Expression, Purification, and Crystallization of CTB-MPR$_{649-684}$, a Candidate Mucosal Vaccine Component Against HIV-1

by

Ho-Hsien Lee

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2015 by the
Graduate Supervisory Committee:

Petra Fromme, Chair
Tsafrir Mor
Alexandra Ros

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

$CTB\text{-}MPR_{649\text{-}684}$ is a translational fusion protein consisting of the cholera toxin B subunit (CTB) and the conserved residues 649-684 of gp41 membrane proximal region (MPR). It is a candidate vaccine component aimed at early steps of the HIV-1 infection by blocking viral mucosal transmission. Bacterially produced CTB-MPR was previously shown to induce HIV-1 transcytosis-blocking antibodies in mice and rabbits. However, the induction of high-titer MPR specific antibodies with HIV-1 transcytosis blocking ability remains a challenge as the immuno-dominance of CTB overshadows the response to MPR. X-ray crystallography was used to investigate the structure of CTB-MPR with the goal of identifying potential solutions to improve the immune response of MPR. Various CTB-MPR variants were designed using different linkers connecting the two fusion proteins. The procedures for over-expression *E. coli* and purification have been optimized for each of the variants of CTB-MPR. The purity and oligomeric homogeneity of the fusion protein was demonstrated by electrophoresis, size-exclusion chromatography, dynamic light scattering, and immuno-blot analysis. Crystallization conditions for macroscopic and micro/nano-crystals have been established for the different variants of the fusion protein. Diffraction patterns were collected by using both conventional and serial femto-second crystallography techniques. The two crystallography techniques showed very interesting differences in both the crystal packing and unit cell dimensions of the same CTB-MPR construct. Although information has been gathered on CTB-MPR, the intact structure of fusion protein was not solved as the MPR region showed only weak electron density or was cleaved during crystallization of macroscopic crystals. The MPR region is present in micro/nano-crystals, but due to the severe limitation of the Free Electron Laser beamtime, only a partial data set was obtained and is insufficient for structure determination. However, the work of this thesis has established methods to purify large quantities of CTB-MPR and has established procedures to grow crystals for X-ray structure analysis. This has set the foundation for future structure determination experiments as well as immunization studies.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 The fusion protein CTB and MPR

HIV and AIDS

Human immunodeficiency virus (HIV) shown in Figure 1.1, is the causal agent of acquired immune deficiency syndrome (AIDS). The virus targets cells expressing CD4 receptors, namely $CD4^+$ T cells and some dendritic cells and macrophages that also express CD4. As $CD4^+$ T cells are inactivated and killed during the course of the infection, the number of $CD4^+$ T cells decreases and causes a gradual deterioration of the immune system that eventually leads to AIDS. The disease is characterized by a susceptibility to infection with opportunistic pathogens and the development of cancer at the end stages of the disease.

HIV remains as one of the leading causes of death worldwide. In 2012 there were an estimated 35.3 million infected individuals, with 2.3 million new infections, and 1.6 million AIDS related death (UNAIDS, 2013). Although the advancements in antiretroviral therapies and access to treatment has curbed the growth of infections in many areas, it is clear that the HIV/AIDS pandemic will continue until an effective pathway of preventing viral infection and transmission is found.

There are currently no cures or working vaccines for HIV. There are several factors that make HIV treatment difficult such as the rapid reproduction, the ability to persist in a latent form as a transcriptionally silent provirus and the high antigenic variability resulting from the high mutation rates in HIV-1. The cause of the high

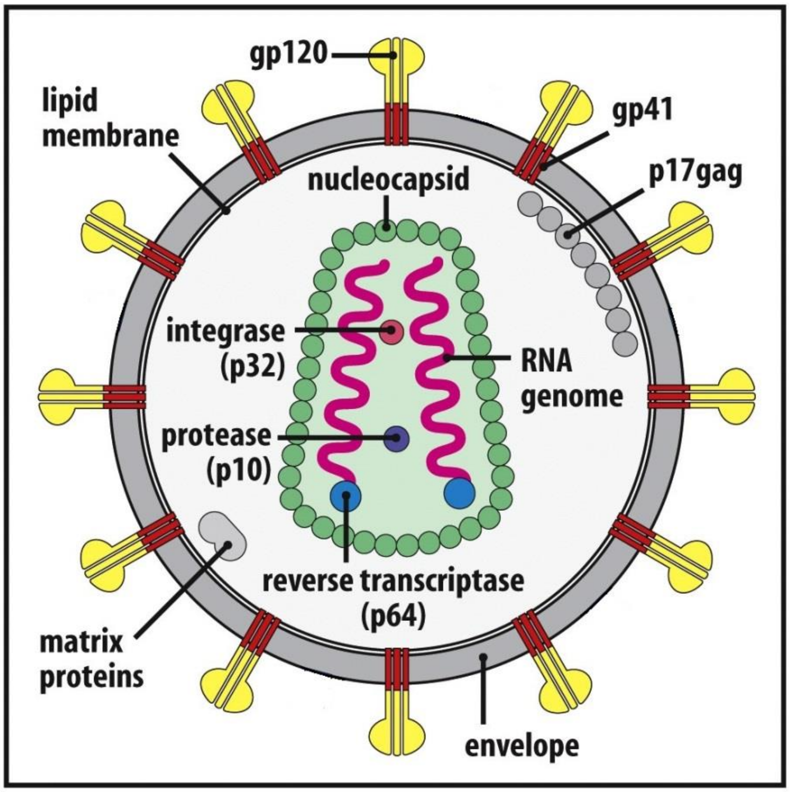


Figure 1.1: **Illustration of the HIV-1** HIV is a retrovirus that contains an RNA genome which is reverse transcribed into DNA in the infected cell by viral reverse transcriptase and integrated into the host-cell DNA with the viral integrase. A viral protease is involved in the processing of many viral proteins. Two envelope glycoproteins gp120 (yellow) and gp41 (red) form the “spikes” on the surface of the HIV and have shown to play significant role in viral infection. Image adapted from (Murphy et al., 2008)

mutation rate is often attributed to the low fidelity of the viral reverse transcriptase (Roberts et al., 1988, Preston et al., 1988). The virus evolves by the high variability and selective proliferation of mutant viruses that are both resistant to antiviral drugs and have the ability to escape recognition by the immune system then followed by the outgrowth of these variants (Murphy et al., 2008). So far, the majority of the drugs developed have only been targeting and inhibiting the reverse transcriptase and protease and the combination therapy of using multiple drugs together have been shown to be more effective than using a single drug (Gortmaker et al., 2001, Murphy et al., 2008).

Therefore an effective treatment or vaccine for HIV should consist of multiple components that target various steps of the transmission and infection processes of HIV expanding beyond the current drug target. It is especially important to target the early stages of the viral cycle to take advantage of viral vulnerabilities of the low founder viral population, lower number of variants and to minimize the establishment of systematic infection (Haase, 2010).

Membrane proximal region of envelope glycoprotein 41 in HIV-1

The envelope proteins (Env) and the viral surface proteins of HIV have been one of the major focuses of research and development for drugs and vaccines against HIV. The *env* gene codes for a precursor glycoprotein known as gp160, which is later processed by protease cleavage to form gp120 and gp41 (numbers in the names correspond to the molecular weight of the protein). The “surface subunit” gp120 is exposed on the exterior of the viral envelope and associates, non-covalently, with the “transmembrane subunit” gp41 that contains an ectodomain largely responsible for the trimizeration of the Env (Wyatt & Sodroski, 1998). Together, the two glycoproteins form the protrusion on the virion that are often referred to as “spikes” and are involved in essential steps in the viral transmission and infection processes (Murphy et al., 2008). Env is heavily glycosylated with the vast majority of the highly variable glycans attached to gp120 and only few are found on gp41 (Wang et al., 2013, Leonard et al., 1990).

The gp120 subunit is very immunogenic and had been the main target of vaccine research during the first two decades of HIV exploration since its discovery (Zolla-

Pazner, 2004). However, the heavy glycosylation of gp120, its ever-shifting glycosylation patterns and the particularly high mutation rate of HIV virus leading to high variability of the protein sequence are undoubtedly major contributing factors in the immune-evasion strategy of the virus, making gp120 less than ideal immunogen (Jeang, 2007, Steckbeck et al., 2011, Montero et al., 2008).

In contrast to gp120, the amino acid sequence of gp41, is more highly conserved, indicating that the structural and functional attributes of gp41 are under more exacting selective pressure than those of the gp120 (Soudeyns et al., 1999, Bouvin-Pley et al., 2014, Travers et al., 2005). Indeed, gp41 has several regions that are crucial to viral functions identified to be potential drug targets against HIV-1 (Zolla-Pazner, 2004). One of these more conserved regions of gp41 is the membrane proximal external region ($MPER_{662–684}$), which has been known to play a key role in viral fusion with host cells (Buzon et al., 2010, Cai et al., 2011) and is also the site the few broadly neutralizing monoclonal antibodies (Abs) discovered to date such as 2F5, 4E10, and 10E8 (Zwick et al., 2001, Zolla-Pazner, 2004, Huang et al., 2012).

Over 90% of global HIV-1 transmissions occur across mucosal surfaces, for example through the mucosa lining of the genital and lower intestinal tracts during sexual transmission or oropharyngeal mucosa in breastfeeding (Kresina & Mathieson, 1999, Overbaugh et al., 1999). The mucosal surfaces of the cervix, rectum and colon are characteristically lined by a single layer of epithelial cells that are joined by tight junctions, preventing, when intact, the passage of HIV-1 between the cells. Instead, the virus co-opts the cellular process known as transcytosis to cross through the cells (without infecting them). Once passed across the epithelial barrier, the virus is likely to encounter its target CD4$^+$ host cells.

The transcytosis process depends on a 36-amino acid-long domain of gp41 encompassing the MPER and extending partially into the C-terminal heptad repeat domain lying immediately upstream to it (residues 649-684). We term this extended-MPER region “MPR” (also referred to as the P1 peptide in previously published work)(Alfsen & Bomsel, 2002, Alfsen et al., 2005). MPR was shown to be crucial for the viral crossing of the epithelial membrane (Alfsen & Bomsel, 2002, Zolla-Pazner, 2004, Matoba et al., 2004). The domain is vital for the viral epithelial transcytosis of HIV using a non-fusogenic mechanism by binding to the glycosphingolipid galatosylcermide (GalCer) and the heparan sulfate proteoglycan agrin on the apical surface of epithelial cells to initiate endocytosis and transcytosis (Bomsel, 1997, Alfsen et al., 2001, Alfsen & Bomsel, 2002, Alfsen et al., 2005). This is significant as it is the first step in one of the principle route of HIV infection. If the virus can be prevented from entering the body to reach its host cell this early in the transmission process, the chances of the virus establishing a chronic infection would be lowered.

Support for the above notion comes from studies conducted with highly HIV-exposed persistently seronegative (HEPS) individuals. HEPS, are people with high exposure risk to HIV, who despite documented high potential for exposure to the virus (e.g. multiple and frequent unprotected sexual activities) still remain negative for anti-HIV serum IgG and lack of detectable HIV RNA or DNA (Devito et al., 2000a, Fowke et al., 1996). Still, a subset of these HEPS exhibit gp41-specific secretory IgAs in their genital secretions. These antibodies were shown to be efficient in blocking viral transcytosis across the epithelium and were suggested to be correlated to their protection against HIV infections (Devito et al., 2000c, Devito et al., 2002, Tudor et al., 2009, Miyazawa et al., 2009). These findings suggest that an HIV-1 vaccine component that consists of an immunogen that is mucosally targeted and capable of eliciting Abs against MPR which can both block the epithelial transcytosis process and neutralize the infection of CD4$^+$ cells could be relatively effective. MPR was also shown to be able to bind to dendritic cells, an important class of antigen presenting cell (Bomsel & Magerus-Chatinet, 2004, Magerus-Chatinet et al., 2007). Interaction with antigen presenting cells is beneficial to elicit immune response against the peptide.

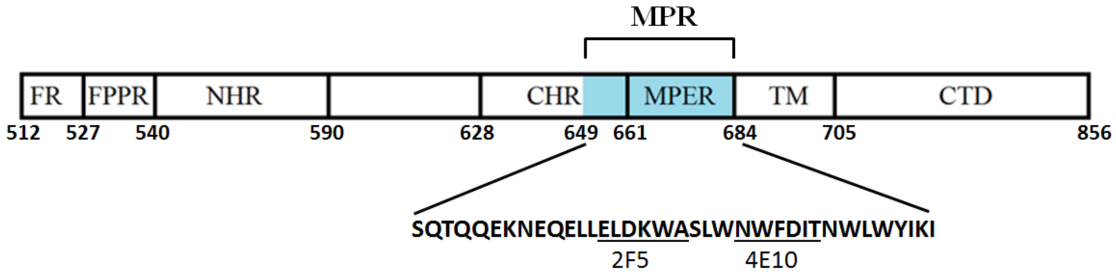


Figure 1.2: **Schematic diagram of the part of the HIV-1 *env* gene that encodes for the gp41 portion.** Fusion peptide (FR. Residues 512-539); fusion peptide proximal region (FPPR residues 528–539; N-terminal heptad-repeat region (NHR, residues 540–590); C-terminal heptad-repeat region (CHR, residues 628–661); membrane proximal external region (MPER, residues 662–684); membrane proximal region (MPR, residues 647–684, blue), transmembrane domain (TM, residues 685–705); cytoplasmic C-terminal domain (CTD, residues 706–856). The epitopes of 2 broadly neutralizing antibodies 2F5 and 4E10 are marked.

Cholera toxin B subunit as mucosal adjuvant

The MPR peptide by itself is a rather poor immunogen (Denner, 2011). To boost the immunogenicity at the mucosal surface, an adjuvant is needed to enhance the immune response of MPR. Cholera toxin B subunit (CTB) was chosen as the candidate to be the fusion partner.

Secreted by *Vibrio cholera*, the cholera toxin is a protein complex that is responsible for many of the symptoms of cholera infections. It is an oligomeric protein made up of six subunits, a single A subunit and five copies of B subunit that form a pentameric ring around the tail of the A subunit. The primary function of CTB is to serve as a delivery system for the A subunit. In the intestines, pentameric CTB binds to the pentasaccharide moiety of G_{M1} gangliosides on the mucosal (luminal) side of epithelial cells. The AB_5 complex is internalized through endocytosis followed by the release of the A subunit through the cleavage of disulfide bonds (Merritt et al., 1994).

Many immunological studies have been done on using CTB as a mucosal adjuvant and as a carrier for foreign peptide epitopes (Quiding et al., 1991, Peltola et al., 1991, Malley et al., 2004, Qu et al., 2005). The fusion protein retains the important structural and functional characteristics of the native CTB such as the pentamerization and binding to the G_{M1} ganglioside receptors (Holmgren et al., 1993, Sun et al., 1994). CTB has also been used successfully to produce human vaccines against cholera itself (Quiding et al., 1991) and a vaccine against enterotoxigenic *Escherichia coli* (Peltola et al., 1991). It also proved to be a good adjuvant in immunizing mice against *Streptococcus*

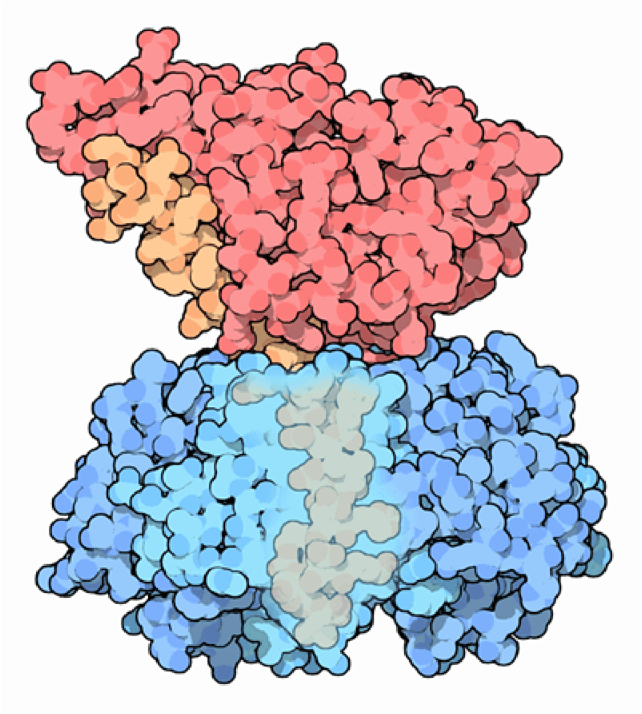


Figure 1.3: **Space-filling model of cholera toxin.** Cholera toxin is an oligomeric complex made up of six protein subunits. One single copy of subunit A and five copies of subunit B. Subunit A has two important segments, A1 (red) is a globular enzyme, while A2 (orange) is an alpha helical chain that fits in the center of the five subunit B ring (blue) (Goodsell, 2005).

pneumonia (Malley et al., 2004) and the severe acute respiratory syndrome-associated coronavirus (SARS) (Qu et al., 2005). CTB fusion protein has also been used to elicit anti-HIV-1 Abs before, namely the V3 domain of gp120 using V3-CTB fusion proteins although was not successful in acting as an effective vaccine against HIV-1 (Backstrom et al., 1994, Zolla-Pazner et al., 2011).

These findings suggested that CTB may be a suitable adjuvant for the purpose of eliciting anti-MPR Abs and can serve as a carrier and for the enhancement of the immune response that targets the mucosal membrane. Fusion proteins consisting of CTB and $\text{MPR}_{649\text{-}684}$ were created (named CTB-MPR). The proposed mechanism of CTB-MPR inducing the production of IgA against $\text{MPR}_{649\text{-}684}$ of gp41 in the mucosal immune system is shown in Figure 1.3. In published experiments of Matoba and co-workers, CTB-MPR has been shown to elicit HIV-1 transcytosis blocking Abs (Matoba et al., 2004, Matoba et al., 2006). However, not all trials had successfully shown anti-transcytosis activity and the titer of MPR specific antibodies caused by CTB-MPR remained low. Subsequent booster immunizations in rabbits with CTB-MPR did not result in further improvement of the anti-MPR Ab responses because anti-CTB Abs continued to increase and thereby overshadowed the response to MPR (Matoba et al., 2008). This indicated a need for the improvement of the MPR-based immunogen.

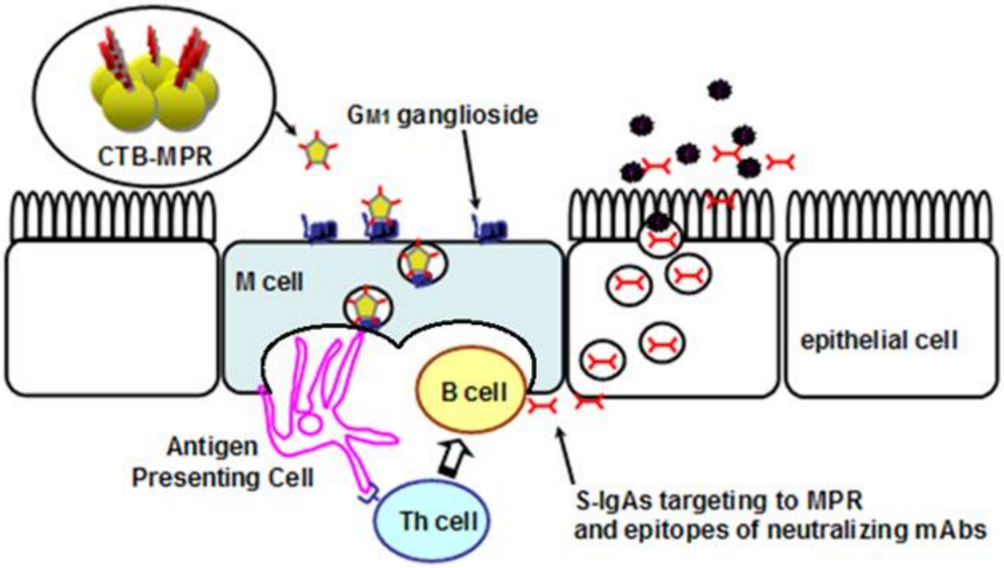


Figure 1.4: **Proposed mechanism of the fusion protein of CTB-MPR** The pentameric CTB domain binds to G_{M1} gangliosides after which the fusion protein complex is taken up by the cell via endocytosis as a whole. The complex is then presented to antigen presenting cells which promotes a T helper cell response that signals B cells to produce Abs against the presented antigen, in this case MPR. (Adapted from N. Matoba)

Motivations and objectives

To understand the function of MPR and the membrane associated processes it is involved in, such as the epithelial transcytosis and the membrane fusion to the host cells, as well as interactions with the immune system, it is imperative to obtain the structural information. The molecular structure defines the properties and function of proteins as well as their complex macromolecular assembles (Rupp, 2010). Part of the structure of gp41, that may represent the post-fusion state of gp41 after the virus has fused with the host cell, has been solved by X-ray crystallography. It formed a six helices bundle of a trimer with each domain of gp41 containing two anti-parallel helices shown in Figure 1.5A (Chan et al., 1997, Melikyan et al., 2000). However, these structures provided only limited information for understanding of the structure-function relationship of gp41 as they contained only part of the MPER domain and the sites of the broadly neutralizing Abs (2F5, 4E10 and Z13) were buried inside the bundle of helices and were not exposed for Ab access shown in Figure 1.5B and 1.5C (Shi et al., 2010, Buzon et al., 2010). It was not until recently, when there has been published work on a proposed ‘pre-fusion’ conformation of gp41. A ‘pre-fusion’ NMR dynamics study showed structures that contained large portions of gp41, however it is questionable if these structures really represent a “pre-fusion” conformation as they are similar to the two anti-parallel helices conformation reported previously for the post fusion conformation (Lakomek et al., 2014). Another NMR structure of pre-fusion state included MPER attached to a trimerization domain of bacteriophage T4 fibritin foldon domain that forces the MPER in a conformation in which the Ab sites are accessible was solved recently and is shown in

Figure 1.5D (Reardon et al., 2014). Interestingly, the gp120 and gp41 crystal structure shown in Figure 1.5E showed a different conformation. In this structure, gp41 is arranged in a four helices collar that is very different from the two anti-parallel helices. However the part of gp41 used for this structural study did not contain the MPER domain that bind the neutralizing Ab epitopes (Pancera et al., 2014). The structures from these publications also differ from each other, making it difficult to determine which (if any) would represent the native pre and post fusion conformation of gp41.

CTB-MPR has shown promising results in inducing anti-HIV Abs that could block transcytosis, thereby hinting that the Ab sites are exposed in CTB-MPR; it would be very important for further improvement of CTB-MPR as a potential vaccine against HIV to determine the structure of MPR in its CTB-MPR fusion conformation as none of the published structures of gp41 may represent the immunogenic form of MPR.

The unraveling of the structure of CTB-MPR would surely help in improving the vaccine design as CTB-MPR has shown already promise in anti-HIV activities. To achieve this goal, we have decided to pursue the path of protein X-ray crystallography for structure determination.

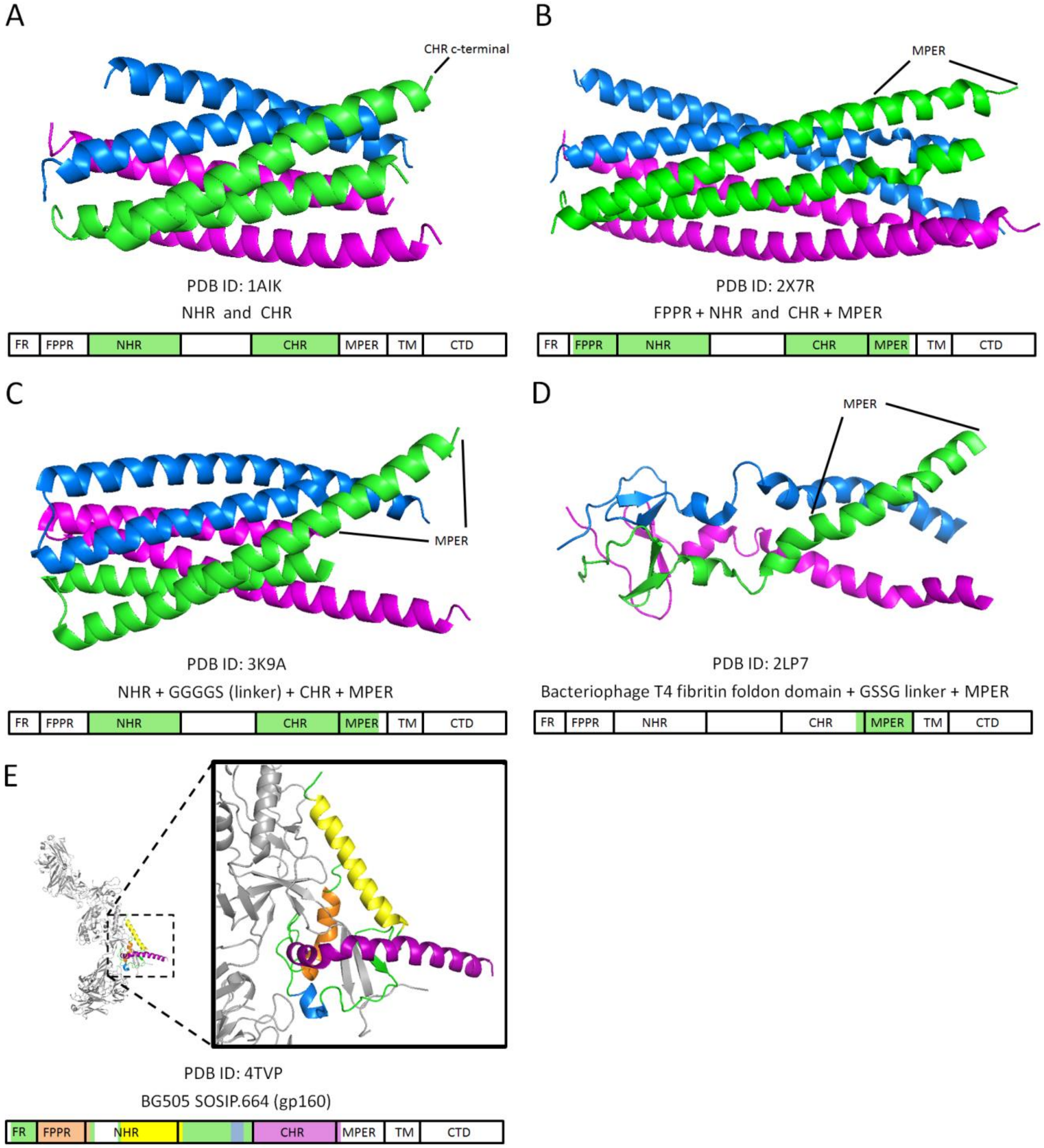


Figure 1.5: **Published gp41 and MPER structures.** The structures were obtained from the PDB and the pictures were generated by PyMol (Delano, 2002). (A) PDB ID: 1AIK. Structure of gp41 post fusion 6-helices bundle that only contained the helices NHR and CHR (Chan et al., 1997) (B) PDB ID: 2X7R. Structure of gp41 6-helices bundle that also included parts of FPPR and MPER (Buzon et al., 2010) (C) PDB ID: 3K9A. Structure of gp41 6-helices bundle that included overhanging MPER without the FPPR (Shi et al., 2010) (D) PDB ID: 2LP7. NMR structure of MPER linked with the foldon domain of bacteriophage T4 fibritin (Reardon et al., 2014) (E) PDB ID: 4TVP. Structure of gp120 and gp41 pre-fusion confirmation with gp41 arranged in a 4-helices collar (MPER not included) (Pancera et al., 2014)

1.2 X-ray Protein Crystallography

Overview and theory

The structure and function of molecules are related to each other at the atomic and molecular level. Therefore, in order to completely understand the properties and functions of proteins and their interaction with other molecules, it is necessary to determine their structure. Nearly 100,000 protein structures have been solved and deposited in the protein database (PDB) by X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy (EM), with the vast majority (~90%) of the structures determined by protein X-ray crystallography.

In conventional x-ray crystallography, a single crystal is placed into a finely focused x-ray beam and the x-rays scattered by the electrons of the atoms in the crystal are recorded as diffraction patterns by a detector. The diffraction patterns are analyzed by X-ray structure analysis. The amplitude of the structure factors is determined from the intensity for the reflection, the wavelength corresponds to the wavelength of the X-ray beam but phase information for each structure factor is lost and has to be solved by molecular replacement (for homologous structures) or de-novo by MIR (multiple isomorphous replacement), MAD (multiple anomalous dispersion) or SAD (single anomalous dispersion). After retrieval of the phases, the electron density of the crystallized protein is reconstructed. An atomic model fitting the electron density is built using computational methods and refined into the resultant protein structure (Woolfson, 1997, Rupp, 2010). Although it may appear quite straightforward and simple, there are various challenges that must be overcome to obtain a crystal structure.

When an X-ray photon strikes an atom, it interacts with the electron cloud of the atom and is either absorbed or scattered. The scattering from a single protein molecule by the X-rays is too small to be measured with the current technology (Glaeser, 1999, Spence & Doak, 2004). In order to increase the signal to detectable levels, it requires the protein to be arranged in an orderly three-dimensional repeating lattice of equivalent unit cells to amplify the signal (Figure 1.6). The amplification can be of the order N^2, where N is the number of unit cells in the crystal (Miao et al., 2004).

