occupancy method (Colonna-Cesari and Sander, 1990). Indeed, FoldX is ideal for predicting hydrophobic core-packing mutations because it uses atomic occupancy method to estimate the burial of residues . Atomic occupancy reflects the packing density around atoms because it calculates the volume of the atoms surrounding the residue in question, thus implying a degree of hydrophobic core-packing and how a mutation may disturb it. For the rest of terms in **Eq. 4.1.**, the van der Waals contributions (ΔG_{vdw}) are derived in a similar manner as desolvation, using instead the experimental energy transfers from water to vapour (Radzicka and Wolfenden, 1988). Hydrogen bonds are geometrically estimated and energetic contributions of forming hydrogen bonds within the protein (ΔG_{hbond}) and with the solvent (ΔG_{wb}) are computed. Electrostatic contributions (ΔG_{el}) are calculated using an implementation of the Coulomb's law. For proteins that form complexes, the rate of complex formation is accounted in the term ΔG_{kon}. For a comprehensive explanation of the FoldX algorithm and its energy terms see the

works by Guerois, Nielsen and Serrano (2002) and Schymkowitz *et al.* (2005). For the generation of a rational approach library, the mutations designed for the construction of the semi-rational approach library were used for stability predictions with FoldX, as follows. First, 3-D structure 3TEW was optimised with the FoldX RepairPDB function. This function produces a model with minimum energy by rearranging the protein's side chains. Then, the BuildModel function was used on the repaired 3TEW model for ΔΔG calculations. Mutations that yielded ΔΔG < – 0.5 kcal/mol were selected to generate a small rational approach library (**Figure 4.1**).

4.3.2 Development of a semi-high throughput screening method for thermostable PA(+++) variants

The development of an effective screening method is essential to explore the effect of all possible mutations in a library. The field of directed evolution has developed a plethora of high-throughput screening methods because the iterative nature of the evolutionary process and generation of large libraries (10^5 - 10^6) demand efficient identification of beneficial mutations to maximise the potential of finding improved variants (Firth and Patrick, 2005; Packer and Liu, 2015). Evolution is hereby directed towards the protein function that the method is screening for by linking genotype to phenotype. In consequence, the development of a stability screening method is particularly difficult due to the inherent complexity of protein conformational stability and thus difficult to link to a specific functionality trait in a high-throughput manner (Ota *et al.*, 2018; Kurahashi, Tanaka and Takano, 2019). Typically, a high-throughput method for thermostability screening involves thermal treatment of the mutant library followed by measurement of the residual activity of a protein's function, such as catalytic activity, that can be readily evaluated using spectrophotometry to estimate the

consumption of coloured substrates or the titration of chromogenic reactions, thus the thermostability of a variant is indirectly inferred as the residual activity relative to the wild-type (Dror *et al.*, 2014; Zhang *et al.*, 2016). However, residual activity assays are difficult to develop for recombinant protein-based vaccines because these antigens are typically proteins without catalytic activity, found on the outer layer of the infectious agent, or toxin components whose catalytic activity is an undesired feature for vaccine development. Moreover, there is growing evidence from directed evolution experiments that stability and activity are usually decoupled during the evolutionary process (Kurahashi, Tanaka and Takano, 2019).

Recent advancements in high-throughput drug discovery have developed methods to directly evaluate the thermostability of thousands of protein formulations using thermal shift assays like differential scanning fluorimetry (DSF) (Erlanson *et al.*, 2016). Such methods follow the unfolding transition of proteins as the temperature is gradually ramped. DSF uses polarity-sensitive fluorescent dyes whose fluorescence increase in hydrophobic environments, thereby following protein unfolding as the hydrophobic core is gradually solvent exposed until the protein is irreversibly denatured (He *et al.*, 2010). This type of assay assume thermodynamic equilibrium thus can only be applied on purified proteins. Protein purification is often time- and resource-consuming, unfeasible for screening of semi-rational approach libraries.

Therefore, the challenge is to craft an effective stability screening method for semi-rational protein engineering approach without relying on catalytic activity and to indirectly measure thermal stability bypassing protein purification. Inspired by thermal shift assays in drug discovery, we took advantage of the established system for periplasmic protein

expression (**Chapter 3**) and adapted it into a 96-well plate configuration. The mutant library can be expressed in 96-well plate configuration and periplasmic extracts containing overexpressed PA(+++) variants can be recovered for subsequent thermal treatment. Periplasmic extracts are then transferred into a 96-well microtiter plate and the library is exposed to the pre-determined aggregation temperature of the wild-type (PA(+++)). The variants are screened for protein thermostability indirectly, based on the detection of thermally-induced aggregates using the polarity-sensitive fluorescent dye Nile red, whose fluorescence increases in presence of the solvent-exposed hydrophobic regions of denatured proteins (Sackett and Wolff, 1987). Thus, if a variant is less prone to aggregate during heat treatment, it will show weaker fluorescence in presence of Nile red related to parental PA(+++), suggesting improved thermostability.

First, the temperature for library screening was established at the aggregation temperature of PA(+++) contained in periplasmic extracts (**Figure 4.2**). Then, PA(+++) expression and recovery by osmotic shock was optimised for 96-well plates format (**Figure 4.3**), and the optimal Nile red concentration for the assay was established by staining PA(+++) aggregates obtained after expression and recovery in 96-well plate format (**Figure 4.4a**). Finally, the relationship between initial periplasmic extract concentration and fluorescence of stained aggregates was established using different concentrations of PA(+++) periplasmic extracts obtained after 96-well plate expression (**Figure 4.4b**). The ultimate thermostability screening for the semi-rational approach library is described in **Figure 4.5**.

4.3.2.1 Defining PA(+++) aggregation temperature

As demonstrated in **Chapter 3**, the osmotic shock recovery method yields a periplasmic extract with overexpressed PA(+++). Yet, as shown in SDS-PAGE analysis, there are other

proteins present in the extract that could affect the overall aggregation propensity of the periplasmic extract. Since these "impurity proteins" are of smaller size and present in lower concentration, it was assumed that the overall aggregation propensity of the periplasmic extract is directly dependent on PA(+++) thermostability. Consequently, for library screening, it was assumed that the difference on aggregation propensity between periplasmic extracts containing the WT (PA(+++)) and mutant variants is directly dependent on the effect that the mutation has on PA(+++) thermostability. Therefore, a threshold for positive variants was established at the aggregation temperature of PA(+++) contained in periplasmic extracts.

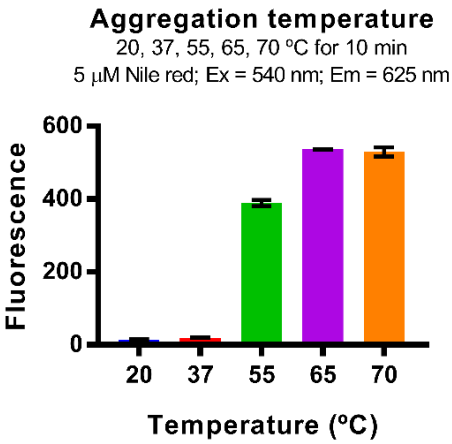


Figure 4.2 PA(+++) aggregation temperature in periplasmic extracts. PA(+++) was expressed in 50 mL 2xTY media in flask culture for 24 hrs and then 2 mL aliquots of culture were transferred into 2 mL microcentrifuge tubes. Cells were harvested and periplasmic extracts containing overexpressed PA(+++) were obtained by osmotic shock method. PA(+++) extracts were then diluted 1:2 with 100 mM MOPS buffer (100 mM MOPS, 200 mM NaCl; pH 7.0). PA(+++) aggregation temperature was determined after incubating the diluted periplasmic extracts at different temperatures (room temperature (20), 37, 55, 65 and 70 °C) for 10 min in a clear-bottom 96-well microtiter plate in presence of 5 µM Nile red. Aggregates were detected by fluorescence at 625 nm after 540 nm excitation with a SpectraMax M2e microplate reader. Corresponding blank fluorescence was subtracted from data points. Bars showing mean with SEM (n = 3).

Periplasmic extracts were obtained from 50 mL flask culture expression of PA(+++) (grown under the culturing conditions defined in **Chapter 3**). After incubation for 10 min at 20, 37, 55, 65 and 70 °C, aggregation was indirectly measured following Nile red staining. The emission maxima of Nile red bound to PA(+++) aggregates was experimentally determined at 625 nm after excitation with 540 nm, and as reported by others (Soliakov *et al.*, 2012). PA(+++) aggregates were detected after treatment at 55 °C and reached saturation at 65 °C (**Figure 4.2**). Thus, it was determined that PA(+++) is irreversibly denatured after incubation for 10 min at 65 °C, reaching maximum aggregation.

4.3.2.2 Assay optimisation in 96-well microtiter plate

The growth conditions and induction of protein expression in 96-well plates were established using a Master Plate (MP) containing only *E. coli* BL21 (DE3) expressing PA(+++). A 96-well plate containing 100 µL of 2xTY media with kanamycin was inoculated with a replica of an overnight culture (30 °C, 1050 rpm) of the MP. The plate was incubated at 37 °C, 1050 rpm, and growth was monitored with a plate reader until mid-log phase was reached at OD_{600} 0.5 – 0.6 within the plate (**Figure 4.3a**), showing a coefficient of variation of 11.15% at induction time. Protein expression was induced with 0.1 mM IPTG (final induction concentration) by adding 50 µL of fresh 2xTY media supplemented with kanamycin and IPTG. Expression was performed at 20 °C, 900 rpm for 24 hrs in a thermoblock. Cells were harvested by centrifuge and PA(+++) was found overexpressed in periplasmic extracts after osmotic shock, as shown in SDS-PAGE (**Figure 4.3b)**.

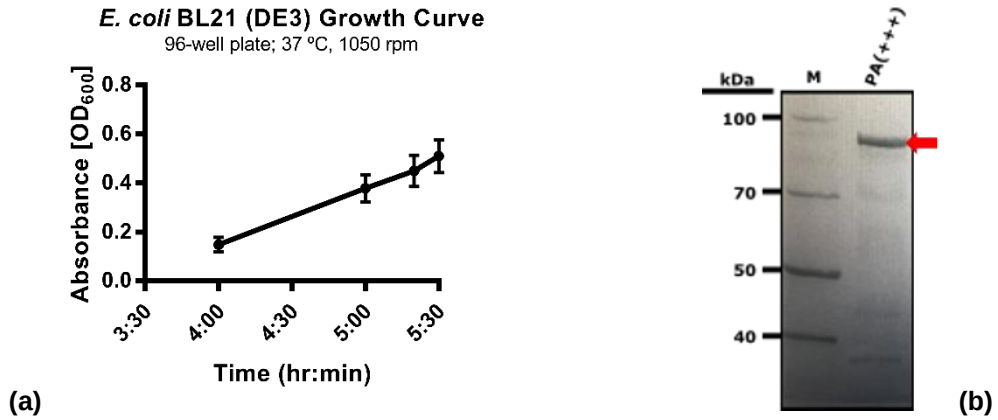


Figure 4.3 Optimisation of protein expression in 96-well plates and recovery of periplasmic extracts by osmotic shock (a) Growth curve of *E. coli* BL21 (DE3) in 96-well plates prior induction of PA(+++) expression. Plate was incubated at 37 °C, 1050 rpm. OD_{600} was measured with a SpectraMax M2e microplate reader. Each point represents the mean OD_{600} within the 96-wells and SD (CV at induction time = 11.15%). **(b)** SDS-PAGE analysis of periplasmic extract containing PA(+++) (red arrow) expressed in 96-well plates. Protein expression was induced with 0.1 mM IPTG (final concentration), cultured at 20 °C, 900 rpm for 24 hrs. Cells were harvested and resuspended in ice-cold water following incubation for 15 min, then ice-cold hypertonic solution (100 mM Tris, 0.5 mM EDTA, 20% (*w*/*v*) sucrose; pH 8.0; supplemented with protease inhibitors) was added and incubated for 15 min. The plate was then centrifuged, and supernatants are periplasmic extracts.

Figure 4.3 Optimisation of protein expression in 96-well plates and recovery of periplasmic extracts by osmotic shock

Another experiment was set to determine the optimal concentration of Nile red for detection of PA(+++) aggregates. PA(+++) was produced in 96-well plate culture as described above. Periplasmic extracts were incubated for 10 min at 65 °C in presence of different Nile red concentrations in clear-bottom 96-well plates. Nile red working solutions were prepared in 50 mM MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0) with 50% (*v*/*v*) ethanol (final ethanol content per sample was 4.5%). A saturation curve was plotted after measuring Nile red fluorescence intensity at 625 nm, 540 nm excitation (**Figure 4.4a**). Nile red reached saturation at 25 µM.

The relationship between initial protein concentration and thermally induced aggregates detected by Nile red fluorescence was also explored. It is well known that aggregation is highly influenced by protein concentration (Morris, Watzky and Finke, 2009; Novo, Freire and Al-Soufi, 2018) and since it is technically difficult to manually adjust all the periplasmic extracts in a 96-well plate to the same initial concentration it would be advantageous to find a way to consider initial protein concentration for estimation of aggregation. We opted for a direct method using a plate reader to measure OD_{280} absorbance of the periplasmic extract before Nile red staining. PA(+++) was expressed in a 96-well plate as described above and the periplasmic extracts were transferred to a 96-well clear-bottom microtiter plate. The extracts were diluted using 50 mM MOPS buffer and topped to 200 µL with DDI water to obtain a range of protein concentrations. The concentration of the diluted extracts was estimated by OD_{280} using a SpectraMax M2e microplate reader and then protein aggregation was induced as described above, by incubation at 65 °C for 10 min, in presence of 25 µM Nile red. The OD_{280} absorbance is reported as instrument reading (without subtracting blanks) in arbitrary units. It is important to mention that the type of clear-bottom microtiter plate used absorbs UV light, thus this is only an approximation of the periplasmic extract concentrations. As expected, the relationship of protein concentration and aggregation (measured indirectly as fluorescence intensity) fits a linear regression and shows a strong-positive correlation (**Figure 4.4b**). This relationship can be used to normalize aggregation/protein concentration during library screening.

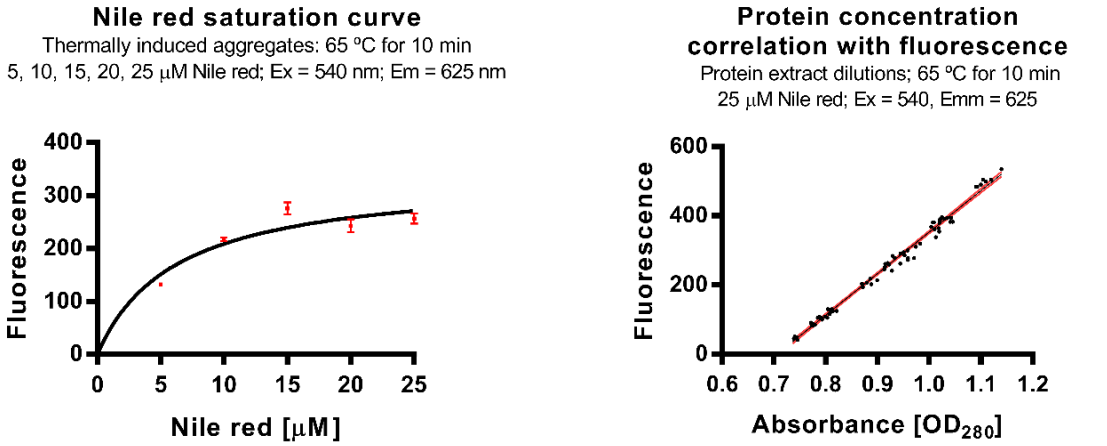


Figure 4.4 Nile red saturation curve and relationship of periplasmic concentration and fluorescence of stained aggregates. **(a)** Nile red saturation curve. Periplasmic extracts were diluted 1:3 with MOPS buffer (50 mM MOPS, 100 mM NaCl; pH 7.0) and incubated at 65 °C for 10 min in presence of different Nile red concentrations. Fluorescence intensity was measured with a SpectraMax M2e microplate reader (excitation = 540 nm; emission = 625 nm), and corresponding blanks were subtracted. Each point represents mean with SEM (n = 7). Curve was fitted using one-site specific binding equation by GraphPad 7.0. **(b)** The relationship between protein concentration (OD_{280}) and aggregation (fluorescence) was established. Periplasmic extracts were diluted to achieve a range of concentrations. Protein concentration was estimated by OD_{280} absorbance using a SpectraMax M2e microplate reader. Thermally induced aggregates were stained with the previously established Nile red optimal concentration (25 μM). OD_{280} and fluorescence are reported as instrument reading (without subtracting blanks) in arbitrary units. The relationship was fitted to a linear regression ($R^2 = 0.991$). The red line represents 95% confidence interval. A strong positive correlation was obtained (r = 0.995). Fits and coefficients were calculated with GraphPad 7.0.

4.3.3 Semi-rational approach library screening

The semi-rational approach library was arrayed in a 96-well plate. Library screening was conducted using the previously established optimal conditions (**Figure 4.5**).

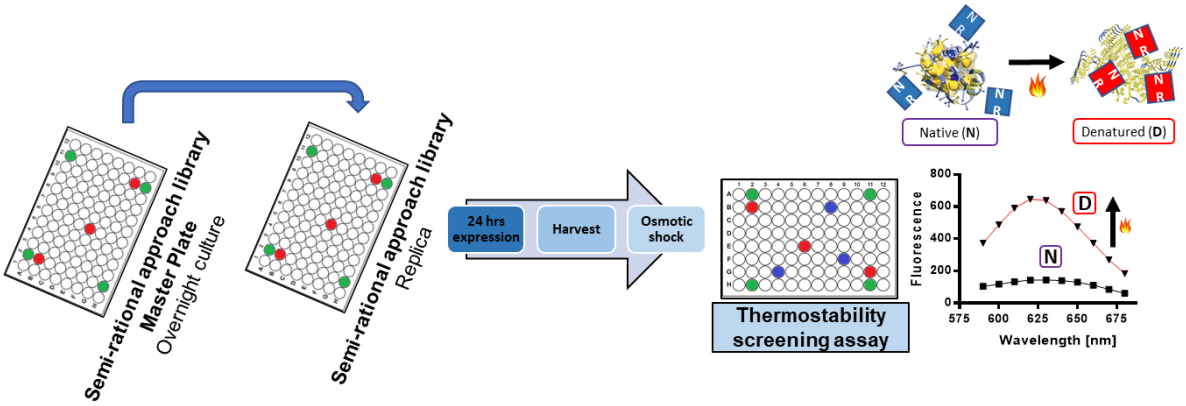


Figure 4.5. Semi-rational approach library screening method. The semi-rational approach library was arrayed in a 96-well plate (**Master Plate**). Each well was inoculated with one colony of *E. coli* BL21 (DE3) harbouring WT PA(+++) or mutant construct. Wells B2, E6 and H11 (red) were inoculated with PA(+++) (WT) as internal control. Wells A2, H2, A11, H11 were inoculated with PA(+++) for mock expression control $^{(-)}$IPTG (green). Wells A1:H1 and A12:H12 were not inoculated. The Master Plate (MP) was incubated overnight at 30 °C, 1050 rpm. Next day, the MP was replicated into fresh 100 µL 2xTY media, incubated at 37 °C, 1050 rpm until OD_{280} reached 0.5 – 0.7 average withing the plate. Protein expression was induced with 0.1 mM IPTG (final concentration) (except for $^{(-)}$IPTG wells) and proteins were expressed at 20 °C, 900 rpm for 24 hrs. Cells were harvested by centrifuge and periplasmic extracts were obtained by osmotic shock. For the thermostability screening assay, periplasmic extracts were transferred into a clear-bottom 96-well plate and diluted 1:1 with 50 mM MOPS buffer. Periplasmic concentrations were estimated with OD_{280} using a SpectraMax M2e microplate reader and then proteins were stained with 25 µM Nile red (final concentration). Fluorescence was measured with a SpectraMax M2e microplate reader ($\mathbf{F_0}$ (**N**)). Then the plate was sealed with a lid and incubated at 65 °C for 10 min. After cooling-down, fluorescence was measured to detect thermally induced aggregates ($\mathbf{F_{agg}}$ (**D**)) (Fluorescence measurements: spectra 590 – 680 nm, step 10 nm, after excitation with 540 nm; emission at 625 nm after excitation with 540 nm).

The screening results are presented as relative stability (**Figure 4.6a**), as calculated in equation 4.2 (**Eq. 4.2.**), where F_o is fluorescence of stained proteins before heat treatment (excitation = 540 nm; emission = 625 nm), F_{agg} is fluorescence after heat treatment (i.e. aggregates fluorescence) (excitation = 540 nm; emission = 625 nm), OD_{280} is absorbance. It reports the change of fluorescence upon heating ($\frac{F_{agg}}{F_o}$), adjusted by the relation aggregates/protein concentration ($\frac{F_{agg}}{OD_{280}}$) established in **Figure 4.4b**, of a variant relative to the wild-type (*WT*) (i.e. PA(+++)). The mean from $\frac{F_{agg}}{F_o}$ and $\frac{F_{agg}}{OD_{280}}$ corresponding to wells B2, E6 and H11 are $WT\,\frac{F_{agg}}{F_o}$ and $WT\,\frac{F_{agg}}{OD_{280}}$, respectively. A positive variant threshold was established at -1SD from the WT mean.

Eq. 4.2. $$Relative\ stability = \frac{variant\frac{F_{agg}}{F_o}}{WT\frac{F_{agg}}{F_o}} \times \frac{variant\frac{F_{agg}}{OD_{280}}}{WT\frac{F_{agg}}{OD_{280}}}$$

Positive variants with mutations in positions T81, L340 and V644 were picked for validation. These variants were expressed in 50 mL flask culture, recovered by osmotic shock and the thermostability assay was performed as described for library screening. SDS-PAGE analysis showed that variant V644 was not present in the periplasmic extract (**Figure 4.6b**). Moreover, emission spectrum of variant V644 prior library screening (**Figure 4.6a**) shows a red-shift related to the WT, and displays an emission maxima similar to that of the $^{(-)}$IPTG control (~ 655 nm). It was later confirmed by sequencing that there was a frame shift mutation in the V644 construct. The validation assay on the other two variants showed that position L340 had a neutral effect related to PA(+++), whereas T81 showed improved thermostability (**Figure 4.6b**). Sequencing confirmed mutation T81I. Thus, variant PA(+++)-T81I was selected for characterisation.

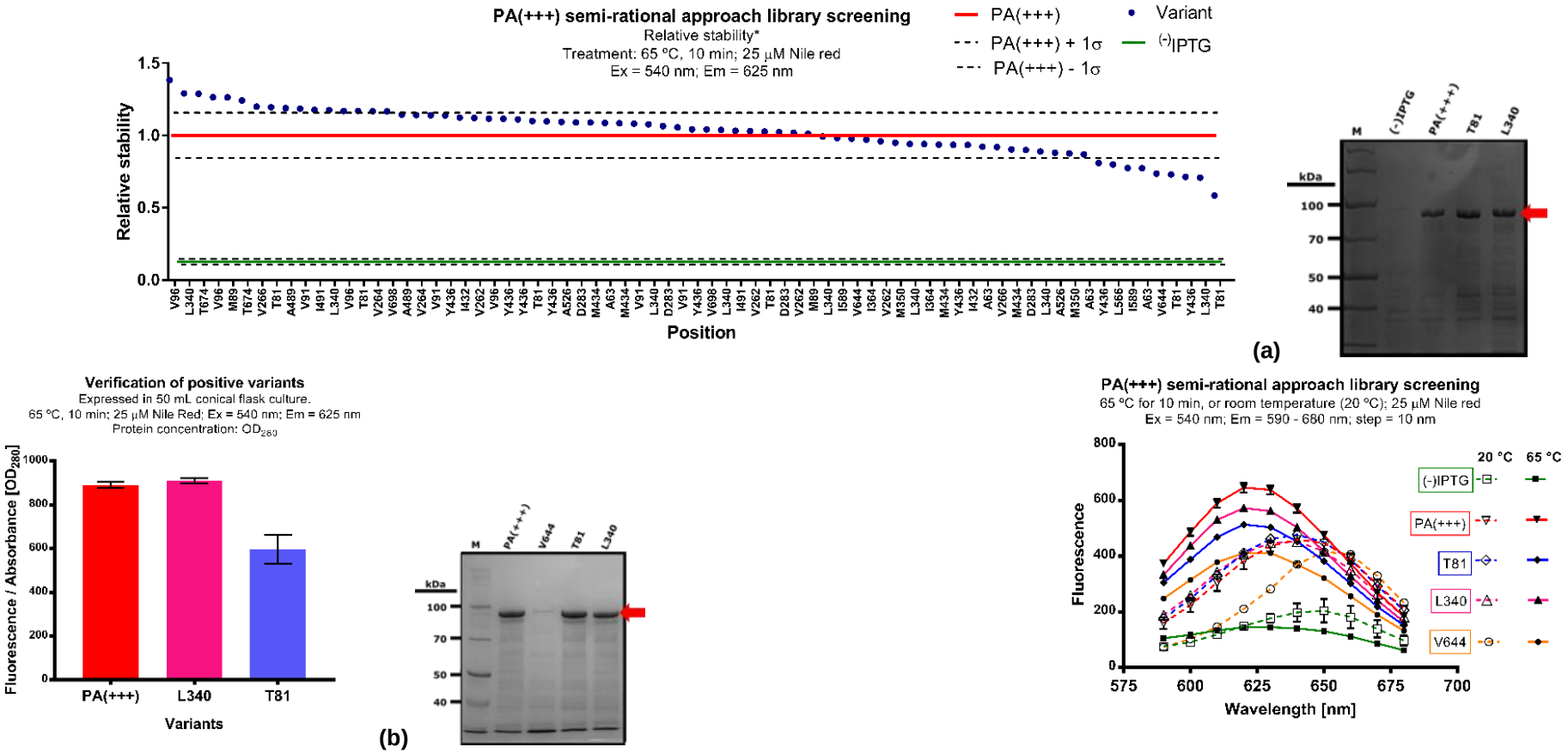


Figure 4.6 Semi-rational approach library screening. (a) Screening results are shown as relative stability, as calculated by ***Eq. 4.2**. The red solid line is denatured PA(+++). Green solid line is $^{(-)}$IPTG, negative control; black dotted lines are +1SD and -1SD from the mean. Variants (blue circles) with a relative stability below -1SD of the WT mean are considered hits. Top right: SDS-PAGE analysis of periplasmic extracts. The red arrow indicates WT and variants. Bottom right: emission spectra (540 nm excitation) before heat treatment (dotted line, open shapes) and after heat treatment at 65 °C (solid line, solid shapes), with the following colour code: $^{(-)}$IPTG: green; PA(+++): red; T81: blue; L340: pink; V644: orange. $^{(-)}$IPTG and PA(+++) data points represent mean with SEM (n = 3). Spectra reported as instrument reading (without subtracting blanks). **(b)** Validation assay of positive variants using periplasmic extracts from 50 mL expression cultures and SDS-PAGE analysis. Periplasmic extracts were transferred into a clear-bottom 96-well plate and OD_{280} was measured with a SpectraMax M2e microplate reader. Aggregation was induced in presence of Nile red and their fluorescence was normalised with OD_{280} using instrument readings (without subtracting blanks). Bars showing mean with SEM (n = 6).

4.3.4 Rational approach library screening

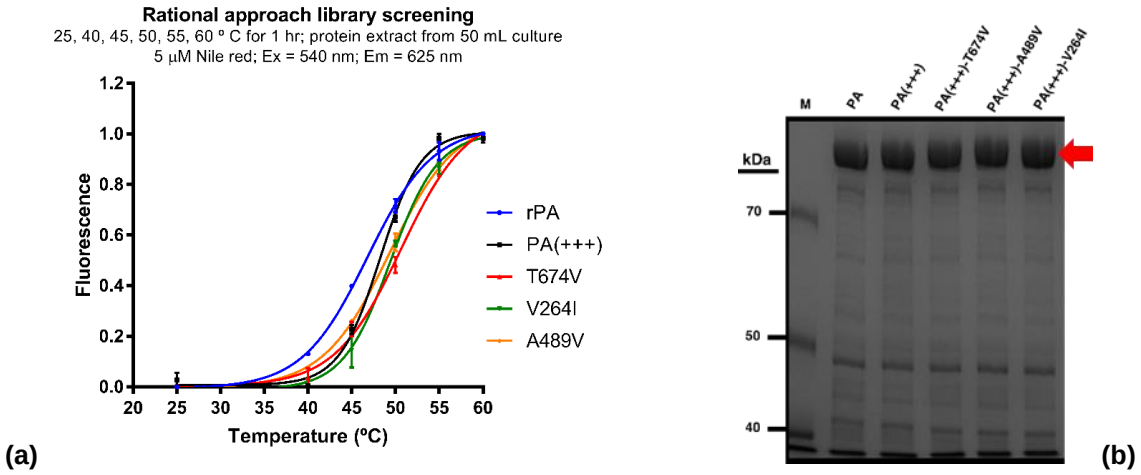


Figure 4.7 Rational approach library screening. (a) The concentration of periplasmic extracts, recovered from 50 mL expression culture, was adjusted to OD_{280} 0.9 using 50 mM MOPS buffer with cuvettes and a BioPhotometer Plus UV/Vis photometer. Periplasmic extracts were incubated at 25, 40, 45, 50, 55 and 60 °C for 1 hr. Proteins were then transferred into a clear-bottom 96-well plate and stained with 5 µM Nile red (final concentration). Data points showing mean and SEM (n = 3). Curves were fitted using the Boltzmann sigmoid equation and Normalize by GraphPad 7.0. **(b)** SDS-PAGE analysis of periplasmic extracts after OD_{280} was adjusted to 0.9. Red arrow: rPA and variants.

Figure 4.7 Rational approach library screening

The rational approach library was screened for thermostability using periplasmic extracts obtained from 50 mL cultures in conical flasks. The original rPA, PA(+++) and variants PA(+++)-T674V, PA(+++)-V264I and PA(+++)-A489V were expressed in 50 mL and periplasmic extracts were obtained by osmotic shock as described in **Chapter 3**. The concentration of the periplasmic extracts was adjusted to 0.9 OD_{280} with 50 mM MOPS buffer using a spectrophotometer. Then, protein extracts in microcentrifuge tubes were incubated at 20, 40, 45, 50, 55 and 60 °C for 1 hr, and immediately cool down in ice. Aggregates were thoroughly mixed and transferred into a 96-well clear-bottom microtiter plate. Aggregates

were then stained with 5 μM Nile red (final concentration) and fluorescence was recorded at 625 nm, 540 nm excitation.

Aggregation is highly dependent on protein concentration. Even though the initial concentration of all extracts was adjusted to an OD_{280} of 0.9 (measured using a UV cuvette), the content of impurity proteins may vary and change the absorption coefficient of the different extracts. Yet, SDS-PAGE analysis showed that protein background does not vary appreciably among periplasmic extracts (**Figure 4.7b**). As shown in **Figure 4.7a**, aggregation measured this way fits a sigmoid curve. It would be misleading to model kinetical parameters based on this fit because there are other undefined proteins present in the periplasmic extract that may influence thermal denaturation of rPA (and variants) but one can point out some qualitative statements. As suggested by the sigmoidal fit, protein aggregation occurs after a lag phase, until a nucleus is formed to trigger a growth phase of aggregation whereby soluble monomers are promptly incorporated into the insoluble aggregate-mass until maximum aggregation is reached (plateau). The lag phase would involve temperature-dependent destabilisation of the native state until partially unfolded intermediates forming intermolecular interactions trigger the aggregation reaction. Then, during growth phase, equilibrium shifts towards aggregate formation whereby partially unfolded monomers are more likely to be incorporated into aggregates than returning to their native conformation. This process is typically observed during *in vitro* formation of amyloid fibrils and other types of aggregates (Meisl *et al.*, 2016). Since this process accelerates as temperature increases, the differences observed on aggregation propensity can be directly related to protein thermostability, provided protein aggregation is irreversible under these conditions. In consequence, variants PA(+++)-T674V, PA(+++)-V264I and PA(+++)-

A489V showed improved thermostability related to PA(+++) and were selected for characterisation.

4.4 Conclusion

Semi-rational and rational approach protein engineering aiming to improve the thermostability of PA(+++), a protease- aggregation-resistant rPA variant designed in **Chapter 3**, was performed. Libraries were generated by site-directed mutagenesis using PA(+++) as parental template. A semi-rational approach library was designed in collaboration with Dr. Antonina Andreeva (MRC Laboratory of Molecular Biology, Cambridge) using structure-based analysis of a high-resolution PA X-ray crystal structure (PDB: 3TEW) to design mutations aiming to improve hydrophobic core-packing. The stability of these mutations was predicted *in silico* using FoldX. Mutations with predicted $\Delta\Delta G < -0.5$ kcal/mol were considered beneficial and were used to construct a small rational approach library.

A semi-high throughput method was developed for thermostability screening of the semi-rational approach library. The assay was based on indirect evaluation of thermostability by estimation of aggregation propensity after subjecting the mutant library to the predetermined aggregation temperature of PA(+++) and then detecting thermally-induced aggregates using the polarity-sensitive fluorescent dye Nile red. The conditions for the screen were optimised to determine the optimal Nile red concentration (25 µM), the aggregation temperature of PA(+++) in periplasmic extracts (65 °C) as well as the optimal culturing conditions for library expression and recovery in 96-well plate format. The semi-rational approach library was arrayed and expressed in a 96-well plate and variants were recovered in periplasmic extracts using a miniaturised adaptation of the osmotic shock method

established in **Chapter 3**. After screening, three variants resulted positive, however only the variant PA(+++)-T81I showed improved thermostability in the validation assay. The major limitation of this screening method is the difficulty to accurately estimate initial protein concentration of the periplasmic extracts. It was decided to use OD_{280} to estimate protein concentration because it is a non-invasive technique that enables direct measurement of the extracts in the assay plate prior screening. However, the presence of impurity proteins could overestimate the concentration of the protein variants as well as to cause variation among different extracts due to inherent variation in the extraction efficiency for each expressed variant. Moreover, the screen was conducted in polystyrene clear-bottom plates that absorb UV light. We intended to validate this approach by estimating the correlation between initial protein concentration, estimated by OD_{280}, and aggregate formation, estimated by Nile red fluorescence, using a range of protein extract dilutions, resulting in a strong correlation ($r = 0.995$). Using this relationship, aggregates fluorescence was normalised against initial protein concentration for relative stability estimations (as described in **Eq. 4.2**). This semi-high throughput method can be improved using UV-plates for a more accurate estimation of protein concentration.

Mutations V264I, A489V and T674V were predicted by FoldX to improve conformational stability. Variants PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V, as well as PA(+++) and original rPA, were screened for thermostability using a similar method for detection of thermally-induced, except that periplasmic extracts were recovered from 50 mL expression cultures and were incubated under a range of temperatures (25, 40, 45, 50, 55 and 60 °C). For this screening method, the initial protein concentration of all

periplasmic extracts was adjusted to the same OD_{280}. The three new variants showed improved thermostability related to both PA(+++) and rPA.

By comparing the results of both approaches, the question of which one is more effective can be raised. The positions selected by structure-based design for improvement of hydrophobic core-packing were used for the construction of the semi-rational approach library. The same mutations were then evaluated with FoldX to generate the rational approach library. So, essentially both libraries were guided by the same principle of structure-based analysis seeking to improve hydrophobic core-packing. FoldX only reduced the library size by predicting which mutations were beneficial according to its energy function. The semi-high throughput screening assay uses PA(+++) aggregation temperature as a threshold for positive hits, which is arguably a condition too stringent as most of the variants were found above or near the threshold. In fact, mutations predicted by FoldX to be stabilising were overlooked by the semi-high throughput screening. Moreover, the only variant that resulted positive from screening of the semi-rational approach library, PA(+++)-T81I, was found highly destabilising by FoldX ($\Delta\Delta G$ = 5.9 kcal/mol). Protein thermostability is a complex phenomenon and the effect of mutations can only be confirmed after characterisation, but we believe that the combination of both approaches can be complementary as variants can access higher thermostability by different mechanisms. This will be discussed in the next chapter.

Chapter 5. Biophysical characterisation of thermostable variants of recombinant Protective Antigen

5.0 Abstract

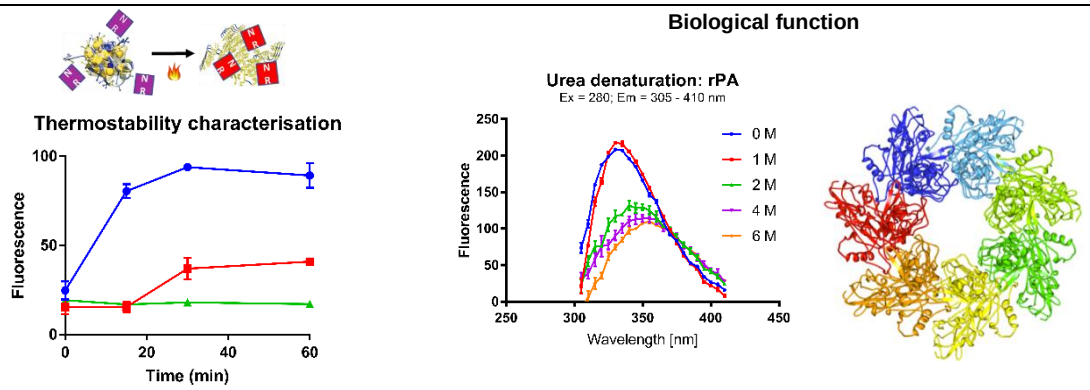


Figure 5.0. Graphical abstract. Biophysical characterisation of rPa and thermostable variants.

Recombinant vaccines must elicit the same correlates of protection as antigens during natural infection in order to ensure effectiveness. Protective Antigen (PA) elicits production of neutralising antibodies that protect against Anthrax disease, thus next-generation vaccines are developed based on recombinant production of PA (rPA). rPA correlates of protection are tightly related to its biological function and, hence, conformational structure. In **Chapter 4**, we designed rPA variants with improved thermostability as candidates for the development of a thermostable next-generation Anthrax vaccine. Variants PA(+++)-T81I, PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V resulted positive for improved thermostability after screening and were characterised in this Chapter. Characterisation of rPA, PA(+++), PA(+++)-T81I, PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V was performed to analyse if mutations altered the physicochemical properties and biological function of rPA. Thermostability was indirectly characterised using the polarity sensitive dye

Nile red to investigate the conformational state of the variants after heat treatment. Structural studies were performed using intrinsic tryptophan fluorescence and protein unfolding was monitored after urea denaturation. Biological function was characterised by limited proteolysis with trypsin and by detecting the formation of a naturally occurring oligomer structure called the pre-pore using native-PAGE. It was found that variants PA(+++)-T81I and PA(+++)-T674V exhibit improved thermostability related to the other variants. Also, all proteins showed a biphasic unfolding transition during urea denaturation assays, in agreement with the literature, evidence that rPA and the engineered variants can bind the Ca^{2+} ions essential for biological function, indicating conservation of PA native folding. Limited proteolysis showed that all variants can be cleaved in a furin-like site in domain 1 essential for oligomerisation. Interestingly, PA(+++)-T81I showed signals of proteolysis resistance suggesting that mutation T81I increased the rigidity of domain 1. Pre-pore formation was induced in rPA, PA(+++), PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V following by its detection by native-PAGE. The native-PAGE also revealed that native rPA and engineered variants are present in three distinct isoforms, described in the literature as signals of deamidation. Further characterisation of presumed deamidated proteins and their potential effect on biological function and vaccine efficacy should be conducted in the future. Additional future work includes the elucidation of thermodynamic parameters of the thermostable variants using differential scanning calorimetry and spectroscopic techniques.

5.1 Introduction

Next-generation recombinant Anthrax vaccine candidates based on rPA are effective at generating correlates of protection, however issues regarding their long-term stability have

halted regulatory approval (Baillie, 2009; Kaur, Singh and Bhatnagar, 2013). The hypothesis of this thesis is that rPA stability can be improved by protein engineering approach aiming to obtain a thermostable variant with improved hydrophobic core-packing. On **Chapter 4**, semi-rational and rational approach protein engineering was performed on PA(+++), a protease- aggregation-resistant variant (see **Chapter 3**), to generate two mutant libraries. The variant PA(+++)-T81I resulted positive after screening the semi-rational approach library, as well as variants PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V from the rational approach library. In this Chapter, characterisation of these thermostable variants was conducted.

Engineered recombinant vaccines must generate the same correlates of protection as of the original antigen during natural infection in order to ensure effectiveness. Recombinant vaccines are often designed to generate neutralising antibodies against a recombinant protein that mimics an antigen during natural infection to elicit protection against disease. In the case of Anthrax, protection against disease is achieved by antibodies that neutralise PA functionality (Ngundi *et al.*, 2010). Therefore, an effective next-generation recombinant vaccine candidate must display the same biological function as original PA. PA is an 83 kDa protein comprising 4 domains whose activity are central to the biological function of the Anthrax toxin, functioning as delivery system of the effector proteins into the cell (Fabre *et al.*, 2016). Domain 4 is involved in receptor binding. Domain 1 binds a pair of Ca^{2+} ions and contains a furin-like proteolytic site ($_{164}$RKKR$_{167}$). PA is cleaved *in vivo* by furin-like proteases after binding its receptors on the cell surface, as well as *in vitro* by trypsin, generating a C-terminal 63 kDa protein (PA63) that serves as binding site for the toxin effector proteins Lethal Factor (LF) and Oedema Factor (EF). The Ca^{2+} ions remain bound

in PA63 and are believed to further stabilise this conformation (Gupta *et al.*, 2003). After cleavage, domain 3 mediates the formation of PA63 oligomers comprised of 7 protomers (heptamer). This complex exhibits a ring shape that is known as the pre-pore (Bann, 2012). The complex is then endocytosed and in the mid-endosome the pre-pore suffers a pH-mediated conformational rearrangement whereby domain 2 from each protomer contribute to the formation a transmembrane pore that serves as translocase for LF and EF (Jiang *et al.*, 2015).

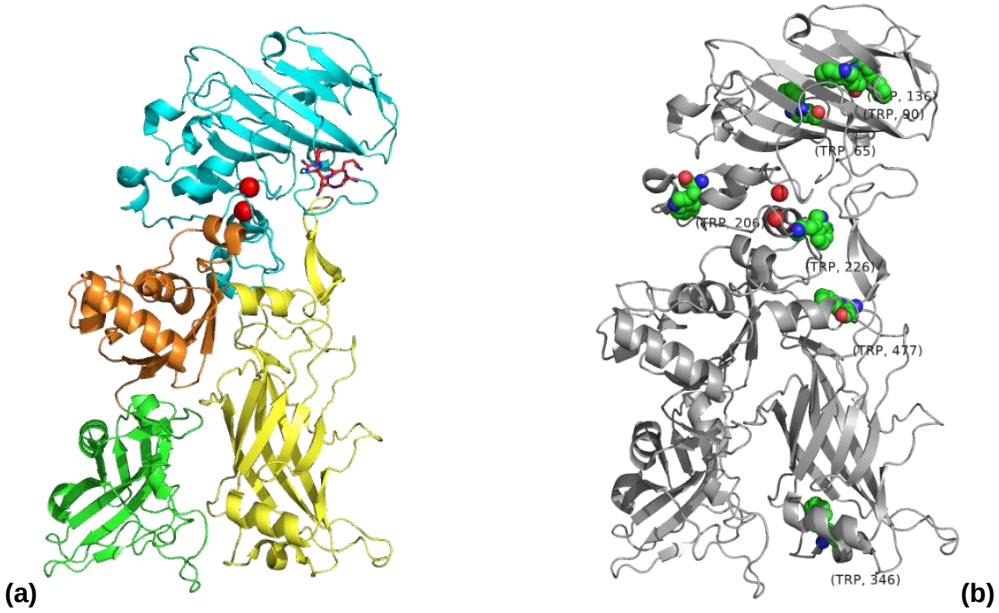


Figure 5.1. PA X-ray crystal structure (PDB: 3TEW). (a) PA structure coloured by Domains: Domain 1 is in cyan; Domain 2 is in yellow; Domain 3 is in orange; Domain 4 is in green. Red spheres in Domain 1 are Ca^{2+} ions; Red sticks in Domain 1 represent the furin-like proteolytic site ($_{164}$RKKR$_{167}$). **(b)** PA structure displaying its 7 tryptophan residues (green spheres) and the 2 Ca^{2+} ions (red spheres).

Figure 5.1 PA X-ray crystal structure (PDB: 3TEW)

Bioanalytical techniques were applied to characterise the thermostability, structure and biological function of our engineered rPA variants. A successful candidate for the development of a thermostable next-generation Anthrax vaccine should demonstrate

improved thermostability while maintaining biological function. As biological function is tethered to structure, we took advantage of the 7 tryptophan residues contained in rPA to analyse the variants' tertiary structure by intrinsic tryptophan fluorescence and characterised protein unfolding by urea denaturation. Thermostability was characterised indirectly by detecting thermal denaturation using the polarity sensitive dye Nile red. Biological function was further analysed by limited proteolysis with trypsin and by evaluating oligomerisation using native-PAGE.

5.2 Aims

- To characterise the variants obtained by semi-rational and rational approach protein engineering using bioanalytical techniques and computational tools.
- To characterise the thermal stability of the variants using the polarity-sensitive fluorescence dye Nile red to detect thermal denaturation.
- To analyse the tertiary conformation and unfolding of the proteins using intrinsic tryptophan fluorescence spectroscopy following urea-denaturation.
- To analyse the biological function of the variants by limited proteolysis with trypsin and investigate oligomerisation using native-PAGE.

5.3 Results and discussion

5.3.1 Prediction of interatomic interactions using computational methods: engineered variants present improved hydrophobic interactions.

Arpeggio webserver (Jubb *et al.*, 2017) was used to calculate the changes in local interatomic interactions produced by mutated residues in variants PA(+++)-T81I, PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V (**Figure 5.2**). The X-ray crystal structure 3TEW was

repaired using FoldX RepairPDB function and mutant models were then created using the BuildModel function. The generated PDB files were cleaned using PyRosetta function clean.atom (Chaudhury, Lyskov and Gray, 2010). Arpeggio calculates interatomic interactions based on atom type, distance and angles, using the rules set by the database CREDO (Schreyer and Blundell, 2013). Visualising the effect of mutations on interatomic interactions can provide valuable insight about the engineered variants. As shown in **Figure 5.2**, the mutant variants display improved hydrophobic networks.