Bragg's Law describes the coherent scattering of light (in this case x-rays) from a crystal lattice (Figure 1.7):

$$2d \sin \theta = n\lambda$$

where d is the distance between the reflective planes, θ is the angle of incidence of the x-ray which will reflect back with a same angle of scattering, n is an integer, and λ is the wavelength of the X-ray. When X-rays strike a protein crystal, the waves are scattered from the planes separated by a distance d. The waves will interfere constructively and destructively depending on the path difference between the two waves $2d \sin \theta$. The waves that remain in phase (a phase shift that is a multiple of 2π) will interfere constructively resulting in measureable reflections. Scattering from each set of planes in a crystal leads to a reflection which corresponds to a structure factor in the reciprocal space.

In order to reconstruct the electron density of the scattering molecule, the Fourier reconstruction from reciprocal diffraction space of the diffraction pattern back into a three-dimensional direct molecular space is needed (Rupp, 2010). However, to carry out

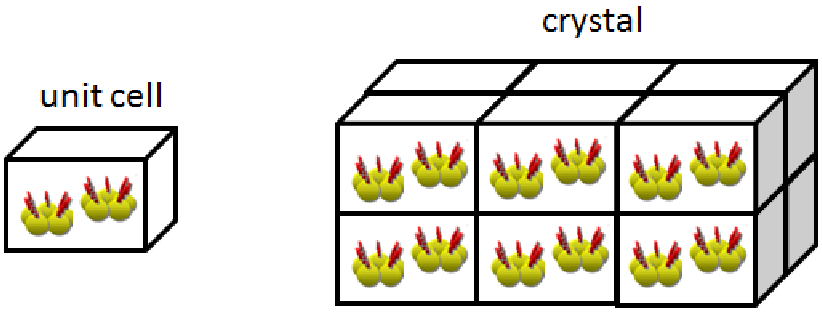


Figure 1.6: **Unit cell and crystal.** A unit cell is the smallest unit of a repeating lattice that contains the structure and symmetry information. A crystal is an ordered 3-dimensional repeating array of unit cells.

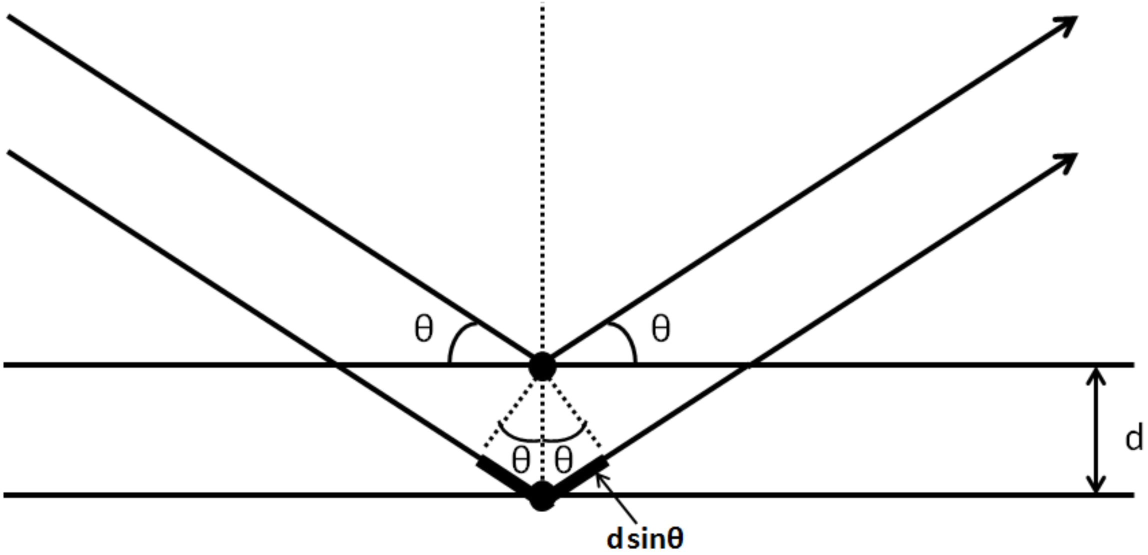


Figure 1.7: **Bragg's Law** The graphical interpretation of Bragg's Law shows conditions for the constructive interference of the X-ray diffraction from the reflective planes of a crystal.

Fourier transform, the properties of the scattered X-rays are needed in order to reconstruct the wave; wavelength, amplitude and phase angle. The wavelength of the scattered X-rays is the same as those the crystal is exposed to. The amplitude is equal to the square root of the intensity the diffraction spot (reflection) and can be calculated based on the measured intensities (Rupp, 2010). The phase angle, which contains most of the structural information, is lost in the measurement and is not readily available; its value must be determined by other experiments. This is known as the phase problem and is a reason why protein structure determination can be quite difficult.

There are several ways to solve the phase problem. The brute force direct method exploits the phase relations between certain sets of structure factors, but it is limited to only relatively small proteins (Rupp, 2010). Another method is using marker atoms such as heavy atom derivatives and comparing it to the native crystal diffraction to produce a difference map to calculate phase (Woolfson, 1997, Rupp, 2010). The introduction of heavy metal atoms is typically done by either soaking the crystal in heavy metal solution or introducing marker atoms such as selenium in form of selenomethionine during expression of the protein in-vivo. The methods summarized above allow for de-novo structure determination and they do not require prior structure information. A third method, molecular replacement requires a structurally similar model as a molecular probe. By using a previously solved structure homologue of the crystallized protein, the phases of the correctly placed models are used as starting phases for map reconstruction (Rupp, 2010).

Once the electron density map is calculated, a model must be generated to fit the electron density, and then further refined. As each atom of the structure contributes to all of the diffraction spots in a nonlinear way, it is necessary to adjust the model parameters to best describe the data. Stereochemical knowledge must generally be incorporated into the refinement in the form of restraints to keep the model within reasonable bounds.

Protein crystallization

In order to obtain a high resolution X-ray structure, a well-diffracting protein crystal is required. The scattering of X-rays is dependent on the atomic number of the atoms in the protein, which corresponds to the number of electrons; the higher the number, the greater the scattering intensity (Woolfson, 1997). However, biological macromolecules are mainly comprised of H, C, N, O, and small amount of S and P. These are atoms with relatively low atomic numbers and low scattering intensity. In addition, the native states of proteins are typically not at solid state but they are surrounded by a fluid environment such as an aqueous solution. It is common for proteins to have a solvent content between 30% and 70%, and even in some cases up to 90% have been reported (McPherson, 2004) creating large unit cells with few crystal contact sites (Rupp, 2010). Therefore, in order to obtain high resolution diffraction, large protein crystals are necessary due to the weak scattering of the elements and the relatively high solvent content of the unit cell.

The first step of protein crystallization is the preparation of the protein in quantities needed for crystallization experiments. Through recombinant DNA engineering, it is possible to overexpress proteins that are synthetically designed or not abundant in nature in bacterial or eukaryotic cells (Rupp, 2010). However, the purification remains a great challenge as high purity samples are required for crystallization. For crystallography, the concept of purity does not only mean the removal unrelated proteins or other undesired cellular components, but also means maintaining the protein in homogenous form in a single conformation and an unique oligomeric state. The presence of contaminants and different conformations or a heterogeneity of oligomeric states may hinder crystal growth in a number of ways. Different forms of the same protein may compete for crystal contacts thereby generating lattice errors leading to mosaicity, dislocation, irregular faces and secondary nucleation, twinning, poor diffraction, and ceasing crystal growth (Ducruix & Giegé, 1992). To probe the purity and homogeneity of protein samples, many different techniques such as electrophoresis, gel filtration, and light scattering methods are used to detect contaminants or structural oligomers based on charge, size, mobility and shape (Ducruix & Giegé, 1992, Rupp, 2010).

Many crystallographers consider protein crystallization to be more of an art than science. It is the least understood step of structure determination of a protein as there is no known method of predicting exactly under what conditions a protein will form crystals (Rupp, 2003). The effects of buffers, salts, pH, detergents, temperature, precipitant, and additives may all have drastic and unpredictable effects on protein crystallization. Therefore, the number of possible crystallization conditions far exceeds the number that may be reasonably tested. Techniques such as the grid screens described in (McPherson, 1982), factorial designs (Carter & Carter, 1979), and sparse matrix sampling (Jancarik & Kim, 1991) have been developed. They sample many conditions that have been previously reported to lead to crystallization of proteins. However, none of the commercial screen systematically investigates the influence of each of the crystallization factors so that the effects of each factor can be determined as there is a limit to the total number of crystallization conditions that can be explored per screen. Many commercial crystallization screens are based on these screening techniques and have been widely used.

Once initial crystallization conditions are found, it is still necessary to further improve the crystal quality in finer screens that screens around the condition that had led to crystal hits in commercial screens by adjusting the concentrations of the components in the crystallization cocktails.

The main objective in protein crystallization is to reach the supersaturated phase to coerce the protein out of the solution gently into a well-ordered lattice. The protein solubility is typically reduced by various methods such as the addition of precipitate to the protein solution, removal of water from the protein-precipitate solution through vapor diffusion, exchange of solvent through dialysis; change of pH, or by use of free interface diffusion. The protein solution becomes supersaturated and when the nucleation zone is reached crystals may form correlating to the schematic phase diagram shown in Figure 1.8 (Ducruix & Giegé, 1992, Chayen & Saridakis, 2008, Rupp, 2010). During a crystallization experiment, crystals do not appear as soon as the supersaturated phase is reached. The crystallization process is not only determined by the thermodynamics but also by the kinetics pf the system. It is very common that no crystals are formed in the supersaturated zone due to kinetic barriers. Nuclei form and dissolve; the size and stability of nuclei depend on the supersaturation. In the metastable zone the nuclei are too small to be stable, therefore crystals cannot form but pre-grown seeding crystals can grow. Once the nucleation zone is reached and activation barriers toward equilibration are overcome, a nucleation event occurs that leads to formation of nuclei that exceed the critical radius, leading to stable nuclei which can further grow into a crystal. By crystal growth the protein concentration in the solution decreases until the supersaturation borderline is reach. Crystals are not always formed in a crystallization experiments, often amorphous precipitate is observed instead of crystals. The reason are manifold and include but are not limited to very high supersaturation, unsuitable conditions for the formation of crystal contacts, partial or complete denaturation of the protein by the supersaturated conditions, flexible domains of the protein prohibit formation of crystal contacts, high dynamics of protein hinder formation or an ordered lattice etc.

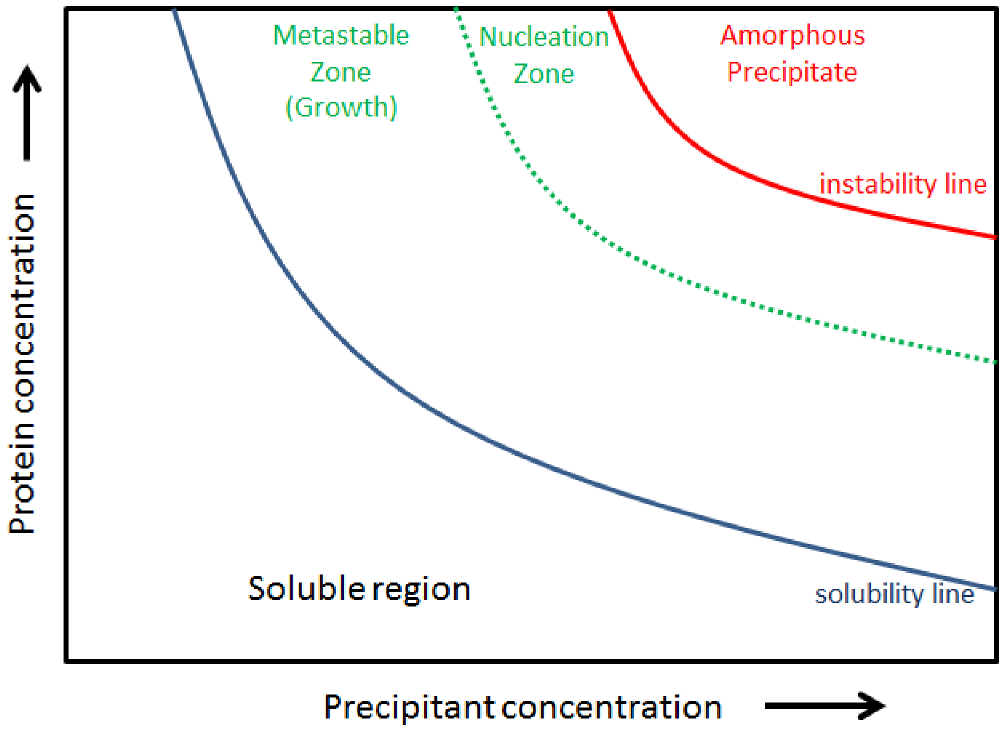


Figure 1.8: **Basic solubility phase diagram** shows the general observation of crystallization. The higher the precipitant concentration, the lowers the maximum concentration for the protein to achieve each phase and vice versa. Between the solubility line and instability line are the metastable and nucleation zones. The metastable zone represents a supersaturated protein solution where pre-formed crystal can grow but no stable nuclei are formed. In the nucleation zone nuclei are stable and crystals are observed.

Though high supersaturation is necessary for the nucleation, it is not desirable for the growth of large, well-ordered single crystals. A high supersaturation promotes the formation of large numbers of nuclei leading to the growth of many protein microcrystals rather than the macroscopic crystals required for conventional protein crystallography, where data are collected from single crystals at synchrotron sources. In addition, at very high supersaturation, the protein will precipitate in form of unordered amorphous precipitates as mentioned previously. While the formation of protein crystals are thermodynamically favored in supersaturated conditions, the amorphous precipitates are kinetically favored under high supersaturation conditions (McPherson et al., 1995).

Serial femtosecond X-ray crystallography

One difficulty with protein crystallography is that crystal growth of protein microcrystals is much more common than the growth of macroscopic protein crystals necessary for data collection at conventional X-ray sources (synchrotrons or rotating anodes) (Cusack et al., 1998). Also, large crystals often suffer from the long range internal disorder or mosaicity. Protein crystals are made up of small mosaic blocks, with each block being well ordered in submicrons dimensions, but the gaps between the blocks create angular misalignments. This long range disorder reduces the intensity and broadens the diffraction spots (Woolfson, 1997). So there are advantages of using smaller crystals with the trade-off of having weaker scattering intensity. These weaker intensities can only be partially enhanced by an increase in X-ray intensity due to the X-ray damage problems in crystallography (Owen et al., 2006, Barty et al., 2012). The micro-focused beamlines at third generation Synchrotron sources, where the beam is focused to 5-25 micrometer size, alleviate some of the weak scattering intensity problem by reducing the focus spot radius and increasing the flux density at the spot (Bilderback et al., 2003, Bilderback et al., 2010). However, this leads to higher radiation damage and limits the lifetime of the crystal for data collection.

With the technology advancement and the introduction of the X-ray free electron laser (XFEL), a new approach to collect diffraction data from much smaller crystals became possible. The XFEL is able to generate a peak X-ray flux that is 10^9 higher than a 3rd generation synchrotron pulse and allows data collection from crystals that contain just a few hundred molecules (Barty et al., 2009, Fromme & Spence, 2011, Chapman et al., 2011). The method, serial femtosecond X-ray crystallography (SFX), is based on the "diffract before destroy" principle. By using short 10-50 fs pulses of X-ray exposures, the diffraction of a nano/microcrystal could be recorded before the molecules in the crystal are destroyed and the crystals disintegrate (Barty et al., 2012). As the beam destroys the sample, fresh new sample is continuously delivered across the pulsed beam in the form of a liquid jet of nano/microcrystals in their mother liquor at ambient temperature, leading to hundreds of thousands of diffraction snapshots of individual crystal in random orientation. SFX eliminates the radiation problem suffered from micro-focused beam data collection at Synchrotrons as each snapshot is collected from a fresh crystal. The general scheme of SFX is illustrated in Figure 1.9.

Data analysis for SFX is a challenge. The diffraction gathered at synchrotrons, depends on a systematic continuous rotation of the crystal by a goniometer, where all planes of the crystal are systematically brought into conditions for diffraction. Thereby it allows for the angular integration across all Bragg reflections. In contrast the SFX snapshot diffraction patterns are all random “still image” which sample a random slice through the Ewald sphere. Furthermore, the SFX beam shows larger intensity variations (of a factor of 4) and the energy profiles are not identical between individual X-ray pulses. Therefore structure determination by SFX requires a larger multiplicity of the data sets (Fromme & Spence, 2011). The merging of thousands or even millions of diffraction data poses a large challenge and requires new tools for the integration of the data, where the intensity of the individual Bragg reflections are determined by Monte Carlo Integration (Kirian et al., 2011). Although the minimum number of the snapshots required is currently unknown and controversially debated, the first near-atomic resolution structure of a protein to be determined using SFX contained more than 12,000 indexed diffraction patterns which is much more than what is needed for conventional crystallography (<1,000) at a synchrotrons (Boutet et al., 2012, Chapman et al., 2011, Redecke et al., 2013, Liu et al., 2013).

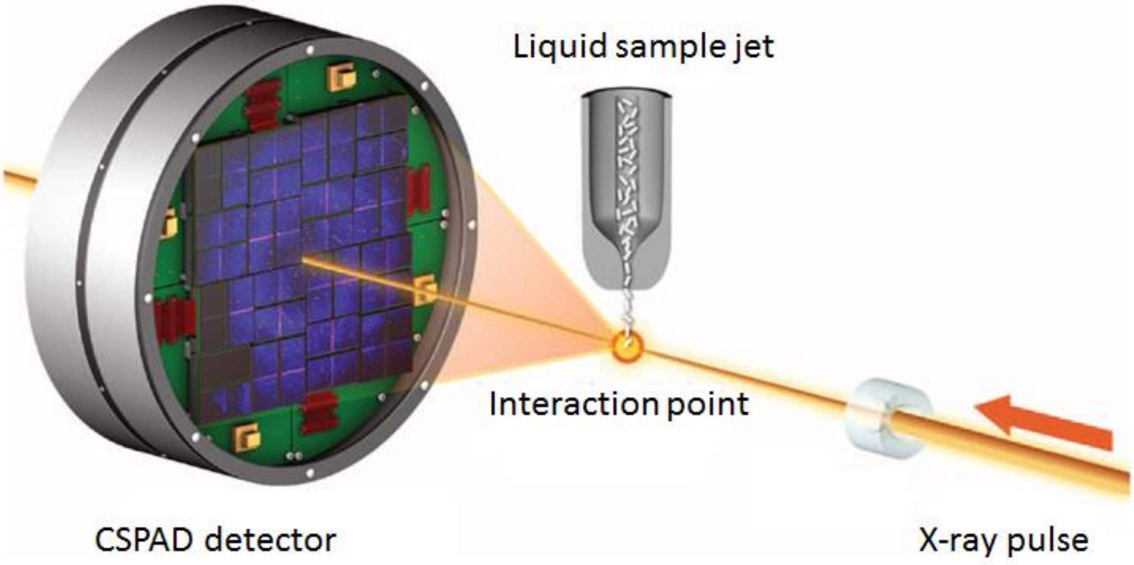


Figure 1.9: **Serial femtosecond crystallography**. Diffraction patterns from single crystals flowing in a liquid jet are recorded on Cornell-SLAC Pixel Array Detectors (CSPAD). Image adapted from (Boutet et al., 2012)

CHAPTER 2

EXPRESSION, PURIFICATION AND CRYSTALLIZATION OF CTB-MPR

Ho-Hsien Lee,[a] Irene Cherni,[b,c] HongQi Yu,[a] Raimund Fromme,[a] Jeffrey D. Doran,[b,c] Ingo Grotjohann,[a] Michele Mittman,[b,c] Shibom Basu,[a] Arpan Deb,[b,c] Katerina Dörner,[a] Andrew Aquila,[d] Anton Barty,[d] Sébastien Boutet,[e] Henry N.Chapman,[d,f] R. Bruce Doak,[g] Mark S. Hunter,[a] Daniel James,[g] Richard A. Kirian,[d,g] Christopher Kupitz,[a] Robert M. Lawrence,[a,c] Haiguang Liu,[g] Karol Nass,[d,f] Ilme Schlichting,[h] Kevin E. Schmidt,[g] M. Marvin Seibert,[e] Robert L. Shoeman,[h] John C. H. Spence,[g] Francesco Stellato,[d] Uwe Weierstall,[g] Garth J. Williams,[e] Chunhong Yoon,[d,i] Dingjie Wang,[g] Nadia A. Zatsepin,[g] Brenda G. Hogue,[b,c] Nobuyuki Matoba,[b,c] Petra Fromme[a*] and Tsafrir S. Mor[b,c*]

[a]Department of Chemistry and Biochemistry, Arizona State University, PO Box 871604, Tempe, AZ 85287-1604, USA, [b]School of Life Sciences, Arizona State University, PO Box 874501, Tempe, AZ 85287-4501, USA, [c]Center for Infectious Diseases and Vaccinology, Biodesign Institute, Arizona State University, PO Box 874501, Tempe, AZ 85287-5401, USA, [d]Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany, [e]Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, [f]University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany, [g]Department of Physics, Arizona State University, PO Box 871504, Tempe, AZ 85287-1504, USA, [h]Max-Planck-Institut für medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany, and [i]European XFEL GmbH, Albert-Einstein-Ring 19, 22761 Hamburg, Germany.

*Correspondence e-mail: petra.fromme@asu.edu, tsafrir.mor@asu.edu

(This chapter has been adapted with permission from a previously published report: Lee, H-H et al. Expression, purification and crystallization of CTB-MPR, a candidate mucosal vaccine component against HIV-1. *IUCrJ 1*(Pt 5): 305-317 (2014))

2.1 Abstract

CTB-MPR is a fusion protein between the B subunit of cholera toxin (CTB) and the membrane-proximal region of gp41 (MPR), the transmembrane envelope protein of *Human immunodeficiency virus 1* (HIV-1), and has previously been shown to induce the production of anti-HIV-1 antibodies with antiviral functions. To further improve the design of this candidate vaccine, X-ray crystallography experiments were performed to obtain structural information about this fusion protein. Several variants of CTB-MPR were designed, constructed and recombinantly expressed in *Escherichia coli*. The first variant contained a flexible GPGP linker between CTB and MPR, and yielded crystals that diffracted to a resolution of 2.3 Å, but only the CTB region was detected in the electron-density map. A second variant, in which the CTB was directly attached to MPR, was shown to destabilize pentamer formation. A third construct containing a polyalanine linker between CTB and MPR proved to stabilize the pentameric form of the protein during purification. The purification procedure was shown to produce a homogeneously pure and monodisperse sample for crystallization. Initial crystallization experiments led to pseudo-crystals which were ordered in only two dimensions and were disordered in the third dimension. Nanocrystals obtained using the same precipitant showed promising X-ray diffraction to 5 Å resolution in femtosecond nanocrystallography experiments at the

Linac Coherent Light Source at the SLAC National Accelerator Laboratory. The results demonstrate the utility of femtosecond X-ray crystallography to enable structural analysis based on nano/microcrystals of a protein for which no macroscopic crystals ordered in three dimensions have been observed before.

2.2 Introduction

The envelope glycoprotein of HIV-1 is a complex composed of three copies of a heterodimer consisting of gp120 and gp41. The latter (Figure 2.1*a*) is embedded in the viral membrane, mediates the fusion between viral and cellular membranes (Teixeira et al., 2011) and plays a major role in viral transmission across the epithelial barrier (Shen et al., 2010, Bomsel et al., 2011, Hessell et al., 2010, Tudor et al., 2009). Mucosal transmission of HIV-1 through monostratified epithelia depends on interactions between the viral envelope membrane protein gp41 and the glycolipid galactosyl ceramide (GalCer) on epithelial cells (Alfsen et al., 2001, Alfsen & Bomsel, 2002, Meng et al., 2002), and also on dendritic cells, the most important class of antigen-presenting cells (Bomsel & Magerus-Chatinet, 2004, Magerus-Chatinet et al., 2007). The GalCer binding domain of gp41 is mediated by a highly conserved membrane-proximal region (MPR) of gp41 consisting of residues 649–684. This region of the protein spans the membrane-proximal external region (MPER; residues 660–683) reviewed by (Zwick, 2005), which includes the epitopes for the broadly neutralizing and transcytosis-blocking monoclonal human antibodies 2F5, 4E10 and 10E8 (Zwick et al., 2001, Huang et al., 2012) and, unlike the MPER itself (residues 650–683), maintains important structural and functional attributes of the native protein, including oligomerization and GalCer binding (Alfsen & Bomsel, 2002).

An effective vaccine against HIV-1 should ideally consist of components that target multiple steps of the viral transmission/infection process. Most importantly, it should engage the virus early in the cycle to minimize the chance of establishing viral reservoirs and subsequent re-dissemination (Valdiserri et al., 2003). From a worldwide perspective, HIV-1 transmission most commonly occurs through the exposure of mucosal surfaces to HIV-positive secretions (Pope & Haase, 2003, Hladik & McElrath, 2008, Haase, 2011). Therefore, the crucial involvement of the MPR in mucosal transmission of HIV and the well characterized, albeit rare, antiviral immune responses directed against this domain make it a prime candidate for an active vaccine.

However, by itself, the MPR was shown to act as a rather poor immunogen and was sensitive to its structural context (Denner, 2011). To overcome these limitations and in particular to boost immunogenicity at the mucosal surface, we have been exploring the MPR through its fusion to the mucosa-targeting cholera toxin B subunit, CTB (Matoba et al., 2004, Matoba et al., 2006, Matoba et al., 2008, Matoba et al., 2009). The CTB pentamer is taken up by mucosal immune cells through endocytosis mediated by binding to G_{M1} gangliosides (Merritt et al., 1994). Thus, a fusion protein comprised of CTB and MPR provides the target epitopes needed to elicit anti-HIV-1 antibodies directed at the MPR and combines the mucosal targeting of CTB and its immunogenicity. However, anti-MPR responses elicited by CTB-MPR were not optimal and indicated a need for an improved MPR-based immunogen (Matoba et al., 2004, Matoba et al., 2006, Matoba et al., 2008, Matoba et al., 2009, Matoba et al., 2011).

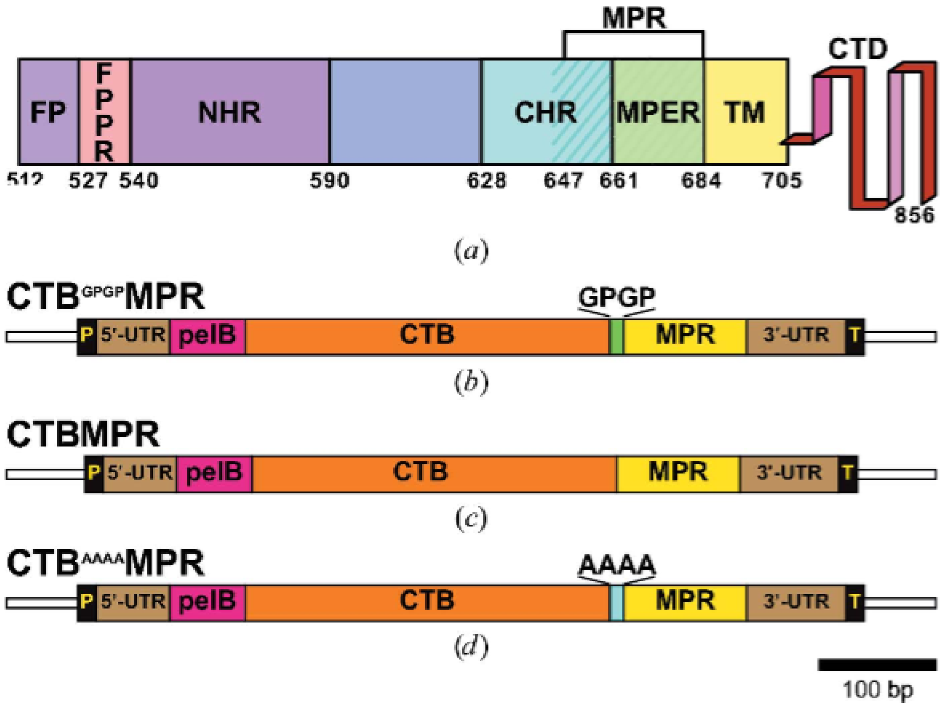


Figure 2.1: **gp41 and CTB-MPR variants** (a) The architecture of gp41. FP (residues 512–527), fusion peptide; FPPR (residues 528–539), fusion peptide proximal region; NHR (residues540–590), N-terminal heptad-repeat region; CHR (residues 628–661), C-terminal heptad-repeat region; MPER (residues 662–684), membrane-proximal external region; MPR (residues 647–684, hatched), membrane-proximal region; TM (residues 685–705), transmembrane domain; CTD (residues 706–856), cytoplasmic C-terminal domain. (b,c,d) DNA constructs for the expression in *E.coli* of the indicated CTB-MPR fusion proteins are based on elements of the pET-22b expression vector. P, T7 bacteriophage promoter; 5'-UTR, upstream untranslated region; pelB, the periplasmic targeting sequence of pectate lyase B of *Erwinia carotovora*; CTB, cholera toxin B subunit; MPR, the membrane-proximal region of the gp41 protein of HIV-1; 3'-UTR, downstream untranslated region; T, T7 terminator. The GPGP and AAAA linkers are indicated above their respective constructs. The three constructs encode the fusion proteins CTB^{GPGP}MPR (b), CTBMPR (c) and CTB^{AAAA}MPR (d) with expected molecular masses (after the processing of the pelB leader sequence) of 16.7, 16.4 and 16.7 kDa, respectively.

Understanding the function of MPR and the membrane-associated processes it takes part in, such as transcytosis and membrane fusion, as well as its interactions with the immune system, requires knowledge of its structure. To better understand the immunogenicity of the fusion protein and to enable us to design even more immunogenic

MPR fusion proteins, we turned to structural investigation. Here, we report on the expression of several novel variants of CTB-MPR with different linkers between the two fusion partners. We further report the purification of these proteins and their biochemical characterization, as well as initial crystallization experiments and X-ray crystallographic analysis.

2.3 Materials and methods

Vectors for bacterial expression of CTB-MPR fusion protein variants

The expression vectors used in this study were all based on the *Escherichia coli* periplasmic targeting vector pET-22b(-) (Novagen; Figure. 2.1b, 2.1c and 2.1d). The cloning of a synthetic gene encoding a fusion protein comprising CTB and the MPR with a flexible GPGP linker between them to obtain the plasmid pTM101 has been described previously (Matoba et al., 2004). To obtain a fusion protein without the C-terminal His tag engineered on the protein product of pTM101, we PCR-amplified the coding sequence with primers oTM066 and oTM123 (see Table 2.1 for a complete list of the oligonucleotides used in this work), and following digestion with *Nco*I and *Blp*I cloned them into the respective sites in the pET-22b(-) vector to obtain pTM199. In this work, the fusion-protein product of this vector is called CTB$^{\mathrm{GPGP}}$MPR.

Table 2.1: **Oligonucleotides used as primers in this study**

No.	Name	5’-Sequence-3’
1	oTM066	AGCCATGGGCACCCCACAAAACATCACTG
2	oTM123	ATTGCTCAGCGGTTCAGATCTTGATATACCAAAGC
3	oTM468	GGCAAATTCCCAAACCCAACAAGAGAAGAATG
4	oTM469	CTTGTTGGGTTTGGGAATTTGCCATGCTAATGGCAGC
5	oTM521	GCGGCCGCGGCCTCCCAAACCCAACAAGAG
6	oTM522	GGCCGCGGCCGCATTTGCCATGCTAATGG

The plasmid pTM199 served as the template to construct two additional variants of the fusion protein by overlap PCR (Aiyar et al., 1996). Briefly, in two separate PCR reactions, the two ‘end’ primers oTM066 and oTM123 were used, respectively, with two ‘mutagenizing’ primers oTM469 and oTM468 to amplify two partially overlapping fragments of the coding region of the fusion gene. The two fragments, now containing the deleted linker region, were gel-purified and used together as templates with the ‘end’ primers to PCR-amplify the complete fusion gene. The fragment was cloned into a pTOPO-TA vector (Invitrogen) to yield pTM545, and the correct sequence was verified. An *Nco*I–*Blp*I fragment from pTM545 was cloned into the corresponding sites of a pET-26b(+) vector to yield pTM556. The periplasmic-directed, linker-less version of the fusion protein encoded by this vector is referred to here as CTBMPR. A similar strategy (employing the ‘end’ primers oTM066 and oTM123 together with the ‘mutagenizing’ primers oTM522 and oTM521) was used to create a vector, pTM646, encoding a variant fusion protein with a tetra-alanine linker dubbed CTB^{AAAA}MPR.

Expression and purification of fusion-protein variants

Bacterial expression of CTB-MPR fusion-protein variants followed our previously published protocol for the $\mathrm{CTB}^{\mathrm{GPGP}}$MPR variant (Matoba et al., 2008). Similarly, we have modified the previously published purification protocol (Matoba et al., 2008) to avoid precipitation of the protein at high pH and to replace the previously used detergents with detergents that would be compatible with crystallization. Briefly, cell pellets from 2 l culture (approximately 5 g) were resuspended in 20 ml ice-cold phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) containing 1 mM phenylmethanesulfonyl fluoride (PMSF), a serine protease inhibitor, to prevent protein degradation. The cells were lysed by passing them twice through a microfluidizer (Microfluidics Microfluidizer) with PMSF added again after the first pass. The lysate was collected in a 40 ml Oak Ridge tube and was centrifuged at 36,000 g for 20 min. The insoluble fraction was washed once by repeated resuspension (in 30 ml ice-cold PBS) and centrifugation. If not immediately used, the pellet was frozen at 80°C.

The pellet, containing the membrane fraction, was resuspended in 30 ml buffer (20 mM bicine pH 8.0, 500 mM NaCl). To fully homogenize the solution, the sample was sonicated at 20% amplitude in 30 s runs (Model 300V/T Ultrasonic Homogenizer, Biologics Inc.) until a homogenous turbid suspension was obtained. The detergent n-dodecyl-β-D-maltoside (βDDM) was used for solubilization. A stock solution of 10% (w/v) was added to a final concentration of 1% (w/v). The protein was solubilized at 4°C overnight with agitation.

The protein solution was centrifuged at 36,000 g for 20 min and the pellet was discarded. A gravity-driven column (Bio-Rad Econo-Column) containing cobalt affinity resin (40 ml bed volume; Talon, Clontech) was equilibrated with binding buffer (resuspension buffer supplemented with 0.05% βDDM). The sample was then loaded onto the column and washed with six bed volumes of binding buffer and ten bed volumes of wash buffer (20 mM bicine pH 8.0, 50 mM NaCl, 5mM imidazole, 0.05% βDDM) to remove weakly bound proteins. Tightly bound proteins were eluted by the application of three bed volumes of elution buffer (20 mM bicine pH 8.0, 50 mM NaCl, 150 mM imidazole, 0.05% βDDM).

The eluted fractions were pooled and then concentrated to approximately 2 mg ml^{-1} using 50 kDa molecular-weight cutoff (MWCO) concentrators (Vivaspin 20 VS2031, Sartorius Stedim Biotech). Concentrated samples were further purified by size-exclusion chromatography (SEC; Superdex 200, GE Healthcare; column volume 24 ml, fluid phase 8 ml) using a high-pressure liquid-chromatography instrument (HPLC; ÄKTAexplorer, Pharmacia). The running buffer consisted of 20 mM HEPES pH 7.5, 10 mM calcium chloride, 0.02% βDDM. For analytical separations, a sample (200 µl) of concentrated CTB- MPR variant was loaded onto the SEC column and chromatography was performed at a flow rate of 0.5 ml min^{-1}. The column was loaded with a maximum of 1 ml sample for preparative separation runs, with only slight broadening of the peaks being observed. The protein elution was detected by absorption at 280 nm. Fractions corresponding to individual peaks were collected and pooled.

The concentrations of CTB-MPR variant preparations were determined spectrophotometrically (A_{280}) using $\varepsilon_{280} = 39,380\ M^{-1}cm^{-1}$ (ε_{280} was calculated with the

ProtParam web application; http://web.expasy.org/protparam/). Assembly of pentamers of the CTB-MPR variants was monitored by ELISA using G_{M1} gangliosides for capture and the MPR-specific human monoclonal antibody 2F5 as described previously (Matoba et al., 2008) and by nondenaturing SDS–PAGE (see below).

SDS-PAGE and immunoblotting

SDS–PAGE using tricine-based buffers in a Bio-Rad Mini-PROTEAN Tetra Cell was performed as previously described by (Lawrence et al., 2011) based on the method of (Schagger, 2006) Following electrophoresis, the gels were stained with Coomassie Brilliant Blue, subjected to silver staining (Lawrence et al., 2011) or processed for immunoblotting.

For immunoblotting, the acrylamide gel was rinsed with water and equilibrated in anode buffer consisting of 60 mM Tris, 40 mM N-cyclohexyl-3-aminopropanesulfonic acid (CAPS), 15% methanol. The PVDF membrane was prepared by soaking in 100% methanol and then equilibrated in cathode buffer consisting of 60 mM Tris, 40 mM CAPS,0.1%SDS.The gel and the membrane were sandwiched between extra-thick blot filter papers (Bio-Rad) soaked in the appropriate electrode buffer and proteins were electroblotted for 30 min at 120 mA (Bio-Rad Transfer-blot SD Semi-dry Transfer Cell). Following blocking for 1 h in PBSTM (PBS, 0.05% Tween 20, 5% dry milk), the PVDF membrane was further incubated in the presence of the 2F5 monoclonal antibody (kindly provided by the NIH’s AIDS Reagent Program; 1:10 000 dilution; (Purtscher et al., 1996)). The membrane was then washed for 3 x 30 min in PBST (PBS, 0.05% Tween 20) prior to incubation (1 h) with rabbit anti-human IgG conjugated to horseradish peroxidase (1:20 000 dilution in PBSTM; Santa Cruz Biotechnology sc-2923). Following three additional 30 min washes, the PVDF membrane was then soaked with Bio-Rad Clarity Western ECL substrate solution and imaged with a UVP BioSpectrum 500C Imaging System.

CTB forms a very stable pentamer that resists dissociation by SDS in a monomer concentration-dependent manner. Nonetheless, CTB pentamers can be denatured by heat and by reduction of the intermolecular disulfide bridges that stabilize the oligomers (Zrimi et al., 2010, Yasuda et al., 1998). Nondenaturing SDS–PAGE was conducted as described above except that DTT was omitted from the loading buffer and the samples were not boiled prior to loading them onto gels (Matoba et al., 2008).

Dynamic light scattering

Dynamic light-scattering (DLS) measurements were performed using a NaBiTec GmbH setup comprising a SpectroSize 302 (Molecular Dimensions) in combination with an S6D microscope (Leica). The purified protein sample (concentrated to 8 mg ml^{-1} as described above) was illuminated in a 3 µl hanging drop using a 24-well crystallization plate (VDX Greased Plate, Hampton Research) covered with siliconized-glass circular cover slides (22 mm; Hampton Research). The well itself was filled with 600 µl SEC running buffer. Prior to the measurement, the protein solution was centrifuged (1,000 g, 30min, 4°C) to remove possible dust particles. During the measurement, the temperature was set to 20°C. Ten consecutive measurements, each with an integration time of 20 s, were averaged. An estimate of the hydrodynamic size was obtained with the instrument software using the following parameters: refractive index 1.33, viscosity 1.006, shape factor 1.0, hydrated shell 0.2 nm.

Crystallization experiments

For crystallization experiments, the fusion-protein preparations were concentrated to a final concentration of 10 mg ml^{-1} using 100 kDa MWCO concentrators (Amicon Centricon YM-100). Initial broad screening for crystallization conditions used NeXtal crystallization kits (The PEGs Suite, The MB Class Suite and The MB Class II Suite) with the vapor-diffusion technique. Screening was performed using 96-well plates (Qiagen CrystalEX 96-well Conical Flat Plate) with the sitting-drop method, where each reservoir well contained 100 µl precipitant solution. The purified protein solution was then mixed in a 1:1 ratio (1 µl:1 µl) with the reservoir solution in the sitting-drop well.

Conditions that produced crystals served to guide us in fine screening by the hanging-drop method using 24-well plates (Hampton Research VDX Greased Plates), with each reservoir well containing 900 µl precipitant solution. The purified protein solution was then mixed with the reservoir solution (3 µl each) on a siliconized glass circle cover slide (22 mm; Hampton Research) and the slide was used to seal the well.

As the broad screening produced crystals in the presence of polyethylene glycol (PEG), our fine screens centered on the addition of PEGs of various defined chain lengths (molecular weights ranging from 300 to 4000) under pH, salt and ionic strength conditions that produced crystals that were hexagonal from one viewing plane and completely round as viewed perpendicularly. Specifically, combinatorial screens involved testing various buffers (50 mM of either sodium acetate pH 4.6, MES pH 6.5 or HEPES pH 7.5) and salts (100 mM of either NH_4Cl, NaCl, $CaCl_2$ or $MgCl_2$).

Fine screens for optimal crystallization conditions of $CTB^{GPGP}MPR$ were conducted with 0.1 M HEPES pH 7.5 and varying concentrations of PEG 400. The best crystals appeared using a reservoir solution consisting of 34% PEG400, 0.2 M $BaCl_2$, 20% ethylene glycol. The hanging drop contained 1.5 µl reservoir solution, 0.5 µl 2M ammonium acetate, 2 µl protein sample and 0.41 µl 10% CYMAL-4 (yielding a final concentration of 0.74% or 2x the critical micelle concentration).

Fine screens for optimal crystallization conditions of CTBMPR were conducted with the choice buffer (50 mM HEPES pH 7.5) and focused on varying concentrations of choice PEGs (20–40% PEG 300, 5–20% PEG 3000 or 5–20% PEG 4000) in the presence of 100 mM NH_4Cl, NaCl or $CaCl_2$. In parallel, we conducted salt-concentration screens (50–200 mM) for NH_4Cl, NaCl and $CaCl_2$ in solutions that contained either 25% PEG 300, 10% PEG 3000 or 10% PEG 4000. Finally, under the choice conditions of buffer, PEG and salt (50 mM HEPES pH 7.5, 25% PEG 300, 200 mM NH_4Cl) we conducted an additive screen (Hampton Research Additive Screen), in which 96 different additives were added (1µl) to the individual drop well in a Qiagen CrystalEX 96-well Conical Flat Plate along with the protein and reservoir drop mixture, which consisted of 50 mM HEPES pH 7.5, 20% PEG 300, 10%(w/v) either glycerol, 2-propanol or CYMAL-4 and 200 mM salt (either NH_4Cl, NaCl or $CaCl_2$).

Fine screens for optimal $CTB^{AAAA}MPR$ crystallization conditions were performed with 100 mM Tris pH 8.5 or 50 mM HEPES pH 7.5 while varying the concentrations of either PEG 1000 (10–30%) or PEG 3350 (5–20%) in the presence of 200 mM of either NH_4Cl, NaCl, $CaCl_2$ or NH_4HCO_2. In parallel, salt-concentration screens of NH_4Cl, NaCl, $CaCl_2$ and NH_4HCO_2 from 0.05 to 0.2 M were set up with 100 mM Tris pH 8.5 or 50 mM HEPES pH 7.5 and either 25% PEG 1000 or 10% PEG 3350.

Nano/microcrystals of $CTB^{AAAA}MPR$ were prepared by the ultrafiltration method. In this method, the supersaturated zone is reached by concentration of the protein by ultra- filtration while salt, precipitant and buffer concentrations remain constant. 300 µl purified protein (10 mg ml^{-1}) was mixed with the same volume of precipitant solution consisting of 200 mM NH_4HCO_2, 30% PEG 3350, 10 mM $CaCl_2$, 20 mM HEPES pH 7.5 in a 100 kDa cutoff concentrator (Amicon Microcon YM-100). The setup was then centrifuged to reduce the retentate volume by half to regain the original protein concentration of 10 mg ml^{-1}. Following overnight incubation, more precipitant solution was added (30µl) to further increase the yield of nano/microcrystals. Crystallization conditions are summarized in Table 2.2.

Table 2.2: **Crystallization conditions**

Construct	Condition
$\mathrm{CTB^{GPGP}MPR}$	34% PEG 400, 0.2 M $BaCl_2$, 20% ethylene glycol, 0.5 M ammonium acetate, 0.74% CYMAL-4
	25–30% PEG 400, 0.2 M $CaCl_2$, 0.1 M HEPES pH 7.5, 0.3M galactose, 80–100 mM NaCl
CTBMPR	25-30% PEG 300, 0.2 M $CaCl_2$, 0.05 M HEPES pH 7.5, 0.02% βDDM
	25–30% PEG 300, 0.2 M NaCl, 0.05 M HEPES pH 7.5, 0.02% βDDM
	25–30% PEG 300, 0.2 M NH_4Cl, 0.05 M HEPES pH 7.5, 0.02% βDDM
$\mathrm{CTB^{AAAA}MPR}$	8-12% PEG 3350, 0.1–0.2 M NH_4HCO_2, 0.01 M $CaCl_2$, 0.05 M HEPES pH 7.5, 0.02% βDDM
$\mathrm{CTB^{AAAA}MPR}$ nano/microcrystals	30% PEG 3350, 0.2 M NH_4HCO_2, 0.01 M $CaCl_2$, 0.05 M HEPES pH 7.5, 0.02% βDDM

Standard X-ray crystallography

Characterization of the $\mathrm{CTB^{GPGP}MPR}$ crystals was performed using synchrotron X-ray radiation on beamline 8.2.2 at the Advanced Light Source (ALS) at a wavelength of 1 Å. The crystals were flash-cooled in liquid nitrogen with a cryoprotectant solution (15% ethylene glycol, 50% PEG 400, 100 mM HEPES, 60 mM NaCl, 200 mM BaCl2, 150 mM imidazole, 0.017% βDDM) and diffraction data were collected at 100 K using an Oxford Cryostream. A total of 520 data frames were collected using 0.25° oscillations and an exposure time of 2.275 s per frame with an ADSC 315 detector.

Serial femtosecond nano/microcrystallography

Nano/microcrystals were grown on-site and were analyzed by DLS prior to serial femtosecond X-ray nano/micro-crystallography using the high-energy free-electron laser at the Coherent X-ray Imaging (CXI) endstation of the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory (Experiment L432, February 2012). This method allows data to be collected from hundreds of thousands of sub-micrometre nano/microcrystals (by spraying them across a pulsed X-ray laser beam) using X-ray snapshots so brief that they outrun radiation damage (for a review of the method, see(Spence et al., 2012)). Data were collected from a stream of fully hydrated nano/microcrystals. Experimental details of the beamline and data collection at CXI have been described by (Boutet & Williams, 2010) and (Boutet et al., 2012). A suspension of nano/microcrystals of CTB^{AAAA}MPR (9.1 mg ml^{-1}, total volume of 330µl) was supplied to the FEL X-ray beam using a gas-focused liquid microjet of 4µm diameter at 20°C, a temperature-controlled antisettling device and a flowrate of 10 µl min^{-1} using a gas dynamic virtual nozzle (Weierstall et al., 2012, DePonte et al., 2008, Weierstall et al., 2008, Lomb et al., 2012). X-ray data were collected from the crystals at an energy of 6.3 keV with a 50 fs pulse duration and an X-ray pulse repetition rate of 120 Hz. Diffraction patterns from protein crystals were identified and selected using the hit-finding program Cheetah (Barty et al., 2014), and indexing and merging was performed using CrystFEL (Kirian et al., 2011, White et al., 2012).

2.4 Results and discussion

CTB^{GPGP}MPR

Previous work suggested that the immunogenicity of the MPR depends on its structural context, especially when fused to other proteins and peptides as is the case for CTB-MPR (Gach et al., 2011, Montero et al., 2012, Matoba et al., 2008, Matoba et al., 2011). Three different CTB-MPR fusion variants were designed that would differ in the linker peptide between the two fusion partners.

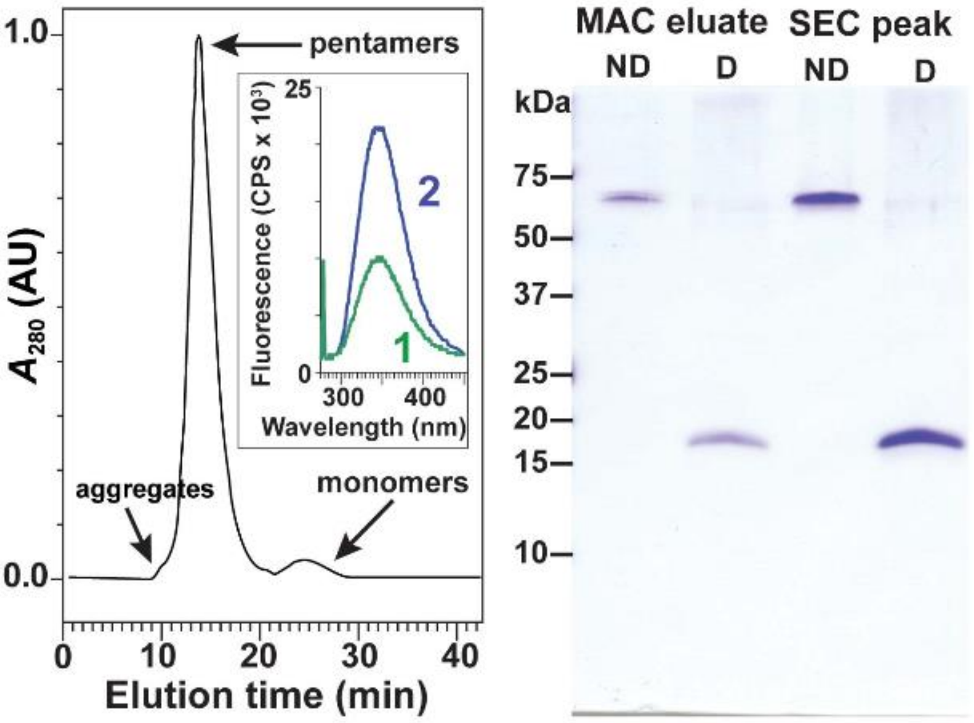
The original fusion protein that was described previously (Matoba et al., 2004) contained a GPGP linker. It is denoted here as CTB^{GPGP}MPR (Figure 2.1b). Two additional variants were created as part of the present study with the GPGP linker either deleted (CTBMPR; Figure 2.1c) or replaced by a tetra-Ala linker (CTB^{AAAA}MPR; Figure 2.1d). To maximize expression levels in bacterial cells, all constructs reported here were devoid of a terminal histidine tag. Instead, we took advantage of a peculiarity of the CTB pentamer, preserved in the context of the fusion proteins, that allows it to specifically bind to metal-affinity resin (Dertzbaugh & Cox, 1998). Importantly, in the absence of a His tag only assembled pentamers can bind to the metal column (Dertzbaugh & Cox, 1998). The fusion proteins were expressed as described by (Matoba et al., 2008) and were purified as described in §2 using the mild detergent βDDM for solubilization and in all further purification steps to facilitate crystallization efforts and biophysical analyses.

The purification scheme described above for CTB^{GPGP}MPR fusion proteins resulted in >99% purity based on silver-stained polyacrylamide gels (Matoba et al., 2008). As previously demonstrated by nondenaturing gel electrophoresis and by G_{M1}

ganglioside ELISA (Matoba et al., 2008, Matoba et al., 2004), such protein preparations were highly homogeneous, consisting of primarily pentameric CTB^{GPGP}MPR and only minor amounts of higher molecular-weight aggregates and monomeric protein. We were able to separate these various molecular forms by SEC–HPLC (Figure 2.2a). Oligomeric state assignment of the peaks was performed based on parallel SEC–HPLC runs with molecular-weight standards. This assignment was confirmed by resolving proteins in the pooled fractions corresponding to the peaks by SDS–PAGE under nonreducing conditions, which allows CTB to retain its pentameric organization (Figure. 2.2b; (Yasuda et al., 1998, Zrimi et al., 2010)). Taken together with the fact that that CTB^{GPGP}MPR binds to the affinity resin, we conclude that the fusion protein is a stable pentamer.

Taking advantage of the presence of five tryptophan residues within the MPR domain (with one more within the CTB moiety), we subjected the proteins in the pooled fractions corresponding to CTB^{GPGP}MPR pentamers to fluorescence spectroscopy (Figure 2.2a, inset). The emission spectrum revealed that the Trp residues in the pentamers were exposed to the aqueous milieu (peak emission at 347 nm; (Ni et al., 2011, Reshetnyak et al., 2001)). The stability of the pentamers was demonstrated by the conservation of the Trp emission profile upon purification and concentration of the protein.

We screened a large number of crystallization conditions which included systematic variation of the protein concentration, pH, precipitant and ionic strength. Furthermore, we tested the reversibility of the crystallization conditions. The initial screens provided important information on the solubility of CTB^{GPGP}MPR. The addition of galactose is essential for crystallization of the protein, while only irreversible precipitation was observed in its absence. Reversible precipitation was observed at pH 7–8 and at medium salt concentrations (50–250 mM). Crystallization was favored by the addition of divalent cations (e.g. Ca^{2+}) over monovalent cations, and shorter-chain polyethylene glycol polymers (PEGs) were the preferred precipitants.



(*a*) (*b*)

Figure 2.2: **CTB^{GPGP}MPR oligomeric states** (*a*) separation of aggregates and monomers from the pentameric CTB^{GPGP}MPR protein by gel filtration on a Superdex 200 column. Assembly status was estimated from parallel resolution of molecular-mass standards (not shown). Inset, tryptophan fluorescence emission spectra of pentameric CTB^{GPGP}MPR in pooled gel-filtration fractions corresponding to the major peak in (*a*). 1 (green), pentamers; 2 (blue), concentrated (Centricon 100) pentamers. Excitation was at 280 nm. (b) Proteins in the unconcentrated metal-affinity chromatography (MAC) eluate and in the size-exclusion chromatography (SEC) fraction corresponding to the main peak of the chromatogram in (*a*) were resolved by SDS–PAGE under nondenaturing (ND; no DTT and no boiling) and denaturing (D) conditions. Molecular-weight standards indicate that CTB^{AAAA}MPR is organized into SDS-stable pentamers. The compact pentamers have a slightly higher electrophoretic mobility than expected based on their mass alone

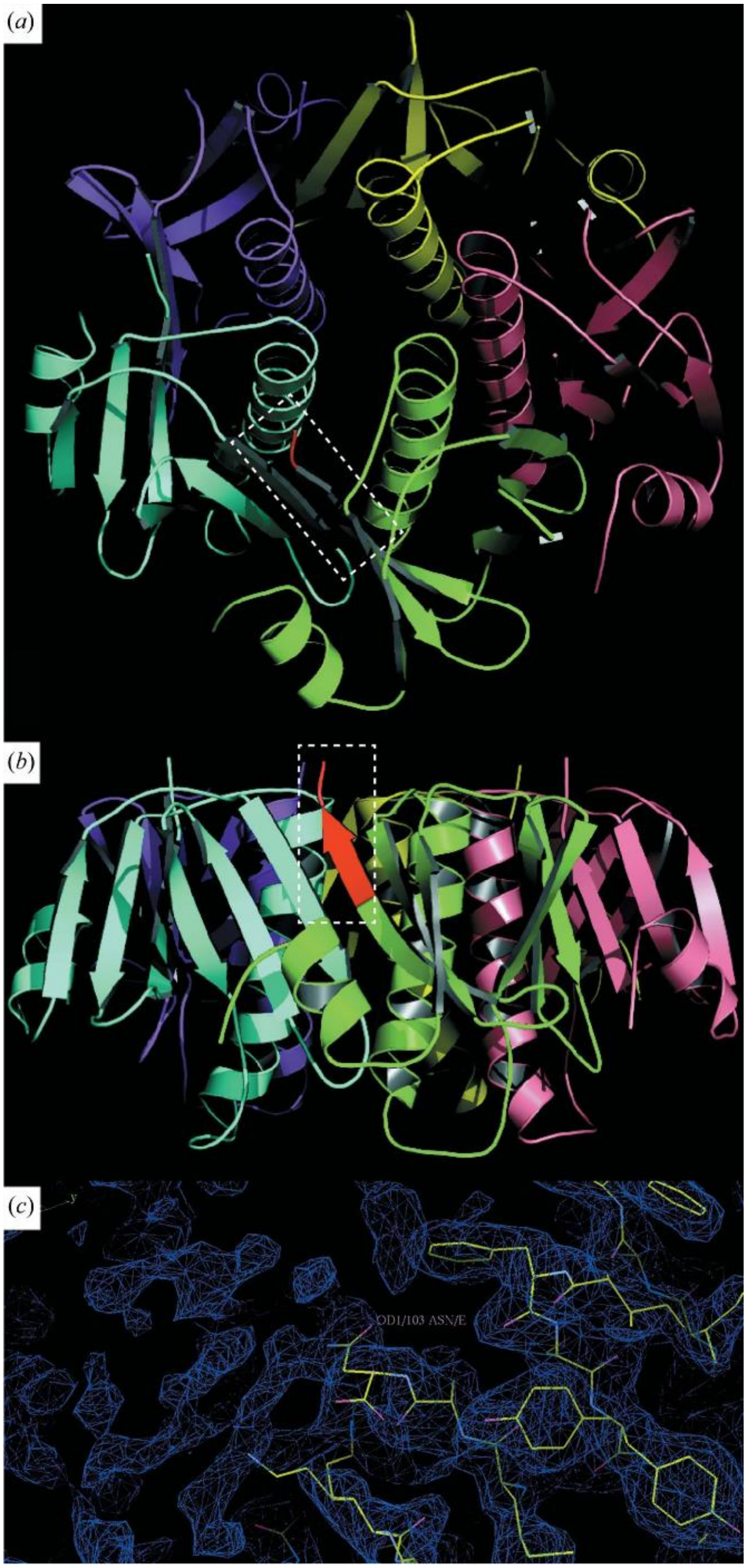


Figure 2.3: **The CTB^{GPGP}MPR structure** reveals the expected pentameric ring arrangement typical of wild-type CTB but not the structure of the MPR. Cartoon representation of the crystal structure of CTB^{GPGP}MPR in two orientations: (*a*) top view, (b) side view. Each subunit is indicated by a different color. The C-terminus of one of the subunits is indicated in red. This region is shown in close-up in (c). (c) 2FoFc electron-density map at a contour level of 1.5 of the C-terminus of CTB in CTB^{GPGP}MPR, which was phased with the pentameric CTB model (PDB entry 1jr0; (Pickens et al., 2002)) using molecular replacement (McCoy et al., 2007) Electron density can be seen beyond the terminal asparagine of CTB where the GPGP linker and MPR connect.

conditions where crystals formed. The crystals were grown in 0.1 M HEPES pH 7.5, 25–30% PEG 400, 0.2M $CaCl_2$, 0.3 M galactose, 80–100 mM NaCl at a protein concentration of 5 mg ml^{-1}. The vapor-diffusion method (sitting drop) using ‘screw-cap’ plates (NeXtal) was used. Isolated crystals were cooled in liquid nitrogen in crystallization buffer containing 36% PEG 400 as a cryoprotectant. X-ray data were collected on beamline 8.3.1 at the Advanced Light Source (ALS). Most of the 50 µm crystals diffracted to about 20 Å resolution. The reflections were broad and anisotropic, indicative of the low order of the crystals in three dimensions. One unit-cell parameter was identified to be 45 Å.