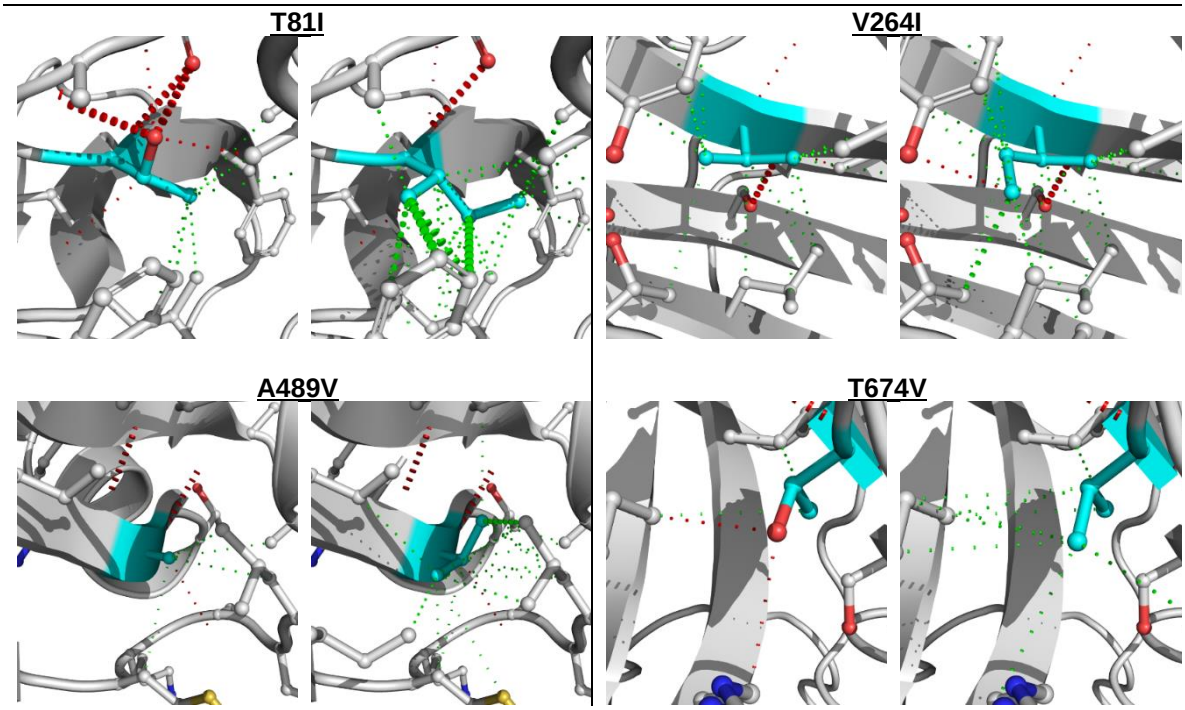


Figure 5.2 Mutation effect on interatomic interactions. The local interatomic atomic interactions involving mutated positions and neighbour residues (ball and stick, showing only side chains) were calculated using the Arpeggio web server using the 3TEW crystal structure and models generated by FoldX. On the left: original PA structure; on the right: engineered variant model. In cyan: residue of interest. For interatomic interactions (dotted lines), reds are polar interactions and hydrogen bonds, greens are hydrophobic interactions. Thicker green dotted lines indicate van der Waals clashes suggesting stronger hydrophobic interactions. Thicker red dotted lines indicate hydrogen bond fulfilling geometrical rules (Schreyer and Blundell, 2013).

5.3.2 Thermostability characterisation using the extrinsic fluorescent dye Nile red

Extrinsic fluorescent dyes are extensively used to study protein structure (Hawe, Sutter and Jiskoot, 2008). Nile red is a polarity sensitive dye whose fluorescence quantum yield increases in apolar environments accompanied by a blue shift in its emission maxima (Sackett and Wolff, 1987). Nile red can bind the exposed hydrophobic regions of proteins and has been used successfully to detect and characterise distinct types of aggregates (Mishra, Sjölander and Hammarström, 2011). We characterised the thermostability of our engineered variants using Nile red to investigate the state of the variants after heat treatment. The assay was optimised using PA(+++). All fluorescence measurements were performed in 96-well plates using SpectraMax M2e microplate reader.

5.3.2.1 Optimisation of thermostability assay: investigating aggregate structure with Thioflavin T

Nile red has poor solubility in water, and it is often prepared using ethanol or other organic solvents (Greenspan and Fowler, 1985). It is thus important to establish the optimal Nile red concentration and the effect that ethanol may have on protein aggregates. In order to determine the effect of ethanol on Nile red fluorescence when bound to aggregates, we prepared PA(+++) aggregates by incubating 120 µg/mL of pure protein at 94 ºC for 5 min. Then, aggregates were stained with Nile red to a final concentration of 10 µM Nile red prepared in different working solutions containing ethanol, obtaining stained aggregates in presence of 10%, 5% and 1 % (*v*/*v*) ethanol, as well as with 1 µM Nile red and 1% (*v*/*v*) ethanol (**Figure 5.3a**). As shown in **Figure 5.3a**, ethanol does show an effect on Nile red fluorescence, as higher ethanol concentration translates into stronger fluorescence intensity.

In order to investigate whether Nile red fluorescence increases in presence of ethanol by somehow improving its quantum yield or due to the disturbance of protein aggregate structure caused by ethanol, Thioflavin T was used to characterise PA(+++) aggregates. Thioflavin T is an amyloid-specific dye whose fluorescence increases when bound to amyloid-like fibrils (Hawe, Sutter and Jiskoot, 2008). These aggregate species have a well-defined structure and are typically formed during late stages of the aggregation process (Roberts, 2007). So, Thioflavin T would not be able to bind PA(+++) aggregates if ethanol is disturbing their structure, provided that PA(+++) aggregates form amyloid fibrils. PA(+++) aggregates were stained with Thioflavin T in absence and presence (10 % (*v*/*v*)) of ethanol (**Figure 5.3b**) and fluorescence was measured. It was concluded that ethanol is indeed disturbing aggregate structure as Thioflavin T binds aggregates in ethanol-free suspensions but not in presence of ethanol suggesting that ethanol disrupts aggregate structure promoting formation of amorphous aggregates, being the reason behind the stronger fluorescence of Nile red bound to aggregates in presence of ethanol. Moreover, as shown in **Figure 5.3a**, Nile red emission maximum is also affected by ethanol concentration, as Nile red in presence of 10 % (*v*/*v*) ethanol has an emission maximum shift from 655 nm when bound to native protein (i.e. without heat treatment) to ~ 650 nm when bound to aggregates, and shifts to ~ 650 nm and ~ 645 nm in presence of 5% and 1% (*v*/*v*) ethanol, respectively. The blue shift in wavelength maximum is dependent on hydrophobicity (Mishra, Sjölander and Hammarström, 2011), thus one can argue that ethanol promotes the formation of structures with reduced aggregate area.

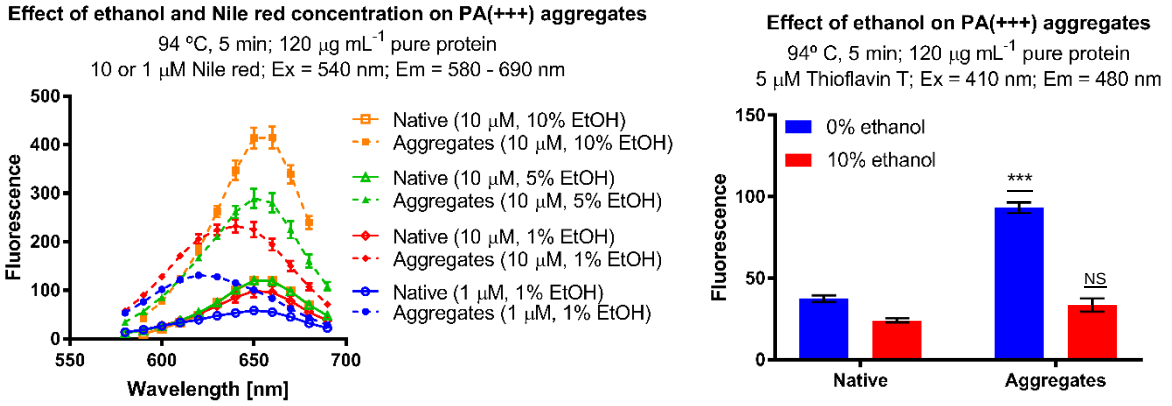


Figure 5.3. Effect of Nile red concentration and ethanol on PA(+++) aggregates. (a) Aggregates were thermally produced using 120 µg/mL of pure PA(+++) after incubation at 94 °C for 5 min. The suspension was mixed thoroughly. Aggregates and native PA(+++) (i.e. PA(+++) without heat treatment) were transferred into a clear-bottom 96-well plate. Aggregates (dotted lines, solid shapes) and native PA(+++) (solid lines, open shapes) were stained to final concentrations of: 10 µM Nile red, 10% ethanol (orange); 10 µM Nile red, 5% ethanol (green); 10 µM Nile red, 1% ethanol (red); 1 µM Nile red, 1% ethanol (blue). Emission spectra were recorded with a SpectraMax M2e microplate reader at 580 – 690 nm, step 10, using a 570 nm filter, after excitation with 540 nm. All measurements were corrected with a corresponding blank for background fluorescence. Data points show mean with SEM (n = 3). **(b)** Aggregates were prepared as described in (a). Native PA(+++) and aggregates were stained with 5 µM Thioflavin T in absence and presence of 10% ethanol in a clear-bottom 96-well plate. Fluorescence was recorded after excitation with 410 nm, 480 nm emission. Columns showing mean with SEM (n = 3). Treatment effect was evaluated using unpaired Student's *t* test (alpha = 0.05), (*** is P value < 0.001; NS = not significant) with GraphPad Prism 7.0.

Figure 5.3. Effect of Nile red concentration and ethanol on PA(+++) aggregates

Interestingly, only aggregates stained with 1 µM Nile red (1 % (*v*/*v*) ethanol) showed a considerable blue shift (**Figure 5.3a**), as emission maximum shifts from ~ 650 nm (native state) to ~ 625 nm. The blue shift and emission maxima are in agreement with another study involving the characterisation of rPA aggregates with Nile red (Soliakov *et al.*, 2012). The incomplete shift under the other conditions (i.e. 10 µM Nile red) may be due to free/unbound dye. Although 10 µM Nile red (1% (*v*/*v*) ethanol) showed better resolution between

aggregates and native state emission spectra than that of 1 µM Nile red (1% (*v*/*v*) ethanol), we decided to use the latter because spectrum shift is more informative for hydrophobicity.

One may consider the use of Thioflavin T instead of Nile red for thermostability characterisation assay, however there is evidence suggesting the formation of a stable partially-unfolded intermediate during thermal denaturation of rPA (Ganesan, Watkinson and Moore, 2012). Thioflavin T binds specifically to amyloid fibrils, thus it would overlook any other partially unfolded or irreversible soluble aggregate species that may be forming after heat treatment (Krebs, Bromley and Donald, 2005; Roberts, 2007; Mishra, Sjölander and Hammarström, 2011). Therefore, we decided to use 1 µM Nile red (1% (*v*/*v*) ethanol) for thermostability characterisation of our engineered variants.

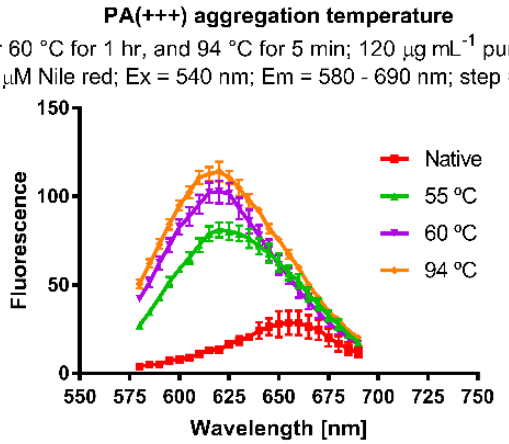


Figure 5.4. PA(+++) aggregation temperature. Pure PA(+++) solutions (120 µg/ml) were incubated at 55 (green) and 60 °C (purple) for 60 min, and at 94 °C for 5 min (orange). Native PA(+++) (i.e. without heat treatment) and aggregates were transferred into a clear-bottom 96-well plate and then stained with 1 µM Nile red, final concentration. Emission spectra were recorded with a SpectraMax M2e microplate reader at 580 – 690 nm, step 5, using a 570 nm filter, after excitation with 540 nm. All measurements were corrected with a corresponding blank for background fluorescence. Data points show mean with SEM (n = 3).

The optimal temperature for the assay was then established. Pure PA(+++) was incubated at 55 and 60 °C for 1 hr and denaturation investigated with 1 µM Nile red (**Figure 5.4**). PA(+++) denaturation showed completion after 1 hr at 60 °C, as Nile red spectrum is almost identical to that of aggregates produced at 94 °C (after 5 min incubation). However, it was decided to use 55 °C because it is in proximity with the experimentally established aggregation temperature of rPA (Ganesan *et al.*, 2016), also Nile red spectrum showed the typical blue shift to emission maximum at 625 nm.

5.3.2.2 Thermostability characterisation: variants PA(+++)-T674V and PA(+++)-T81I showed improved thermostability.

The thermostability of rPA, PA(+++), PA(+++)-T81I, PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V was characterised using the Nile red assay described above. Pure protein solutions containing 60 µg/mL were incubated at 55 °C for 15, 30 and 60 min and then stained with 1 µM Nile red. Results are shown in **Figure 5.5** as fluorescence intensity at 625 nm and emission maximum (barycentric mean). Emission maxima were calculated using **Eq. 5.1**, where λ_{bcm} is barycentric mean; λ is wavelength; F_λ is fluorescence at wavelength λ; *m* is 580 nm; *n* is 690 nm.

Eq. 5.1 $$\lambda_{bcm} = \frac{\sum_{\lambda=m}^{n} F_\lambda \lambda}{\sum_{\lambda=m}^{n} F_\lambda}$$

Barycentric mean was used to estimate protein denaturation. Blue shift in Nile red spectrum was interpreted as the degree of denaturation of the protein caused by heat treatment, that can be Nile red binding solvent-exposed hydrophobic moieties of partially unfolded protein and aggregates. Full protein denaturation was established at 630 nm barycentric mean, that is the emission maximum of Nile red bound to PA(+++) aggregates shown during assay

optimisation and the emission maximum of Nile red bound rPA aggregates reported in the literature (Soliakov *et al.*, 2012)**.** From rPA emission spectra (**Figure 5.5**) we can clearly observe the emission maximum shift from native protein (i.e. without treatment) at ~ 650 nm (649.65 ± 8.84 nm) to fully denatured protein at 630 nm (633.42 ± 0.87 nm) accompanied by increase in fluorescence intensity. Comparing rPA with the other variants, it can be observed that in both native and denatured protein fluorescence intensity is stronger. This might be caused by proteolytic cleavage of rPA (as discussed in **Chapter 3**) and subsequent formation of aggregates with different structure related to the other variants. After 15 min, rPA was fully denatured, suggested by spectral shift to 634.12 ± 0.44 nm accompanied by an increment of fluorescence intensity. PA(+++) showed full denaturation after 30 min (barycentric mean = 628.46 ± 4.74 nm). Analysis of fluorescence data from engineered variants showed improved thermal stability related to PA(+++) and rPA. After 60 min at 55 °C, PA(+++)-A489V was fully denatured (barycentric mean = 633.54 ± 0.52 nm). After 60 min at 55 °C, PA(+++)-T81I and PA(+++)-V264I barycentric mean was 640.38 ± 2.32 and 639.62 ± 0.42 nm, respectively. Variant PA(+++)-T674V did not show signs of denaturation as barycentric mean after 60 min (649.04 ± 0.91 nm) was the same as native protein (649.84 ± 0.20 nm).

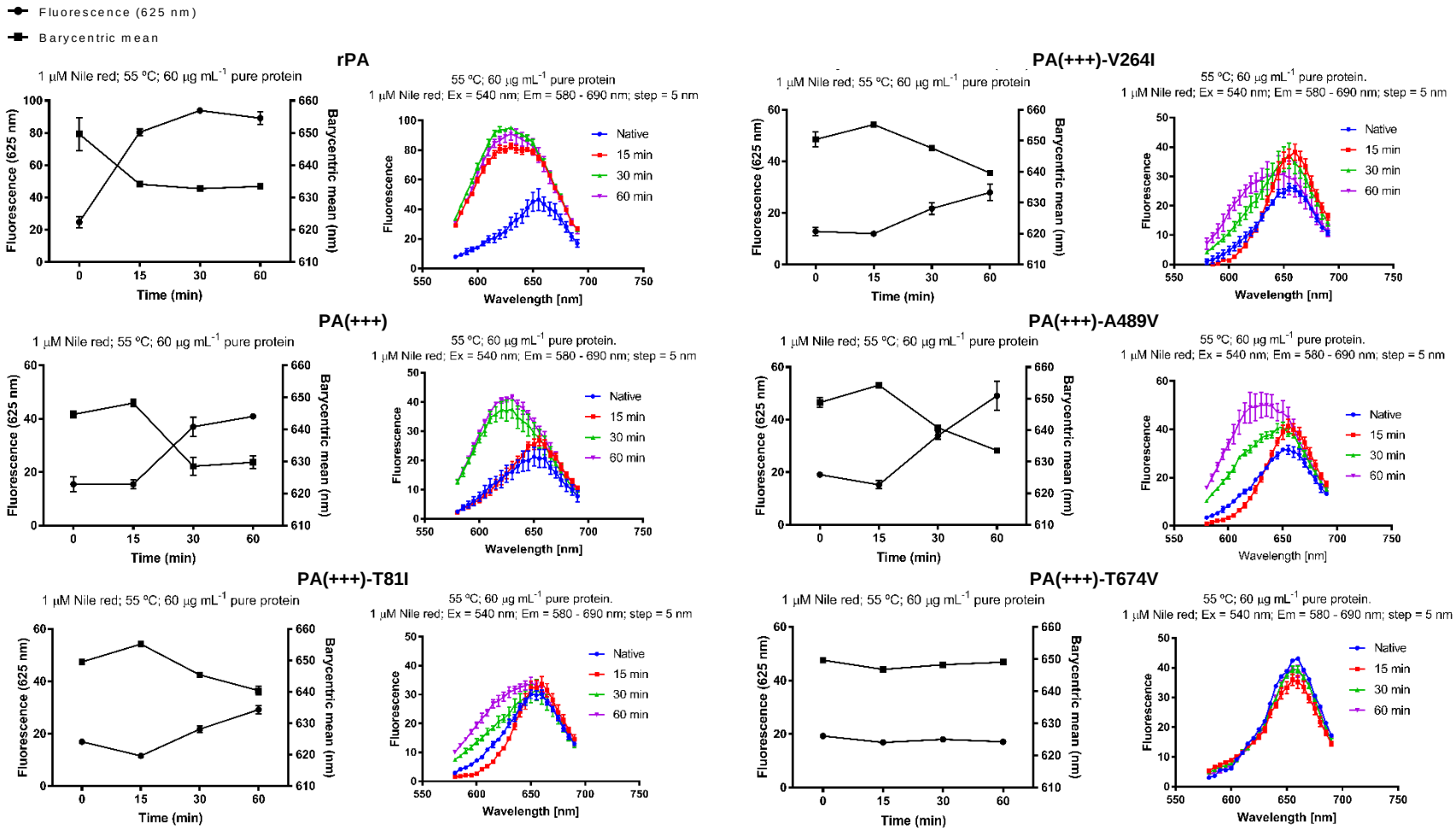


Figure 5.5. Thermostability characterisation. On the right: Nile red emission spectra bound to native (i.e. without heat treatment) and thermally stressed proteins. On the left: barycentric mean (squares) and fluorescence intensity (circles). Emission spectra were recorded at 580 – 690 nm, step 5 nm, using a 570 nm filter, after 540 nm excitation using a SpectraMax M2e microplate reader. All measurements were corrected with corresponding blank for background fluorescence. Data points show mean with SEM (n = 3).

Variant PA(+++)-T674V is clearly the most thermostable variant. However, it is noteworthy that Nile red bound to native PA(+++)-T674V showed stronger fluorescence intensity at emission maxima (~ 650 nm) related to the other variants (except rPA, as discussed above). Examination of the X-ray crystal structure (PDB: 3TEW) shows that residue T674 can form a hydrogen bond with a molecule of water. Using the webserver HotSpot Wizard 3.0 (Sumbalova *et al.*, 2018) to calculate structural pockets applying its integrated tool Fpocket (Le Guilloux, Schmidtke and Tuffery, 2009), it was found that T674 is part of a solvent-accessible pocket in domain 4 (**Figure 5.6**). Mutation T674V is expected to increase the hydrophobicity of this pocket, thus increasing the fluorescence of Nile red bound to native protein.

Thorough examination of PA(+++)-T81I emission spectra shows that fluorescence intensities at emission maxima (650 nm) of Nile red bound to native protein (29.78 ± 2.70 A.U.) and after 15 to 30 min treatment (32.45 ± 4.42 A.U. and 31.58 ± 6.47 A.U., respectively) remain similar, suggesting that PA(+++)-T81I maintains its native conformation after 30 min at 55 °C. The variant shows structural destabilisation after 60 min when emission maximum shifts to ~ 640 nm, however fluorescence intensity at this wavelength (32.34 ± 2.40 A.U.) remains at similar levels. The blue shift in Nile red is produced when hydrophobicity increases due to exposure of hydrophobic residues of misfolded proteins. We considered a protein fully denatured when Nile red emission maximum shift to 630 nm, that is the typical shift observed in stained PA(+++) aggregates, accompanied by an increase in fluorescence intensity. Thus, an “incomplete” spectral shift (~ 640 nm) without increase in fluorescence intensity suggests that structural destabilisation produced aggregate precursors in form of stable unfolded intermediates in equilibrium with

the native state. These type of intermediates are thought to be formed during thermal unfolding of rPA and to be generated due to thermostability differences among rPA domains (Ganesan, Watkinson and Moore, 2012), whereby domain 1 and domain 2 maintain their native conformation at higher temperatures related to domain 3 and domain 4, which rapidly lose their native conformation upon heat stress, as reported by tryptophan fluorescence and circular dichroism analysis of rPA thermal unfolding (Chalton *et al.*, 2007; Ganesan, Watkinson and Moore, 2012). These unfolded intermediates might be further stabilised as result of mutation T81I, preventing full PA(+++)-T81I denaturation in the present work. However, the use of other bioanalytical techniques like circular dichroism and differential scanning colorimetry are necessary to characterise such putative intermediates.

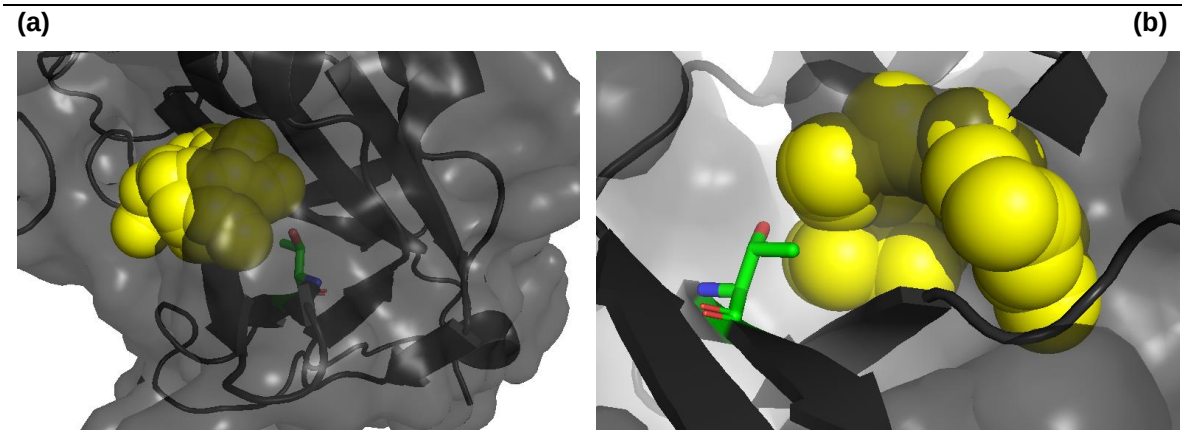


Figure 5.6 Position T674 is part of a solvent accessible pocket. Yellow balls represent the solvent-accessible area that forms the pocket. T674, represented in sticks (green), forms part of the pocket. The pocket shape and forming residues were calculated with Fpocket, using a minimum sphere radius of 2.8 Å (Le Guilloux, Schmidtke and Tuffery, 2009), using the webserver HotSpot Wizard 3.0 and the 3TEW (PDB ID) 3-D structure. **(a)** Surface view of the pocket. **(b)** Buried view of the pocket.

5.3.3 Urea denaturation and intrinsic tryptophan fluorescence

PA binds two Ca^{2+} ions in domain 1 (**Figure 5.1**) that are crucial for structural stabilisation after cleavage by furin-like proteases *in vivo*, allowing the formation of oligomers and toxin assembly (Gao-Sheridan, Zhang and Collier, 2003). Since Ca^{2+} binding is essential for rPA stability and biological function, the chemical unfolding of rPA and our engineered variants was characterised using urea and intrinsic tryptophan fluorescence. Urea denaturation studies by intrinsic tryptophan fluorescence on rPA have described a biphasic unfolding behaviour. The first transition consists of destabilisation of rPA due to the release of the Ca^{2+} ions followed by a second transition of complete protein unfolding (Gupta *et al.*, 2003; Chalton *et al.*, 2007). As shown in **Figure 5.7**, rPA follows the biphasic unfolding transition reported in the literature, and the rest of the variants show similar behaviour. The first transition, corresponding to the detachment of Ca^{2+} ions, can be observed at 1 – 2 M urea, and the second following 4 M urea, in agreement with the literature (Chalton *et al.*, 2007). These results demonstrate that the bioprocess developed in this thesis produces high-quality rPA and that our engineered variants retained native conformational structure.

Unfolding transition signal is dominated by domain 1 because it contains 5/7 tryptophan residues in PA (**Figure 5.1**). Interestingly, variant PA(+++)-T81I shows unfolding resistance at 6 M urea, suggesting that mutation T81I affects the conformational stability of domain 1. As mentioned above, the reason for PA(+++)-T81I improved thermostability might be the stabilisation of a partially folded intermediate produced during thermal denaturation (Ganesan, Watkinson and Moore, 2012), achieved by improved conformational stability of domain 1 via improved hydrophobic core-packing generated by mutation T81I.

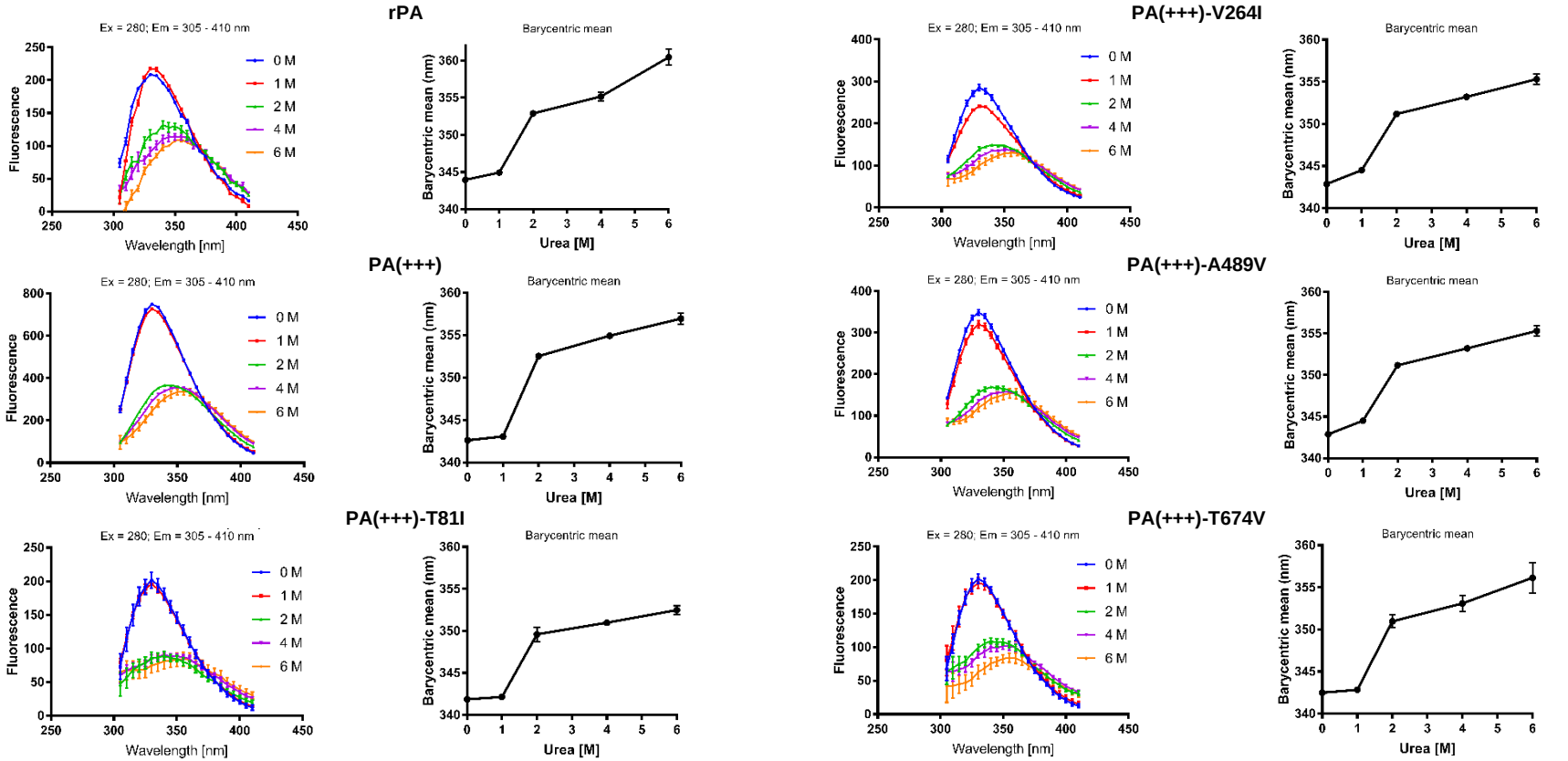


Figure 5.7. Urea denaturation. Proteins were incubated overnight with different urea concentrations and then transferred into a black opaque-bottom 96-well plate. Intrinsic tryptophan emission spectra were recorded from 300 – 410 nm, 5 nm step, after 280 nm excitation using a SpectraMax M2e microplate reader. On the left: intrinsic tryptophan emission spectra. On the right: barycentric mean. All measurements were corrected with corresponding blank for background fluorescence. Data points show mean with SEM (n = 2).

5.3.4 Limited proteolysis: variant PA(+++)-T81I enhanced the rigidity of Domain 1.

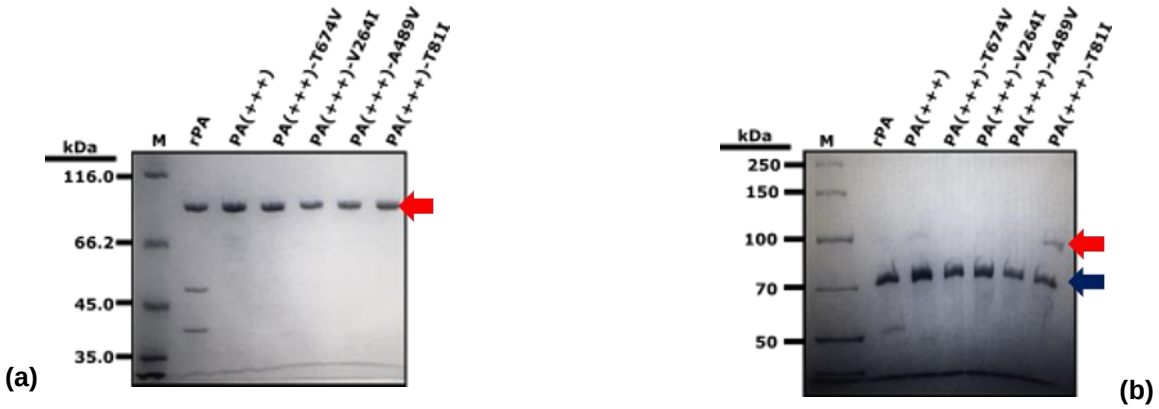


Figure 5.8 Limited proteolysis analysis. (a) SDS-PAGE analysis of untreated proteins. rPA showing its protease propensity (~ 40 and ~ 50 kDa fragments) reported in **Chapter 3**. **(b)** SDS-PAGE analysis of proteins digested by trypsin. Proteolysis was performed with trypsin for 15 min at room temperature and the reaction was stopped with soybean trypsin inhibitor. Red arrows indicate full-size PA (83 kDa), blue arrow indicates PA63 fragments.

Figure 5.8 Limited proteolysis analysis

PA is cleaved by membrane proteases in a furin-sensitive site in domain 1 ($_{164}$RKKR$_{167}$) (**Figure 5.1**) upon binding its receptors *in vivo*, triggering oligomerisation and toxin assembly (Fabre *et al.*, 2016). *In vitro*, rPA (83 kDa) can be cleaved by trypsin into 63 kDa (PA63) and 20 kDa fragments (Singh *et al.*, 1994). As well as being critical for biological function, protease susceptibility is correlated with structural flexibility and is often used as an indirect indicator of protein stability (Young, Skordalakes and Marqusee, 2007). For example, the less stable Ca^{2+}-free rPA is more susceptible to trypsin degradation than natively folded Ca^{2+}-bound rPA (Gupta *et al.*, 2003); also rPA showed resistance to thermolysin degradation when bound to its receptor, demonstrating improved structural stability upon binding (Mullangi *et al.*, 2014). Limited proteolysis analysis of rPA and our engineered variants was performed with trypsin. Pure proteins (150 µg/mL) were incubated with trypsin for 15 min and then the

reaction was stopped by adding soybean trypsin inhibitor. Proteolysis was analysed with SDS-PAGE (**Figure 5.8**).

As discussed in **Chapter 3**, rPA susceptibility to protease degradation can be observed in both **Figure 5.8a** and **b**, showing fragments of 50 kDa and 40 kDa from a chymotrypsin-sensitive site in domain 2 (Singh *et al.*, 1994). This site was removed by site-directed mutagenesis to obtain PA(+++) (see **Chapter 3**), thus these fragments are absent in the other variants. After trypsin treatment, rPA and the other variants formed PA63 as revealed by the ~ 70 kDa fragments in **Figure 5.8b**. Formation of PA63 implies that trypsin can access the $_{164}$RKKR$_{167}$ furin-sensitive site in domain 1, suggesting that the engineered variants retained the structural conformation of original rPA. Interestingly, variant PA(+++)-T81I showed signs of resistance to trypsin degradation, as suggested by the remainder of full-length protein. This result adds evidence to the hypothesis that mutation T81I has increased the rigidity of domain 1, giving variant PA(+++)-T81I improved thermostability via stabilisation of domain 1.

5.3.5 Pre-pore formation analysis

PA63 forms a heptameric ring-shaped complex following cleavage *in vivo* (Petosa *et al.*, 1997; Liu *et al.*, 2015). The heptamer adopts a conformation that prevents early formation of the transmembrane pore that translocate the effector enzymes, EF and LF, into the cytosol when pH drops in the mid-late endosome (Fabre *et al.*, 2016). Hence, the term pre-pore is used to describe the PA63 heptameric ring prior to pH-driven induction of the transmembrane pore. In order to further characterise the biological function of rPA and our engineered variants, the formation of the pre-pore was investigated using native-PAGE.

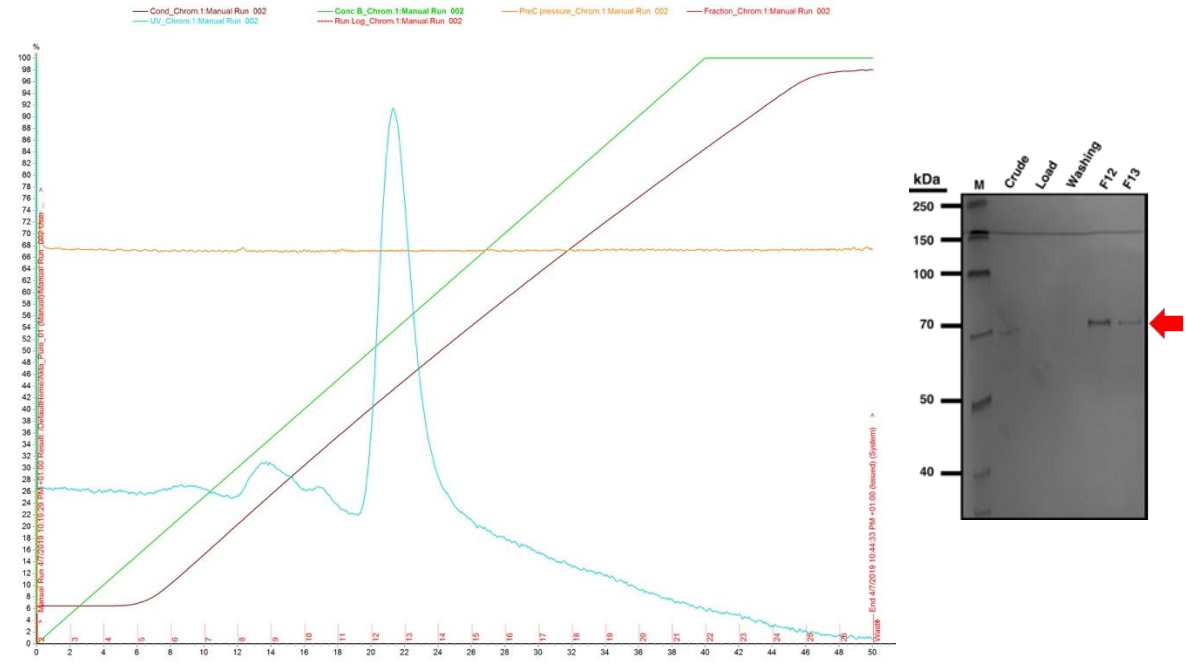
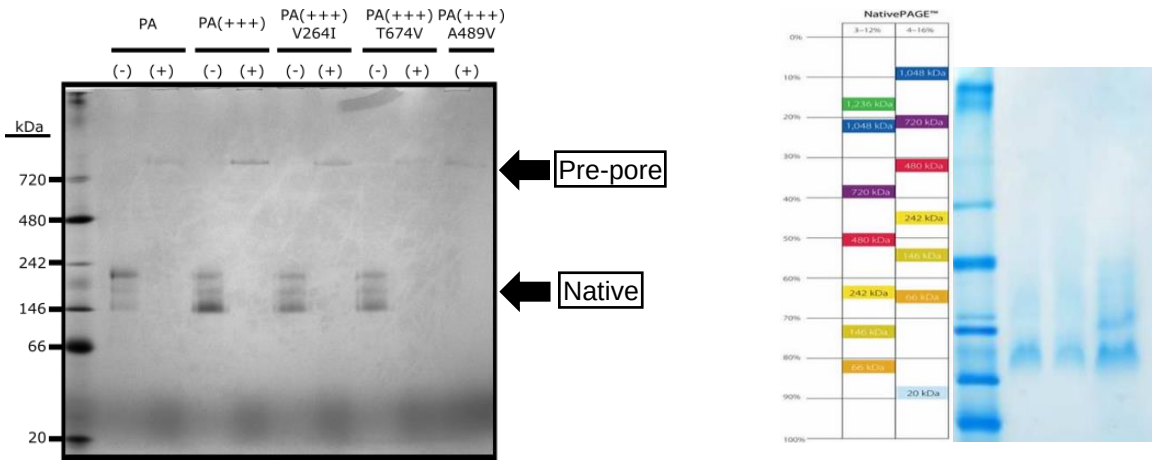


Figure 5.9. The pre-pore was assembled onto a Q-column. Proteins cleaved by trypsin were loaded onto a Q-column to induce formation of the pre-pore. Ion-exchange chromatography was performed using 0 – 1 M NaCl gradient. Protein eluted at ~ 0.5 M NaCl. On the left: chromatogram as displayed by UNICORN Control Software. On the right: SDS-PAGE analysis of fractions containing PA63 (red arrow).

Figure 5.9 The pre-pore was assembled onto a Q-column.

rPA, PA(+++),PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V variants were treated with trypsin to obtain PA63. Since domain 1 is nicked after trypsination (trypsin-site: $_{164}$RKKR$_{167}$), variant PA(+++)-T81I was not part of this experiment. The pre-pore was assembled onto a Q-column (sepharose) and proteins were eluted using a gradient of 0 – 1 M NaCl. The peak in the chromatogram containing PA63 (analysed by SDS-PAGE) showed that proteins eluted at ~ 0.5 M NaCl (**Figure 5.9**). Then, pooled fractions containing PA63 (i.e. purified pre-pore) were stored overnight at 4 °C as well as native protein for control. Native protein and pre-pore (as shown in **Figure 5.10**: (-) and (+), respectively) were loaded onto a NativePAGE^{TM} Bis-Tris 4 – 16% gel and electrophoresis performed at constant

voltage of 150 V for 90 min with light-blue cathode buffer. As shown in **Figure 5.10**, the tested proteins can form the pre-pore as shown by the large size bands (above the 720 kDa standard) in lanes (+). Native-PAGE has been used elsewhere to identify the pre-pore (Feld *et al.*, 2010). For further confirmation, the pre-pore can be directly studied using transmission electron microscopy (Kintzer *et al.*, 2009; Jiang *et al.*, 2015) or by biological assays consisting of measuring the viability of macrophage cells after incubation with PA63 and LF (Friedlander, 1986).



Modified from Wang, *et al*., 2012

Figure 5.10. Pre-pore detection by native-PAGE. (a) Native proteins (lanes (-)) and purified pre-pores (lanes (+)) were loaded onto a NativePAGE™ Bis-Tris 4 – 16% gel and electrophoresis performed at constant voltage of 150 V for 90 min with light-blue cathode buffer. Lane 1 is NativeMark™ Unstained protein standard. **(b)** On the right: native-PAGE analysis of rPA, taken from Wang *et al.*, (2012). They used a 3 – 12% Bis-Tris Native-PAGE™. First lane is NativeMark™ Unstained protein standard; second is rPA in solution; third is reconstituted rPA from a powder formulation; fourth is rPA standard. On the left, a depiction of the molecular weight standard on the different gels, as provided by supplier.

It is worth noting that native proteins (**Figure 5.10**, (-) lanes) showed three distinct bands. These bands are likely rPA isoforms produced by charge heterogeneity, which have been reported in other studies using native-PAGE (Farchaus *et al.*, 1998; Chauhan and Bhatnagar, 2002; Ribot *et al.*, 2006). Further studies on rPA isoforms revealed that charge heterogeneity is caused by deamidation of asparagine residues (Powell *et al.*, 2007). Before further discussion, it is important to highlight relevant differences between the native-PAGE techniques used in the isoform characterisation studies and in this thesis.

We used a commercially available native-PAGE kit (NativePAGE^{TM} Bis-Tris 4 – 16%) inspired in the blue native electrophoresis technique developed by Schägger and von Jagow (1991) that involves the use of Coomassie blue, a dye that binds proteins and confers them a negative net charge while maintaining their native conformation allowing their migration towards the anode. Thus, Coomassie blue resembles the function of SDS in SDS-PAGE and proteins can be separated by size maintaining their native conformation. On the other hand, studies that characterised rPA isoforms were performed using PAGE under non-denaturing conditions without a charge-shift agent (clear-native PAGE), consequently proteins migrate according to their native net charge and size (Wittig and Schägger, 2008). Thus, the appearance of three bands in native rPA and the rest of the variants in our experiment would indicate differences in size rather than charge isoforms. Although deamidation modifies protein mass, native-PAGE does not have the resolution to detect a mass shift of +1 kDa, resulting from the asparagine to iso/aspartic acid mutation (**Figure 5.11**) (Powell *et al.*, 2007). Also, it is unlikely that aggregation caused the appearance of these bands since there was no signs of denaturation of native protein (i.e. without heat treatment) during thermostability characterisation (**Figure 5.5**). Moreover, Wang *et al.*,

(2012) characterised rPA and reconstituted rPA from a dry-powder formulation under similar conditions to our experiment: using a 3 – 12% Bis-Tris Native-PAGE™ and NativeMark™ Unstained protein standard (we used 4 – 16% Bis-Tris Native-PAGE™ and same molecular weight standard), showing that isoform rPA bands appear between the 242 – 146 kDa molecular weight standards, as shown as well in our results (**Figure 5.10**). The possibility that the three bands found in native protein are result of charge heterogeneity are discussed below.

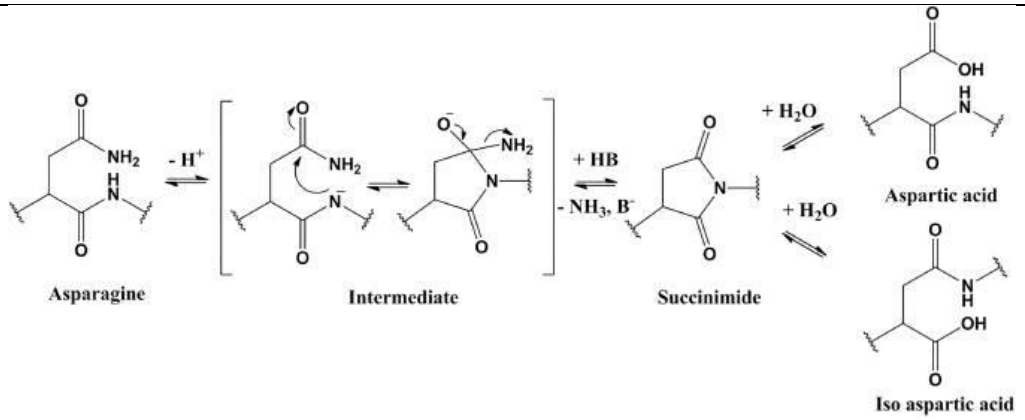


Figure 5.11. Asparagine deamidation reaction. Asn suffers a nucleophilic attack from its backbone nitrogen atom towards its side chain amide group carbon, yielding an intermediate succinimide residue that resolves via hydrolysis into Asp or iso-Asp.

Figure 5.11 Asparagine deamidation reaction

Ribot *et al.*, (2006) demonstrated that isoform heterogeneity can vary depending on the manufacturing process of rPA. They characterised rPA that was produced using a process stablished by the U.S. military (Farchaus *et al.*, 1998) as well as from a manufacturer using GMP. Using clear-native PAGE, four isoforms were identified in rPA produced using the R&D method whereas the GMP grade rPA yielded two isoforms. Thus, the content of rPA isoforms can be altered depending on its bioprocessing. In **Figure 5.10**, we can observe a shift in isoform content between rPA and PA(+++). In rPA, the less charged isoform (top

band) is more populated that the bottom two, whereas in PA(+++) the most negative isoform (bottom band) appears more populated than to the top two. These shifts were likely produced by discrete physicochemical differences between rPA and PA(+++) that our purification method could discriminate. As described in **Chapter 3**, our purification method consists of two subsequent ion-exchange chromatography steps and a polishing step by gel filtration. The second chromatography step involves the use of a weak anion exchanger, thereby the less charged isoforms would not bind the column. In **Figure 3.4** (see **Chapter 3**) it can be observed that an rPA isoform does not bind the DEAE column (weak anion exchanger), as some protein is found in the flow through (FT), shown by SDS-PAGE analysis. On the other hand, this band is absent in the SDS-PAGE analysis of PA(+++) purification (**Figure 3.7**), suggesting that PA(+++) presents different charge heterogeneity related to rPA. In conclusion, **Figure 5.10** suggests that native rPA, PA(+++) and the other variants present some degree of deamidation, displayed in the native-PAGE as charge isoforms. Definite evidence of deamidation should be obtained using other bioanalytical techniques like clear-native electrophoresis and mass-spectrometry (Powell *et al.*, 2007; Verma, Ngundi and Burns, 2016). According to studies by Ribot *et al.*, (2006), rPA isoforms maintained biological activity and vaccine efficacy in rabbits challenged with *B. anthracis* spores.