Under slightly different crystallization conditions that included the presence of Zn^{2+} and lipids, crystals were observed that diffracted to a resolution limit of 2.3 Å. A full data set was collected from these crystals at the Advanced Photon Source (Table 2.3). Unfortunately, only the CTB region was ordered in the electron-density map, definitively demonstrating its pentameric nature (Figure. 2.3a and 2.3b). Weak electron density was observed that extended the C-terminus of CTB, but the structure of the MPR region could not be resolved in the crystals (Figure 2.3c). We hypothesized that this may be caused by the flexibility of the GPGP linker allowing the MPR region to assume multiple positions in the crystals.

Table 2.3: **Crystallographic data for CTB^{GPGP}MPR**

Values in parentheses are for the highest resolution bin.

Wavelength (Å)	1.0
Resolution range (Å)	59.48–2.10 (2.21–2.10)
Space group	R3:H
Unit-cell parameters (Å/°)	a = b = 174.39, c = 64.71, α = β = 90, λ = 120
Multiplicity	3.8 (3.8)
Completeness (%)	99.95 (100.00)
Mean I/σ (I)	6.68 (1.93)
Wilson B factor (Å^2)	30.72
R_{merge}	0.136 (1.302)
R factor	0.214 (0.315)
R_{free}	0.249 (0.388)
No. of atoms	4365
No. of macromolecules	4100
No. of waters	265
No. of protein residues	515
R.m.s.d., bonds (Å)	0.008
R.m.s.d., angles (°)	1.08
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.8
Ramachandran outliers (%)	0.2
Clashscore	8.52
B factors (Å^2)	
Average	40
Macromolecules	39.9
Solvent	42.6

CTBMPR

To test our hypothesis regarding linker flexibility, we created a second fusion protein variant in which the movement of the MPR domain was expected to be restricted by direct fusion of the MPR to the C-terminus of the CTB protein (CTBMPR; Figure 2.1c).

The purification procedure for the linker-less fusion protein CTBMPR followed the same scheme as outlined above except that elution was conducted batchwise with extended incubation periods (from 10 min to 16 h) and higher concentrations of imidazole (300 mM) were required to elute most of the protein from the column (Figure 2.4). The molecular mass of the fusion protein as estimated based on SDS–PAGE resolution (Figure 2.4a) and immunoblotting (Fig. 2.4b) fitted the calculated value based on the sequence of the protein (17 kDa).

The homogeneity of the fusion protein in the pooled eluted fractions was tested by SEC–HPLC. This demonstrated that the preparation can be resolved into various peaks (Fig. 5). The results showed that unlike CTB^{GPGP}MPR, the linker-less fusion protein exists in an equilibrium between several oligomeric molecular forms. Assignment of the oligomeric forms is based on the similarity in the elution volumes of the respective peaks to those of CTB^{GPGP}MPR. Pentamers are not the predominant form of the linker-less CTBMPR protein, at least under our purification conditions. A substantial monomeric population is present alongside the pentamers in preparations obtained under similar purification conditions to those used in the purification of CTB^{GPGP}MPR. In fact, since all of the protein loaded onto the SEC–HPLC column was specifically eluted from the metal-affinity column (and consequently must have been pentameric), it is likely that the

CTBMPR pentamer undergoes (partial) disassembly during manipulation following the metal-affinity chromatography stage.

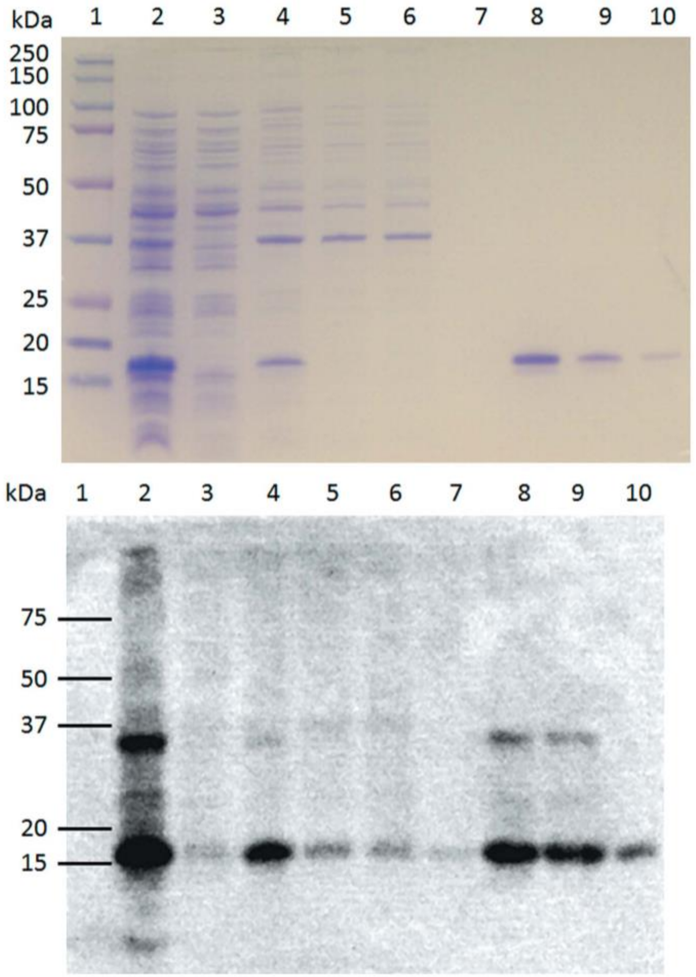


Figure 2.4: **Affinity chromatography purification of CTBMPR.** Protein samples from various steps in the purification process were resolved next to molecular-weight markers (lane 1) by SDS–PAGE and the gel was stained with Coomassie Brilliant Blue (upper panel). The whole cell lysate (lane 2) was spun down and the aqueous fraction (lane 3) was discarded. Membrane proteins were extracted from the pellet with βDDM (lane 4) purified over an affinity chromatography column. The flowthrough was collected (lane 5) and the column was extensively washed as described in the text (lane 6, first wash fraction; lane 7, last wash fraction). Elution required a larger volume of imidazole elution buffer to elute most of the protein bound to the column (lanes 8–10) than expected based on previous results with CTB GPGP MPR (Matoba et al., 2008). Immuno-blotting was performed on the same samples using monoclonal 2F5antibodies (lower panel)

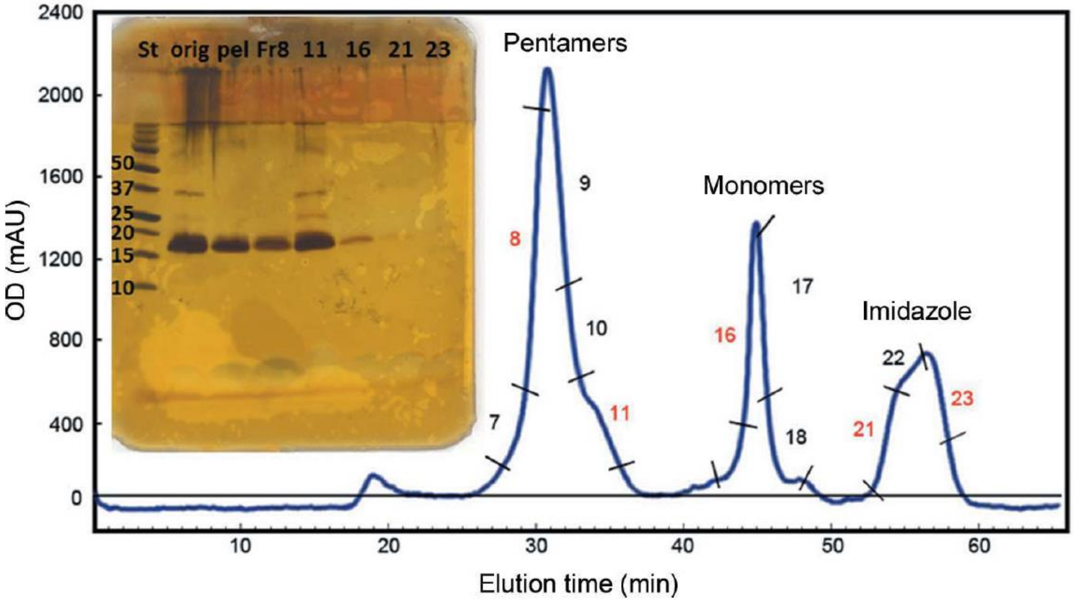
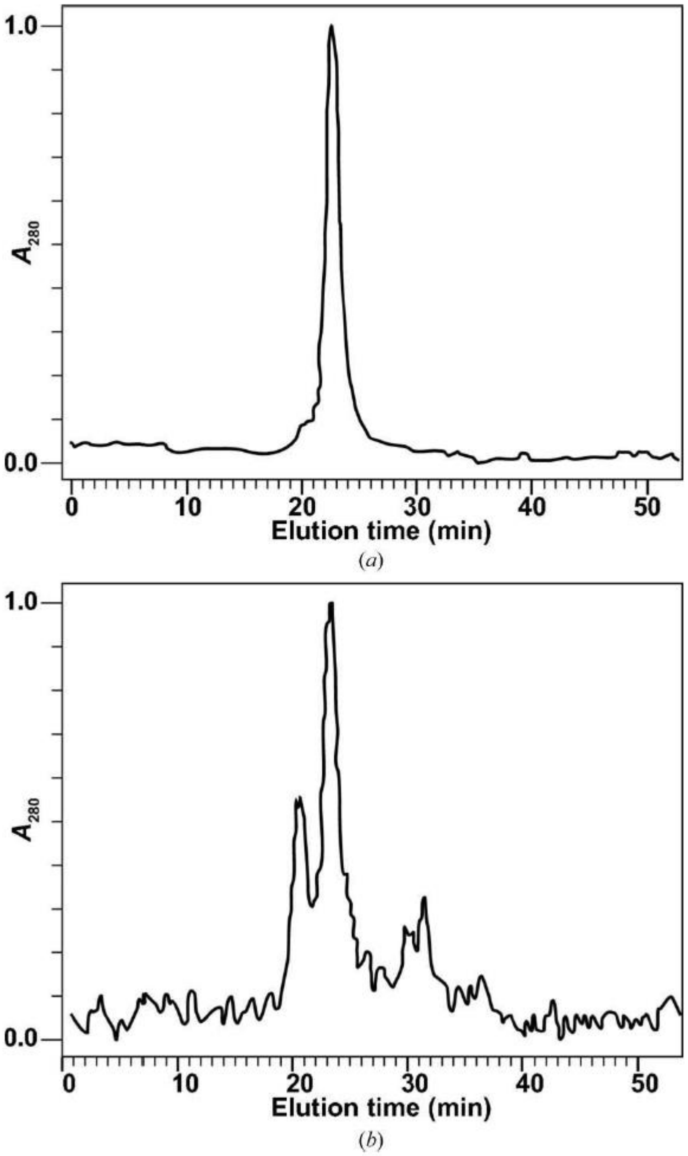


Figure 2.5: **Size-exclusion chromatography of CTBMPR** shows it exists in several metastable oligomeric forms. Affinity-purified CTBMPR was resolved by SEC–HPLC, yielding three major peaks probably corresponding to pentamers (fraction 8) and monomers (fraction 16). Fractions 21 and 23 did not contain appreciable amounts of protein and are likely to contain high concentrations of imidazole. The shoulder at the right of the pentamer peak (fraction 11) may represent the less stable intermediates tetramers and dimers. These fractions (numbered in red), alongside the original sample and a precipitate that formed in the original sample, were analyzed by SDS–PAGE followed by silver staining (inset).

While gp41 is generally assumed to form trimers (Liu et al., 2008, Atilgan et al., 2010) in its pre-fusion form, the involvement of the MPR domain in trimerization is less clear and evidence for alternative associations exist (see, for example,(Alfsen & Bomsel, 2002)). This suggests that the equilibrium between the various oligomeric states is dynamic and may be explained by the competing tendencies of the CTB fusion partner to form pentamers, while the MPR fusion partner may push the equilibrium against pentamerization.

To investigate this hypothesis, we separately pooled the fractions corresponding to the monomeric and the pentameric forms of CTBMPR, concentrated them and analyzed them separately by SEC–HPLC (Figure 6). The pentamer appeared to be stable,



(*a*)

(*b*)

Figure 2.6: **The CTBMPR oligomeric state** is affected by the concentration of the protein. SEC–HPLC fractions corresponding to the pentamer (a) and monomer (b) peaks (Figure 5) were subjected separately to a second SEC–HPLC purification. Absorbance is normalized to the highest peak

leading to a single peak with the same elution time. However, upon concentration of the monomer-containing fractions, most of the fusion protein was shown to elute as a fraction corresponding to the pentamer fraction, suggesting a reorganization of the protein into pentamers. These results provided support for our hypothesis that a dynamic concentration-dependent equilibrium exists between the various oligomeric forms of CTBMPR, where lower concentrations favor monomers and higher concentrations favor pentamer formation.

We carried out crystallization experiments of CTBMPR using the vapor-diffusion method and broad crystal screening, as described earlier, to identify conditions where crystals were able to form. Disappointingly, only a few conditions led to ordered precipitate or pseudo-crystals, and finer screens around the conditions did not produce three-dimensionally ordered crystals. A possible explanation is that the instability of the oligomeric states hinders the formation of crystals.

CTB^{AAAA}MPR

Based on the results with CTBMPR, we designed a third variant of the CTB-MPR fusion protein, CTB^{AAAA}MPR (Figure 2.1d), that links the two fusion partners with a short polyalanine peptide that is expected to assume an α-helical conformation (Oneil & Degrado, 1990). Our aim was to allow the fusion protein to assemble into stable pentamers by facilitating the ability of the MPR moieties to interact with each other while avoiding presumed disorder induced by the flexible GPGP linker. The SEC–HPLC purification profile resembled that for the CTB^{GPGP}MPR variant (Figure 2.7a). The formation of the pentamer, as verified by nondenaturing SDS–PAGE, was still concentration-dependent; however, the pentamer was much more stable for CTB$^{\text{AAAA}}$MPR than for the linker-less construct CTBMPR (Figure 2.7b).

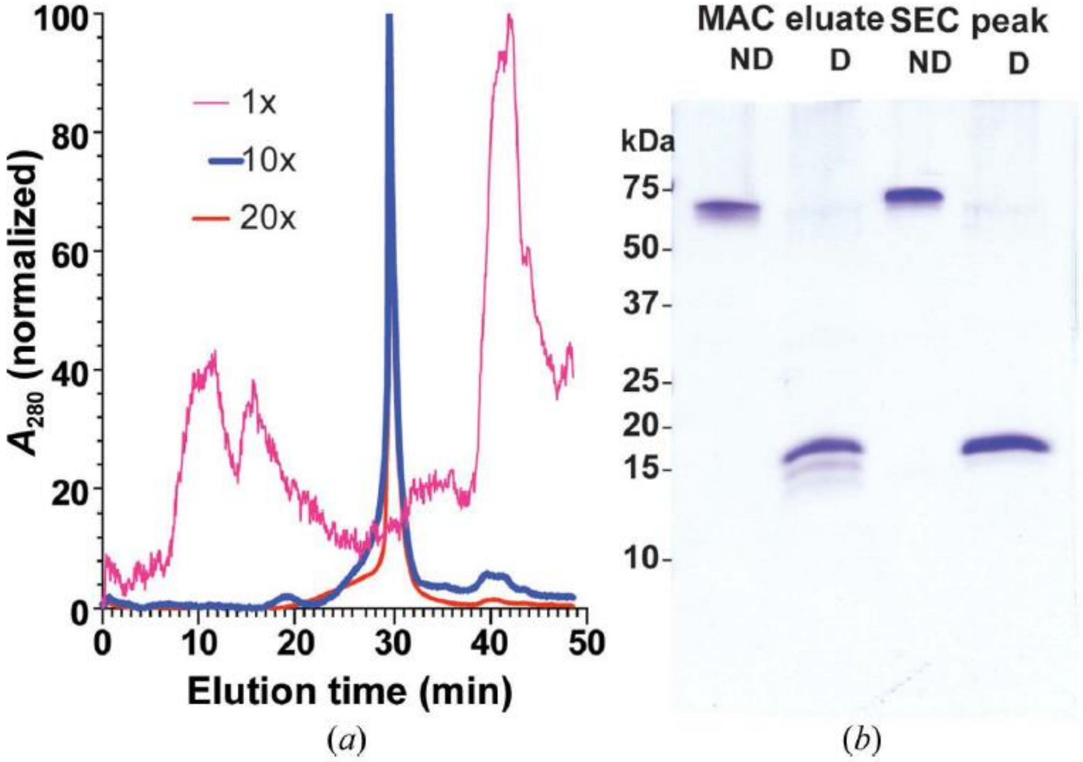
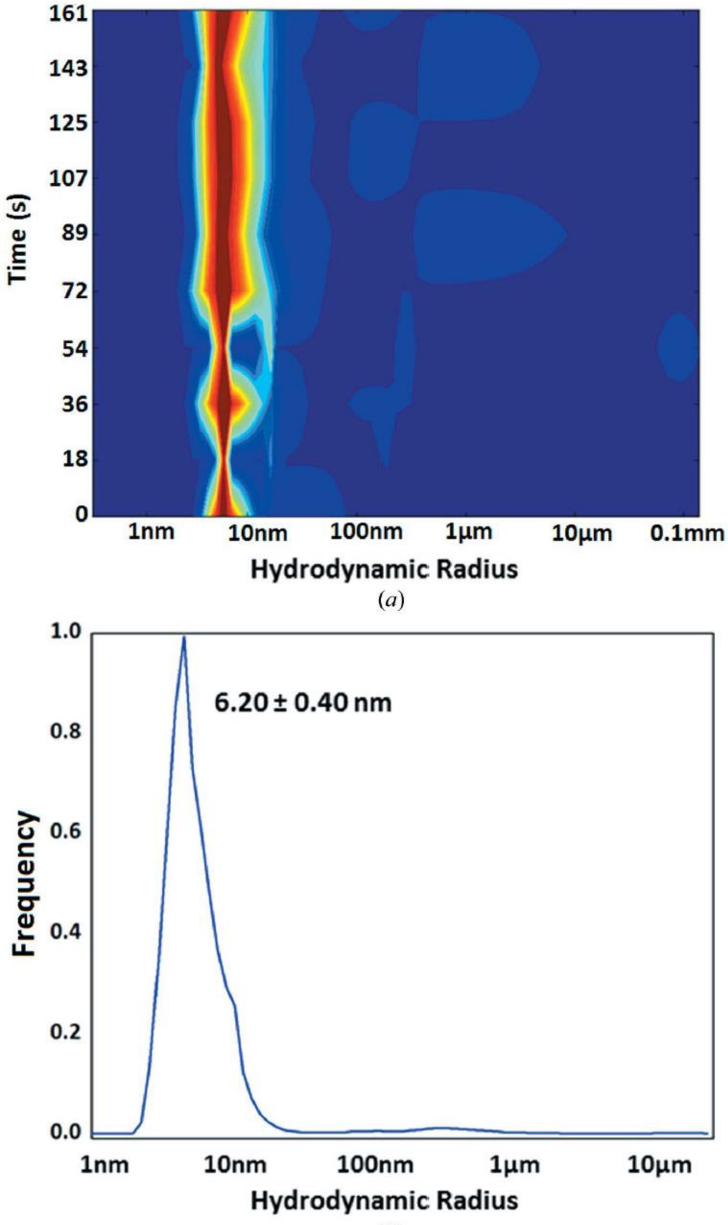


Figure 2.7: **CTB$^{\text{AAAA}}$MPR resolved as an oligomer by SEC–HPLC**. Pink line, the Talon column eluate (not concentrated). Blue line, 10 concentrated eluate sample. Red line, 20 concentrated eluate sample. Spectrograms were normalized to the highest peak. Inset: proteins in fractions corresponding to the main peak of the 20 concentrated eluate chromatogram were resolved by SDS–PAGE under nondenaturing (ND; no DTT and no boiling) and denaturing (D) conditions. Molecular-weight standards indicate that CTBAAAAMPR is organized into SDS-stable pentamers.



(*a*)

(*b*)

Figure 2.8: **DLS data of CTB^{AAAA}MPR** shows that it is monodisperse as a high-order oligomer. (a) DLS measurements were performed so that the size distribution in the sample was analyzed for 20 s and the measurement was repeated consecutively ten times. The moment-to-moment fraction of particles estimated to have a particular hydrodynamic radius is color-coded and shown as a heat plot (red, >90%; blue, none). The narrow vertical and red profile shown indicates high stability over the duration of the measurement and low polydispersity. Time: the total duration of the scanning session (200 s). (b) A distribution curve of particle-size frequencies gives a more quantitative evaluation of the polydispersity, with the mean SD indicated next to the peak. The standard deviation of the size distribution is only 6% of the mean, indicating low polydispersity.

We obtained the size distribution of the purified CTB^{AAAA}MPR by dynamic light scattering (DLS) to determine whether the protein preparation was monodisperse (Figure 2.8). At 8 mg ml^{-1}, the hydrodynamic radius (Stokes radius, r) of the detergent-solubilized protein (i.e. of the protein–detergent micelles) was determined to be 6.2±0.4 nm. The polydispersity was estimated to be 6%, which is well below the 10–15% level considered as monodisperse (Proteau et al., 2010). Note that the DLS measurement in Figure 2.8 shows the direct scattering intensity, which is not corrected for the molecular mass of the particles to detect even traces of aggregates. As the increase in scattered intensity is proportional to r^6, we calculated that the sample was highly monodisperse and contained less than 0.00001% aggregates. Since the exact geometry of CTB^{AAAA}MPR is not known, a generic set of parameters was used assuming that the folded state is spherical with an estimated molecular mass of ~210 kDa, which includes the detergent bound to the protein. The DLS data indicated that CTBAAAAMPR may form a dimer of pentamers, corresponding to a molecular weight of 170 kDa for the protein, while a trimer of pentamers would be 250 kDa larger than the value calculated based on the DLS results. However, it is difficult to determine how much of the estimated molecular mass was associated with the detergent micelles around the hydrophobic regions of the protein.

A large set of crystallization experiments was carried out with purified CTB^{AAAA}MPR similarly to that described above for the linker-less variant CTBMPR. Crystals were observed more frequently for CTB^{AAAA}MPR than for CTBMPR, but despite the fact that CTB^{AAAA}MPR appeared to be more stable and more homogeneous than CTBMPR, the crystal quality was still poor. Under most conditions, pseudo-crystals were observed and were similar in shape to the CTB^{GPGP}MPR crystals (Figure 2.9). The crystals shown in Figure 2.9a feature a hexagonal shape when viewed from the 'top', but are completely round when viewed from the side. X-ray diffraction patterns from these crystals show features of a hexagonal powder diffraction pattern, which may indicate that the crystals consist of stacks of two-dimensional crystals which are disordered in the third dimension. However, we noticed that crystal disorder seemed to be correlated with the size of the crystals, with larger crystals displaying more disorder.

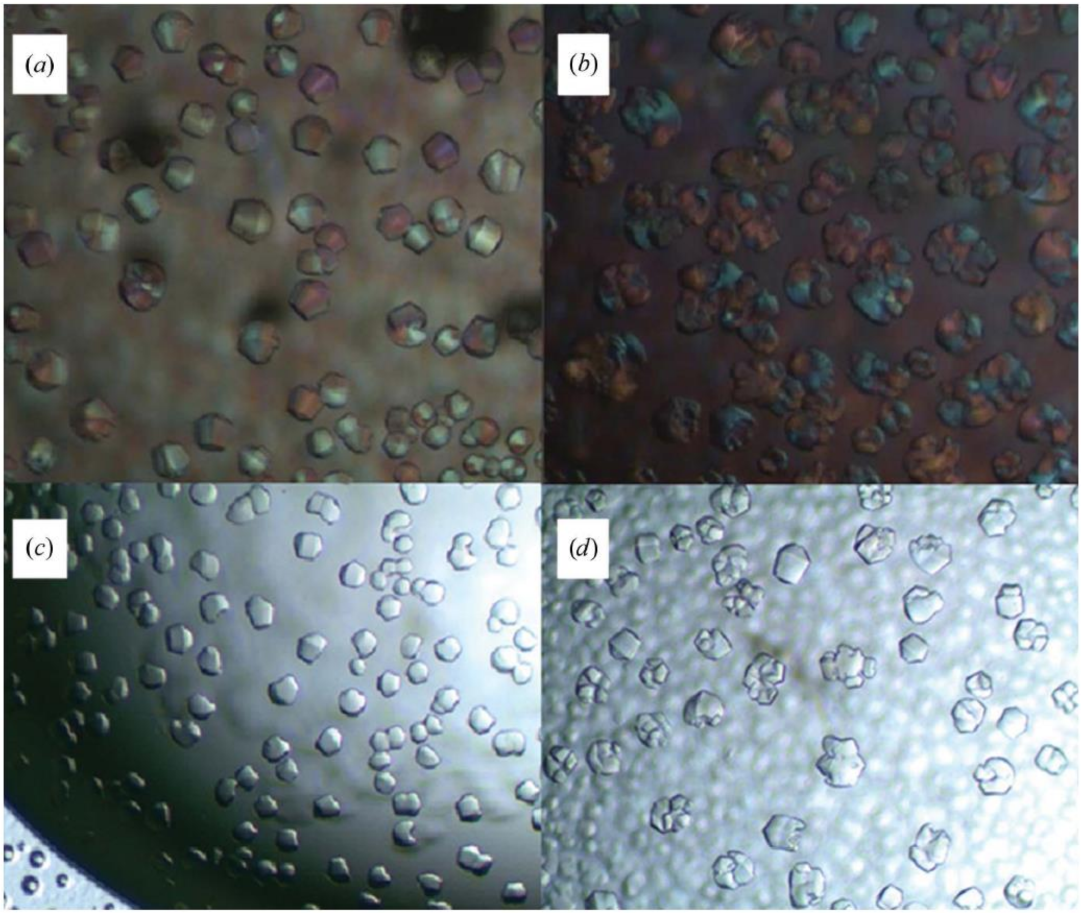


Figure 2.9: **CTB^{AAAA}MPR crystals** form under different conditions of a fine screen. (*a*) 0.2 M ammonium formate, 8% PEG 3350. (b) 0.2 M ammonium formate, 5% PEG 3350. (c) 0.2 M ammonium formate, 12% PEG 3350. (d) 0.1 M ammonium formate, 10% PEG 3350

Taking this into account, crystals were rapidly grown by a fast increase of the supersaturation state using ultrafiltration to concentrate the protein at a constant precipitant concentration (Figure 2.10). Most of the crystals were smaller than the shortest wavelength of visible light; they had the appearance of amorphous precipitates, with very small microcrystals also visible in the sample (Figure 2.10), and this mixture of small (1–2 mm) and very small (<1 mm) crystals will be referred to here as 'nano/microcrystals'. CTB^{AAAA}MPR nano/microcrystals were grown on site at LCLS, characterized by DLS and SONICC and their diffraction quality was tested by the new method of serial femtosecond crystallography (SFX) on the CXI beamline at the LCLS. This beamtime was dedicated to the exploration of the use of SFX for structure elucidation of membrane proteins following the seminal work by (Chapman et al., 2011) and (Boutet et al., 2012). These articles provide detailed description of sample delivery and data collection that will only briefly be recounted here (see the review by(Spence et al., 2012)). Millions of X-ray data diffraction snapshots were collected from a stream of protein nanocrystals or microcrystals in their mother liquor at room temperature as they flow across the beam. Diffraction snapshots of individual crystals of CTB^{AAAA}MPR were collected using X-rays pulses of extremely high intensity (10^9 higher peak brilliance than the brightest third-generation synchrotrons). The 10–50 fs pulses are so brief that the diffraction of each nano/microcrystal is recorded before it is disintegrated. This diffract-before-destroy principle (Barty et al., 2012) overcomes the X-ray damage problem in conventional crystallography and allows data collection from crystals that contain only a few hundred molecules (Chapman et al., 2011). The results from the LCLS beamtime were very promising, as we were able to grow crystals on site and detected the first

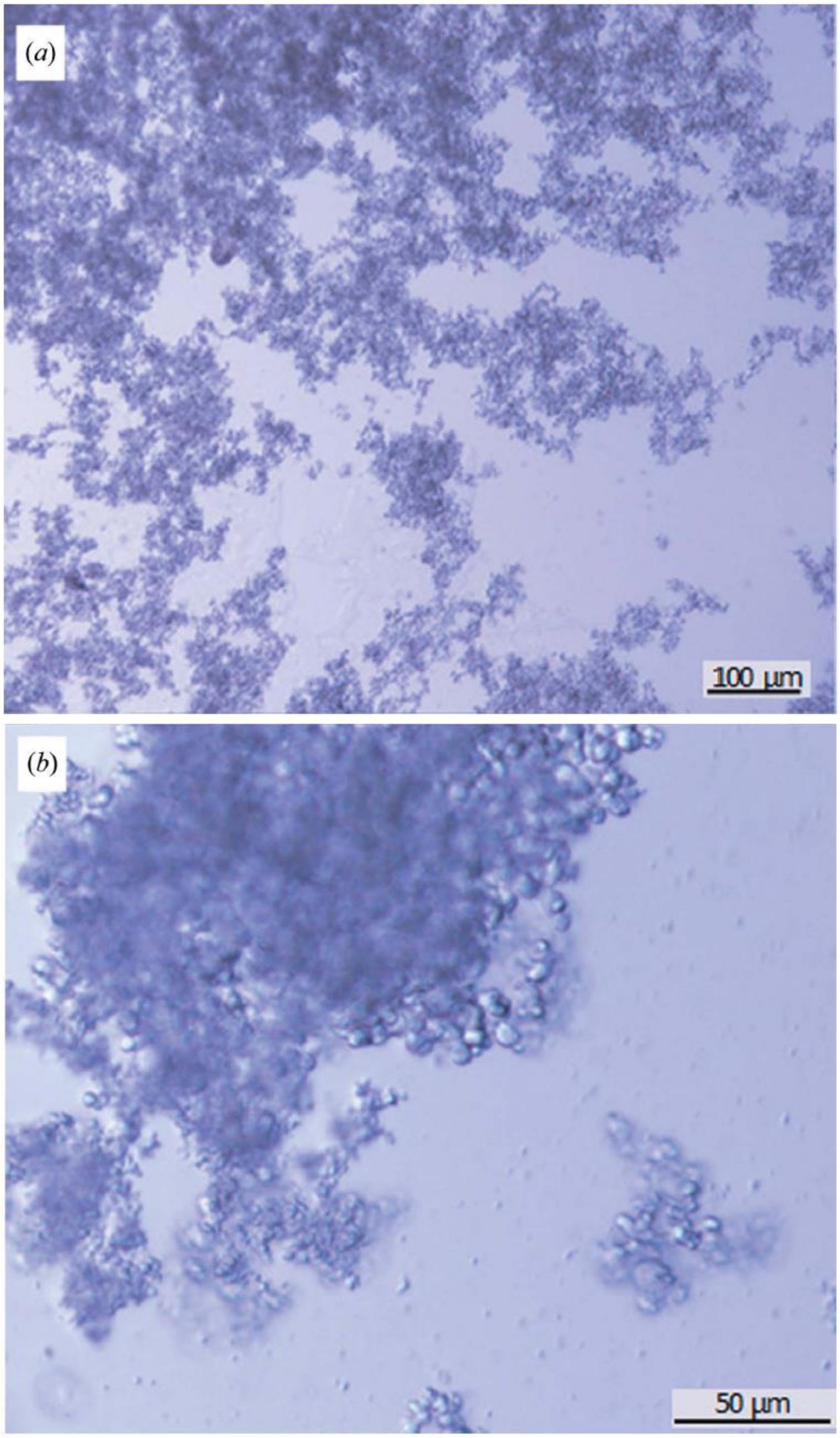
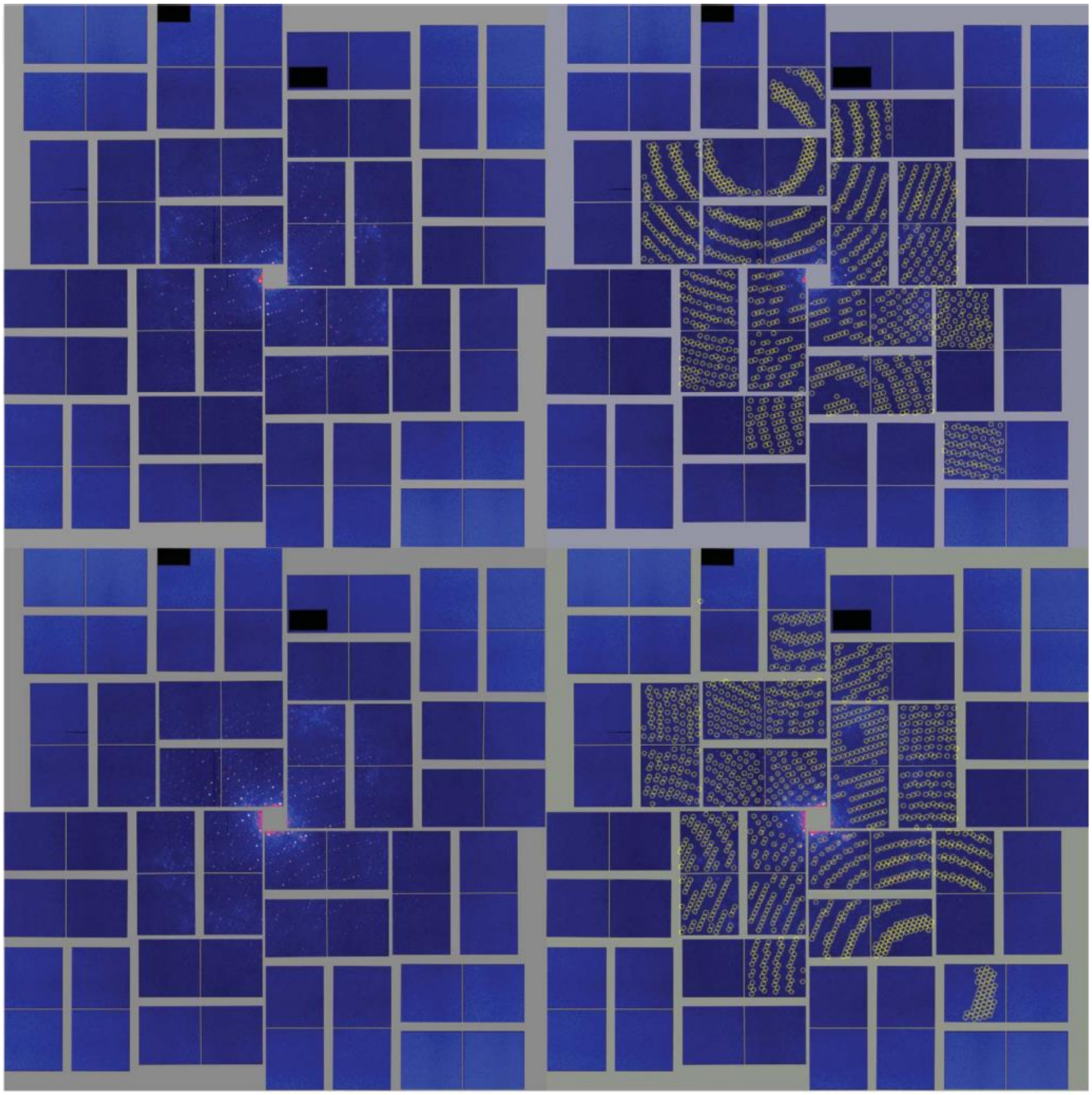


Figure 2.10: **Nano/microcrystals of CTB^{AAAA}MPR** grown in 0.2 M ammonium formate, 30% PEG 3350 before (a) and after (b) filtering through a 20 mm filter. The crystals in (b) are shown at a higher magnification

single-crystal diffraction patterns from CTB^{AAAA}MPR nano/microcrystals. While the larger crystals of CTB^{AAAA}MPR were disordered in the third dimension, the nano/microcrystals are ordered in all three dimensions and show a low degree of disorder. We did not observe any anisotropy of the diffraction patterns even in the third dimension. This is particularly striking since the nano/microcrystals of the protein were grown using the same set of precipitants at initial higher concentration, therefore reaching the supersaturation and nucleation phase much faster than in the vapor-diffusion experiment leading to the larger disordered crystals. A single sort short run of the CTB^{AAAA}MPR nano/microcrystals allowed us to collect 1006 patterns, most of which showed diffraction to 4–6 Å resolution and were successfully indexed (see two typical diffraction patterns and their indexed images in Figure 2.11; Table 2.4). From the indexed patterns, we were able to determine the space group and the unit-cell parameters. The crystals appear to be rhombohedral (consistent with point group R32 with unit-cell parameters $a = b = c = 332$ Å, $\alpha = \beta = \lambda = 60°$). There are only a few published examples of structures with space group R3 and a similar unit-cell parameter to that we observed here for the CTB^{AAAA}MPR fusion protein. Interestingly, the three examples we could find in the PDB happen to be of viral origin. These PDB entries include the structure of *Physalis mottle virus* (PDB entry 1qjz;(Krishna et al., 1999)), with unit-cell parameters $a = b = c = 294$ Å, $\alpha = \beta = \lambda = 59.91°$, and the structures of the *Sesbania mosaic virus* coat protein (PDB entry 1smv; (Bhuvaneshwari et al., 1995)) and its mutant (PDB entry 1x33; (Sangita et al., 2005)), with unit-cell parameters $a = b = c = 291$ Å , $\alpha = \beta = \lambda = 62°$.



(*a*) (*b*)

Figure 2.11: **CTB^{AAAA}MPR FEL diffraction pattern** (*a*) Two CTBAAAAMPR diffraction patterns collected from nano/microcrystals on the CXI beamline at LCLS in February 2012. (b) Indexing of the diffraction patterns in (*a*). The yellow circles indicate the predicted positions of the reflections

Table 2.4: **Crystallographic data for CTBAAAAMPR micro/nanocrystals**

Run time	10 min 40 s
Total No. of raw frames	72767
No. of crystal hits	1006
Hi rate (%)	1.38
No. of indexed patterns	55
Indexing yielding (%)	5.46
Unit cell parameters (Å,°)	a = b = c = 332, $\alpha = \beta = \lambda = 60$
Space group	R32

Since each diffraction pattern is a 'still image' and most reflections are partial, accurate determination of structure requires high redundancy of the data set, i.e. many recordings in the vicinity of each reflection, in order to provide angular integration across the Bragg condition. For example, the first near-atomic resolution structure of a protein to be determined using femtosecond crystallography contained more than 12,000 indexed diffraction patterns (Boutet et al., 2012). While the minimum number of single crystal hits that are required for structure analysis is currently unknown, the thousand reflections that we were able to collect with our very small sample size did not constitute a full native data set that could support structure determination; more data will have to be collected to this end.

It was surprising to see that the nano/microcrystals of CTB^{AAAA}MPR (most of which are <1 mm) are ordered in three dimensions while the larger (100–300 mm) crystals grown with the same set of precipitants are completely disordered in the third dimension. We are currently screening conditions and applying seeding techniques to grow crystals of defined micrometer sizes from the nano/microcrystals for conventional X-ray data collection at synchrotron microfocus beamlines. The plan is to test the diffraction quality of crystals with target sizes ranging from 5 to 100 mm to determine up to which size the crystals are still ordered in three dimensions, with the goal of identifying a 'single-crystal threshold' that may enable data collection at microfocus beamlines. We can then further optimize the crystal quality of the microcrystals by fine screening of the conditions, including the screening of additives.

NMR and crystal structures have been determined of small peptide derivatives of the MPR region that contain binding sites for neutralizing antibodies (Biron et al., 2005, Song et al., 2009, Pejchal et al., 2009) and the consensus is that this peptide can assume an α-helical conformation. Further structural information on the MPR region was obtained by studies involving an in vitro-assembled six-helix bundle consisting of separately produced peptide derivatives of gp41 (Shi et al., 2010) and a chimeric protein consisting of a series of gp41 peptides separated by linkers (Buzon et al., 2010). The conformations observed in these studies are very likely to represent the post-fusion form of MPR. However, the structure shows that the 2F5 binding site is deeply buried inside the three-helix bundle (Shi et al., 2010, Buzon et al., 2010) and therefore these constructs may not induce 2F5-like neutralizing antibodies. During the fusion process, large conformational changes must occur in gp41 that break the interaction between the trimers and expose the 2F5 antibody-binding site, thereby allowing 2F5 to block fusion and transcytosis; thus, a structure of the fusion-active form of MPR is highly desired. Our ultimate goal is to design an optimal CTB-MPR construct that can serve as a vaccine against HIV. Our design of the MPR fusion with CTB is based on the idea of a symmetry mismatch, where the pentameric oligomeric state of CTB hinders the formation of trimers of MPR and thereby stabilizes the MPR region of gp41 in its pre-fusion active form. While we present major strides in this work, further improvement of both the traditional

X-ray crystallography approach (including co-crystallization with neutralizing antibodies) and less-explored innovations such as serial femtosecond crystallography are needed to allow us to meet this goal.

This work presents a proof of principle that three-dimensionally ordered nano/microcrystals can be grown from a protein that had so far resisted growth of any macroscopic crystals that were ordered in three dimensions. Most remarkable is the fact that the SFX diffraction patterns clearly indicate that the nano/macrocrystals were single crystals, while macroscopic crystals grown with the same chemical compounds as precipitants showed the features of two-dimensional crystals stacked nearly randomly in the third dimension. Further enhancement of the quality of the nano/microcrystals by application of improved methods of nanocrystal growth (Kupitz et al., 2014) and the collection of a full data set from these crystals by serial femtosecond nanocrystallography would allow us to determine the structure of CTB^{AAAA}MPR.

2.5 Acknowledgements

We would like to thank Eric Davies (Arizona State University) for technical assistance with the cloning. This work was supported in part by the National Institutes of Health (awards U19 AI062150 to TM; U54 GM094599 to PF, BH and TM; 1R01GM095583 to PF; and R03 AI073157 to NM). We also gratefully acknowledge support by the US National Science Foundation (NSF; award MCB 1120997). Conventional and serial femtosecond crystallography experiments were carried out at the Advanced Light Source (ALS) and the Linac Coherent Light Source (LCLS), respectively, which are national user facilities operated by the University of California and Stanford University, respectively, on behalf of the US Department of Energy (DOE), Office of Basic Energy Sciences. We are grateful for an NSF award (MCB-1021557) to JCHS. We acknowledge support from DOE through the PULSE Institute at the SLAC National Accelerator Laboratory and by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. HL's work was partially supported by Human Frontiers Science Project award No. 024940. We also acknowledge the support of the Helmholtz Association and the Max Planck Society.

CHAPTER 3

SELF-CLEAVGE OF CTB^{AAAA}MPR DURING CRYSTALLIZATION

Ho-Hsien Lee,[a] Raimund Fromme,[a] Irene Cherni,[b,c] Arpan Deb,[b,c] Nobuyuki Matoba,[b,c] Tsafrir S. Mor,[b,c] and Petra Fromme[a]

[a]Department of Chemistry and Biochemistry, and Center for Applied Structural Discovery at the Biodesign Institute, Arizona State University, PO Box 871604, Tempe, AZ 85287-1604, USA, [b]School of Life Sciences, Arizona State University, PO Box 874501, Tempe, AZ 85287-4501, USA, [c]Center for Infectious Diseases and Vaccinology, Biodesign Institute, Arizona State University, PO Box 874501, Tempe, AZ 85287-5401, USA

(This chapter is an adapted version of a manuscript currently being prepared for submission.)

3.1 Abstract

CTB^{AAAA}MPR is a fusion protein consisting of the cholera toxin B subunit (CTB) and the membrane proximal region (MPR) of gp41, which is the transmembrane envelope protein of human immunodeficiency virus 1 (HIV-1). The two proteins are fused together with a tetra-alanine linker. It is a candidate vaccine component aimed at inducing the production

of antibodies targeting the early steps of HIV-1 mucosal transmission. Here we report on crystallization experiments and preliminary X-ray crystallographic analysis of crystals formed from a solution of the fusion protein. The final goal of the study is to investigate the structure of CTB^{AAAA}MPR with MPR in its fusion-active form to guide improvements of the immunogenicity of the fusion protein design. A 1.9 Å data set was obtained from needle-like crystals grown from the monodisperse solution of CTB^{AAAA}MPR. However, X-ray structure analysis revealed that the crystals contained CTB but no electron density was observed in the structure beyond the final residue of CTB suggesting that the MPR domain and alanine linker was cleaved. CTB was found to be arranged in a unique crystal packing and unit cell that has not been observed before for CTB nor any of its fusion proteins. The structure contains two unaligned CTB pentamers in the asymmetrical unit, which suggests that MPR was present during crystallization and influenced the crystal packing. Further investigation by monitoring crystallization on a day by day basis showed that the fusion protein is very stable under conditions that do not lead to formation of crystals, but is cleaved as soon as large well-ordered crystals are formed.

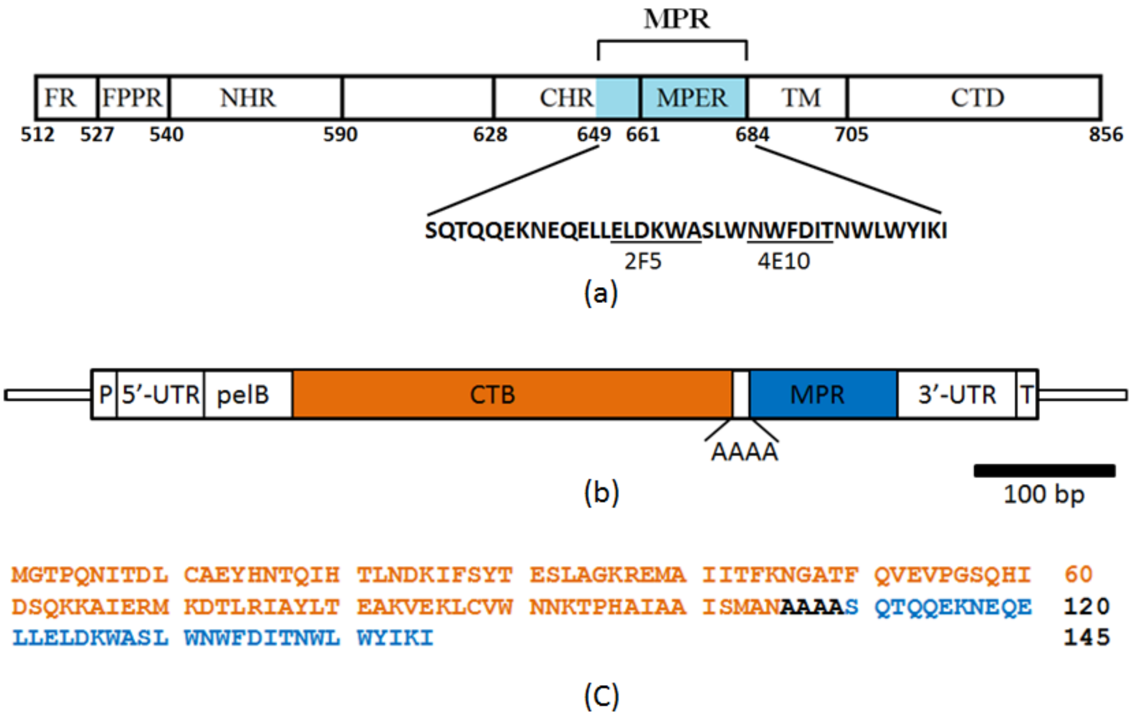
3.2 Introduction

Efforts to develop drugs and vaccines against acquired immune deficiency syndrome (AIDS) have prompted significant research efforts on the envelope glycoproteins of HIV-1. The viral envelope consists of two glycoproteins: 1) gp41 is the transmembrane glycoprotein that contains an ectodomain largely responsible for the trimizeration of the complex and 2) gp120 is an exterior glycoprotein which covers part of gp41. Both glycoproteins play an important role in viral transmission and infection (Wyatt & Sodroski, 1998). Although high mutation rates in HIV-1 lead to high antigenic variability, researchers have identified several well conserved regions of the envelope protein that are crucial to viral functions, which is reviewed in (Zolla-Pazner, 2004). One such region is the transmembrane-proximal region of the gp41. Although it is poorly immunogenic, many of the antibodies that bind to this region, such as 2F5, 4E10, and Z13, are broadly neutralizing against HIV-1 (Zwick et al., 2001, Zolla-Pazner, 2004). Residues 649-684 of gp41, which we termed MPR, play a crucial role in viral epithelial transcytosis by binding to glycosphingolipid galatosylcermide (GalCer) on the surface of epithelial cells and is transcytosed using a non-fusogenic mechanism (Bomsel, 1997, Alfsen et al., 2001, Alfsen & Bomsel, 2002). Antibodies (Abs) that target gp41, such as 2F5 or secretory IgA, are present in mucosal secretions of people that have been highly exposed to HIV, but persistently test negative for HIV in their blood sera; these antibodies were shown to have high efficiency in blocking viral transcytosis (Devito et al., 2000a, Devito et al., 2002, Miyazawa et al., 2009, Tudor et al., 2009). These studies support the hypothesis that inducing Abs against MPR in the mucosal immune system may be an effective method to prevent viral entry into the body, as over 90% of global HIV transmissions occur across the mucosal surface (Kresina & Mathieson, 1999, Overbaugh et al., 1999).

To ensure efficiency, the ideal vaccine should contain components that target various steps of the viral transmission and infection process, especially in the early stages of the viral cycle to take advantage of viral vulnerabilities of the low founder viral population and to minimize the establishment of systematic infection (Haase, 2010). The crucial involvement of MPR in these viral functions makes this domain a prime candidate for vaccine development. These findings led us to construct a fusion protein, consisting of the mucosal adjuvant CTB and MPR as a vaccine component aimed at inducing Abs that would target the step of viral transcytosis in the mucosal epithelial cells. The fusion protein has been shown to induce transcytosis by blocking anti-MPR antibodies in mice and rabbits (Matoba et al., 2004, Matoba et al., 2006, Matoba et al., 2008). However, the immune response of the MPR is overshadowed by the immuno-dominant CTB.

We began structural investigation of CTB MPR with the goal to understand the function of MPR and the viral infection processes it is involved in, and its interaction with the immune system as a fusion protein. Structural information could help us elucidate the immunogenicity of the fusion protein and enable designing even more immunogenic fusion proteins. We have recently published work on structural investigation of various CTB-MPR constructs with different linkers between the fusion partners (Lee et al., 2014). A flexible glycine-proline-glycine-proline linker proved to be too flexible for structure determination using X-ray crystallography as the electron density of the MPR region can be seen but is not resolved well enough for model fitting. Removing of the linker completely, so as to directly attach MPR to CTB, proved difficult for crystallization due to the instability of the pentameric oligomeric state. A third construct was designed using a more rigid tetra-alanine linker, which initially proved difficult for macroscopic crystal formation, however some promising data was collected on micro/nano-crystals using the X-ray free electron laser (XFEL); however the data set was only partial due to sample and XFEL beamtime limitations (Lee et al., 2014).

In this report, we further investigated crystallization conditions that initially produced the micro/nano-crystals and optimized it for obtaining macroscopic crystals for structure analysis at Synchrotron X-ray sources.



```
MGTPQNITDL CAEYHNTQIH TLNDKIFSYT ESLAGKREMA IITFKNGATF QVEVPGSQHI  60
DSQKKAIERM KDTLRIAYLT EAKVEKLCVW NNKTPHAIAA ISMANAAAAS QTQQEKNEQE 120
LLELDKWASL WNWFDITNWL WYIKI                                      145
```

Figure 3.1: **gp41 and CTB^{AAAA}MPR fusion protein** (a) A schematic diagram of HIV-1 gp41. Fusion peptide (FR. Residues 512-539); fusion peptide proximal region (FPPR residues 528–539; N-terminal heptad-repeat region (NHR, residues 540–590); C-terminal heptad-repeat region (CHR, residues 628–661); membrane proximal external region (MPER, residues 662–684); membrane proximal region (MPR, residues 647–684, blue), transmembrane domain (TM, residues 685–705); cytoplasmic C-terminal domain (CTD, residues 706–856). The epitopes of 2 broadly neutralizing antibodies 2F5 and 4E10 are marked. (b) DNA construct for the expression in *E.coli* of the CTB^{AAAA}MPR fusion protein based on elements of the pET-26b(+) expression vector. P, T7 bacteriophage promoter, 5'UTR, upstream untranslated region; pelB, the periplasmic targeting sequence of pectate lyase B of *Erwinia carotovora*; CTB, cholera toxin B subunit; AAAA, tetra-alanine linker; MPR, the membrane-proximal region of gp41 protein of HIV-1; 3'UTR, downstream untranslated region; T, T7 terminator. (c) Amino acid sequence of fusion protein CTB^{AAAA}MPR with expected molecular mass of 16.7kDa. The CTB domains are labeled orange followed by the tetra-alanine linker and the MPR regions labeled in blue

3.3 Materials and Methods

Protein preparation

The cloning of the gene encoding the CTB and the $\mathrm{MPR}_{649\text{-}684}$ with tetra-alanine linker (AAAA) between the two proteins (hereafter dubbed $\mathrm{CTB}^{\mathrm{AAAA}}\mathrm{MPR}$) shown in Figure 3.1 has been described in previously published work (Lee et al., 2014). The gene was inserted into an expression vector derived from pET26B(+) (Novagen), which contains an N-terminal pelB signal sequence for periplasmic localization, to obtain the plasmid pTM646. The construct was verified by DNA sequencing.

The $\mathrm{CTB}^{\mathrm{AAAA}}\mathrm{MPR}$ protein was overexpressed in *Escherichia coli* BL21 (DE3) cells following a protocol derived from previously published work for another variant of the fusion protein (Matoba et al., 2008, Lee et al., 2014). Modifications were made to optimize the protocol to improve the expression as well as reduce the viscosity of the detergent extract for purification purposes.

An overnight starter culture was grown in 50 ml lysogenic broth (Luria-Bertani medium, LB) at 37°C on a shaker set at 200 rpm. A 1:200 dilution of the starter culture was used to inoculate 4 l expression cultures and grown to an $\mathrm{OD}_{600\mathrm{nm}}$ of 0.8 at 37°C. The expression of recombinant protein was then induced with isopropyl β–D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.3 mM at 37°C for 3h. The cells were harvested by centrifugation at 5000 g yielding approximately 2.5 g to 3.0 g of wet cell pellet per liter of culture and frozen at -80°C.

The cell pellets were then thawed on ice before being combined by resuspended in 80 ml of ice-cold phosphate buffer saline (PBS; 137mM NaCl, 2.7 mM KCl, 10mM

Na_2HPO_4, and 1.8 mM KH_2PO_4) supplemented with 1mM phenylmethylsulfonylfluoride (PMSF). Cells were lysed by passing three times through a microfluidizer (Microfluidics Microfluidizer) with the re-supplement of PMSF between the passes. The lysate was collected by centrifugation at 36,000 g and the supernatant was discarded. The insoluble fraction was washed once by resuspension in ice-cold PBS and centrifugation then frozen at -80°C.

The cell pellet, containing the membrane fraction was thawed on ice before resuspending in 60 ml of ice-cold buffer containing 500 mM NaCl, 20 mM bicine pH 8.0. A sonicator (Model 300 V/T Ultrasonic Homogenizer, Biologics, Inc.) was used to fully homogenize the solution by running 30 s on/off cycles at 20% amplitude until a homogenous suspension was obtained. 8 ml of a stock solution containing 10% (w/v) n-dodecyl-β-D-maltoside (βDDM) was added to the homogenized solution and brought to a final volume of 80 ml using more ice-cold solubilization buffer for overnight solubilization at 4°C with agitation.

The detergent extraction solution was centrifuged at 36,000 g and the pellet was discarded. A gravity-driven column (Bio-Rad Econo-column) containing 40 ml of cobalt affinity resin (TALON, Clontech) was equilibrated with binding buffer (solubilization buffer supplemented with 0.05% βDDM). The detergent extraction supernatant was then loaded onto the column and washed with six bed volumes of binding buffer followed by ten bed volumes of wash buffer (50mM NaCl, 20mM bicine pH 8.0, 5 mM imidazole, 0.05% βDDM) removing weakly bound proteins. Three bed volumes of elution buffer containing 50mM NaCl, 20mM bicine pH 8.0, 150 mM imidazole, 0.05% βDDM was used to elute the tightly bound proteins. The eluted fractions were pooled and then concentrated to approximately 2mg ml^{-1} using 50kDa molecular-weight cutoff (MWCO) concentrators (Vivaspin 20, Sartorius Stedim Biotech). The concentrated solution was then subjected to size-exclusion chromatography (Superdex 200, GE Healthcare) in buffer consisting of 10 mM calcium chloride, 20mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) pH 7.5, 0.02% βDDM running at 0.5 ml min^{-1} shown in Figure 3.2. The protein elution was detected by absorption at 280 nm and manually collected to separate shoulders from main peaks. Only the main pentamer peak that eluted at minute 28 was used for crystallization experiments. The concentration of CTB^{AAAA}MPR preparations were determined spectrophotometrically using ε_{280} = 39380 M^{-1}cm^{-1}.