5.4 Conclusion

In **Chapter 4**, semi-rational and rational approach protein engineering approach was applied to design mutations aiming to improve the hydrophobic core-packing of PA(+++), an aggregation- protease-resistant rPA variant, yielding thermostable variants PA(+++)-T81I, PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V. In this Chapter, rPA, PA(+++) and the engineered variants were characterised using biophysical techniques. Thermostability

was indirectly characterised by investigating the conformational state of the proteins after heat treatment using the polarity sensitive dye Nile red. After 60 min at 55 °C, variants PA(+++)-V264I, PA(+++)-T674V and PA(+++)-T81I showed signs of improved thermostability related to PA(+++), with PA(+++)-T674 preserving intact its native conformation as suggested by unchanged Nile red emission spectrum after treatment. Urea denaturation was used to characterise protein unfolding by intrinsic tryptophan fluorescence. All proteins displayed the biphasic unfolding transition reported in the literature for urea denaturation of rPA, indicating that all proteins bind the Ca^{2+} ions essential for structural stability and biological function. Interestingly, variant PA(+++)-T81I showed uncomplete unfolding at 6 M urea. The literature reports that rPA is fully denatured at 6 M (Chalton *et al.*, 2007), in agreement with our results. Intrinsic tryptophan fluorescence is mainly governed by the conformational state of domain 1 since it holds 5/7 tryptophan residues in PA. The absence of spectral shift in emission maximum indicates that the microenvironment of tryptophans in domain 1 remained unchanged at 6 M urea, suggesting that mutation T81I improved the conformational stability of domain 1. Characterised proteins were also susceptible to proteolysis with trypsin yielding the activated form PA63 after cleavage in a furin-like site, demonstrating that the engineered variants display native conformational structure and functionality. Variant PA(+++)-T81I showed signs of protease resistance suggesting increased structural rigidity. Since the furin-like proteolytic site is found in domain 1, this result adds evidence to the hypothesis that mutation T81I improved the conformational stability of domain 1. The formation of the pre-pore, an oligomeric state of PA, was evaluated for further characterisation of biological function. Native-PAGE analysis showed that evaluated proteins (variant PA(+++)-T81I was not tested for pre-pore formation) can form the pre-pore. Surprisingly, native proteins (i.e. pure proteins without trypsin

treatment) subjected to native-PAGE showed the appearance of heterogeneity displayed as three charge isoforms. Such isoforms were found elsewhere during rPA characterisation (Ribot *et al.*, 2006) and concluded that the charge isoforms are result of rPA deamidation (Powell *et al.*, 2007). According to Ribot *et al.*, (2006), charge isoforms are functional and immunogenic. Further studies are needed to characterise these putative isoforms and determine the deamidation effects on protein functionality of rPA and our engineered variants.

These results demonstrate that the protein engineering approach strategy developed in **Chapter 4** generated four thermostable rPA variants that preserved biological function. Variant PA(+++)-T674V showed the highest thermostability as it retained its native conformational structure after heat treatment. Variant PA(+++)-T81I showed signs of achieving improved thermostability by increasing the rigidity of domain 1, suggesting the stabilisation of a misfolded intermediate that prevents aggregation. In future work, the thermodynamic parameters of these thermostable variants should be investigated with differential scanning calorimetry and spectroscopic techniques in order to fully elucidate their stabilising mechanisms.

Chapter 6. Discussion and conclusion

This thesis explored the development of a robust bioprocess for laboratory scale production of high-quality recombinant Protective Antigen (rPA) and the application of protein engineering approach to improve its thermostability to obtain a candidate for the development of a thermostable next-generation Anthrax vaccine.

High-quality soluble rPA was produced in *E. coli* BL21 (DE3) by periplasmic expression using a codon-optimised construct and by optimisation of culture conditions using low temperature of expression (20 °C) and low inducer concentration (0.1 mM IPTG). rPA was purified from clarified periplasmic extracts, obtained by osmotic shock, following three liquid chromatography steps: (1) rPA was captured using Q-column, a strong anion-exchanger, and proteins were separated by step-wise elution, (2) then fractions containing rPA were concentrated after loaded onto a DEAE-column, a weak anion-exchanger, and (3) rPA was obtained pure after isocratic gel filtration. For comparison, biopharmaceutical companies Emergent Biosolutions Inc. (EBI) and Altimmune have developed rPA-based vaccine candidates that have shown promising results in early clinical trials (Brown *et al.*, 2010). EBI produces rPA using a bioprocess originally described by Farchaus *et al.*, (1998), using a non-encapsulated non-toxigenic genetically engineered *B. anthracis* strain for production of rPA, enabling synthesis using its native secretory pathway and readily recovery of native folded protein. On the other hand, Altimmune produces rPA using a codon-optimised genetic construct (GenBank Accession no. AX353770) using the pET system in *E. coli* UT5600 (DE3) (Williamson *et al.*, 2005). Both processes present caveats: for EBI, rPA yields are affected due to extracellular proteases produced by *B. anthracis*, whereas Altimmune rPA must be recovered from inclusion bodies following cytoplasmic expression

in *E. coli*. Recombinant expression in *B. anthracis* has been significantly improved recently using a genetically engineered strain lacking extracellular proteases in combination with promoter engineering that can induce overexpression of rPA (Pomerantsev *et al.*, 2017; Sharma *et al.*, 2018). The results presented in this thesis reaffirm evidence that soluble rPA can be obtained in *E. coli* only by periplasmic expression (see **Table 3.1**). Moreover, while *E. coli* inclusion bodies can result in a cost-effective way to produce high protein yields at industrial scale, it is troublesome and time consuming to solubilise inclusion bodies for laboratory scale use. Consequently, the bioprocess here described provides an option for production of soluble rPA for characterisation studies at laboratory scale and potential for industrial scaling.

Since recombinant vaccines are not as immunogenic as inactivated or live-attenuated vaccines they need the addition of adjuvants to potentiate immune response. Alum is a widely used adjuvant and is considered the golden-rule for regulatory approval of vaccines (Awate, Babiuk and Mutwiri, 2013). However, rPA is a thermolabile protein whose poor stability is exacerbated when bound to alum adjuvants (Soliakov *et al.*, 2012; Wagner *et al.*, 2012; D'Souza *et al.*, 2013). In fact, the regulatory approval of the abovementioned rPA-based vaccine candidates was halted due to issues regarding antigen's stability (Kaur, Singh and Bhatnagar, 2013), which fuelled the investigation on novel adjuvants and lyophilised formulations to improve vaccine stability (Wang *et al.*, 2012; Kachura *et al.*, 2016; Savransky *et al.*, 2019; Weir *et al.*, 2019). It was later demonstrated that rPA misfolding upon binding alum can be significantly reduced by improving the rPA-alum interface with the addition of optimal concentrations of phosphate buffer (Watkinson *et al.*, 2013). The work by Watkinson *et al.*, (2013) thus suggests that an rPA-based vaccine has the potential to be stabilised by

other means apart from adjuvant engineering and lyophilisation, such as the improvement of rPA intrinsic conformational stability. Consequently, we proposed the development of a thermostable rPA variant using protein engineering approach to develop a well-defined thermostable next-generation vaccine. The hypothesis was that rPA thermostability can be improved by mutations designed to improve its hydrophobic-core packing.

First, the variant PA(+++) was generated by site-directed mutagenesis (SDM) on rPA to remove a chymotrypsin-sensitive site, found to cause proteolysis during bioprocessing, and to improve its solubility by the addition of mutations T576E and S559L, reported in the literature to supress an aggregation-prone region (APR) (Ganesan *et al.*, 2016). PA(+++) was used as starting point for the protein engineering campaign because improved solubility is a desirable feature for therapeutic proteins. Moreover, a soluble protein is more robust to explore mutational effects at higher temperature and suitable for designing a semi- or high-throughput screening.

A semi-rational approach was applied designing mutations guided by structure-based design using a high-resolution (1.45 Å) X-ray crystal structure of PA (PDB: 3TEW). Recent advances in recombinant vaccine technology are increasingly using structure-based design of antigens and epitope focused vaccines with the aim to develop rationally designed vaccines (Correia *et al.*, 2014; Malito, Carfi and Bottomley, 2015). Hotspots were selected for mutagenesis aiming to improve the hydrophobic core-packing of PA(+++): hydrophobic residues forming cavities were mutated to bulkier hydrophobic amino acids; polar residues found in the protein core were substituted to hydrophobic amino acids. Mutations were introduced in PA(+++) with SDM to generate a semi-rational approach library. These same mutations were then analysed with FoldX to predict their effect on protein stability. FoldX is

an empirical force-field algorithm widely used for stability predictions (Schymkowitz *et al.*, 2005). FoldX is ideal for mutational predictions on hydrophobic core-packing because its energetic function calculates the volume occupied by a residue to estimate solvent accessibility, so it implicitly accounts the effect of a mutation on packing (Colonna-Cesari and Sander, 1990; Guerois, Nielsen and Serrano, 2002), although it has shown better accuracy predicting stability effects on surface residues probably due to the inherent complexity of the network interactions in the protein core (Thiltgen and Goldstein, 2012). FoldX input is a high-resolution crystal structure of the wildtype (PDB: 3TEW) and the output is the difference of the free energy of unfolding (ΔΔG) between the wildtype (PA) (ΔG_{wt}) and the mutant variant (ΔG_{v}). A threshold of $\Delta\Delta G \leq -0.5$ kcal/mol was set to select beneficial mutations, which is the approximate error reported by Guerois, Nielsen and Serrano (2002). The predicted beneficial mutations were introduced in PA(+++) to produce a small rational approach library. This approach to narrow possible successful variants using computational predictions has been applied effectively in other thermostability studies (Wijma *et al.*, 2014; Broom *et al.*, 2017; Han *et al.*, 2018).

A semi-high throughput method was developed for screening of the semi-rational approach library. The library was arrayed in a 96-well plate. Mutant variants were expressed in 96-well format and then recovered in periplasmic extracts using a miniaturised version of the osmotic shock method. Periplasmic extracts containing the mutant variants were used for thermostability screening. Thermostability screens are typically developed by measuring the residual function of a protein after heat treatment (Eijsink *et al.*, 2005). Indeed, a screening for compounds that improve rPA thermostability was developed by heat treatment of rPA in presence of different osmolytes and then measuring the cell viability of macrophages in

presence of heat-treated rPA and lethal factor (Singh *et al.*, 2004). However, this type of assay is not suitable for high or semi-high throughput screening because the maintenance of macrophage cultures would make it expensive and time-consuming. Sorrell (*et al.*, 2010) used differential scanning fluorimetry (DSF) to develop a high-throughput method for screening rPA anti-toxin compounds. DSF uses fluorescent probes that bind hydrophobic regions of proteins that are gradually exposed as temperature ramps, monitoring protein unfolding through fluorescence intensity. This type of thermal shift assay is attractive because thermodynamic parameters can be directly extracted using low amounts of protein (Wright *et al.*, 2017), yet it is only suitable for purified proteins because it assumes two-state equilibrium unfolding. We developed a method that uses the polarity-sensitive fluorescent probe Nile red to determine the fluorescence of PA(+++) variants contained in periplasmic extracts before and after heat-treatment to indirectly screen protein thermostability. The assay temperature was set at the aggregation temperature of PA(+++) contained in periplasmic extracts, experimentally established at 65 ºC (after 10 min). Consequently, thermostable variants are screened for their aggregation propensity relative to PA(+++). Three variants resulted positive and were expressed in 50 mL culture using conical flasks to perform a validation assay, one resulting successful: variant PA(+++)-T81I. The rational approach library, consisting of variants PA(+++)-V264I, PA(+++)-A489V and PA(+++)-T674V, was expressed in 50 mL culture in conical flasks. Variants contained in periplasmic extracts were incubated at a range of temperatures and their thermostability was indirectly measured by staining denatured proteins with Nile red, following the same principle as the abovementioned screening. The three rational approach variants showed improved thermostability related to both PA(+++) and rPA. Interestingly, the screening assay for the rational approach library showed signs of following a sigmoid behaviour typical to that of

aggregation models based on a lag-phase following rapid aggregate growth upon nucleation (Meisl *et al.*, 2016). Thus, one can propose for future work the development of a high-throughput screening method for PA(+++) mutant libraries based on DSF methodology to estimate the aggregation temperature of variants contained in periplasmic extracts.

It was surprising to find that the stabilising mutations predicted by FoldX did not result positive during the semi-rational approach screening and, conversely, mutation T81I was found highly destabilising by FoldX ($\Delta\Delta G$ = 5.9 kcal/mol). Characterisation of these variants revealed that mutants reached improved thermostability by different mechanisms. The conformational state of the variants after heat treatment at 55 °C for 60 min was probed with Nile red fluorescence. The blue shift that Nile red emission spectrum suffers upon binding to rPA aggregates was used to estimate protein denaturation. A fully denatured protein was considered when Nile red emission maxima shifted from ~ 650 nm, bound to native protein, to 630 nm after heat treatment. This condition is in agreement with rPA aggregates stained with Nile red reported in the literature (Soliakov *et al.*, 2012) and was corroborated experimentally for PA(+++). The four mutant variants, PA(+++)-T81I, PA(+++)-V264I , PA(+++)-A489V and PA(+++)-T674V, showed improved stability related to both PA(+++) and rPA. Variant PA(+++)-T674V was the most thermostable as Nile red spectrum remained unchanged after 60 min. PA(+++)-T81I showed a blue-shift to ~ 640 nm after 60 min, with no change in fluorescence intensity at emission maxima compared to Nile red bound to native protein (i.e. without heat treatment). The rest of the variants showed some degree of denaturation after 60 min. Variants were further characterised after chemical denaturation with urea using intrinsic tryptophan fluorescence to monitor unfolding transitions. PA unfolds in urea following a biphasic transition (Chalton *et al.*, 2007). The first

transition comprises the release of the two Ca^{2+} bound to domain 1, and the second to the global unfolding of rPA. All proteins, including mutant variants, PA(+++) and rPA showed biphasic unfolding transitions. The Ca^{2+} ions are essential for PA functionality (Gao-Sheridan, Zhang and Collier, 2003; Lacy *et al.*, 2004) thus these results are an indication that the mutant variants display the native state conformation and functionality of the original rPA. Interestingly, variant PA(+++)-T81I showed unfolding resistance at 6 M urea while the rest of the proteins are present unfolded. Domain 1 govern the intrinsic fluorescence of rPA because it contains 5/7 tryptophan residues, thus tryptophans in variant PA(+++)-T81I are not fully solvent-exposed suggesting that mutation T81I conferred domain 1 improved stability. The biological function of the variants was characterised by limited proteolysis of a furin-like site in domain 1 that upon cleavage triggers conformational rearrangements in PA that leads to toxin assembly (Liu *et al.*, 2015). All proteins left the fragments corresponding to the furin-like site after proteolysis with trypsin. Nonetheless, variant PA(+++)-T81I showed proteolysis-resistance, suggesting increased rigidity of domain 1 and showing further evidence for improvement of the thermostability of domain 1.

Non-cooperative thermal unfolding, a common phenomenon in multi-domain proteins (Young, Skordalakes and Marqusee, 2007; Bhaskara and Srinivasan, 2011; Chakraborty *et al.*, 2012; Lamazares *et al.*, 2015), has been proposed for rPA in other works, however, it has been difficult to conclude that rPA thermally unfolds following a non-two-state transition due to low enthalpy change between the putative unfolding transitions reported in the literature (Chalton *et al.*, 2007; Ganesan, Watkinson and Moore, 2012; Ganesan *et al.*, 2016). Based on these studies, the current rPA thermal unfolding model suggests an initial global destabilisation of the protein backbone due to the loss of the Ca^{2+}

ions bound to domain 1 followed by a second unfolding transition. The second transition is thought to be non-cooperative since tryptophan fluorescence analysis shows evidence that domain 1 and domain 2 can maintain their native conformation during prolonged heat stress at 45 °C. Of the 7 tryptophan residues in rPA, 3 are in the core of domain 1 (W65, W90 and W316), 2 in proximity to the Ca^{2+} ions (W226 and W206) and 2 in domain 2 (W346 and W447). In the work by Ganesan, *et al.* (2012), isothermal unfolding analysis at 45 °C by tryptophan fluorescence shows that after the initial transition there is minimal change in tryptophan fluorescence spectrum, suggesting that these residues remain in an environment shielded from the solvent, while domain 3 and domain 4 showed signs of loss of secondary structure as reported by circular dichroism analysis. Thus, they proposed the occurrence of a thermal unfolding intermediate consisting of a molten globule state whereby domain 1 and domain 2 maintain most of their native structure and interactions intact while domain 3 and domain 4 structure is further destabilised. The partially-folded intermediates eventually produce intermolecular interactions that trigger aggregation. Using this model to analyse our results, one can propose how variants PA(+++)-T674V and PA(+++)-T81I reached thermodynamic stability by different mechanisms.

First of all, variant PA(+++) contains mutation T576E in domain 3 that silences an aggregation prone region (APR) known to trigger aggregation in rPA at room temperature (Ganesan *et al.*, 2016). Therefore, as the results reported by Ganesan *et al.*, (2016) suggest, PA(+++) aggregation is caused by increasing concentration of partially-folded intermediates that accumulate until aggregation is triggered by intermolecular interactions of their solvent-exposed hydrophobic regions, this process is often referred as nucleation (Roberts, 2007). Nucleation occurs when unfolded intermediates reach a concentration that shifts the

thermodynamic equilibrium towards irreversible aggregate formation. A variant with improved conformational stability can maintain its native state conformation at equilibrium at higher temperatures, preventing the accumulation of unfolded intermediates thus avoiding nucleation. Mutation T674V resides in domain 4. This domain has been reported to have poor interface contacts with the other domains, generating conformational flexibility that is thought to be needed to promote receptor binding (Feld *et al.*, 2012). Moreover, there is evidence showing that rPA stability significantly improves upon receptor binding, implying stabilisation of domain 4 (Mullangi *et al.*, 2014). We propose that the improved hydrophobic core-packing promoted by mutation T674V increased domain 4 rigidity and folding cooperativity with the rest of domains, conferring PA(+++)-T674V improved thermostability. On the other hand, the thermostabilising path followed by variant PA(+++)-T81I consists of improving hydrophobic core-packing of domain 1, that consequently stabilises the misfolding intermediate formed during the thermal unfolding model described above. So, while PA(+++)-T674V improved rigidity prevents the formation of misfolded intermediates, PA(+++)-T81I prevents aggregation by improving the rigidity of domain 1 structure forming intermediates, rising the energetic barrier for nucleation. Future analysis of thermal unfolding by circular dichroism and tryptophan fluorescence are needed to validate these suggestions. Also, differential scanning calorimetry is needed to estimate enthalpy changes as well as to determine if the unfolding transition has indeed shifted to higher temperatures.

The biological function of rPA variants was further explored by investigating the formation of a naturally occurring heptamer, named the pre-pore, that serves as delivery system for the effector proteins into the cell (Fabre *et al.*, 2016). Variant PA(+++)-T81I was

not tested because mutation T81I resides in the part of domain 1 that is cleaved by trypsin prior pre-pore assembly (Petosa *et al.*, 1997). Pre-pore formation was analysed by native-PAGE and it was found present in all proteins. Surprisingly, native proteins (i.e. pure protein not treated with trypsin) displayed three charge isoforms that have been reported in the literature as protein heterogeneity related to deamidation of rPA (Powell *et al.*, 2007). Although it was found that these isoforms retain biological function and immunogenicity (Ribot *et al.*, 2006), long-term storage can increase the deamidation rate of rPA (D'Souza *et al.*, 2013; Verma *et al.*, 2013; Verma, Ngundi and Burns, 2016). Site-directed mutagenesis have been applied to substitute the Asn residues thought to cause deamidation, yielding mixed results regarding rPA biological activity and immunogenicity (Verma *et al.*, 2013; Verma and Burns, 2018). Long-term stability studies should be performed on our engineered variants to test if their improved thermostability can decrease the deamidation rate of the protein, following the hypothesis that chemical stability issues of biopharmaceutical products can be reduced if the protein conformational stability is improved (Manning *et al.*, 2010).

Future work should explore the combination of mutations designed in this thesis to further improve rPA thermostability via additive effects. Mutations should be added to variant PA(+++)-T674V because it showed the highest thermostability improvement and the expected outcome of combinatory effects would be to increase temperature of unfolding and prevent protein aggregation at higher temperature. Mutation T81I improved the rigidity of domain 1 and partially inhibited proteolysis in the furin-like cleavage site. This proteolytic site has been reported to be part of a neutralising epitope (McComb and Martchenko, 2016), thus relevant for vaccine effectivity. Mutation T81I conferred proteolytic resistance without altering the amino acid sequence of the furin-like cleavage site, enabling conservation of this

protective epitope, thus in combination with T674V can be used to obtain a thermostable proteolytic-resistant vaccine candidate. Also, other strategies can be applied to rationally design mutations to further improve PA(+++)-T674V thermostability. The use of engineered disulphide bonds designed to restrict protein motions has proven useful to generate thermostable proteins (Wijma *et al.*, 2014; Yu and Huang, 2014; Li *et al.*, 2018). This is achieved by replacing a pair of WT residues with Cys, aiming to rigidify the protein through the formation of a covalent bond between Cys side chains. This approach has been used to engineer a stable recombinant vaccine candidate against the human immunodeficiency virus (HIV) (Khayat *et al.*, 2013) and in attempts to produce a universal influenza vaccine (Lu, Welsh and Swartz, 2014). In the case of rPA, disulphide bonds can be designed to improve the domain interface interactions between domain 4 and domain 2, aiming to hinder domain 4 hinge-like motions, described in **Chapter 1** (**Section 1.3**) and elsewhere (Bann, 2012; Feld *et al.*, 2012). Another type of structure-guided approach involves locking the protein into a stable conformation by substituting key residues with Pro, which is a residue known to restrict backbone flexibility (Yu and Huang, 2014). This strategy has been applied for the development of a vaccine against the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 vaccines are designed to generate neutralising antibodies against the spike protein, which is the viral surface protein that binds cellular receptors and mediates entry into the cells (Walls *et al.*, 2020). As with PA, the dramatic conformational rearrangements that the spike protein undergoes to accomplish its functions translate into poor stability when recombinantly produced, promoting conformational heterogeneity that may reduce vaccine effectivity (Walls *et al.*, 2020). The substitution of 2 subsequent residues to Pro in a loop that triggers conformational rearrangements locked the spike protein into a conformation that reduced conformational heterogeneity, improved recombinant production

yields and promoted a stronger immune response (Pallesen *et al.*, 2017; Wrapp *et al.*, 2020). The work by Jiang *et al.* (2015) elucidates the conformational rearrangements that PA undergoes to perform its biological function and the key motifs that trigger such conformational changes. These motifs can be used as a guide to rationally select residues for Pro substitutions, aiming to lock rPA into a stable native conformation. A possible candidate could be the motif 2α1 (**Figure 1.7**), whose zipper-like rearrangements promote conformational flexibility of domain 2 that enables PA transition to form the transmembrane pore.