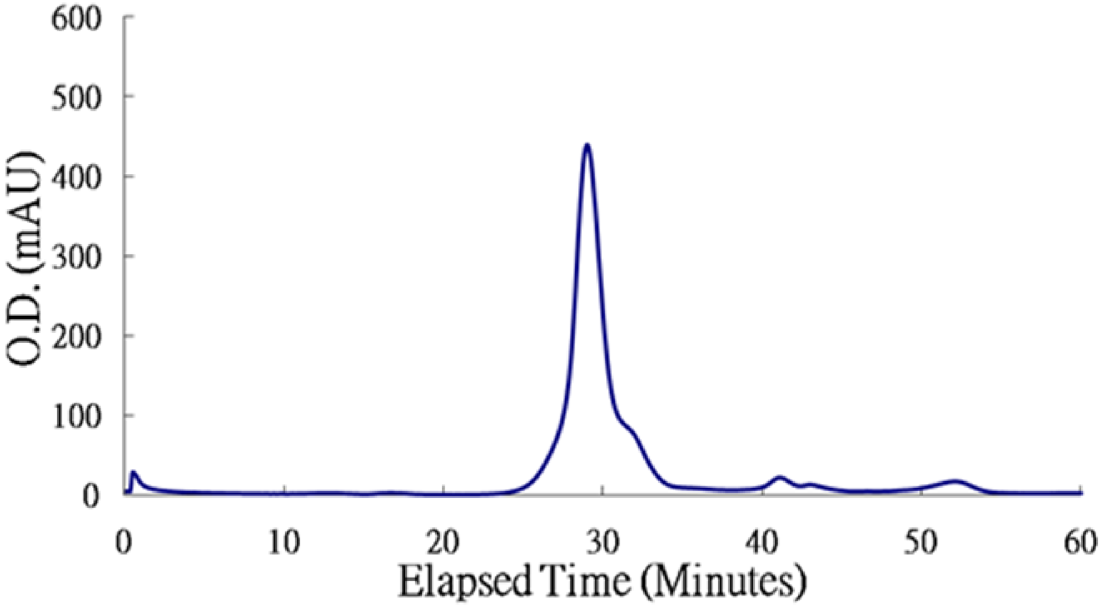


Figure 3.2: **Size exclusion chromatograph of CTB^{AAAA}MPR** shows a dominate pentameric peak at minute 28 with a trailing shoulder that is discarded. The small peak at 42 minute corresponds to the monomeric fusion protein and imidazole from the metal affinity column elution buffer at minute 52.

SDS-PAGE and immuno-blot-analysis

SDS-PAGE gels for protein samples were done using tricine-based buffers and Bio-Rad Mini-PROTEAN Tetra Cell following procedures based on the method developed by (Schagger, 2006) and adapted by Lawrence et al. (2011). SDS-PAGE for dissolved crystals was performed using the PhastSystem-High Speed Electrophoresis System (Pharmacia) using High Density PhastGels (Pharmacia) and PhastGel buffer strips (Pharmacia). After electrophoresis, the gels were subjected to silver staining (Lawrence et al., 2011) or processed for immunoblotting with monoclonal antibody 2F5 or 4E10 (AIDS Reagent Program) or polyclonal anti-CTB antibodies (List Biological Laboratories, Inc.) following the protocol described in (Lee et al., 2014). The assembly of pentamers of the CTB^{AAAA}MPR was monitored by enzyme-linked immunosorbent assay (ELIZA) using G_{M1} gangliosides to capture CTB pentamers and detected with monoclonal antibody 2F5 as previously described in (Matoba et al., 2008).

Crystallization experiments

CTB^{AAAA}MPR protein in buffer from the size-exclusion chromatography was concentrated to 10 mg ml^{-1} for crystallization experiments using 100kDa MWCO concentrators (Millipore, Centricon).

Broad matrix screening was done using the Phoenix crystallization robot (Rigaku) with commercial screening kits from Qiagen and Hampton Research. The experiments were set up by sitting-drop vapor-diffusion method using MRC Crystallization Plates

(Hampton Research) that contained 96 wells and 2 drop chamber per well. The reservoirs were filled with 65 µl of precipitate solutions from the screening kits. In the drop well, the protein was mixed with the reservoir solution in 1:1 and 2:1 ratios, 100 nl protein plus 100 nl precipitant solution and 200 nl protein plus 100 nl precipitant solution respectively.

Additional crystallization screening and optimization experiments were performed based on the condition that had previously produced micro/nanocrystals (Lee et al., 2014), which contained 200 mM ammonium formate, 30% polyethylene glycol (PEG) 3350, 10 mM calcium chloride, 20 mM HEPES pH 7.5 with the batch method described in (Lee et al., 2014).

The goal was to grow larger crystals suitable for conventional X-ray structure analysis using Synchrotron Radiation. These fine screening and optimization experiments were performed by the hanging-drop vapor-diffusion method using 24-well plates (Hampton Research VDX greased plates). The reservoir well contained 900 µl of precipitate solution and 3 µl of the protein sample was mixed with an equal volume of reservoir solution on siliconized glass covers slides before the wells were sealed. The plates were incubated in a 20°C incubator. The concentration of ammonium formate tested ranged from 50 mM to 500 mM, and the PEG 3350 concentration ranged from 5% to 30%. Calcium chloride was present to maintain the oligomeric stability of the $\text{CTB}^{\text{AAAA}}\text{MPR}$ pentameric fusion protein and HEPES for maintaining the pH, the concentration of both chemicals were used at low concentrations of 5-15 mM calcium chloride and 10-30 mM HEPES pH 7.5.

After an incubation time of seven days at 20°C, thin needles were observed in several crystallization drops at PEG concentrations between 8-15% PEG 3350. After further optimization, large needle shaped crystals were grown. The condition that produced the best diffracting crystals contained 200 mM ammonium formate, 10 mM calcium chloride, 20mM HEPES pH 7.5, and 8% PEG. Please note that all the various concentration of PEG 3350 that have been used for the vapor diffusion experiments were with 10-25% lower than the initial conditions of 30% PEG established for the growth of the micro/nano-crystals for the serial femtosecond crystallography experiments. Figure 3.3 shows results from a set of crystallization experiments where the PEG concentration was systematically varied. The results indicated that the concentration of PEG 3350 has a direct effect on the size of the needle-like crystals, where the increase in PEG concentration results in the decrease in crystal size.

The crystals were cryo-protected in a solution containing 0.3M ammonium formate, 15 mM calcium chloride, 30 mM HEPES pH 7.5, 50% PEG3350 and flash frozen in liquid nitrogen prior to X-ray data collection at the Advanced Light Source (ALS).

CTB^{AAAA}MPR crystal seeding experiments were conducted using micro/nano-crystals grown with batch method described in (Lee et al., 2014) using the microseeding method base on (Bergfors, 2003). The experiments were set up on a 24-well plate that contained 900 µl reservoir solutions and the protein solution was mixed in a 1:1 ratio (3 µl to 3 µl) on the siliconized glass cover. A human hair washed with ethanol and rinsed repeatedly in deionized water was streaked through the micro/nano-crystal solution and

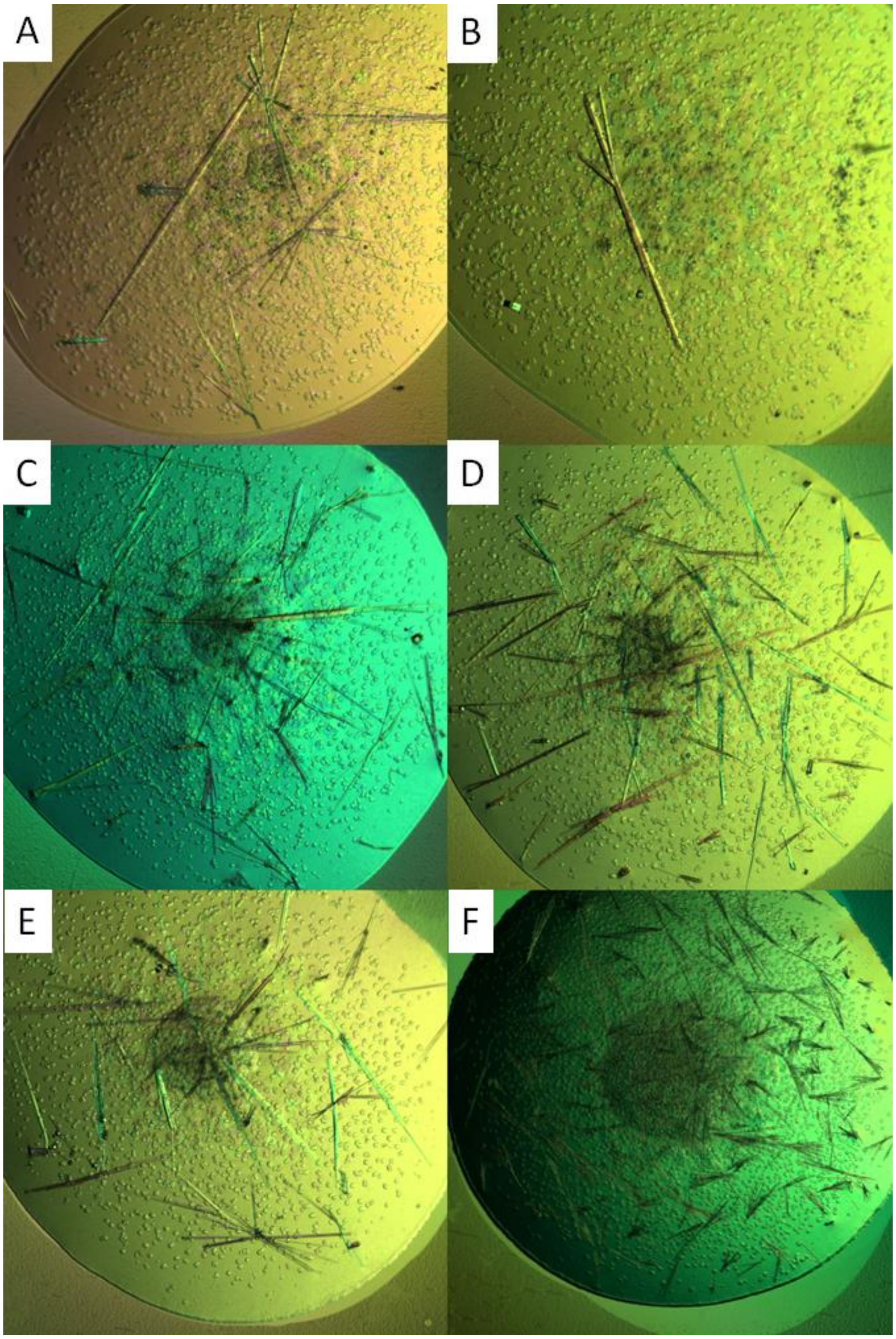


Figure 3.3: **CTB^{AAAA}MPR crystals** of various sizes formed under different PEG3350 concentrations. All conditions contained 200 mM ammonium formate, 10 mM calcium chloride, 0.02M HEPES pH 7.5, (A) 8% PEG3350; (B) 9% PEG3350; (C) 10% PEG3350; (D) 11% PEG3350; (E) 12% PEG3350; (F) 13% PEG3350

placed in hanging drops before the wells were sealed. The condition that produced the crystal clusters contained 200 mM ammonium formate, 10 mM calcium chloride, 20mM HEPES pH 7.5, and 15% PEG3350.

X-ray crystallography

X-ray data on the needle-like crystals grown from the solution of CTB^{AAAA}MPR was performed using synchrotron X-ray radiation on beamline 8.2.1 at the Advanced Light Source (ALS) in Berkeley at a wavelength of 1 Å. The best data set consisted of 390 frames collected using 0.33° oscillation with an ADSC 315 detector (see table 3.1 for the data statistics).

Data integrating and merging was done using **XDS** (Kabsch, 2010). Scaling was performed using the CCP4 program suite **AIMLESS** (Evans & Murshudov, 2013). The space group was determined to be $P2_12_12_1$ with unit cell dimensions of a=43.35Å, b=114.63Å, c=213.20Å and α=β=γ=90°. The structure was solved by molecular replacement using **Phaser** (McCoy et al., 2007) with a model derived by molecular replacement from the CTB structure in the PDB entry 1JR0 as the search model. The best solution showed that there are two CTB pentamers in the asymmetric unit (TFZ = 22.9 and LLG = 20201). Primary refinement was done using CCP4 program suite **REFMAC5** (Murshudov et al., 2011) and further refinement was performed with **PHENIX** (Adams et al., 2010). All residues were in the in the allowed regions of the Ramachandran plot with 98% in favored regions. Results of the data-processing and refinement statistics are shown in Table 3.1. Structural representations were prepared with **PyMOL** (Delano, 2002) and views from various angles are shown in Figure3.4.

3.4 Results and discussion

Cleavage of CTB^{AAAA}MPR

The X-ray structure analysis and data statistics of the needle-shaped crystals grown from the solution containing CTB^{AAAA}MPR was very encouraging. The crystal featured the space group $P2_12_12_1$ with unit cell constants of a = 43.35 Å, b = 114.63 Å, c= 213.20 Å, $\alpha = \beta = \lambda = 90°$. The data sets could be evaluated to 1.9 Å resolution, with high completeness even in the highest resolution shell. The R-values were low with R-work of 0.183 (0.252) and an R-free of 0.219 (0.308) (see Table 3.1 for full data statistics). The structure revealed the typical ring of wild-type CTB pentamer. The crystals contain an unusual packing of CTB not observed in any CTB crystals so far with two pentamers molecules in the asymmetric unit. However, no electron density was observed beyond the CTB domain. The structure is shown in Figure 3.5 (A-C). The packing of CTB in the crystals, shown in Figure 3.5 (D), visualize the extremely tight packing of the CTB molecules leaving very little room for MPR. To verify whether cleavage of the fusion protein had occurred, the crystals were retrieved from the cryo-loops after returning from ALS and analyzed with SDS-PAGE using the PhastSystem (Pharmacia). The crystals were dissolved 5 µl SDS sample buffer and compared with pre-crystallized sample from the same sample preparation batch. The expected size of CTB^{AAAA}MPR is 16.7kDa and this is the molecular weight identified for the fusion protein in the pre-crystallization

Table 3.1: **Crystallographic data for data collected from crystals grown from a solution containing for CTB^{AAAA}MPR at ALS 8.2.1(Berkeley)**

Values in parentheses are for the highest resolution bin.

Wavelength (Å)	1
Resolution range (Å)	48.33-1.9 (1.97-1.9)
Space group	P 2_1 2_1 2_1
Unit cell parameters (Å / °)	a = 43.35, b = 114.63, c = 213.20, $\alpha = \beta = \gamma = 90$
Total reflections	164002 (16588)
Unique reflections	84204 (8404)
Multiplicity	1.9 (2.0)
Completeness	0.99 (1.00)
Mean I/sigma(I)	6.42 (1.90)
Wilson B-factor	24.42
R-merge	0.058 (0.408)
R-meas	0.082 (0.577)
CC1/2	0.992 (0.687)
CC*	0.998 (0.902)
Reflections used for R-free	4116 (425)
R-work	0.183 (0.252)
R-free	0.219 (0.308)
CC (work)	0.937 (0.841)
CC (free)	0.927 (0.811)
Number of atoms	8968
Number of Macromolecules	8140
Number of water	828
Protein residues	1030
RMS (bonds - Å)	0.008
RMS (angles - degree)	1.09
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.6
Ramachandran outliers (%)	0
Rotamer outliers (%)	3.3
Clashscore	4.17
Average B-factor	28.2
Macromolecules B-factor	27.4
Solvent B-factor	36.0

sample in Figure 3.4. However, the protein from the dissolved crystals shows a lower molecular weight band around 11.5kDa, which is the expected molecular weight of CTB without the MPR domain.

To confirm the result, we performed an immunoblotting experiment with 4E10 antibodies (anti-MPR) which confirmed that the dissolved crystal sample does not contain MPR, while the polyclonal Abs against CTB showed activity with both the crystal and the pre-crystallization sample. We concluded that the fusion protein was cleaved between CTB and the alanine linker during the crystallization process so that the crystals do not contain the MPR domain and the alanine-linker.

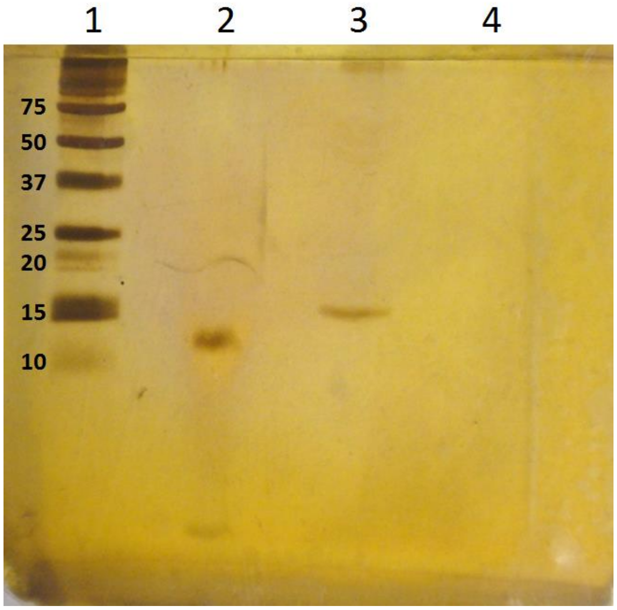


Figure 3.4: **SDS-PAGE of dissolved CTB^{AAAA}MPR crystal** using the Phastsystem and a high density PhastGel. A molecular weight marker (Bio-Rad, Precision Plus Protein^{TM} Dual Color Standards) is shown in lane 1, followed by the dissolved CTB^{AAAA}MPR crystals that were used for data collection at ALS in lane 2. Lane 3 contained the positive control of the sample before crystallization and lane 4 is the negative control containing only the crystallization buffer.

Even though the MPR has been cleaved and is not present in the crystals, we hypothesize that it contributed to the formation of the new crystal form of CTB. The majority of the CTB structures in the PDB are from crystals that feature the monoclinic C2 space group (Figure 3.6A) while the CTB structure in our crystal grown from a solution of CTB^{AAAA}MPR showed an orthorhombic $P2_12_12_1$ space group (Figure 3.6B). Cholera toxin has shown to form crystals in orthorhombic space groups, but these cases involved either mutations, crystallization with the A-subunit as AB_5 complex (van den Akker et al., 1997) or feature CTB with ligands or inhibitors bound to the receptor sites of CTB shown in Figure 3.6C and 3.6D (Aman et al., 2001, Fan et al., 2001). Furthermore, the packing of CTB in our crystals is unique. Nearly all the CTB structures in the PDB contain only one CTB pentamer per asymmetric unit, while there are two CTB pentamer molecules per asymmetric unit side by side present in our crystals. The two pentamers are inverted and unaligned facing different planes as shown in Figure 3.6B.

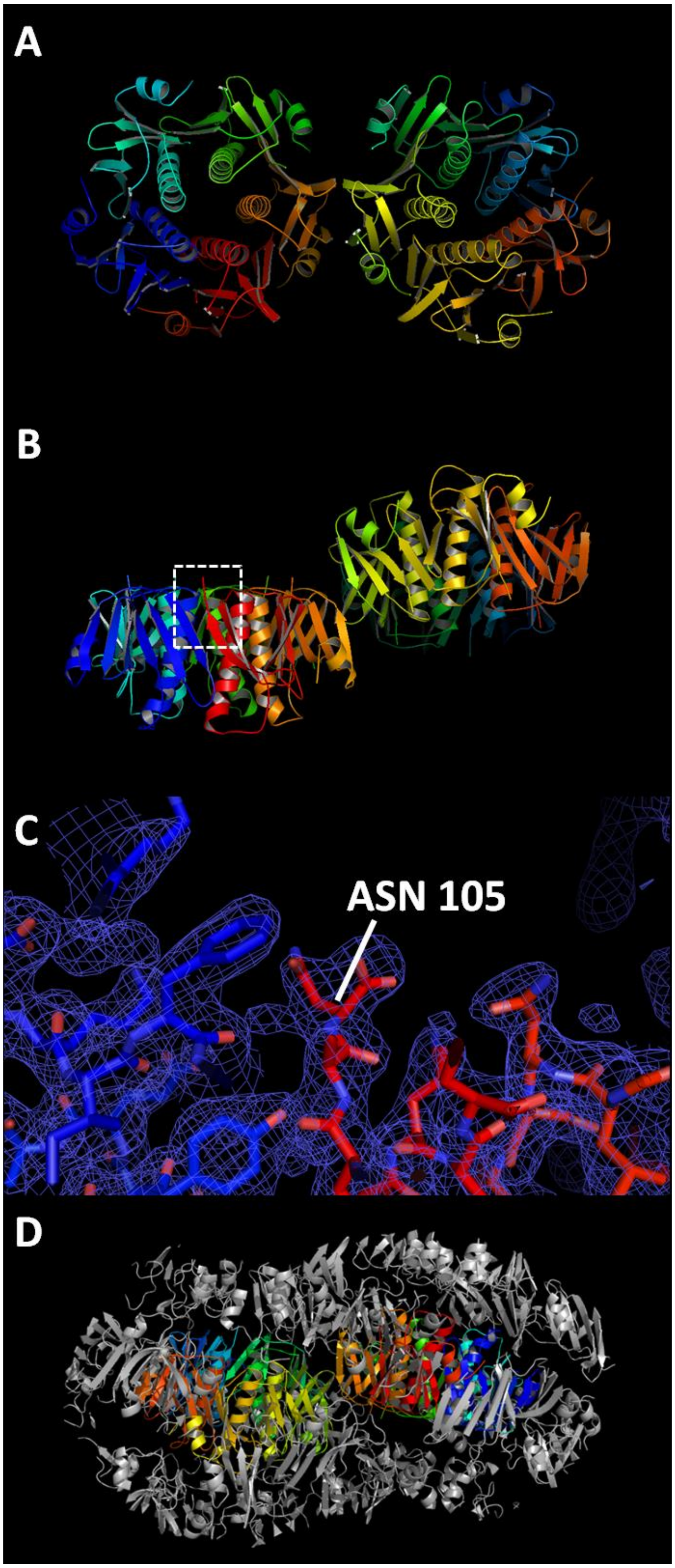
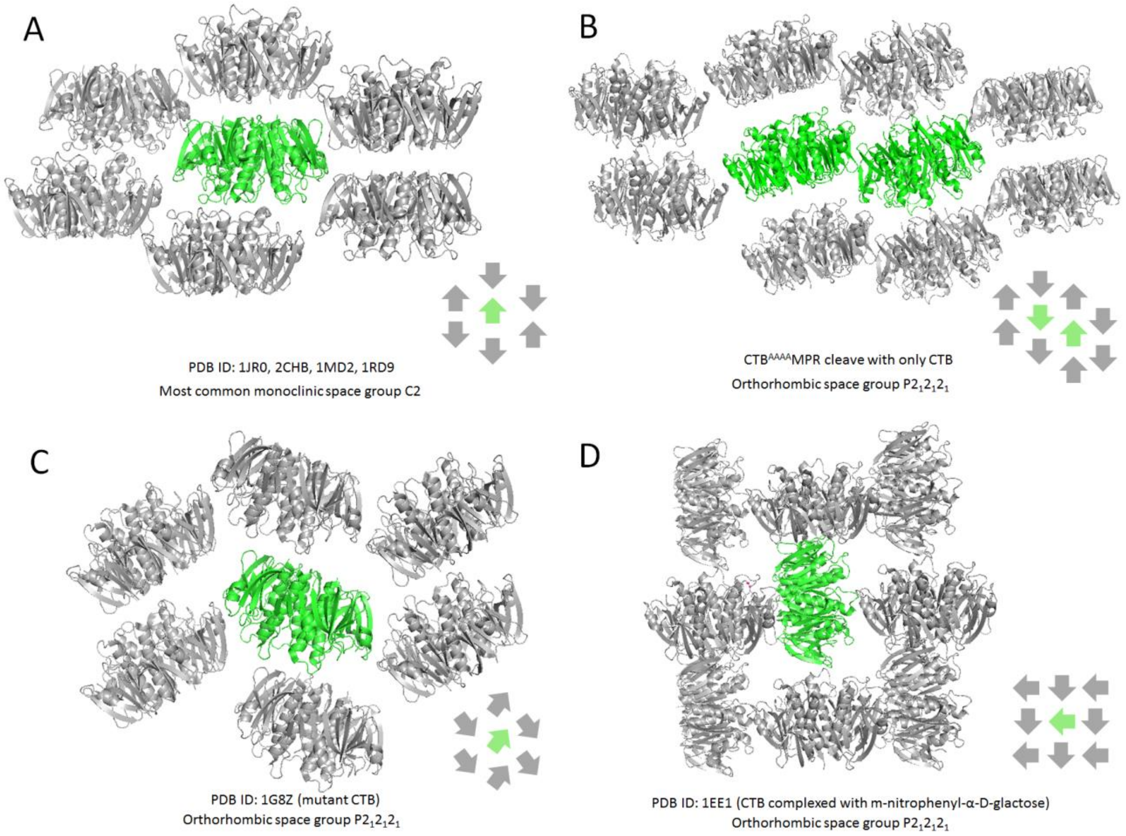


Figure 3.5: **Structural representation of the protein-structure in the crystals grown from the solution containing CTB^{AAAA}MPR.** The structure was phased with the pentameric CTB model (derived from PDB entry 1JR0) using molecular replacement. The structure shows two CTB pentamer molecules per asymmetrical unit. (A) Top view. (B) Side view with the C-terminus of one of the subunit marked, which is shown in close-up view in (C). (C) Electron density map at a contour level of 1δ of the C-terminus of CTB. No electron density can be seen beyond the terminal asparagine residue of CTB, which indicates that the crystal contain only the fusion partner CTB. (the figure was produced using the program PyMOL (Delano, 2002), (D) CTB asymmetric unit shown with 20 Å surrounding in the crystal packing. Adjacent CTB molecules are shown in gray.



PDB ID: 1JR0, 2CHB, 1MD2, 1RD9
Most common monoclinic space group C2

CTB^{AAAA}MPR cleave with only CTB
Orthorhombic space group $P2_12_12_1$

PDB ID: 1G8Z (mutant CTB)
Orthorhombic space group $P2_12_12_1$

PDB ID: 1EE1 (CTB complexed with m-nitrophenyl-α-D-glactose)
Orthorhombic space group $P2_12_12_1$

Figure 3.6: **CTB crystal packing.** Structures obtained from PDB and generated using PyMol (Delano, 2002). Each asymmetric unit is shown in green and symmetric mates in gray. (A) The most common CTB space group is C2. CTB molecules in the same column alternate facing opposite directions. The structure is generated using 1JR0 (Merritt et al., 1994) All the CTB molecules crystallized with C2 space group show extremely similar packing arrangements. (B) CTB packing from CTB^{AAAA}MPR crystals featuring the $P2_12_12_1$ space group. There are two CTB molecules per asymmetric unit and they are facing opposite directions and are not aligned. (C) A CTB mutant with $P2_12_12_1$ space group PDB ID: 1G8Z (Aman et al., 2001). CTB molecules in the same column faces the same direction however the molecule is turned nearly 90 degrees between the columns. (D) CTB complexed with m-nitrophenyl-α-D-galactose shows the $P2_12_12_1$ space group PDB ID: 1EE1 (Fan et al., 2001) but with a different packing than our crystals grown from the protein CTB^{AAAA}MPR. The CTB molecules next to each other are perpendicular to each other.

Daily observation of crystallization and crystal seeding

In order to determine if the crystallization process induces the cleaving of the fusion protein, experiments were set up to observe the crystallization process day by day. With these experiments we tried to determine whether freshly grown crystals still retained the MPR domain and if the cleavage of the fusion protein was just an artifact from crystal age and freezing. Crystallization trials were set up using 24-well plates with all wells containing the same condition that produced the best diffracting needle-like crystals and the experiment was performed at two different temperatures, 20°C and 10°C. Whole crystallization droplets on the siliconized glass slide from two of the wells were harvested and frozen at -80°C each day.

For the experiment performed at 20°C, small round crystals were observed after three days and needle-like crystals appeared after the fourth day. A photo of the well on day 6 (last day) is shown in Figure 3.8. The analysis of the whole crystal drop on SDS-PAGE showed cleavage of the fusion protein after three days coinciding with the formation of the first crystal. Figure 3.7 show the results of western-blot analysis with anti-MPR and anti-CTB antibodies. The results confirm that the uncleaved fusion protein binds to both types of antibodies while the cleaved product showed only binding to anti-CTB antibodies, confirming that the alanine linker and MPR has been cleaved off.

The experiment was repeated at 10°C. Here crystal formation was observed after two weeks and showed the same results on silver-stain in Figure 3.9; the fusion protein is cleaved when crystals are forming in the hanging droplet. No cleavage takes place when the fusion protein is incubated for the same amount of time at the corresponding temperature in the crystallization buffer where the PEG concentration has been reduced to 0 and 2% or the protein is just incubated in the isolation buffer. We concluded that the cleavage of the fusion protein is coupled with the crystallization of the fusion protein and it is not an artifact from crystal handling.

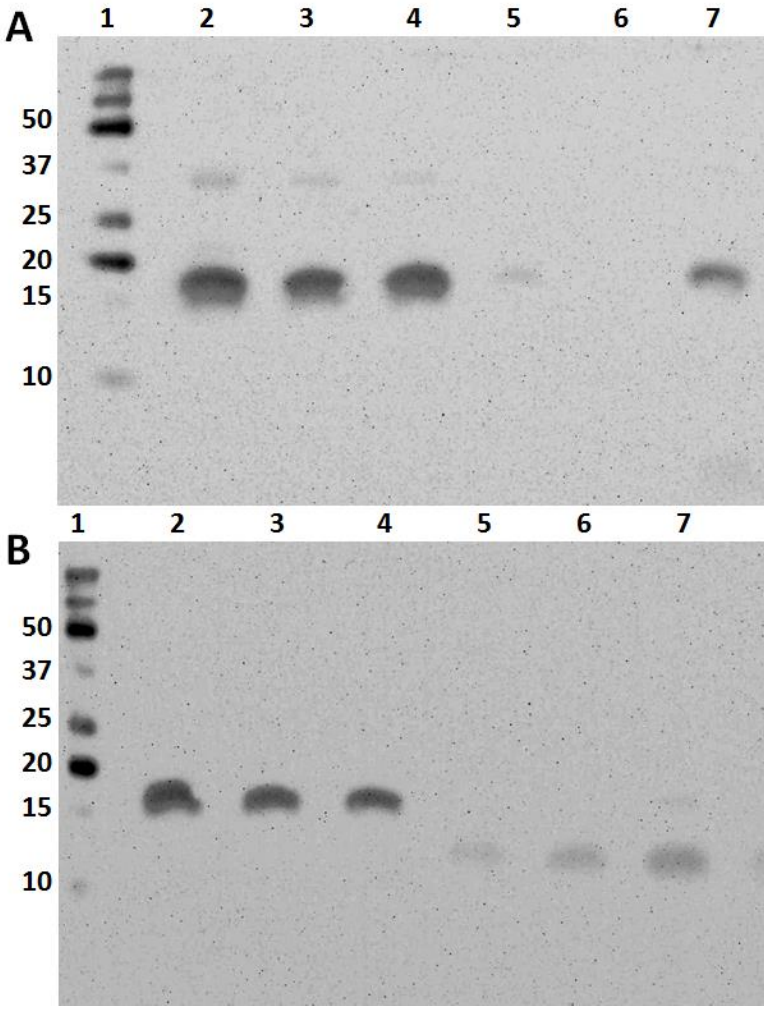


Figure 3.7: **Immunoblotting of day to day crystallization experiments.** Lane 1 represents the molecular weight marker (Bio-Rad, Precision Plus Protein^{TM} Western C Standards), followed by sample at time = 0 in lane 2; t = 1 day in lane 3; t = 2 days in lane 4; t = 3 days in lane 5; t = 4 days in lane 6; t = 5 days in lane 7; (A) Samples tested against monoclonal 4E10 (anti-MPR) antibodies. Signal was only detected at the band corresponding to uncleaved CTB^{AAAA}MPR at 17kDa. (B) Samples tested against polyclonal anti-CTB antibodies. The signal was detected for both the uncleaved CTB^{AAAA}MPR at 17kDa and the cleaved product at 11.5kDa.

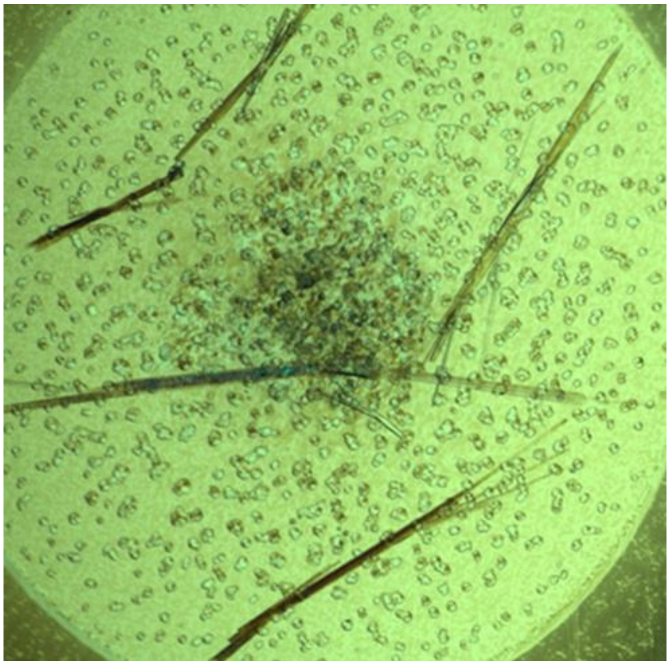


Figure 3.8: **Day 6 of daily observation crystallization experiment at 20°C.** Small round crystals were formed on the third day while long needle-like crystals were observed on the fourth day.

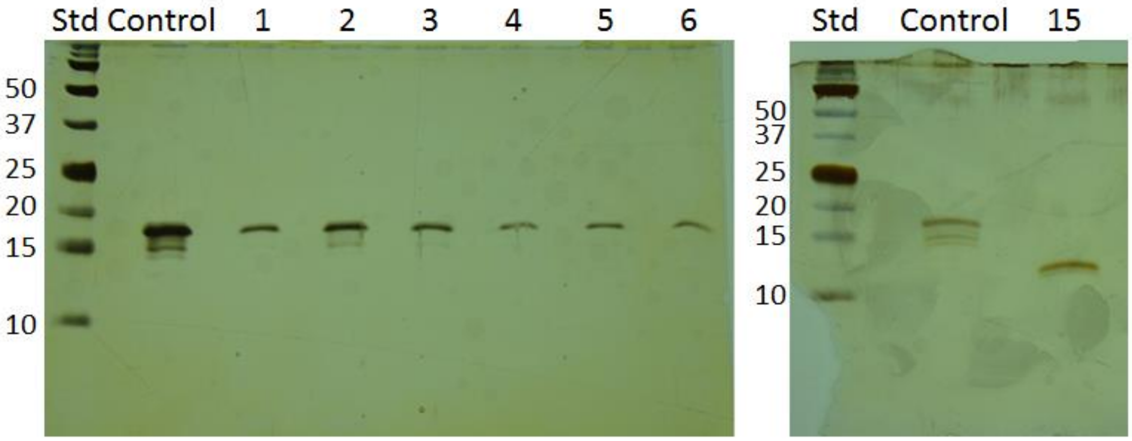


Figure 3.9: **SDS-PAGE of daily observation crystallization experiment at 10°C.** Std represents the molecular weight standard. Control represents the pre-crystallization sample solution. Lane numbers correspond to the day number since crystallization experiment was set up. Small round crystals appeared after two weeks shown in the lane labeled 15 with the cleaved product.

In previous work on CTB^{AAAA}MPR, data on the micro/nano-crystals of the fusion protein were collected at the Linac Coherent Light Source (Lee et al., 2014). Although there was not enough data for structure determination, the space group and unit cell dimensions were completely different from the needle shaped crystals described here. The nanocrystals crystallized in the rhombic space group $R3_2$ with huge unit cell constants of a = b = c = 332 Å, $\alpha = \beta = \lambda = 60°$. This unit cell is very rare in the pdb with only three virus structures reporting similar unit cell constants and space groups. We have reproduced the nanocrystals under exactly the same conditions as described in (Lee et al., 2014) and analyzed the nanocrystals with SDS-PAGE. The result in Figure 3.10 shows that the MPR domain is present in these showers of micro/nano-crystals, i.e. the nanocrystals with the large rhombic space group contain the intact fusion protein.

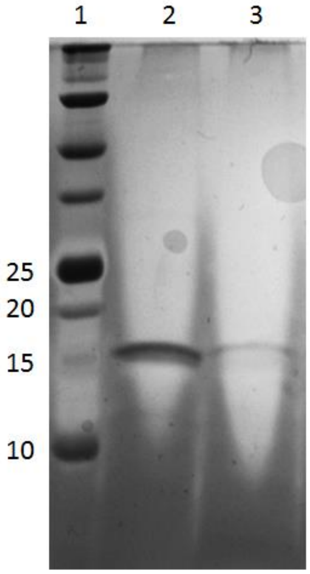


Figure 3.10: **SDS-PAGE silver-stain of CTB^{AAAA}MPR nanocrystal.** Lane 1 represents a molecular weight marker. Lane 2 represents the pre-crystallization solution. Lane 3 represents the nano-crystals. The signal is weak due to the low volume used and possible loss of protein on the concentrator during the batch method.

We therefore explored if seeding with the nanocrystals of the intact fusion protein could induce the formation of larger crystals of the intact fusion protein with the rhombic space group. When these micro/nano-crystals were used for microseeding, it produced larger crystals in a clustered formation shown in Figure 3.11. However, upon examining the protein content of these clustered-crystals, the protein also showed auto-cleavage of the MPR domain.

In conclusion, we have determined that the cleavage of the fusion protein CTB^{AAAA}MPR is coupled with the crystallization process leading to the formation of large well-ordered crystals of CTB in a new crystal packing. Rapid crystallization of CTB^{AAAA}MPR in form of nanocrystals through the batch method lead to crystals of a large rhombic space group which contained the full length fusion protein CTB^{AAAA}MPR, but required a XFEL that produce X-rays more than a billion time brighter than a synchrotron(Fromme & Spence, 2011) to obtain diffraction (Lee et al., 2014).

Even though MPR is not present in the crystal structure of CTB in the space group $P2_12_12_1$, it may have contributed to the crystallization as the crystals show a unique packing of CTB with two pentamers in the asymmetrical unit not aligned to each other.

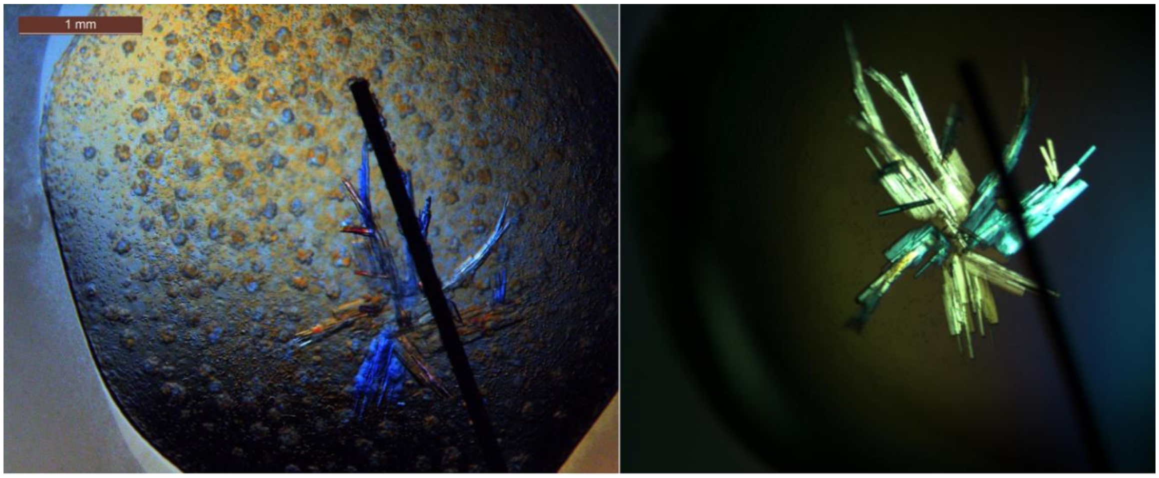


Figure 3.11: **CTB^{AAAA}MPR crystal cluster** obtained from seeding with micro/nano-crystals view with two different polarized lens microscopes.

3.5 Acknowledgements

We would like to thank Eric Davies (Arizona State University) for technical assistance with the cloning. This work was supported in part by the National Institutes of Health (awards U19 AI062150 to TM; U54 GM094599 to PF and TM; and R03 AI073157 to NM). X-ray crystallography experiments were carried out at the Advanced Light Source (ALS), which is a national user facility operated by the University of California on the behalf of the US Department of Energy (DOE), Office of Basic Energy Sciences.

CHAPTER 4

ADDITIONAL EXPERIMENTS, CONCLUSION, AND OUTLOOK

4.1: N105A CTB^{AAAA}MPR

Unable to receive additional beamtime to complete our partial SFX data sets for CTB^{AAAA}MPR at LCLS, we opted to create a new construct with the goal to make the fusion protein more resistant to cleavage. Self-splicing of proteins has been reported through the cyclization of asparagine residues into succinimide, a reaction that lead to the cleavage of the peptide bond (Clarke, 1994, Mathys et al., 1999, Shemella et al., 2007). Indeed, the last amino acid of CTB is an asparagine. The structure determined by X-ray crystallography of CTB crystallized from the solution of CTB^{AAAA}MPR (see chapter 3) confirmed that the last amino acid that is visible in the electron density map is the terminal residue of CTB, which provided further evidence that the self-cleavage have occurred there.

Testing the hypothesis, we decided to design a new construct where we changed the asparagine residue into an alanine by site directed mutagenesis there by extending the poly-alanine linker, hoping to hinder the cleavage of the fusion protein.

Mutation of Expression Vector

We have attempted to perform site directed mutagenesis, where the pTM646 (expression vector for CTB^{AAAA}MPR) was used as the template, using site directed mutagenesis kits (QuikChange and QuikChange II, Agilent), However the cloning has failed. We

Table 4.1: **Oligonucleotides used as primers**

No.	Name	5’-Sequence-3’
1	oTM850	GCCATTAGCATGGCAGCTGCGGCCGCGGCCTCCC
2	oTM851	GGGAGGCCGCGGCCGCAGCTGCCATGCTAATGGC
3	oTM856	ATGAAATACCTGCTGCCGACCGCTGC
4	oTM857	CCCATTCGCCAATCCGGATATAGTTCCTCC

contribute the problem to the extremely high GC-content of the flanking alanine residues (GCC and GCG codons) in the sequence region, as we could not obtain PCR products using the primers of the mutated sequence. A different set of primers (see table 4.1) were tested but the experiments were not successful, as the binding of the primers to each other was irreversible due to the high GC-content of the primers.

Switching to overlapping PCR (Aiyar et al., 1996), two PCR reactions were conducted to generate the mutation using oTM850/oTM851 mutational primers and oTM856/oTM857 end primers that are located upstream and downstream of the gene. In reaction I, oTM851 and oTM856 were used to generate the PCR fragment upstream of the mutation, and reaction II using oTM850 and oTM857 to generate the PCR fragment downstream of the mutation with overlapping at the mutation. The two fragments were then extracted from the gels and joined together in a 3rd PCR reaction using the end primers oTM856 and oTM857 to produce the full length mutational gene. The resulting PCR product was treated with *Nco*I–*Blp*I restriction digestion then ligated into *Nco*I–*Blp*I digested pET-26b(+) vector (see Figure 4.1). The vector was electroporated into DH5α cell lines and grown overnight. The colonies were screened for CTB-MPR by colony screening PCR. The resultant positive colonies were further tested for the desired mutation by isolating plasmids followed by DNA sequencing of the region of interest

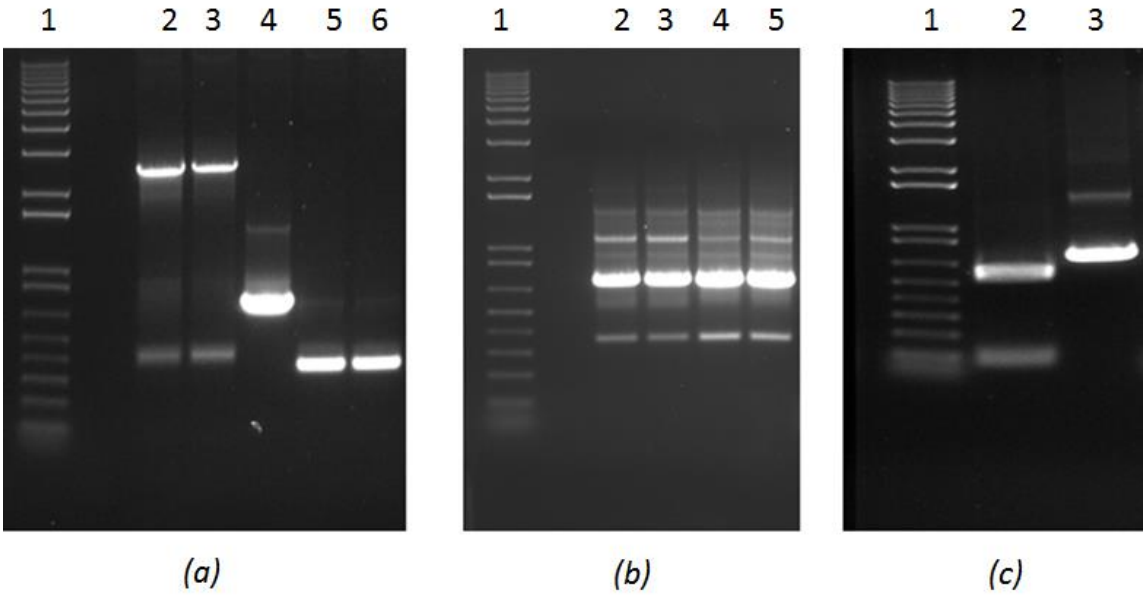


Figure 4.1: **PCR products of N105A mutation** (*a*) DNA standard (lane 1), Reaction I product (lane 2-3), positive control template (lane 4) Reaction II product (lane 5-6). (b) DNA standard (lane 1), PCR product with 50ng of each Reaction I and II (lane 2-3), PCR product with 100ng of each Reaction I and II (lane 4-5). (c) DNA standard (lane 1), *Nco*I–*Blp*I digestion reaction (lane 2), undigested control (lane 3)

using primers oTM 856/oTM 857. Upon confirmation of the desired clone, the plasmid (named as pTM 922), was electroporated into BL21 cell lines.

Expression and purification

The initial expression and purification of N105A CTB^{AAAA}MPR construct following procedures similar to the protocols established for CTB^{AAAA}MPR and described in Chapter 3.

However, problems occurred during the metal affinity chromatography purification step. The protein does not bind strongly to the cobalt-affinity column and the vast majority of the protein was lost in the flow through and washes. Only extremely small amounts of the protein N105A CTB^{AAAA}MPR was detected in the fractions that were eluted with imidazole elution. The protein band was not even detected with silver staining and was barely visible on immunoblots against CTB and MPR (Figure 4.2 a).

The binding of fusion protein to the Talon metal affinity column solely depends on the natural pentameric formation of CTB. This is the reason why CTB and its fusion proteins can be isolated with cobalt-affinity chromatography without an HIS-tag (Dertzbaugh & Cox, 1998). The finding that N105A CTB^{AAAA}MPR does not bind to the cobalt-affinity column indicates that the additional alanine may destabilize the pentameric structure of CTB. The low concentration of the eluted protein will further destabilize the pentameric structure as the formation of CTB oligomers is dependent on the concentration of the protein (Yasuda et al., 1998).

The protocol was modified by using terrific broth (TB) for cell growth instead of the LB medium. TB is a rich growth medium and it has been shown in the literature to increase the cell density and protein concentration in over-expression experiments (Tartoff & Hobb, 1987). Cells were grown at 37°C until an OD_{600nm} of 0.6 was reached. Expression was induced at 25°C with 1 mM IPTG and cells were grown overnight after induction. The wet cell mass yielded from a one-liter TB culture was increased by a factor of 4. Approximately 10 grams were harvested in contrast to a yield of only 2.5 grams per liter for the LB culture. The cells were lysed with a microfluidizer (Microfluidics) and the isolated membrane pellet was frozen at -80°C.

The detergent extraction buffer volume was also decreased from 80 ml to 50ml per 10 grams of wet cell mass. We anticipated to obtain higher concentration of N105A

$CTB^{AAAA}MPR$ protein using this procedure, but we were aware of the fact that there might be a trade-off as with a lower volume of the detergent containing solubilization buffer we may not solubilize all the proteins possible. A second detergent extraction step was performed to examine the cell pellet of the first detergent extraction step. The goal was to check if it is possible to extract additional N105A $CTB^{AAAA}MPR$ protein. Although additional N105 $CTB^{AAAA}MPR$ can be detected in the second extraction, the concentration proved to be not high enough for significant binding to the metal affinity column.

Further purification continued using the detergent extracts described above. The target protein was traced through purification using immunoblot as shown in Figure 4.2. The higher protein concentration led to a significant reduction in the protein loss in the Talon flow through and washes (Figure 4.2 lanes 4 and 5) and the protein could be eluted from the column with imidazole (lane 7). The large improvement in binding and isolation is evident from the comparison of the protein yields achieved in the original protocol (*a,* lanes 7-9) and the modified protocol (b, lane 7).

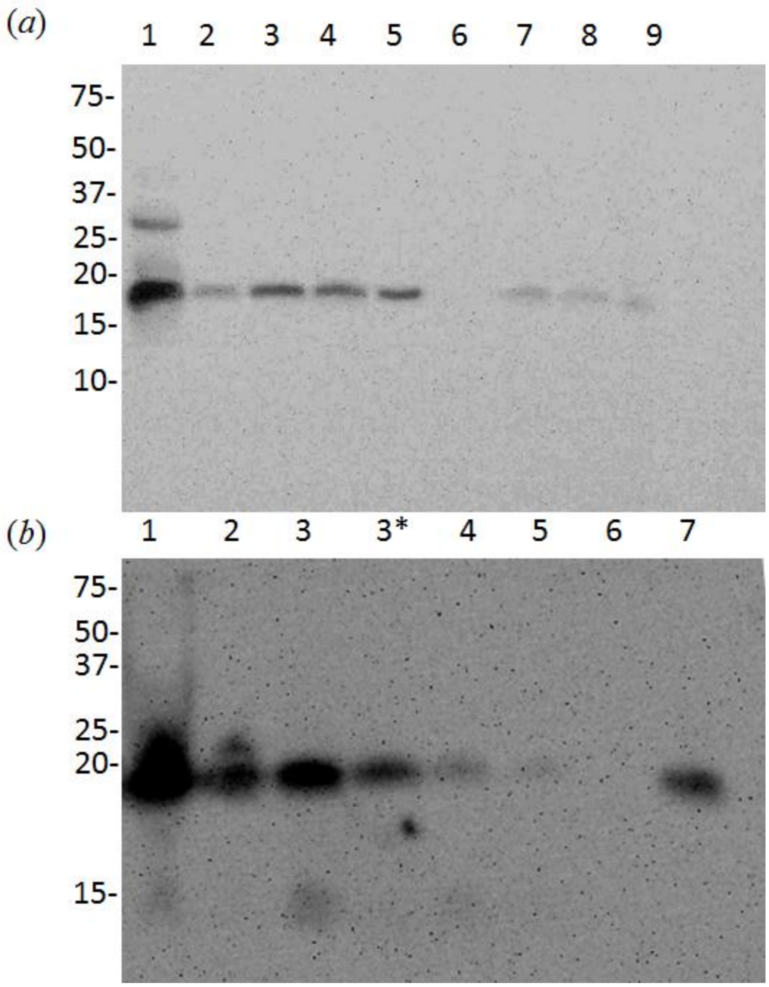


Figure 4.2: **Extraction and Talon purification of N105A Western blots** with 2F5 monoclonal Ab against gp41. (*a*) Purification using same procedure as CTB^{AAAA}MPR in LB media. Showing whole cell lysate (lane 1), aqueous fraction (lane 2), detergent extraction (lane 3), Talon flowthrough (lane 4), washes (lane 5 and 6), elutions (lane 7-9). (b) Purification using TB media and reduced detergent extraction volume. Whole cell lysate (lane 1), aqueous fraction (lane 2), first detergent extraction (lane 3), second detergent extraction (lane 3*), Talon flowthrough (lane 4), washes (lane 5 and 6), elution (lane 7)

The sample from the Talon elution was concentrated 20x using 50 kDa molecular-weight cutoff (MWCO) concentrators (Vivaspin 20 VS2031, Sartorius Stedim Biotech) and loaded on a SuperDex200 column for size-exclusion chromatography (SEC) with the same running buffer as described for the CTB^{AAAA}MPR experiments in chapter 3. The chromatogram is shown in Figure 4.3. The peak intensity for N105A CTB^{AAAA}MPR was considerably lower than CTB^{AAAA}MPR and the chromatogram indicated the presence of aggregates. The protein compositions of the pentamer peak along with its shoulders were examined by SDS-PAGE. The results (insert in Figure 4.3) confirmed that the main peak and its shoulders all contained the fusion protein.

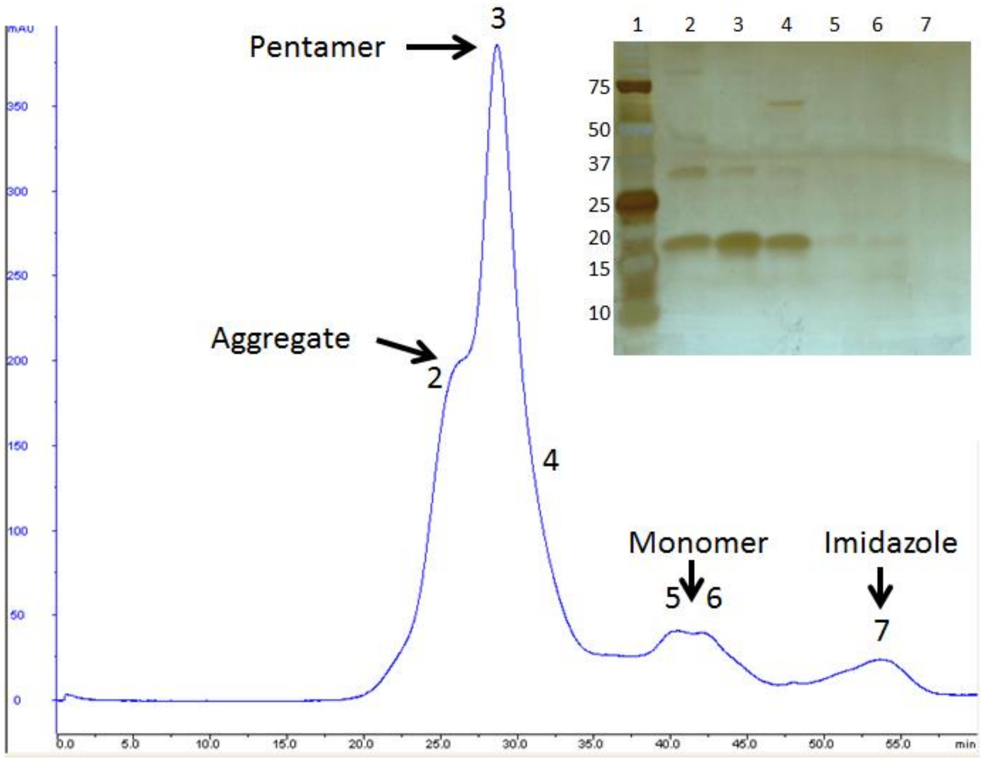


Figure 4.3: **Size-exclusion chromatography of N105A** shows an aggregation peak shoulder before the elution of the main pentamer peak. The peaks fractions were examined by SDS-PAGE shown in the inlet. Lane 1 is a molecular standard and each of the lane numbers corresponds with the marked fraction number indicated on the chromatogram.

The pentamer peak was collected from four runs of SEC and the corresponding fractions containing out fusion protein were pooled together. The concentration was determined spectrophotometrically using the molar extinction coefficient of ε_{280} = 39380 $M^{-1}cm^{-1}$ (calculated with the ProtParam web application; web.expasy.org/protparam/), and concentrated to a final concentration of 10 mg/ml using 100 kDa MWCO concentrators (Amicon Centricon YM-100). Dynamic light-scattering measurements were performed using a NaBiTec GmbH setup. The DLS instrumental setup combines the SpectroSize 302 DLS instrument (Molecular Dimensions) with an S6D microscope (Leica) and an IR camera and allows DLS measurements of concentrated solution in drops as small as 0.5 microliter. The DLS measurements of the protein sample were performed in a 3 µl hanging drop using a 24-well crystallization plate (VDX Greased Plate, Hampton Research) covered with siliconized-glass circular cover slides (22 mm; Hampton Research). The well itself was filled with 600 µl SEC running buffer. Prior to the measurement, the protein solution was centrifuged (1,000 g, 10 minutes at 4°C) to remove any dust particles. During the measurement, the temperature was set to 20°C. Ten consecutive measurements, with an integration time of 20 seconds each, were averaged. An estimate of the hydrodynamic size was obtained with the instrumental software using the following parameters: refractive index 1.33, viscosity 1.006, shape factor 1.0, hydrated shell 0.2 nm.