The thermostable variants generated in this and future work are to be investigated for immunogenicity. Protection against Anthrax disease is mediated by neutralising antibodies against PA, thus immunogenicity can be evaluated by performing an enzyme-linked immunosorbent assay (ELISA) using antiserum raised by animal immunisation with mutant rPA and evaluating binding towards WT rPA, expecting to find comparable response titre results (Ganesan *et al.*, 2016). This assay demonstrates that antibodies produced by immunisation with rPA variants can also bind WT rPA, indicating comparable immune responses. The standard to demonstrate Anthrax vaccine efficacy is the toxin-neutralising antibody (TNA) assay. TNA assay evaluates the ability of antibodies to neutralise the activity of the Lethal toxin in a cytotoxic assay (*in vitro*) (Ngundi *et al.*, 2010). Then, immunized animals are challenged with Anthrax spores and survival is correlated with pre-challenge TNA levels to determine a protective threshold (Ionin *et al.*, 2013; Savransky *et al.*, 2017). The most recent regulatory approval of an Anthrax vaccine is a new indication for AVA, for its use in post-exposure prophylaxis (PEP) treatment (Ionin *et al.*, 2013). Since Anthrax PEP studies in humans are unfeasible and unethical, a validated TNA assay in rabbits and non-

human primates was used to correlate protection and efficacy of the vaccine in humans, following a special regulatory approval scheme (Burns, 2012; Beasley, Brasel and Comer, 2016). In summary, validated ELISA and TNA assay are essential immunological studies for Anthrax vaccines and should be included in future work to determine if thermostable rPA variants show immunogenicity equivalent to that of original rPA.

In conclusion, this thesis demonstrated that the thermostability of rPA can be improved by increasing hydrophobic core-packing using semi-rational and rational approach engineering. It also demonstrated the use of variant PA(+++) for the development of a semi-high throughput screening of a semi-rational approach library. This assay demonstrated that PA(+++) and its periplasmic expression in 96-well plates have the potential to establish in the future a platform to explore mutational effects using thermal shift assays in a high throughput format, which was not viable for rPA due to its aggregation-prone behaviour. The variants that showed the highest thermostability were PA(+++)-T81I and PA(+++)-T674V, and it was proposed that they reached improved thermostability by different thermodynamic mechanisms that should be elucidated in future work using calorimetry assays and spectroscopic techniques.

Ultimately this research contributed to the field of rational design of vaccines by exploring the improvement of an antigen's thermostability through protein engineering approach. As computational methods for energetic predictions and 3-D modelling advance, the development of recombinant-based vaccines can be established as a benchmark for swift response and fast development of effective and stable vaccines for diseases that represent a latent risk of outbreak, such as the recent Ebola crisis and the current coronavirus pandemic. In the case of bioterrorism, research should be aimed to explore multivalent vaccines that

confer broader protection to different agents, following recent efforts (Majumder *et al.*, 2019).

Appendix

Appendix I. Synthetic rPA sequence by GenScript.

Synthetic rPA contains the *pagA* gene codon optimised by GenScript. Underlined are restriction sites NdeI and EcoRI. In bold are stop codons. In lowercase is the *pelB* leader sequence.

CATATG*aaatacctgctgccgaccgctgctgctggtctgctgctgctggctgctcaaccggcgatggca*GAAG
TGAAACAGGAAAACCGTCTGCTGAATGAATCTGAAAGCTCTAGTCAAGGCCTGCTGGGTTATTACTTCAGTGA
TCTGAATTTTCAGGCCCCGATGGTGGTTACCAGCTCAACCACGGGTGACCTGTCTATTCCGTCGAGCGAACTG
GAAAATATCCCGAGTGAAAACCAGTATTTCCAATCTGCGATTTGGAGTGGCTTTATCAAAGTGAAAAAGTCGG
ATGAATACACCTTCGCGACGAGCGCCGACAATCATGTGACCATGTGGGTTGATGACCAGGAAGTTATCAACAA
AGCCTCCAACTCAAACAAGATCCGTCTGGAAAAGGGTCGCCTGTATCAGATTAAAATCCAGTACCAACGCGAA
AACCCGACCGAAAAAGGCCTGGATTTTAAGCTGTATTGGACGGACAGCCAGAACAAGAAGGAAGTGATCTCTA
GTGATAACCTGCAGCTGCCGGAACTGAAACAAAAGTCCTCAAATAGCCGTAAAAAGCGCTCGACCAGCGCAGG
TCCGACGGTCCCGGATCGTGACAACGATGGTATTCCGGATAGCCTGGAAGTGGAAGGCTACACCGTTGACGTC
AAAAATAAGCGCACGTTCCTGTCCCCGTGGATTTCAAACATCCACGAAAAGAAAGGTCTGACCAAATATAAGT
CGAGCCCGGAAAAATGGTCCACCGCTTCAGACCCGTACTCCGATTTTGAAAAAGTCACCGGCCGTATTGATAA
GAATGTGTCACCGGAAGCACGTCATCCGCTGGTTGCAGCATATCCGATCGTGCACGTTGATATGGAAAACATC
ATCCTGTCGAAGAACGAAGATCAGAGCACCCAAAATACGGACTCTGAAACCCGTACGATTAGTAAAAACACCT
CTACGAGTCGCACCCATACGTCCGAAGTCCACGGTAATGCCGAAGTGCATGCATCATTTTTCGATATTGGCGG
TTCCGTGTCAGCCGGCTTTTCGAACAGCAATTCTAGTACCGTTGCAATCGACCACTCTCTGAGTCTGGCCGGT
GAACGTACCTGGGCAGAAACGATGGGCCTGAACACCGCAGATACGGCTCGTCTGAACGCTAATATTCGCTATG
TGAATACCGGTACGGCGCCGATCTACAACGTTCTGCCGACCACGAGTCTGGTCCTGGGCAAAAATCAGACCCT
GGCTACGATTAAAGCGAAGGAAAACCAGCTGTCGCAAATCCTGGCCCCGAACAATTATTACCCGAGCAAAAAT
CTGGCGCCGATTGCCCTGAACGCACAGGATGACTTCTCCTCAACCCCGATCACGATGAACTATAATCAGTTTC
TGGAACTGGAAAAAACCAAGCAGCTGCGTCTGGACACGGATCAAGTGTATGGTAATATTGCCACCTACAACTT
CGAAAATGGTCGTGTCCGCGTGGATACCGGCTCGAACTGGAGCGAAGTTCTGCCGCAGATTCAAGAAACCACG
GCACGCATTATCTTTAATGGCAAAGATCTGAACCTGGTGGAACGTCGCATCGCAGCTGTTAATCCGTCTGATC
CGCTGGAAACCACGAAACCGGACATGACCCTGAAAGAAGCCCTGAAGATTGCATTTGGCTTCAATGAACCGAA
CGGTAATCTGCAGTATCAAGGCAAGGATATCACCGAATTTGACTTCAACTTCGATCAGCAAACGAGCCAGAAC
ATCAAAAATCAACTGGCTGAACTGAACGCGACCAACATCTACACGGTTCTGGATAAGATCAAGCTGAACGCTA
AGATGAACATTCTGATCCGTGACAAGCGCTTCCATTACGATCGTAACAATATTGCTGTTGGTGCGGATGAATC
TGTCGTGAAAGAAGCACACCGCGAAGTCATTAATTCGAGCACCGAAGGCCTGCTGCTGAACATCGACAAAGAT
ATTCGCAAGATCCTGAGTGGTTATATTGTTGAAATCGAAGATACCGAAGGCCTGAAAGAAGTCATTAATGACC
GTTACGATATGCTGAACATCTCTAGTCTGCGCCAGGACGGCAAGACCTTCATCGATTTCAAGAAGTACAACGA
CAAACTGCCGCTGTACATCTCCAACCCGAATTATAAAGTTAACGTCTACGCGGTGACCAAGGAAAACACGATT
ATCAATCCGTCCGAAAACGGTGATACCTCAACGAACGGCATTAAAAAGATTCTGATCTTTAGCAAAAAGGGCT
ATGAAATCGGT**TAATGA**ATTC

References

Abrami, L. *et al.* (2010) 'Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors', *PLoS Pathogens*, 6(3), p. e1000792. doi: 10.1371/journal.ppat.1000792.

Abrami, L., Leppla, S. H. and Gisou Van Der Goot, F. (2006) 'Receptor palmitoylation and ubiquitination regulate anthrax toxin endocytosis', in *Journal of Cell Biology*, pp. 309–320. doi: 10.1083/jcb.200507067.

Ahuja, N., Kumar, P. and Bhatnagar, R. (2001) 'Rapid purification of recombinant anthrax-protective antigen under nondenaturing conditions', *Biochemical and Biophysical Research Communications*. 2001/08/04, 286(1), pp. 6–11. doi: 10.1006/bbrc.2001.5337.

Alfaro-Chávez, A. L. *et al.* (2019) 'Improving on nature's shortcomings: Evolving a lipase for increased lipolytic activity, expression and thermostability', *Protein Engineering, Design and Selection*, 32(1), pp. 13–24. doi: 10.1093/protein/gzz024.

Anaraki, S. *et al.* (2008) 'Investigations and control measures following a case of inhalation anthrax in East London in a drum maker and drummer, October 2008', *Eurosurveillance*, 13(51). doi: https://doi.org/10.2807/ese.13.51.19076-en.

Angarica, V. E. and Sancho, J. (2012) 'Protein Dynamics Governed by Interfaces of High Polarity and Low Packing Density', *PLoS ONE*. 2012/10/26, 7(10), pp. e48212–e48212. doi: 10.1371/journal.pone.0048212.

Ashenberg, O., Gong, L. I. and Bloom, J. D. (2013) 'Mutational effects on stability are largely conserved during protein evolution', *Proceedings of the National Academy of Sciences of the United States of America*. 2013/12/09, 110(52), pp. 21071–21076. doi:

10.1073/pnas.1314781111.

Awate, S., Babiuk, L. A. and Mutwiri, G. (2013) ‘Mechanisms of action of adjuvants’, *Frontiers in Immunology*, 4(MAY), p. 114. doi: 10.3389/fimmu.2013.00114.

Bachran, C. and Leppla, S. H. (2016) ‘Tumor targeting and drug delivery by anthrax toxin’, *Toxins*, 8(7), p. 197. doi: 10.3390/toxins8070197.

De Baets, G. *et al.* (2014) ‘A genome-wide sequence-structure analysis suggests aggregation gatekeepers constitute an evolutionary constrained functional class’, *Journal of Molecular Biology*, 426(12), pp. 2405–2412. doi: 10.1016/j.jmb.2014.04.007.

Baillie, L. W. (2009) ‘Is new always better than old? The development of human vaccines for anthrax’, *Human Vaccines*, 5(12), pp. 806–816. doi: 10.4161/hv.9777.

Baillie, L. W. J., Johnson, M. and Manchee, R. J. (1994) ‘Evaluation of Bacillus subtilis strain IS53 for the production of Bacillus anthracis protective antigen’, *Letters in Applied Microbiology*. 1994/10/01, 19(4), pp. 225–227. doi: 10.1111/j.1472-765X.1994.tb00949.x.

Bann, J. G. (2012) ‘Anthrax toxin protective antigen - Insights into molecular switching from prepore to pore’, *Protein Science*, pp. 1–12. doi: 10.1002/pro.752.

Beasley, D. W. C., Brasel, T. L. and Comer, J. E. (2016) ‘First vaccine approval under the FDA Animal Rule’, *npj Vaccines*, 1, p. 16013. doi: 10.1038/npjvaccines.2016.13.

Belton, F. C. and Strange, R. E. (1954) ‘Studies on a protective antigen produced in vitro from Bacillus anthracis: medium and methods of production.’, *British journal of experimental pathology*, 35(2), pp. 144–152. Available at: https://www.ncbi.nlm.nih.gov/pubmed/13149735.

Berger, T., Kassirer, M. and Aran, A. A. (2014) ‘Injectional anthrax - New presentation of an old disease’, *Eurosurveillance*, 19(32). doi: 10.2807/1560-7917.ES2014.19.32.20877.

Besschetnova, T. Y. *et al.* (2015) ‘Regulatory mechanisms of anthrax toxin receptor 1-dependent vascular and connective tissue homeostasis’, *Matrix biology : journal of the International Society for Matrix Biology*. 2015/01/05, 42, pp. 56–73. doi: 10.1016/j.matbio.2014.12.002.

Bhaskara, R. M. and Srinivasan, N. (2011) ‘Stability of domain structures in multi-domain proteins’, *Scientific Reports*, 1, p. 40. doi: 10.1038/srep00040.

Blaustein, R. O. *et al.* (1989) ‘Anthrax toxin: channel-forming activity of protective antigen in planar phospholipid bilayers’, *Proceedings of the National Academy of Sciences of the United States of America*, 86(7), pp. 2209–2213. doi: 10.1073/pnas.86.7.2209.

Bloom, J. D. *et al.* (2006) ‘Protein stability promotes evolvability’, *Proceedings of the National Academy of Sciences of the United States of America*. 2006/04/04, 103(15), pp. 5869–5874. doi: 10.1073/pnas.0510098103.

Boël, G. *et al.* (2016) ‘Codon influence on protein expression in E. coli correlates with mRNA levels’, *Nature*. 2016/01/13, 529(7586), pp. 358–363. doi: 10.1038/nature16509.

Bourgogne, A. *et al.* (2003) ‘Global effects of virulence gene regulators in a Bacillus anthracis strain with both virulence plasmids’, in *Infection and Immunity*, pp. 2736–2743. doi: 10.1128/IAI.71.5.2736-2743.2003.

Bradley, K. A. *et al.* (2003) ‘Binding of Anthrax Toxin to Its Receptor Is Similar to α Integrin-Ligand Interactions’, *Journal of Biological Chemistry* , 278(49), pp. 49342–49347. Available at: http://www.jbc.org/content/278/49/49342.abstract.

Broom, A. *et al.* (2017) 'Computational tools help improve protein stability but with a solubility tradeoff', *Journal of Biological Chemistry*, 292(35), pp. 14349–14361. doi: 10.1074/jbc.M117.784165.

Brown, B. K. *et al.* (2010) 'Phase I Study of Safety and Immunogenicity of an Escherichia coli-Derived Recombinant Protective Antigen (rPA) Vaccine to Prevent Anthrax in Adults', *PLoS ONE*. 2010/11/17, 5(11), p. e13849. doi: 10.1371/journal.pone.0013849.

Brown, M. J., Thoren, K. L. and Krantz, B. A. (2015) 'Role of the α Clamp in the Protein Translocation Mechanism of Anthrax Toxin', *Journal of Molecular Biology*. 2015/09/05, 427(20), pp. 3340–3349. doi: 10.1016/j.jmb.2015.08.024.

Burns, D. L. (2012) 'Licensure of vaccines using the animal rule', *Current Opinion in Virology*, 2(3), pp. 353–356. doi: 10.1016/j.coviro.2012.01.004.

Camilloni, C. *et al.* (2016) 'Towards a structural biology of the hydrophobic effect in protein folding', *Scientific Reports*, 6(1), p. 28285. doi: 10.1038/srep28285.

Candela, T. and Fouet, A. (2005) 'Bacillus anthracis CapD, belonging to the γ-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan', *Molecular Microbiology*, 57(3), pp. 717–726. doi: 10.1111/j.1365-2958.2005.04718.x.

Candela, T., Mock, M. and Fouet, A. (2005) 'CapE, a 47-amino-acid peptide, is necessary for Bacillus anthracis polyglutamate capsule synthesis', *Journal of Bacteriology*, 187(22), pp. 7765–7772. doi: 10.1128/JB.187.22.7765-7772.2005.

Carter, K. C. (1988) 'The Koch-Pasteur dispute on establishing the cause of anthrax.', *Bulletin of the history of medicine*, 62(1), pp. 42–57. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/3285924.

Chadegani, F. *et al.* (2014) '19F nuclear magnetic resonance and crystallographic studies of 5-fluorotryptophan-labeled anthrax protective antigen and effects of the receptor on stability', *Biochemistry*. 2014/01/15, 53(4), pp. 690–701. doi: 10.1021/bi401405s.

Chakraborty, A. *et al.* (2012) 'Anthrax outbreaks in Bangladesh, 2009-2010', *American Journal of Tropical Medicine and Hygiene*, 86(4), pp. 703–710. doi: 10.4269/ajtmh.2012.11-0234.

Chalton, D. A. *et al.* (2007) 'Unfolding transitions of Bacillus anthracis protective antigen', *Archives of Biochemistry and Biophysics*. 2007/05/29, 465(1), pp. 1–10. doi: 10.1016/j.abb.2007.04.030.

Charlton, S. *et al.* (2007) 'A study of the physiology of Bacillus anthracis Sterne during manufacture of the UK acellular anthrax vaccine', *Journal of Applied Microbiology*, 103(5), pp. 1453–1460. doi: 10.1111/j.1365-2672.2007.03391.x.

Chaudhury, S., Lyskov, S. and Gray, J. J. (2010) 'PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta', *Bioinformatics*. 2010/01/07, 26(5), pp. 689–691. doi: 10.1093/bioinformatics/btq007.

Chauhan, V. and Bhatnagar, R. (2002) 'Identification of amino acid residues of anthrax protective antigen involved in binding with lethal factor', *Infection and Immunity*, 70(8), pp. 4477–4484. doi: 10.1128/IAI.70.8.4477-4484.2002.

Chen, K. H. *et al.* (2016) 'Anthrax toxin protective antigen variants that selectively utilize either the CMG2 or TEM8 receptors for cellular uptake and tumor targeting', *Journal of Biological Chemistry*. 2016/08/23, 291(42), pp. 22021–22029. doi:

10.1074/jbc.M116.753301.

Chi, X. *et al.* (2015) 'Generation and characterization of human monoclonal antibodies targeting anthrax protective antigen following vaccination with a recombinant protective antigen vaccine', in *Clinical and Vaccine Immunology*, pp. 553–560. doi: 10.1128/CVI.00792-14.

Cleret, A. *et al.* (2014) 'Lung Dendritic Cells Rapidly Mediate Anthrax Spore Entry through the Pulmonary Route', *The Journal of Immunology*. 2007/06/06, 178(12), pp. 7994–8001. doi: 10.4049/jimmunol.178.12.7994.

Colonna-Cesari, F. and Sander, C. (1990) 'Excluded volume approximation to protein-solvent interaction. The solvent contact model', *Biophysical Journal*, 57(5), pp. 1103–1107. doi: 10.1016/S0006-3495(90)82630-8.

Correia, B. E. *et al.* (2014) 'Proof of principle for epitope-focused vaccine design', *Nature*. 2014/02/05, 507(7491), pp. 201–206. doi: 10.1038/nature12966.

St. Croix, B. *et al.* (2000) 'Genes expressed in human tumor endothelium', *Science*, 289(5482), pp. 1197–1202. doi: 10.1126/science.289.5482.1197.

Cromwell, M. E. M., Hilario, E. and Jacobson, F. (2006) 'Protein aggregation and bioprocessing', *AAPS Journal*. 2006/10/10, 8(3), pp. E572-9. doi: 10.1208/aapsj080366.

Cunningham, K. *et al.* (2002) 'Mapping the lethal factor and edema factor binding sites on oligomeric anthrax protective antigen', *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), pp. 7049–7053. doi: 10.1073/pnas.062160399.

D'Souza, A. J. M. *et al.* (2013) 'Rapid deamidation of recombinant protective antigen when

adsorbed on aluminum hydroxide gel correlates with reduced potency of vaccine', *Journal of Pharmaceutical Sciences*. 2012/12/18, 102(2), pp. 454–461. doi: 10.1002/jps.23422.

Dai, Z. *et al.* (1995) 'The atxA gene product activates transcription of the anthrax toxin genes and is essential for virulence', *Molecular Microbiology*, 16(6), pp. 1171–1181. doi: 10.1111/j.1365-2958.1995.tb02340.x.

De, S. and Olson, R. (2011) 'Crystal structure of the Vibrio cholerae cytolysin heptamer reveals common features among disparate pore-forming toxins', *Proceedings of the National Academy of Sciences of the United States of America*. 2011/04/18, 108(18), pp. 7385–7390. doi: 10.1073/pnas.1017442108.

Delany, I., Rappuoli, R. and De Gregorio, E. (2014) 'Vaccines for the 21st century', *EMBO Molecular Medicine*. 2014/04/06, 6(6), pp. 708–720. doi: 10.1002/emmm.201403876.

Dobson, C. M. (2003) 'Protein folding and misfolding', *Nature*. 2003/12/20, 426(6968), pp. 884–890. doi: 10.1038/nature02261.

Dror, A. *et al.* (2014) 'Protein engineering by random mutagenesis and structure-guided consensus of Geobacillus stearothermophilus lipase T6 for enhanced stability in methanol', *Applied and Environmental Microbiology*. 2013/12/20, 80(4), pp. 1515–1527. doi: 10.1128/AEM.03371-13.

Drysdale, M., Bourgogne, A. and Koehler, T. M. (2005) 'Transcriptional analysis of the Bacillus anthracis capsule regulators', *Journal of Bacteriology*, 187(15), pp. 5108–5114. doi: 10.1128/JB.187.15.5108-5114.2005.

Dubendorf, J. W. and Studier, F. W. (1991) 'Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor', *Journal of*

Molecular Biology, 219(1), pp. 45–59. doi: 10.1016/0022-2836(91)90856-2.

Van Durme, J. *et al.* (2016) 'Solubis: A webserver to reduce protein aggregation through mutation', *Protein Engineering, Design and Selection*. 2016/06/11, 29(8), pp. 285–289. doi: 10.1093/protein/gzw019.

Edelheit, O., Hanukoglu, A. and Hanukoglu, I. (2009) 'Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies', *BMC Biotechnology*, 9(1), p. 61. doi: 10.1186/1472-6750-9-61.

Eijsink, V. G. H. *et al.* (2005) 'Directed evolution of enzyme stability', *Biomolecular Engineering*, 22(1–3), pp. 21–30. doi: 10.1016/j.bioeng.2004.12.003.

Emanuele Campese, A. P. (2015) 'Genotyping of Bacillus anthracis Strains Circulating in Albania', *Journal of Bioterrorism & Biodefense*. Edited by L. Serrecchia. OMICS International., pp. 1–6. doi: 10.4172/2157-2526.1000131.

Erlanson, D. A. *et al.* (2016) 'Twenty years on: The impact of fragments on drug discovery', *Nature Reviews Drug Discovery*, 15(9), pp. 605–619. doi: 10.1038/nrd.2016.109.

Van Ert, M. N. *et al.* (2007) 'Global genetic population structure of Bacillus anthracis', *PLoS ONE*, 2(5), p. e461. doi: 10.1371/journal.pone.0000461.

Fabre, L. *et al.* (2016) 'Structure of anthrax lethal toxin prepore complex suggests a pathway for efficient cell entry', *Journal of General Physiology*, 148(4), pp. 313–324. doi: 10.1085/jgp.201611617.

Fang, H. *et al.* (2005) ' Anthrax Lethal Toxin Blocks MAPK Kinase-Dependent IL-2 Production in CD4 + T Cells ', *The Journal of Immunology*. 2005/04/09, 174(8), pp. 4966–

4971. doi: 10.4049/jimmunol.174.8.4966.

Farchaus, J. W. *et al.* (1998) ‘Fermentation, purification, and characterization of protective antigen from a recombinant, avirulent strain of Bacillus anthracis’, *Applied and environmental microbiology*, 64(3), pp. 982–991. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9501438.

Feinen, B. *et al.* (2014) ‘Advax-adjuvanted recombinant protective antigen provides protection against inhalational anthrax that is further enhanced by addition of murabutide adjuvant’, *Clinical and Vaccine Immunology*, 21(4), pp. 580–586. doi: 10.1128/CVI.00019-14.

Feld, G. K. *et al.* (2010) ‘Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers’, *Nature Structural and Molecular Biology*. 2010/11/03, 17(11), pp. 1383–1390. doi: 10.1038/nsmb.1923.

Feld, G. K. *et al.* (2012) ‘Domain flexibility modulates the heterogeneous assembly mechanism of anthrax toxin protective antigen’, *Journal of Molecular Biology*. 2011/10/31, 415(1), pp. 159–174. doi: 10.1016/j.jmb.2011.10.035.

Fernandez-Escamilla, A. M. *et al.* (2004) ‘Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins’, in *Nature Biotechnology*. United States, pp. 1302–1306. doi: 10.1038/nbt1012.

Firoved, A. M. *et al.* (2005) ‘Bacillus anthracis edema toxin causes extensive tissue lesions and rapid lethality in mice’, *American Journal of Pathology*, 167(5), pp. 1309–1320. doi: 10.1016/S0002-9440(10)61218-7.

Firth, A. E. and Patrick, W. M. (2005) ‘Statistics of protein library construction’,

Bioinformatics, 21(15), pp. 3314–3315. doi: 10.1093/bioinformatics/bti516.

Friedlander, A. M. (1986) 'Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process', *Journal of Biological Chemistry*, 261(16), pp. 7123–7126.

Ganesan, A. *et al.* (2016) 'Structural hot spots for the solubility of globular proteins', *Nature Communications*, 7, p. 10816. doi: 10.1038/ncomms10816.

Ganesan, A., Watkinson, A. and Moore, B. D. (2012) 'Biophysical characterisation of thermal-induced precipitates of recombinant anthrax protective antigen: Evidence for kinetically trapped unfolding domains in solid-state', *European Journal of Pharmaceutics and Biopharmaceutics*. 2012/06/12, 82(3), pp. 475–484. doi: 10.1016/j.ejpb.2012.05.019.

Gao-Sheridan, S., Zhang, S. and Collier, R. J. (2003) 'Exchange characteristics of calcium ions bound to anthrax protective antigen', *Biochemical and Biophysical Research Communications*, 300(1), pp. 61–64. doi: 10.1016/S0006-291X(02)02771-7.

Gasteiger, E. *et al.* (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in Walker, J. M. (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607. doi: 10.1385/1-59259-890-0:571.

Giuliani, M. M. *et al.* (2006) 'A universal vaccine for serogroup B meningococcus', *Proceedings of the National Academy of Sciences of the United States of America*. 2006/07/06, 103(29), pp. 10834–10839. doi: 10.1073/pnas.0603940103.

Goel, A. K. (2015) 'Anthrax: A disease of biowarfare and public health importance', *World Journal of Clinical Cases*, 3(1), p. 20. doi: 10.12998/wjcc.v3.i1.20.

Goldman, D. L. (2014) 'Molecular determinants for a cardiovascular collapse in anthrax',

Frontiers in Bioscience, E6(1), pp. 139–147. doi: 10.2741/e697.

Greenspan, P. and Fowler, S. D. (1985) 'Spectrofluorometric studies of the lipid probe, nile red', *Journal of Lipid Research*, 26(7), pp. 781–789. Available at: http://www.jlr.org/content/26/7/781.abstract.

Guerois, R., Nielsen, J. E. and Serrano, L. (2002) 'Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations', *Journal of Molecular Biology*, 320(2), pp. 369–387. doi: 10.1016/S0022-2836(02)00442-4.

Le Guilloux, V., Schmidtke, P. and Tuffery, P. (2009) 'Fpocket: An open source platform for ligand pocket detection', *BMC Bioinformatics*, 10(1), p. 168. doi: 10.1186/1471-2105-10-168.

Gupta, P. K. *et al.* (2003) 'Conformational fluctuations in anthrax protective antigen: A possible role of calcium in the folding pathway of the protein', *FEBS Letters*, 554(3), pp. 505–510. doi: 10.1016/S0014-5793(03)01226-2.

Gupta, P., Waheed, S. M. and Bhatnagar, R. (1999) 'Expression and purification of the recombinant protective antigen of Bacillus anthracis', *Protein Expression and Purification*. 1999/07/30, 16(3), pp. 369–376. doi: 10.1006/prep.1999.1066.

Han, N. *et al.* (2018) 'Enhancing thermal tolerance of Aspergillus niger PhyA phytase directed by structural comparison and computational simulation', *BMC Biotechnology*, 18(1), p. 36. doi: 10.1186/s12896-018-0445-y.

Hassett, K. J. *et al.* (2015) 'Glassy-state stabilization of a dominant negative inhibitor anthrax vaccine containing aluminum hydroxide and glycopyranoside lipid a adjuvants', *Journal of Pharmaceutical Sciences*, 104(2), pp. 627–639. doi: 10.1002/jps.24295.

Hawe, A., Sutter, M. and Jiskoot, W. (2008) ‘Extrinsic fluorescent dyes as tools for protein characterization’, in *Pharmaceutical Research*, pp. 1487–1499. doi: 10.1007/s11095-007-9516-9.

He, F. *et al.* (2010) ‘High throughput thermostability screening of monoclonal antibody formulations’, *Journal of Pharmaceutical Sciences*, 99(4), pp. 1707–1720. doi: 10.1002/jps.21955.

Head, B. M., Rubinstein, E. and Meyers, A. F. (2016) ‘Alternative pre-approved and novel therapies for the treatment of anthrax’, *BMC Infect Dis*. 2016/11/05, 16(1), p. 621. doi: 10.1186/s12879-016-1951-y.

Holty, J. E. C. *et al.* (2006) ‘Systematic review: A century of inhalational anthrax cases from 1900 to 2005’, *Annals of Internal Medicine*, 144(4), pp. 270–280. doi: 10.7326/0003-4819-144-4-200602210-00009.

Hsueh, Y. H. *et al.* (2017) ‘Poly-γ-glutamic acid synthesis, gene regulation, phylogenetic relationships, and role in fermentation’, *International Journal of Molecular Sciences*, 18(12), p. 2644. doi: 10.3390/ijms18122644.

Hughes (2002) ‘Anthrax bioterrorism: Lessons learned and future directions’, in *Emerging Infectious Diseases*. Centers for Disease Control and Prevention, Atlanta, GA, USA, pp. 1013–1014. doi: 10.3201/eid0810.020466.

Ionin, B. *et al.* (2013) ‘Evaluation of immunogenicity and efficacy of anthrax vaccine adsorbed for postexposure prophylaxis’, *Clinical and Vaccine Immunology*. 2013/05/08, 20(7), pp. 1016–1026. doi: 10.1128/CVI.00099-13.

Ivins, B. E. *et al.* (1998) ‘Comparative efficacy of experimental anthrax vaccine candidates

against inhalation anthrax in rhesus macaques', *Vaccine*, 16(11–12), pp. 1141–1148. doi: 10.1016/S0264-410X(98)80112-6.

Ivins, B. E. and Welkos, S. L. (1986) 'Cloning and expression of the Bacillus anthracis protective antigen gene in Bacillus subtilis', *Infection and Immunity*, 54(2), pp. 537–542. doi: 10.1128/iai.54.2.537-542.1986.

Iyer, V. *et al.* (2013) 'Biophysical characterization and immunization studies of dominant negative inhibitor (DNI), a candidate anthrax toxin subunit vaccine', in *Human Vaccines and Immunotherapeutics*, pp. 2362–2370. doi: 10.4161/hv.25852.

Janowiak, B. E., Fischer, A. and Collier, R. J. (2010) 'Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen', *Journal of Biological Chemistry*. 2010/01/08, 285(11), pp. 8130–8137. doi: 10.1074/jbc.M109.093195.

Jiang, J. *et al.* (2015) 'Atomic structure of anthrax protective antigen pore elucidates toxin translocation', *Nature*. 2015/03/18, 521(7553), pp. 545–549. doi: 10.1038/nature14247.

Joo, J. C. *et al.* (2010) 'Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity', *Journal of Biotechnology*, 146(1–2), pp. 31–39. doi: 10.1016/j.jbiotec.2009.12.021.

Jubb, H. C. *et al.* (2017) 'Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures', *Journal of Molecular Biology*, 429(3), pp. 365–371. doi: 10.1016/j.jmb.2016.12.004.

Kachura, M. A. *et al.* (2016) 'A CpG-Ficoll Nanoparticle Adjuvant for Anthrax Protective Antigen Enhances Immunogenicity and Provides Single-Immunization Protection against Inhaled Anthrax in Monkeys', *The Journal of Immunology*. 2015/11/25, 196(1), pp. 284–

297. doi: 10.4049/jimmunol.1501903.

Kaur, M., Singh, S. and Bhatnagar, R. (2013) ‘Anthrax vaccines: Present status and future prospects’, *Expert Review of Vaccines*. 2013/08/30, 12(8), pp. 955–970. doi: 10.1586/14760584.2013.814860.

Kelly-Cirino, C. D. and Mantis, N. J. (2009) ‘Neutralizing monoclonal antibodies directed against defined linear epitopes on domain 4 of anthrax protective antigen’, *Infection and Immunity*. 2009/08/26, 77(11), pp. 4859–4867. doi: 10.1128/IAI.00117-09.

Khayat, R. *et al.* (2013) ‘Structural Characterization of Cleaved, Soluble HIV-1 Envelope Glycoprotein Trimers’, *Journal of Virology*. 2013/07/03, 87(17), pp. 9865–9872. doi: 10.1128/jvi.01222-13.

Kintzer, A. F. *et al.* (2009) ‘The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes’, *Journal of Molecular Biology*, 392(3), pp. 614–629. doi: 10.1016/j.jmb.2009.07.037.

Krebs, M. R. H., Bromley, E. H. C. and Donald, A. M. (2005) ‘The binding of thioflavin-T to amyloid fibrils: Localisation and implications’, *Journal of Structural Biology*. 2005/01/05, 149(1), pp. 30–37. doi: 10.1016/j.jsb.2004.08.002.

Kurahashi, R., Tanaka, S. ichi and Takano, K. (2019) ‘Activity-stability trade-off in random mutant proteins’, *Journal of Bioscience and Bioengineering*, 128(4), pp. 405–409. doi: 10.1016/j.jbiosc.2019.03.017.

Lacy, D. B. *et al.* (2004) ‘Structure of heptameric protective antigen bound to an anthrax toxin receptor: A role for receptor in pH-dependent pore formation’, *Proceedings of the National Academy of Sciences*. 2004/08/25, 101(36), pp. 13147–13151. doi:

10.1073/pnas.0405405101.

Laforce, F. M. (1978) ‘Woolsorters’ disease in England’, *Bulletin of the New York Academy of Medicine: Journal of Urban Health*, 54(10), pp. 956–963. Available at: https://www.ncbi.nlm.nih.gov/pubmed/361121.

Laird, M. W. *et al.* (2004) ‘Production and purification of Bacillus anthracis protective antigen from Escherichia coli’, *Protein Expression and Purification*. 2004/10/13, 38(1), pp. 145–152. doi: 10.1016/j.pep.2004.08.007.

Lamazares, E. *et al.* (2015) ‘Rational stabilization of complex proteins: A divide and combine approach’, *Scientific Reports*, 5, p. 9129. doi: 10.1038/srep09129.

Laws, T. R. *et al.* (2016) ‘A comparison of the adaptive immune response between recovered anthrax patients and individuals receiving three different anthrax vaccines’, *PLoS ONE*, 11(3), pp. e0148713–e0148713. doi: 10.1371/journal.pone.0148713.

Lei, S. P. *et al.* (1987) ‘Characterization of the Erwinia carotovora pelB gene and its product pectate lyase.’, *Journal of bacteriology*, 169(9), pp. 4379–4383. doi: 10.1128/jb.169.9.4379-4383.1987.

Leppla, S. H. (1988) ‘Production and Purification of Anthrax Toxin’, in *Methods in Enzymology*. Academic Press, pp. 103–116. doi: 10.1016/S0076-6879(88)65019-1.

Li, G. *et al.* (2018) ‘Enhancing the thermostability of Rhizomucor miehei lipase with a limited screening library by rational-design point mutations and disulfide bonds’, *Applied and Environmental Microbiology*, 84(2), pp. e02129-17. doi: 10.1128/AEM.02129-17.

Liang, X. *et al.* (2016) ‘Involvement of the pagR gene of pXO2 in anthrax pathogenesis’,

Scientific Reports, 6, p. 28827. doi: 10.1038/srep28827.

Liang, X. *et al.* (2017) 'The pag Gene of pXO1 Is Involved in Capsule Biosynthesis of Bacillus anthracis Pasteur II Strain', *Frontiers in Cellular and Infection Microbiology*, 7, p. 203. doi: 10.3389/fcimb.2017.00203.

Liu, H. *et al.* (2016) 'A single residue substitution accounts for the significant difference in thermostability between two isoforms of human cytosolic creatine kinase', *Scientific Reports*, 6, p. 21191. doi: 10.1038/srep21191.

Liu, S. *et al.* (2009) 'Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin in vivo', *Proceedings of the National Academy of Sciences*, 106(30), pp. 12424–12429. doi: 10.1073/pnas.0905409106.

Liu, S. *et al.* (2010) 'Anthrax toxin targeting of myeloid cells through the CMG2 receptor is essential for establishment of bacillus anthracis infections in mice', *Cell Host and Microbe*, 8(5), pp. 455–462. doi: 10.1016/j.chom.2010.10.004.

Liu, S. *et al.* (2013) 'Key tissue targets responsible for anthrax-toxin-induced lethality', *Nature*, 501(7465), pp. 63–68. doi: 10.1038/nature12510.

Liu, S. *et al.* (2015) 'Bacillus anthracis toxins', in Alouf, J., Ladant, D., and Popoff, M. R. B. T.-T. C. S. of B. P. T. (Fourth E. (eds) *The Comprehensive Sourcebook of Bacterial Protein Toxins*. Boston: Academic Press, pp. 361–396. doi: 10.1016/B978-0-12-800188-2.00013-6.

Liu, S., Moayeri, M. and Leppla, S. H. (2014) 'Anthrax lethal and edema toxins in anthrax pathogenesis', *Trends in Microbiology*. 2014/04/02, 22(6), pp. 317–325. doi: 10.1016/j.tim.2014.02.012.

Lu, J. *et al.* (2009) 'High-level expression and single-step purification of recombinant Bacillus anthracis protective antigen from Escherichia coli', *Biotechnology and Applied Biochemistry*. 2008/04/18, 52(2), p. 107. doi: 10.1042/ba20070245.

Lu, Y., Welsh, J. P. and Swartz, J. R. (2014) 'Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines', *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), pp. 125–130. doi: 10.1073/pnas.1308701110.

Lutz, S. (2010) 'Beyond directed evolution-semi-rational protein engineering and design', *Current Opinion in Biotechnology*. 2010/09/24, 21(6), pp. 734–743. doi: 10.1016/j.copbio.2010.08.011.

Majumder, S. *et al.* (2019) 'A bivalent protein R-PabXPB comprising PA domain IV and exosporium protein BxpB confers protection against B. Anthracis Spores and Toxin', *Frontiers in Immunology*, p. 498. doi: 10.3389/fimmu.2019.00498.

Makino, S. *et al.* (2002) ' Effect of the Lower Molecular Capsule Released from the Cell Surface of Bacillus anthracis on the Pathogenesis of Anthrax ', *The Journal of Infectious Diseases*, 186(2), pp. 227–233. doi: 10.1086/341299.

Malito, E., Carfi, A. and Bottomley, M. J. (2015) 'Protein crystallography in vaccine research and development', *International Journal of Molecular Sciences*, 16(6), pp. 13106–13140. doi: 10.3390/ijms160613106.

Manning, M. C. *et al.* (2010) 'Stability of protein pharmaceuticals: An update', *Pharmaceutical Research*. 2010/02/10, 27(4), pp. 544–575. doi: 10.1007/s11095-009-0045-6.

Marston, C. K. *et al.* (2011) ‘Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States’, *PLoS ONE*. 2011/12/09, 6(12), pp. e28274–e28274. doi: 10.1371/journal.pone.0028274.

Maugini, E. *et al.* (2009) ‘Structural adaptation of the subunit interface of oligomeric thermophilic and hyperthermophilic enzymes’, *Computational Biology and Chemistry*, 33(2), pp. 137–148. doi: 10.1016/j.compbiolchem.2008.08.003.

McComb, R. C. and Martchenko, M. (2016) ‘Neutralizing antibody and functional mapping of Bacillus anthracis protective antigen-The first step toward a rationally designed anthrax vaccine’, *Vaccine*. 2015/11/28, 34(1), pp. 13–19. doi: 10.1016/j.vaccine.2015.11.025.

Meisl, G. *et al.* (2016) ‘Molecular mechanisms of protein aggregation from global fitting of kinetic models’, *Nature Protocols*, 11(2), pp. 252–272. doi: 10.1038/nprot.2016.010.

Merkel, T. J. *et al.* (2013) ‘Protective-antigen (PA) based anthrax vaccines confer protection against inhalation anthrax by precluding the establishment of a systemic infection’, *Human Vaccines and Immunotherapeutics*, 9(9), pp. 1841–1848. doi: 10.4161/hv.25337.

Mishra, R., Sjölander, D. and Hammarström, P. (2011) ‘Spectroscopic characterization of diverse amyloid fibrils in vitro by the fluorescent dye Nile red’, *Molecular BioSystems*, 7(4), pp. 1232–1240. doi: 10.1039/c0mb00236d.

Missiakas, D. and Schneewind, O. (2017) ‘ Assembly and Function of the Bacillus anthracis S-Layer ’, *Annual Review of Microbiology*, 71(1), pp. 79–98. doi: 10.1146/annurev-micro-090816-093512.

Moayeri, M., Wiggins, J. F. and Leppla, S. H. (2007) ‘Anthrax protective antigen cleavage and clearance from the blood of mice and rats’, *Infection and Immunity*, 75(11), pp. 5175–

5184. doi: 10.1128/IAI.00719-07.

Morris, A. M., Watzky, M. A. and Finke, R. G. (2009) 'Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1794(3), pp. 375–397. doi: 10.1016/j.bbapap.2008.10.016.

Mourez, M. *et al.* (2003) 'Mapping dominant-negative mutations of anthrax protective antigen by scanning mutagenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 100(SUPPL. 2), pp. 13803–13808. doi: 10.1073/pnas.2436299100.

Mullangi, V. *et al.* (2014) 'Long-range stabilization of anthrax protective antigen upon binding to CMG2', *Biochemistry*, 53(38), pp. 6084–6091. doi: 10.1021/bi500718g.

Ngundi, M. M. *et al.* (2010) 'Comparison of three anthrax toxin neutralization assays', *Clinical and Vaccine Immunology*. 2010/04/07, 17(6), pp. 895–903. doi: 10.1128/CVI.00513-09.

Northey, J. G. B., Di Nardo, A. A. and Davidson, A. R. (2002) 'Hydrophobic core packing in the sh3 domain folding transition state', *Nature Structural Biology*, 9(2), pp. 126–130. doi: 10.1038/nsb748.

Novo, M., Freire, S. and Al-Soufi, W. (2018) 'Critical aggregation concentration for the formation of early Amyloid-β (1-42) oligomers', *Scientific Reports*, 8(1), p. 1783. doi: 10.1038/s41598-018-19961-3.

Okello, A., Welburn, S. and Smith, J. (2015) 'Crossing institutional boundaries: Mapping the policy process for improved control of endemic and neglected zoonoses in sub-Saharan Africa', *Health Policy and Planning*, 30(6), pp. 804–812. doi: 10.1093/heapol/czu059.

Omersa, N., Podobnik, M. and Anderluh, G. (2019) ‘Inhibition of pore-forming proteins’, *Toxins*, 11(9), p. 545. doi: 10.3390/toxins11090545.

Omland, K. S. *et al.* (2008) ‘Interlaboratory comparison of results of an anthrax lethal toxin neutralization assay for assessment of functional antibodies in multiple species’, *Clinical and Vaccine Immunology*. 2008/04/16, 15(6), pp. 946–953. doi: 10.1128/CVI.00003-08.

Ota, N. *et al.* (2018) ‘The direction of protein evolution is destined by the stability’, *Biochimie*, 150, pp. 100–109. doi: 10.1016/j.biochi.2018.05.006.

Pace, C. N. *et al.* (1996) ‘Forces contributing to the conformational stability of proteins’, *FASEB Journal*. 1996/01/01, 10(1), pp. 75–83. doi: 10.1096/fasebj.10.1.8566551.

Packer, M. S. and Liu, D. R. (2015) ‘Methods for the directed evolution of proteins’, *Nature Reviews Genetics*. 2015/06/10, 16(7), pp. 379–394. doi: 10.1038/nrg3927.

Pallesen, J. *et al.* (2017) ‘Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen’, *Proceedings of the National Academy of Sciences of the United States of America*, 114(35), pp. E7348–E7357. doi: 10.1073/pnas.1707304114.

Petosa, C. *et al.* (1997) ‘Crystal structure of the anthrax toxin protective antigen’, *Nature*. 1997/02/27, 385(6619), pp. 833–838. doi: 10.1038/385833a0.

Pilpa, R. M. *et al.* (2011) ‘A receptor-based switch that regulates anthrax toxin pore formation’, *PLoS Pathogens*. 2011/12/08, 7(12), pp. e1002354–e1002354. doi: 10.1371/journal.ppat.1002354.

Plaut, R. D. *et al.* (2018) ‘Avirulent bacillus anthracis strain with molecular assay targets as surrogate for irradiation-inactivated virulent spores’, *Emerging Infectious Diseases*, 24(4),

pp. 691–699. doi: 10.3201/eid2404.171646.

Pomerantsev, A. P. *et al.* (2006) ‘Genome engineering in Bacillus anthracis using cre recombinase’, *Infection and Immunity*, 74(1), pp. 682–693. doi: 10.1128/IAI.74.1.682-693.2006.

Pomerantsev, A. P. *et al.* (2011) ‘A Bacillus anthracis strain deleted for six proteases serves as an effective host for production of recombinant proteins’, *Protein Expression and Purification*, 80(1), pp. 80–90. doi: 10.1016/j.pep.2011.05.016.

Pomerantsev, A. P. *et al.* (2017) ‘Genome engineering in Bacillus anthracis using tyrosine site-specific recombinases’, *PLoS ONE*, 12(8), p. e0183346. doi: 10.1371/journal.pone.0183346.

Powell, B. S. *et al.* (2007) ‘Multiple asparagine deamidation of Bacillus anthracis protective antigen causes charge isoforms whose complexity correlates with reduced biological activity’, *Proteins: Structure, Function and Genetics*. 2007/05/01, 68(2), pp. 458–479. doi: 10.1002/prot.21432.

Pucci, F. and Rooman, M. (2017) ‘Physical and molecular bases of protein thermal stability and cold adaptation’, *Current Opinion in Structural Biology*, 42, pp. 117–128. doi: 10.1016/j.sbi.2016.12.007.

Pullan, S. T. *et al.* (2015) ‘Whole-genome sequencing investigation of animal-skin-drum-associated UK anthrax cases reveals evidence of mixed populations and relatedness to a US case’, *Microbial Genomics*, 1(5), p. doi: 10.1099/mgen.0.000039.

Puziss, M. *et al.* (1963) ‘Large-scale production of protective antigen of Bacillus anthracis in anaerobic cultures.’, *Applied microbiology*, 11(4), pp. 330–334. doi:

10.1128/aem.11.4.330-334.1963.

Quinn, C. P. *et al.* (2004) ‘ Immune Responses to Bacillus anthracis Protective Antigen in Patients with Bioterrorism-Related Cutaneous or Inhalation Anthrax ’, *The Journal of Infectious Diseases*, 190(7), pp. 1228–1236. doi: 10.1086/423937.

Radestock, S. and Gohlke, H. (2011) ‘Protein rigidity and thermophilic adaptation’, *Proteins: Structure, Function and Bioinformatics*, 79(4), pp. 1089–1108. doi: 10.1002/prot.22946.

Radha, C. *et al.* (1996) ‘Thermostabilization of protective antigen - The binding component of anthrax lethal toxin’, *Journal of Biotechnology*, 50(2–3), pp. 235–242. doi: 10.1016/0168-1656(96)01569-6.

Radzicka, A. and Wolfenden, R. (1988) ‘Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous Solution’, *Biochemistry*, 27(5), pp. 1664–1670. doi: 10.1021/bi00405a042.

Rappuoli, R. *et al.* (2016) ‘Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design’, *Journal of Experimental Medicine*. 2016/03/28, 213(4), pp. 469–481. doi: 10.1084/JEM.20151960.

Rathi, P. C. *et al.* (2016) ‘Application of Rigidity Theory to the Thermostabilization of Lipase A from Bacillus subtilis’, *PLoS Computational Biology*, 12(3), pp. e1004754–e1004754. doi: 10.1371/journal.pcbi.1004754.

Razvi, A. and Scholtz, J. M. (2006) ‘Lessons in stability from thermophilic proteins’, *Protein Science*. 2006/07/04, 15(7), pp. 1569–1578. doi: 10.1110/ps.062130306.

Reeves, C., Charles-Horvath, P. and Kitajewski, J. (2013) ‘Studies in mice reveal a role for

anthrax toxin receptors in matrix metalloproteinase function and extracellular matrix homeostasis', *Toxins*, 5(2), pp. 315–326. doi: 10.3390/toxins5020315.

Ribot, W. J. *et al.* (2006) 'Comparative vaccine efficacy of different isoforms of recombinant protective antigen against Bacillus anthracis spore challenge in rabbits', *Vaccine*, 24(17), pp. 3469–3476. doi: 10.1016/j.vaccine.2006.02.013.

Roberts, C. J. (2007) 'Non-native protein aggregation kinetics', *Biotechnology and Bioengineering*, 98(5), pp. 927–938. doi: 10.1002/bit.21627.

Rose, G. D. *et al.* (1985) 'Hydrophobicity of amino acid residues in globular proteins', *Science*, 229(4716), pp. 834–838. doi: 10.1126/science.4023714.

Rosovitz, M. J. *et al.* (2003) 'Alanine-scanning mutations in domain 4 of anthrax toxin protective antigen reveal residues important for binding to the cellular receptor and to a neutralizing monoclonal antibody', *Journal of Biological Chemistry*. 2003/05/29, 278(33), pp. 30936–30944. doi: 10.1074/jbc.M301154200.

Roth, T. *et al.* (2017) 'Thermostabilization of the uronate dehydrogenase from Agrobacterium tumefaciens by semi-rational design', *AMB Express*. 2017/05/23, 7(1), p. 103. doi: 10.1186/s13568-017-0405-2.

Sackett, D. L. and Wolff, J. (1987) 'Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces', *Analytical Biochemistry*. 1987/12/01, 167(2), pp. 228–234. doi: 10.1016/0003-2697(87)90157-6.

Saile, E. and Koehler, T. M. (2006) 'Bacillus anthracis multiplication, persistence, and genetic exchange in the rhizosphere of grass plants', in *Applied and Environmental Microbiology*. Department of Microbiology and Molecular Genetics, The University of

Texas—Houston Health Science Center, Houston, Texas 77030, pp. 3168–3174. doi: 10.1128/AEM.72.5.3168-3174.2006.

Sammond, D. W. *et al.* (2016) 'Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms', *PLoS ONE*, 11(1), pp. e0145848–e0145848. doi: 10.1371/journal.pone.0145848.

Samples, J., Willis, M. and Klauber-DeMore, N. (2013) 'Targeting Angiogenesis and the Tumor Microenvironment', *Surgical Oncology Clinics of North America*, 22(4), pp. 629–639. doi: 10.1016/j.soc.2013.06.002.

Sarvas, M. *et al.* (2004) 'Post-translocational folding of secretory proteins in Gram-positive bacteria', *Biochimica et Biophysica Acta - Molecular Cell Research*. 2004/11/18, 1694(1-3 SPEC.ISS.), pp. 311–327. doi: 10.1016/j.bbamcr.2004.04.009.

Savransky, V. *et al.* (2017) 'Correlation between anthrax lethal toxin neutralizing antibody levels and survival in guinea pigs and nonhuman primates vaccinated with the AV7909 anthrax vaccine candidate', *Vaccine*. 2017/07/31, 35(37), pp. 4952–4959. doi: 10.1016/j.vaccine.2017.07.076.

Savransky, V. *et al.* (2019) 'Repeat-Dose Toxicity Study of a Lyophilized Recombinant Protective Antigen-Based Anthrax Vaccine Adjuvanted With CpG 7909', *International Journal of Toxicology*, 38(3), pp. 163–172. doi: 10.1177/1091581819848722.

Schägger, H. and von Jagow, G. (1991) 'Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form', *Analytical Biochemistry*, 199(2), pp. 223–231. doi: 10.1016/0003-2697(91)90094-A.

Scheiblhofer, S. *et al.* (2017) 'Influence of protein fold stability on immunogenicity and its

implications for vaccine design', *Expert Review of Vaccines*. 2017/03/24, 16(5), pp. 479–489. doi: 10.1080/14760584.2017.1306441.

Schiffmiller, A., Anderson, D. and Finkelstein, A. (2015) 'Ion selectivity of the anthrax toxin channel and its effect on protein translocation', *Journal of General Physiology*, 146(2), pp. 183–192. doi: 10.1085/jgp.201511388.

Schreyer, A. M. and Blundell, T. L. (2013) 'CREDO: A structural interactomics database for drug discovery', *Database*, 2013. doi: 10.1093/database/bat049.

Schymkowitz, J. *et al.* (2005) 'The FoldX web server: An online force field', *Nucleic Acids Research*. 2005/06/28, 33(SUPPL. 2), pp. W382-8. doi: 10.1093/nar/gki387.

Scobie, H. M. *et al.* (2005) 'A Soluble Receptor Decoy Protects Rats against Anthrax Lethal Toxin Challenge', *The Journal of Infectious Diseases*, 192(6), pp. 1047–1051. doi: 10.1086/432731.

Sharma, A. K. *et al.* (2018) 'Effect of over expressing protective antigen on global gene transcription in Bacillus anthracis BH500', *Scientific Reports*, 8(1), p. 16108. doi: 10.1038/s41598-018-34196-y.

Sharma, M. *et al.* (1996) 'Expression and purification of anthrax toxin protective antigen from Escherichia coli', *Protein Expression and Purification*. 1996/02/01, 7(1), pp. 33–38. doi: 10.1006/prep.1996.0005.

Simonson, T. S. *et al.* (2009) 'Bacillus anthracis in China and its relationship to worldwide lineages', *BMC Microbiology*, 9(1), p. 71. doi: 10.1186/1471-2180-9-71.

Singh, P. *et al.* (2013) 'Effect of Signal Peptide on Stability and Folding of Escherichia coli

Thioredoxin', *PLoS ONE*, 8(5), p. e63442. doi: 10.1371/journal.pone.0063442.

Singh, S. *et al.* (2004) 'Thermal inactivation of protective antigen of Bacillus anthracis and its prevention by polyol osmolytes', *Biochemical and Biophysical Research Communications*, 322(3), pp. 1029–1037. doi: 10.1016/j.bbrc.2004.08.020.

Singh, Y. *et al.* (1994) 'The chymotrypsin-sensitive site, FFD315, in anthrax toxin protective antigen is required for translocation of lethal factor', *Journal of Biological Chemistry*. 1994/11/18, 269(46), pp. 29039–29046. Available at: http://dx.doi.org/.

Smith, K. *et al.* (2012) 'Human monoclonal antibodies generated following vaccination with AVA provide neutralization by blocking furin cleavage but not by preventing oligomerization', *Vaccine*. 2012/03/20, 30(28), pp. 4276–4283. doi: 10.1016/j.vaccine.2012.03.002.

Soliakov, A. *et al.* (2012) 'Anthrax sub-unit vaccine: The structural consequences of binding rPA83 to Alhydrogel®', *European Journal of Pharmaceutics and Biopharmaceutics*, 80(1), pp. 25–32. doi: 10.1016/j.ejpb.2011.09.009.

Sorrell, F. J. *et al.* (2010) 'Development of a differential scanning fluorimetry based high throughput screening assay for the discovery of affinity binders against an anthrax protein.', *Journal of pharmaceutical and biomedical analysis*. 2010/04/09, 52(5), pp. 802–808. doi: 10.1016/j.jpba.2010.02.024.

Sternbach, G. (2003) 'The history of anthrax', *Journal of Emergency Medicine*, 24(4), pp. 463–467. doi: 10.1016/S0736-4679(03)00079-9.

Storm, L. *et al.* (2018) 'Anthrax protective antigen is a calcium-dependent serine protease', *Virulence*, 9(1), pp. 1085–1091. doi: 10.1080/21505594.2018.1486139.

Sumbalova, L. *et al.* (2018) 'HotSpot Wizard 3.0: Web server for automated design of mutations and smart libraries based on sequence input information', *Nucleic Acids Research*, 46(W1), pp. W356–W362. doi: 10.1093/nar/gky417.

Sumi, T. and Koga, K. (2019) 'Theoretical analysis on thermodynamic stability of chignolin', *Scientific Reports*, 9(1), p. 5186. doi: 10.1038/s41598-019-41518-1.

Suryanarayana, N. *et al.* (2016) ' Soluble Expression and Characterization of Biologically Active Bacillus anthracis Protective Antigen in Escherichia coli ', *Molecular Biology International*. 2016/03/12, 2016, pp. 1–11. doi: 10.1155/2016/4732791.

Tee, K. L. and Wong, T. S. (2013) 'Polishing the craft of genetic diversity creation in directed evolution', *Biotechnology Advances*, 31(8), pp. 1707–1721. doi: 10.1016/j.biotechadv.2013.08.021.

Thiltgen, G. and Goldstein, R. A. (2012) 'Assessing Predictors of Changes in Protein Stability upon Mutation Using Self-Consistency', *PLoS ONE*, 7(10), p. e46084. doi: 10.1371/journal.pone.0046084.

Timofeev, V. *et al.* (2018) 'Insights from Bacillus anthracis strains isolated from permafrost in the tundra zone of Russia', *bioRxiv*, 14(5), p. 486290. doi: 10.1101/486290.

Tokuriki, N. *et al.* (2007) 'The Stability Effects of Protein Mutations Appear to be Universally Distributed', *Journal of Molecular Biology*, 369(5), pp. 1318–1332. doi: 10.1016/j.jmb.2007.03.069.

Toyomane, K. *et al.* (2019) ' Upstream sequence-dependent suppression and AtxA-dependent activation of protective antigens in Bacillus anthracis ', *PeerJ*, 7, p. e6718. doi: 10.7717/peerj.6718.

Tuan-Anh, T. *et al.* (2017) 'Novel methods to optimize gene and statistic test for evaluation - an application for Escherichia coli', *BMC Bioinformatics*, 18(1), p. 100. doi: 10.1186/s12859-017-1517-z.

Turnbull, P. C. B. (1991) 'Anthrax vaccines: past, present and future', *Vaccine*, 9(8), pp. 533–539. doi: https://doi.org/10.1016/0264-410X(91)90237-Z.

Turnbull, P. C. B. (2000) 'Current status of immunization against anthrax: Old vaccines may be here to stay for a while', *Current Opinion in Infectious Diseases*. 2002/04/20, 13(2), pp. 113–120. doi: 10.1097/00001432-200004000-00004.

Uchida, I. *et al.* (1997) 'Cross-talk to the genes for Bacillus anthracis capsule synthesis by atxA, the gene encoding the trans-activator of anthrax toxin synthesis', *Molecular Microbiology*, 23(6), pp. 1229–1240. doi: 10.1046/j.1365-2958.1997.3041667.x.

Uchida, M. *et al.* (2012) 'Protective effect of Bacillus anthracis surface protein EA1 against anthrax in mice', *Biochemical and Biophysical Research Communications*. 2012/04/18, 421(2), pp. 323–328. doi: 10.1016/j.bbrc.2012.04.007.

Vázquez-Laslop, N. *et al.* (2001) 'Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked Escherichia coli', in *Journal of Bacteriology*, pp. 2399–2404. doi: 10.1128/JB.183.8.2399-2404.2001.

Verma, A. *et al.* (2013) 'Use of site-directed mutagenesis to model the effects of spontaneous deamidation on the immunogenicity of Bacillus anthracis protective antigen', *Infection and Immunity*. 2012/11/02, 81(1), pp. 278–284. doi: 10.1128/IAI.00863-12.

Verma, A. and Burns, D. L. (2018) 'Improving the stability of recombinant anthrax protective antigen vaccine', *Vaccine*, 36(43), pp. 6379–6382. doi: 10.1016/j.vaccine.2018.09.012.

Verma, A., Ngundi, M. M. and Burns, D. L. (2016) ‘Mechanistic analysis of the effect of deamidation on the immunogenicity of anthrax protective antigen’, *Clinical and Vaccine Immunology*. 2016/02/26, 23(5), pp. 396–402. doi: 10.1128/CVI.00701-15.

Vieille, C. and Zeikus, G. J. (2001) ‘Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability’, *Microbiology and Molecular Biology Reviews*. 2001/03/10, 65(1), pp. 1–43. doi: 10.1128/mmbr.65.1.1-43.2001.

Vodkin, M. H. and Leppla, S. H. (1983) ‘Cloning of the protective antigen gene of Bacillus anthracis’, *Cell*, 34(2), pp. 693–697. doi: 10.1016/0092-8674(83)90402-6.

Wagner, L. *et al.* (2012) ‘Structural and immunological analysis of anthrax recombinant protective antigen adsorbed to aluminum hydroxide adjuvant’, in *Clinical and Vaccine Immunology*. Center for Biologics Evaluation and Research, Food and Drug Administration Bethesda, Maryland, USAMeade Biologics, Hillsborough, North Carolina, USALaboratory of Malaria Immunology and Vaccinology, National Institute of Allergy and Infectious Diseases, Na, pp. 1465–1473. doi: 10.1128/CVI.00174-12.

Walls, A. C. *et al.* (2020) ‘Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein’, *Cell*, 181(2), pp. 281-292.e6. doi: 10.1016/j.cell.2020.02.058.

Wan, C. *et al.* (2019) ‘Rational Design of a Chimeric Derivative of PcrV as a Subunit Vaccine Against Pseudomonas aeruginosa’, *Frontiers in immunology*, 10, p. 781. doi: 10.3389/fimmu.2019.00781.

Wang, S. H. *et al.* (2012) ‘Stable dry powder formulation for nasal delivery of anthrax vaccine’, *Journal of Pharmaceutical Sciences*. 2011/09/08, 101(1), pp. 31–47. doi: 10.1002/jps.22742.

Warburton, M. *et al.* (2015) 'OneClick: A Program for Designing Focused Mutagenesis Experiments', *AIMS Bioengineering*, 2(3), pp. 126–143. doi: 10.3934/bioeng.2015.3.126.

Watkinson, A. *et al.* (2013) 'Increasing the Potency of an Alhydrogel-Formulated Anthrax Vaccine by Minimizing Antigen-Adjuvant Interactions', *Clinical and Vaccine Immunology*, 20(11), pp. 1659–1668. doi: 10.1128/cvi.00320-13.

Weir, G. M. *et al.* (2019) 'Single dose of DPX-rPA, an enhanced-delivery anthrax vaccine formulation, protects against a lethal Bacillus anthracis spore inhalation challenge', *npj Vaccines*, 4(1), p. 6. doi: 10.1038/s41541-019-0102-z.

Weiss, S. *et al.* (2006) 'Immunological correlates for protection against intranasal challenge of Bacillus anthracis spores conferred by a protective antigen-based vaccine in rabbits', *Infection and Immunity*. 2005/12/22, 74(1), pp. 394–398. doi: 10.1128/IAI.74.1.394-398.2006.

Whiting, G. C. *et al.* (2004) 'Characterisation of adsorbed anthrax vaccine by two-dimensional gel electrophoresis', *Vaccine*, 22(31–32), pp. 4245–4251. doi: 10.1016/j.vaccine.2004.04.036.

Whiting, G., Wheeler, J. X. and Rijpkema, S. (2014) 'Identification of peptide sequences as a measure of Anthrax vaccine stability during storage', *Human Vaccines and Immunotherapeutics*. 2014/03/19, 10(6), pp. 1669–1681. doi: 10.4161/hv.28443.

Wigelsworth, D. J. *et al.* (2004) 'Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen', *Journal of Biological Chemistry*, 279(22), pp. 23349–23356. doi: 10.1074/jbc.M401292200.

Wijma, H. J. *et al.* (2014) 'Computationally designed libraries for rapid enzyme

stabilization', *Protein Engineering, Design and Selection*. 2014/01/08, 27(2), pp. 49–58. doi: 10.1093/protein/gzt061.

William Studier, F. *et al.* (1990) 'Use of T7 RNA polymerase to direct expression of cloned genes', in *Methods in Enzymology*. Academic Press, pp. 60–89. doi: 10.1016/0076-6879(90)85008-C.

Williams, R. C. *et al.* (2003) 'Production of Bacillus anthracis protective antigen is dependent on the extracellular chaperone, PrsA', *Journal of Biological Chemistry*, 278(20), pp. 18056–18062. doi: 10.1074/jbc.M301244200.

Williamson, E. D. *et al.* (2005) 'Immunogenicity of Recombinant Protective Antigen and Efficacy against Aerosol Challenge with Anthrax', in *Infect Immun*. Defence Science and Technology Laboratory Porton Down, Salisbury, Wilts. SP4 0JQ, United Kingdom, Avecia Biotechnology, Billingham, Cleveland, TS23 1YN, United Kingdom, Battelle Medical Research and Evaluation Facility, Columbus, Ohio, Centers for Disease, pp. 5978–5987. doi: 10.1128/iai.73.9.5978-5987.2005.

Williamson, E. D. and Dyson, E. H. (2015) 'Anthrax prophylaxis: Recent advances and future directions', *Frontiers in Microbiology*. 2015/10/07, 6(SEP), p. 1009. doi: 10.3389/fmicb.2015.01009.

Wittig, I. and Schägger, H. (2008) 'Features and applications of blue-native and clear-native electrophoresis', *Proteomics*, 8(19), pp. 3974–3990. doi: 10.1002/pmic.200800017.

Worsham, P. L. and Sowers, M. R. (1999) ' Isolation of an asporogenic (spoOA) protective antigen-producing strain of Bacillus anthracis ', *Canadian Journal of Microbiology*, 45(1), pp. 1–8. doi: 10.1139/w98-108.

Wrapp, D. *et al.* (2020) 'Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation', *Science*, 367(6483), pp. 1260 LP – 1263. doi: 10.1126/science.abb2507.

Wright, T. A. *et al.* (2017) 'Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry', *Journal of Physical Chemistry Letters*, 8(3), pp. 553–558. doi: 10.1021/acs.jpclett.6b02894.

Wu, G. *et al.* (2010) 'Soluble expression and purification of the anthrax protective antigen in E. coli and identification of a novel dominant-negative mutant N435C', *Applied Microbiology and Biotechnology*. 2010/03/10, 87(2), pp. 609–616. doi: 10.1007/s00253-010-2495-5.

Wu, I. and Arnold, F. H. (2013) 'Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures', *Biotechnology and Bioengineering*. 2013/02/14, 110(7), pp. 1874–1883. doi: 10.1002/bit.24864.

Wu, X. *et al.* (2004) 'Codon optimization reveals critical factors for high level expression of two rare codon genes in Escherichia coli: RNA stability and secondary structure but not tRNA abundance', *Biochemical and Biophysical Research Communications*, 313(1), pp. 89–96. doi: 10.1016/j.bbrc.2003.11.091.

Wu, X. C. *et al.* (1991) 'Engineering a Bacillus subtilis expression-secretion system with a strain deficient in six extracellular proteases', *Journal of Bacteriology*, 173(16), pp. 4952–4958. doi: 10.1128/jb.173.16.4952-4958.1991.

Wurm, D. J. *et al.* (2016) 'The E. coli pET expression system revisited—mechanistic correlation between glucose and lactose uptake', *Applied Microbiology and Biotechnology*. 2016/05/27, 100(20), pp. 8721–8729. doi: 10.1007/s00253-016-7620-7.

Yoon, S., Kim, S. and Kim, J. (2009) ‘Secretory Production of Recombinant Proteins in Escherichia coli’, *Recent Patents on Biotechnology*, pp. 23–29. doi: 10.2174/187220810790069550.

Young, T. A., Skordalakes, E. and Marqusee, S. (2007) ‘Comparison of Proteolytic Susceptibility in Phosphoglycerate Kinases from Yeast and E. coli: Modulation of Conformational Ensembles Without Altering Structure or Stability’, *Journal of Molecular Biology*, 368(5), pp. 1438–1447. doi: 10.1016/j.jmb.2007.02.077.

Yu, H. and Huang, H. (2014) ‘Engineering proteins for thermostability through rigidifying flexible sites’, *Biotechnology Advances*, 32(2), pp. 308–315. doi: 10.1016/j.biotechadv.2013.10.012.

Zarschler, K. *et al.* (2013) ‘High-yield production of functional soluble single-domain antibodies in the cytoplasm of Escherichia coli’, *Microbial Cell Factories*, 12(1), p. 97. doi: 10.1186/1475-2859-12-97.

Zhang, W. *et al.* (2018) ‘Development an effective system to expression recombinant protein in E. coli via comparison and optimization of signal peptides: Expression of Pseudomonas fluorescens BJ-10 thermostable lipase as case study’, *Microbial Cell Factories*, 17(1), p. 50. doi: 10.1186/s12934-018-0894-y.

Zhang, X. F. *et al.* (2016) ‘A general and efficient strategy for generating the stable enzymes’, *Scientific Reports*, 6, p. 33797. doi: 10.1038/srep33797.

Zhou, B., Carney, C. and Janda, K. D. (2008) ‘Selection and characterization of human antibodies neutralizing Bacillus anthracis toxin’, *Bioorganic and Medicinal Chemistry*, 16(4), pp. 1903–1913. doi: 10.1016/j.bmc.2007.11.001.

Zieliński, M. *et al.* (2019) 'Expression and purification of recombinant human insulin from E. coli 20 strain', *Protein Expression and Purification*, 157, pp. 63–69. doi: 10.1016/j.pep.2019.02.002.

Web references

United Nations, (2015). Transforming our world: The 2030 Agenda for Sustainable Development [online]. *United Nations General Assembly Resolution A/RES/70/1*. [Viewed 6 May 2019]. Available from: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

World Health Organization, (2006). Temperature sensitivity of vaccines. World Health Organization. https://apps.who.int/iris/handle/10665/69387

Altimmune, (2018). Altimmune Announces Pre-Clinical Data From its SparVax-L Anthrax Vaccine Program. [Viewed 30 June 2019]. Available from: https://altimmune.com/altimmune-announces-pre-clinical-data-from-its-sparvax-l-anthrax-vaccine-program/