At 10 mg ml^{-1}, the hydrodynamic radius (Stokes radius, r) of the detergent-solubilized protein was determined to be 7.56±0.45 nm. The results of 7 DLS measurements are shown in Figure 4.4. The left plot shows a graphic representation of the size distribution of particles from each of the 7 runs (each time point on the Y-axis represents one run, the X-axis indicates the Stokes radius determined from each of the runs). The right plot shows the direct scattering intensity, which is not corrected for the molecular mass of the particles to detect even traces of aggregates. Both plots indicate that the sample is highly monodisperse as a high molecular weight oligomer. The molecular weight determined from the DLS experiment is 343.3 kDa .

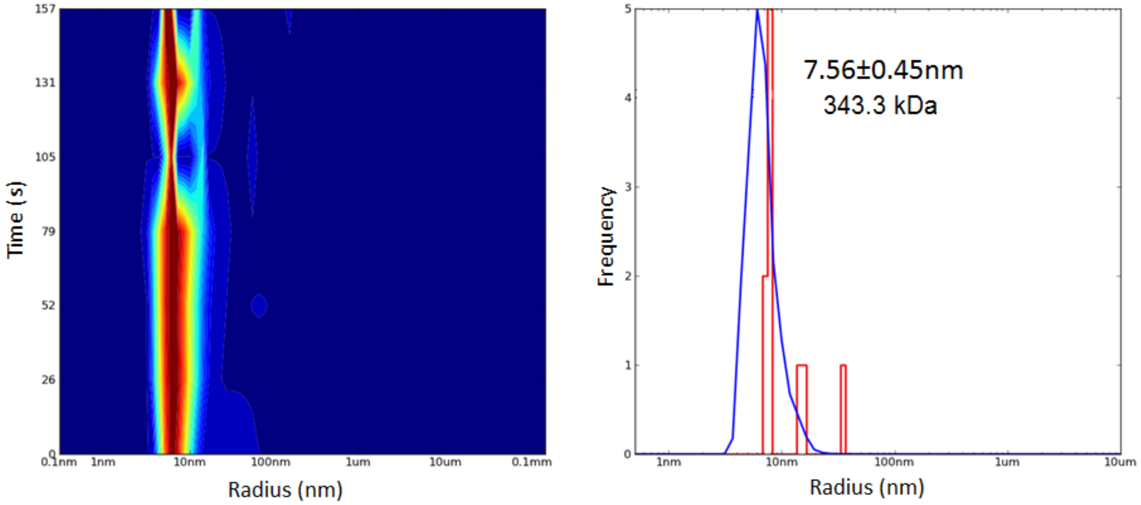


Figure 4.4: **DLS data of N105A CTB^{AAAA}MPR** shows that it is monodisperse at 10mg/ml and forms an high-order oligomer. (left) The size distribution of particles determined from each of the 7 DLS runs. The plot shows the particular hydrodynamic radius determined from each run (indicated by the time line on the Y-axis). The fractions of particles of a particular size are shown color-coded and as a heat plot (red, >90% of particles; dark blue, none). The narrow vertical red profile indicates high stability of size distribution over the duration of all 7 measurements indicative of a low polydispersity. (right) The distribution curve of particle-size frequencies gives a more quantitative evaluation of the polydispersity, with the mean Stokes radius and its error margin indicated next to the peak.

Crystallization experiments

Several broad combinatorial crystallization screens were setup using commercial screening kits from NeXtal and Hampton Research with the vapor diffusion technique. Screening was performed using 96-well plates (Qiagen CrystalEX 96-well Conical Flat Plate) with the sitting-drop method, where each reservoir well contained 100 µl precipitant solutions. The purified protein solution was then mixed in a 1:1 ratio (1 µl:1 µl) with the reservoir solution in the sitting-drop well.

However, no crystal formation was observed and the majority of the wells showed amorphous precipitates. Precipitate was also observed after five days storing at 4°C in the concentrated protein sample at 10mg ml^{-1} without the addition of precipitate solution. We examined the precipitates that formed without the addition of precipitate solution with SONICC for nano-crystals but have detected no signal, we therefore assumed that the precipitates were very likely amorphous and not nano-crystalline. The same precipitates were further examined with SDS-PAGE and the SDS gel. The results shown in Figure 4.5 showed that the precipitate showed cleavage of the N105A CTB^{AAAA}MPR fusion protein. The concentrated sample from SEC clearly shows the molecular weight of the protein to be approximately at the expected 17kDa (Figure 4.3 insert), however, the precipitate formed without addition of precipitate solution ran well below the 15kDa marker and correspond to the molecular weight of CTB only at around 11.5kDa.

This result indicates that the cleavage of the fusion protein is NOT auto-catalyzed by the terminal asparagine residue.

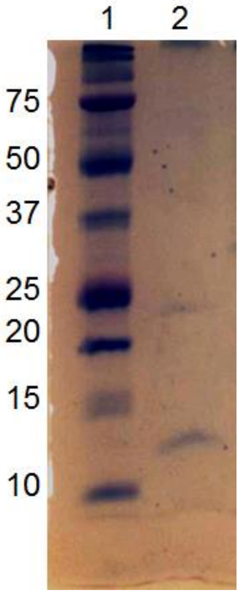


Figure 4.5: **SDS-PAGE of N105A Precipitate:** Molecular weight standard (lane 1) and the protein precipitate (lane 2). It shows the molecular weight of the precipitate to be below 15kDa, corresponding to the approximate molecular of a cleaved product that only contains CTB at 11.5kDa.

4.2: Conclusion

Our studies of the different variants of the CTB-MPR fusion protein with different linkers, we have revealed differences in their biophysical characteristics. Crystallization of the different fusion protein was very challenging possibly due to the symmetry mismatch between the pentameric CTB and the trimeric MPR. Protocols for the purification of all constructs has been established and optimized for each of the individual variants of CTB-MPR. Crystallization experiments led to the determination of growth conditions for the different variants of the fusion protein and several diffraction data sets and protein structures were obtained.

While the CTB^{GPGP}MPR formed crystals of the intact fusion, the electron density beyond the CTB was not well defined, most likely due to the flexible linker, leading to difficulties to fit a structure model of MPR into the electron density map. The removal of the linker compromised the pentameric state of CTB and crystallization proved to be difficult to achieve. Only round pseudo-crystals that only produced powdered diffraction were observed. A more rigid tetra-alanine linker seems to have re-stabilized the CTB pentamer. However, the macroscopic crystals grown from CTB^{AAAA}MPR showed cleavage of the poly-alanine linker and MPR from CTB during the formation of the crystals.

The micro/nanocrystals of CTB^{AAAA}MPR contained the intact fusion protein, but only a partial data set has been collected at the Free Electron Laser and we are unable to complete the data set due to the severe limitations in beamtime at the X-ray FEL at LCLS. We will apply again in May 2015 for protein screening beamtime to collect a complete data set for structure determination of CTB^{AAAA}MPR.

The large difference in the space group and unit cell size between the data collected from the large crystals that contained only CTB and the nanocrystals of the intact fusion protein for which data have been collected by serial femtosecond crystallography highlight significance of serial micro/nanocrystallography for difficult to crystallize proteins especially membrane proteins.

The diffraction data that were gathered from the micro/nanocrystal contained the intact CTB^{AAAA}MPR while macroscopic crystals were either unordered and gave only powder diffraction or contained only the cleaved product, i.e. the structure consisted of CTB-only.

The cleavage was originally attributed to the presence of the asparagine residue as the last amino acid of CTB as it has been reported in the literature that asparagine can self-cyclize into succinimide which is a common “protein splicing” technique in bacteria (Clarke, 1994, Mathys et al., 1999, Shemella et al., 2007). However, the asparagine proved to be not the cause of cleavage as the mutant protein N105A CTB^{AAAA}MPR that was isolated after a long elaborate process of cloning and protein isolation optimization showed similar cleavage as the original CTB^{AAAA}MPR construct.

Independent results with a different fusion protein (MBP-MPR-TM) indicate that MPR with an alanine linker might be highly prone to self-cleavage (Gong, 2014). The cleavage of both CTB^{AAAA}MPR and N105A CTB^{AAAA}MPR is highly similar to the situation where cleavage of a fusion protein that consisted of maltose binding protein (MBP) fused with the membrane proximal region and transmembrane domain of gp41 (MPR-TM) was observed. In this construct, cleavage was observed when MPR-TM and MBP were linked with a triple alanine linker (Gong, 2014). The cleavage could be triggered when MPR is being attached to a large soluble protein with a rigid alanine linker.

The cleavage problem might be overcome by the use of micro/nanocrystals, which contain the full length fusion protein. This showed that the MPR is initially present in the tiny crystals with the large rhombic unit cell. Macroscopic crystal formation is then triggering cleavage of the fusion protein leading to the formation of larger crystals with a small orthorhombic unit cell that only contains CTB.

However, the initial presence of MPR may contribute to the crystal formation as the structures of the cleaved CTB protein shows a different space group and unit cell parameters compared to the native fusion partner alone, both for the CTB and MBP fusion proteins.

Although the structure of the fusion protein of MPR with CTB was not solved, this work has elucidated new information on the fusion protein and very valuable knowledge was gained on self-cleaving process induced by crystallization that may have a huge impact on future studies not only of MPR but also other fusion proteins of membrane or membrane attached proteins. The work sets the foundations for the structure determination of various CTB-MPR variants of potential HIV vaccines and shows the significance of the new method of SFX for difficult to crystallize proteins.

4.3: Outlook

The final medicinal goal of the project is to develop a HIV-1 vaccine component that induces the production of Abs that can block the transcytosis of HIV-1 across the epithelial membrane. For the past few years our work has been focused on the structure determination of fusion proteins of the membrane proximal region of the HIV membrane protein gp41. New CTB-MPR constructs were produced with the purpose of structure determination, however so far only few new immunization experiments were performed with these constructs. Although the different variants proved to be difficult for structural determination with crystallography, they could be very useful for immunology.

It would be interesting to see the effects of the different linkers on CTB-MPR have on the immunization response as previous immunization experiments only explored the first $\mathrm{CTB}^{\mathrm{GPGP}}\mathrm{MPR}$ construct. The constructs CTBMPR and N105A $\mathrm{CTB}^{\mathrm{AAAA}}\mathrm{MPR}$

might be especially interesting targets for immunization. The results have shown that it forms a stable pentamer at high concentrations but the pentameric oligomeric state is unstable at lower concentrations. The major difficulty in the previous immunization experiments with CTB-MPR was the fact that after initial antibody production against MPR and CTB, boosting only further increased the immune response against CTB. CTBMPR and N105A CTB^{AAAA}MPR open the possibility to destabilize the CTB part of the construct in vivo as the dilution of the protein after injection may lead to the disassembly of the CTB pentamer which may then be less immunogenic thereby increasing the chances for formation of Abs against MPR.

The goal of structure determination of CTB-MPR should also continue using the SFX method. The diffraction from LCLS showed great promise and should be further pursued. With the availability of additional XFEL instruments worldwide and additional beamlines at the current two XFELs in Stanford and at SACLA in Japan, it may be possible to receive additional beam time for the CTB-MPR project and collect a full data set.

From the structural point of view, we should also explore the structure of gp41 beyond the MPR region. Currently most of the published structures of gp41 contained only parts of the ectodomain of gp41 (Chan et al., 1997, Melikyan et al., 2000, Shi et al., 2010, Pancera et al., 2014, Reardon et al., 2014), and limited structure information of the transmembrane domain or c-terminal domain has been published. These domains of gp41 may offer additional information in the viral infection cycle of HIV-1.

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, *et al.*, 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21.

Aiyar A, Xiang Y, Leis J, 1996. Site-directed mutagenesis using overlap extension PCR. *Methods Mol Biol* **57**, 177-91.

Alfsen A, Bomsel M, 2002. HIV-1 gp41 envelope residues 650-685 exposed on native virus act as a lectin to bind epithelial cell galactosyl ceramide. *J Biol Chem* **277**, 25649-59.

Alfsen A, Iniguez P, Bouguyon E, Bomsel M, 2001. Secretory IgA specific for a conserved epitope on gp41 envelope glycoprotein inhibits epithelial transcytosis of HIV-1. *J Immunol* **166**, 6257-65.

Alfsen A, Yu HF, Magerus-Chatinet A, Schmitt A, Bomsel M, 2005. HIV-1-infected blood mononuclear cells form an integrin- and agrin-dependent viral synapse to induce efficient HIV-1 transcytosis across epithelial cell monolayer. *Molecular Biology of the Cell* **16**, 4267-79.

Aman AT, Fraser S, Merritt EA, *et al.*, 2001. A mutant cholera toxin B subunit that binds GM1-ganglioside but lacks immunomodulatory or toxic activity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8536-41.

Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR, 2010. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys J* **99**, 933-43.

Backstrom M, Lebens M, Schodel F, Holmgren J, 1994. Insertion of a HIV-1-neutralizing epitope in a surface-exposed internal region of the cholera toxin B-subunit. *Gene* **149**, 211-7.

Barty A, Caleman C, Aquila A, *et al.*, 2012. Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat Photonics* **6**, 35-40.

Barty A, Kirian RA, Maia FR, *et al.*, 2014. : software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J Appl Crystallogr* **47**, 1118-31.

Barty A, Soufli R, Mccarville T, *et al.*, 2009. Predicting the coherent X-ray wavefront focal properties at the Linac Coherent Light Source (LCLS) X-ray free electron laser. *Opt Express* **17**, 15508-19.

Bergfors T, 2003. Seeds to crystals. *J Struct Biol* **142**, 66-76.

Bhuvaneshwari M, Subramanya HS, Gopinath K, Savithri HS, Nayudu MV, Murthy MR, 1995. Structure of sesbania mosaic virus at 3 A resolution. *Structure* **3**, 1021-30.

Bilderback DH, Bazarov IV, Finkelstein K, *et al.*, 2003. Energy-recovery linac project at Cornell University. *J Synchrotron Radiat* **10**, 346-8.

Bilderback DH, Brock J, Dale D, Finkelstein K, Pfeifer M, Gruner SM, 2010. Energy recovery linac (ERL) coherent hard x-ray sources. *New Journal of Physics* **12**.

Biron Z, Khare S, Quadt SR, Hayek Y, Naider F, Anglister J, 2005. The 2F5 epitope is helical in the HIV-1 entry inhibitor T-20. *Biochemistry* **44**, 13602-11.

Bomsel M, 1997. Transcytosis of infectious human immunodeficiency virus across a tight human epithelial cell line barrier. *Nat Med* **3**, 42-7.

Bomsel M, Magerus-Chatinet A, 2004. Galactosyl ceramide is a common mucosal dendritic and epithelial cell receptor for HIV-1 and mediates raft dependent entry of HIV-1 across mucosae. *Molecular Biology of the Cell* **15**, 440a-a.

Bomsel M, Tudor D, Drillet AS, *et al.*, 2011. Immunization with HIV-1 gp41 subunit virosomes induces mucosal antibodies protecting nonhuman primates against vaginal SHIV challenges. *Immunity* **34**, 269-80.

Boutet S, Lomb L, Williams GJ, *et al.*, 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* **337**, 362-4.

Boutet S, Williams GJ, 2010. The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New Journal of Physics* **12**.

Bouvin-Pley M, Morgand M, Meyer L, *et al.*, 2014. Drift of the HIV-1 envelope glycoprotein gp120 toward increased neutralization resistance over the course of the epidemic: a comprehensive study using the most potent and broadly neutralizing monoclonal antibodies. *J Virol* **88**, 13910-7.

Buzon V, Natrajan G, Schibli D, Campelo F, Kozlov MM, Weissenhorn W, 2010. Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog* **6**, e1000880.

Cai L, Gochin M, Liu K, 2011. Biochemistry and biophysics of HIV-1 gp41 - membrane interactions and implications for HIV-1 envelope protein mediated viral-cell fusion and fusion inhibitor design. *Curr Top Med Chem* **11**, 2959-84.

Chan DC, Fass D, Berger JM, Kim PS, 1997. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-73.

Chapman HN, Fromme P, Barty A, *et al.*, 2011. Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-7.

Chayen NE, Saridakis E, 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**, 147-53.

Clarke ND, 1994. A proposed mechanism for the self-splicing of proteins. *Proc Natl Acad Sci U S A* **91**, 11084-8.

Cusack S, Belrhali H, Bram A, Burghammer M, Perrakis A, Riekel C, 1998. Small is beautiful: protein micro-crystallography. *Nat Struct Biol* **5 Suppl**, 634-7.

Delano W, 2002. The PyMOL Molecular Graphics System (2002) on the World Wide Web www.pymol.com. In.: Schrödinger, LLC.

Denner J, 2011. Towards an AIDS vaccine: the transmembrane envelope protein as target for broadly neutralizing antibodies. *Hum Vaccin* **7 Suppl**, 4-9.

Deponte DP, Weierstall U, Schmidt K, *et al.*, 2008. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *Journal of Physics D-Applied Physics* **41**.

Dertzbaugh MT, Cox LM, 1998. The affinity of cholera toxin for Ni2+ ion. *Protein Eng* **11**, 577-81.

Devito C, Broliden K, Kaul R, *et al.*, 2000a. Mucosal and plasma IgA from HIV-1-exposed uninfected individuals inhibit HIV-1 transcytosis across human epithelial cells. *J Immunol* **165**, 5170-6.

Devito C, Hinkula J, Kaul R, *et al.*, 2002. Cross-clade HIV-1-specific neutralizing IgA in mucosal and systemic compartments of HIV-1-exposed, persistently seronegative subjects. *J Acquir Immune Defic Syndr* **30**, 413-20.

Devito C, Hinkula J, Kaul R, *et al.*, 2000c. Mucosal and plasma IgA from HIV-exposed seronegative individuals neutralize a primary HIV-1 isolate. *AIDS* **14**, 1917-20.

Ducruix A, Giegé R, 1992. *Crystallization of nucleic acids and proteins : a practical approach*. Oxford England ; New York: IRL Press at Oxford University Press.

Evans PR, Murshudov GN, 2013. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-14.

Fan E, Merritt EA, Zhang Z, *et al.*, 2001. Exploration of the GM1 receptor-binding site of heat-labile enterotoxin and cholera toxin by phenyl-ring-containing galactose derivatives. *Acta Crystallogr D Biol Crystallogr* **57**, 201-12.

Fowke KR, Nagelkerke NJ, Kimani J, *et al.*, 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* **348**, 1347-51.

Fromme P, Spence JCH, 2011. Femtosecond nanocrystallography using X-ray lasers for membrane protein structure determination. *Current Opinion in Structural Biology* **21**, 509-16.

Gach JS, Leaman DP, Zwick MB, 2011. Targeting HIV-1 gp41 in close proximity to the membrane using antibody and other molecules. *Curr Top Med Chem* **11**, 2997-3021.

Glaeser RM, 1999. Review: electron crystallography: present excitement, a nod to the past, anticipating the future. *J Struct Biol* **128**, 3-14.

Gong Z, 2014. *Structural Studies of the Transmembrane and Membrane Proximal Domains of HIV-1 gp41 by X-Ray Crystallography*: Arizona State University, Ph.D. Thesis.

Goodsell D, 2005. Cholera Toxin - September 2005 Molecule of the Month. In. *RCSB Protein Data Bank*. http://www.rcsb.org/pdb/101/motm.do?momID=69.

Gortmaker SL, Hughes M, Cervia J, *et al.*, 2001. Effect of combination therapy including protease inhibitors on mortality among children and adolescents infected with HIV-1. *New England Journal of Medicine* **345**, 1522-8.

Haase AT, 2010. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature* **464**, 217-23.

Haase AT, 2011. Early events in sexual transmission of HIV and SIV and opportunities for interventions. *Annu Rev Med* **62**, 127-39.

Hessell AJ, Rakasz EG, Tehrani DM, *et al.*, 2010. Broadly neutralizing monoclonal antibodies 2F5 and 4E10 directed against the human immunodeficiency virus type 1 gp41 membrane-proximal external region protect against mucosal challenge by simian-human immunodeficiency virus SHIVBa-L. *J Virol* **84**, 1302-13.

Hladik F, Mcelrath MJ, 2008. Setting the stage: host invasion by HIV. *Nature Reviews Immunology* **8**, 447-57.

Holmgren J, Lycke N, Czerkinsky C, 1993. Cholera toxin and cholera B subunit as oral-mucosal adjuvant and antigen vector systems. *Vaccine* **11**, 1179-84.

Huang J, Ofek G, Laub L, *et al.*, 2012. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406-12.

Jeang K-T, 2007. *HIV-1 : molecular biology and pathogenesis*. San Diego, Calif.: Elsevier.

Kabsch W, 2010. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32.

Kirian RA, White TA, Holton JM, *et al.*, 2011. Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallogr A* **67**, 131-40.

Kresina TF, Mathieson B, 1999. Human immunodeficiency virus type 1 infection, mucosal immunity, and pathogenesis and extramural research programs at the National Institutes of Health. *J Infect Dis* **179 Suppl 3**, S392-6.

Krishna SS, Hiremath CN, Munshi SK*, et al.*, 1999. Three-dimensional structure of physalis mottle virus: implications for the viral assembly. *J Mol Biol* **289**, 919-34.

Kupitz C, Basu S, Grotjohann I*, et al.*, 2014. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261-5.

Lakomek NA, Kaufman JD, Stahl SJ, Wingfield PT, 2014. HIV-1 envelope protein gp41: an NMR study of dodecyl phosphocholine embedded gp41 reveals a dynamic prefusion intermediate conformation. *Structure* **22**, 1311-21.

Lawrence RM, Varco-Merth B, Bley CJ, Chen JJ, Fromme P, 2011. Recombinant production and purification of the subunit c of chloroplast ATP synthase. *Protein Expr Purif* **76**, 15-24.

Lee HH, Cherni I, Yu H*, et al.*, 2014. Expression, purification and crystallization of CTB-MPR, a candidate mucosal vaccine component against HIV-1. *IUCrJ* **1**, 305-17.

Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ, 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* **265**, 10373-82.

Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S, 2008. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109-13.

Liu W, Wacker D, Gati C*, et al.*, 2013. Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521-4.

Lomb L, Steinbrener J, Bari S*, et al.*, 2012. An anti-settling sample delivery instrument for serial femtosecond crystallography. *J Appl Crystallogr* **45**, 674-8.

Magerus-Chatinet A, Yu H, Garcia S, Ducloux E, Terris B, Bomsel M, 2007. Galactosyl ceramide expressed on dendritic cells can mediate HIV-1 transfer from monocyte derived dendritic cells to autologous T cells. *Virology* **362**, 67-74.

Malley R, Morse SC, Leite LC*, et al.*, 2004. Multiserotype protection of mice against pneumococcal colonization of the nasopharynx and middle ear by killed nonencapsulated cells given intranasally with a nontoxic adjuvant. *Infect Immun* **72**, 4290-2.

Mathys S, Evans TC, Chute IC*, et al.*, 1999. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**, 1-13.

Matoba N, Geyer BC, Kilbourne J, Alfsen A, Bomsel M, Mor TS, 2006. Humoral immune responses by prime-boost heterologous route immunizations with CTB-MPR(649-684), a mucosal subunit HIV/AIDS vaccine candidate. *Vaccine* **24**, 5047-55.

Matoba N, Griffin TA, Mittman M, *et al.*, 2008. Transcytosis-blocking abs elicited by an oligomeric immunogen based on the membrane proximal region of HIV-1 gp41 target non-neutralizing epitopes. *Curr HIV Res* **6**, 218-29.

Matoba N, Kajiura H, Cherni I, *et al.*, 2009. Biochemical and immunological characterization of the plant-derived candidate human immunodeficiency virus type 1 mucosal vaccine CTB-MPR. *Plant Biotechnol J* **7**, 129-45.

Matoba N, Magerus A, Geyer BC, *et al.*, 2004. A mucosally targeted subunit vaccine candidate eliciting HIV-1 transcytosis-blocking Abs. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13584-9.

Matoba N, Shah NR, Mor TS, 2011. Humoral immunogenicity of an HIV-1 envelope residue 649-684 membrane-proximal region peptide fused to the plague antigen F1-V. *Vaccine* **29**, 5584-90.

Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-74.

Mcpherson A, 2004. Introduction to protein crystallization. *Methods* **34**, 254-65.

Melikyan GB, Markosyan RM, Hemmati H, Delmedico MK, Lambert DM, Cohen FS, 2000. Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. *J Cell Biol* **151**, 413-23.

Meng G, Wei X, Wu X, *et al.*, 2002. Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells. *Nat Med* **8**, 150-6.

Merritt EA, Sarfaty S, Van Den Akker F, L'hoir C, Martial JA, Hol WG, 1994. Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci* **3**, 166-75.

Miao J, Chapman HN, Kirz J, Sayre D, Hodgson KO, 2004. Taking X-ray diffraction to the limit: macromolecular structures from femtosecond X-ray pulses and diffraction microscopy of cells with synchrotron radiation. *Annu Rev Biophys Biomol Struct* **33**, 157-76.

Miyazawa M, Lopalco L, Mazzotta F, *et al.*, 2009. The 'immunologic advantage' of HIV-exposed seronegative individuals. *AIDS* **23**, 161-75.

Montero M, Gulzar N, Klaric KA, *et al.*, 2012. Neutralizing epitopes in the membrane-proximal external region of HIV-1 gp41 are influenced by the transmembrane domain and the plasma membrane. *J Virol* **86**, 2930-41.

Montero M, Van Houten NE, Wang X, Scott JK, 2008. The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: dominant site of antibody neutralization and target for vaccine design. *Microbiol Mol Biol Rev* **72**, 54-84, table of contents.

Murphy K, Travers P, Walport M, Janeway C, 2008. *Janeway's immunobiology*. New York: Garland Science.

Murshudov GN, Skubak P, Lebedev AA, *et al.*, 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-67.

Ni D, Zook J, Klewer DA, Nieman RA, Soll J, Fromme P, 2011. Isolation, folding and structural investigations of the amino acid transporter OEP16. *Protein Expr Purif* **80**, 157-68.

Oneil KT, Degrado WF, 1990. A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino-Acids. *Science* **250**, 646-51.

Overbaugh J, Kreiss J, Poss M, *et al.*, 1999. Studies of human immunodeficiency virus type 1 mucosal viral shedding and transmission in Kenya. *J Infect Dis* **179 Suppl 3**, S401-4.

Owen RL, Rudino-Pinera E, Garman EF, 2006. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc Natl Acad Sci U S A* **103**, 4912-7.

Pancera M, Zhou T, Druz A, *et al.*, 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*.

Pejchal R, Gach JS, Brunel FM, *et al.*, 2009. A conformational switch in human immunodeficiency virus gp41 revealed by the structures of overlapping epitopes recognized by neutralizing antibodies. *J Virol* **83**, 8451-62.

Peltola H, Siitonen A, Kyronseppa H, *et al.*, 1991. Prevention of travellers' diarrhoea by oral B-subunit/whole-cell cholera vaccine. *Lancet* **338**, 1285-9.

Pickens JC, Merritt EA, Ahn M, Verlinde CL, Hol WG, Fan E, 2002. Anchor-based design of improved cholera toxin and E. coli heat-labile enterotoxin receptor binding antagonists that display multiple binding modes. *Chem Biol* **9**, 215-24.

Pope M, Haase AT, 2003. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med* **9**, 847-52.

Preston BD, Poiesz BJ, Loeb LA, 1988. Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168-71.

Proteau A, Shi R, Cygler M, 2010. Application of dynamic light scattering in protein crystallization. *Curr Protoc Protein Sci* **Chapter 17**, Unit 17 0.

Purtscher M, Trkola A, Grassauer A*, et al.*, 1996. Restricted antigenic variability of the epitope recognized by the neutralizing gp41 antibody 2F5. *AIDS* **10**, 587-93.

Qu D, Zheng B, Yao X*, et al.*, 2005. Intranasal immunization with inactivated SARS-CoV (SARS-associated coronavirus) induced local and serum antibodies in mice. *Vaccine* **23**, 924-31.

Quiding M, Nordstrom I, Kilander A*, et al.*, 1991. Intestinal immune responses in humans. Oral cholera vaccination induces strong intestinal antibody responses and interferon-gamma production and evokes local immunological memory. *J Clin Invest* **88**, 143-8.

Reardon PN, Sage H, Dennison SM*, et al.*, 2014. Structure of an HIV-1-neutralizing antibody target, the lipid-bound gp41 envelope membrane proximal region trimer. *Proc Natl Acad Sci U S A* **111**, 1391-6.

Redecke L, Nass K, Deponte DP*, et al.*, 2013. Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227-30.

Reshetnyak YK, Koshevnik Y, Burstein EA, 2001. Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between fluorescence and microenvironment parameters of individual tryptophan residues. *Biophys J* **81**, 1735-58.

Roberts JD, Bebenek K, Kunkel TA, 1988. The Accuracy of Reverse-Transcriptase from Hiv-1. *Science* **242**, 1171-3.

Rupp B, 2010. *Biomolecular crystallography : principles, practice, and application to structural biology*. New York: Garland Science.

Sangita V, Lokesh GL, Satheshkumar PS*, et al.*, 2005. Structural studies on recombinant T = 3 capsids of Sesbania mosaic virus coat protein mutants. *Acta Crystallogr D Biol Crystallogr* **61**, 1402-5.

Schagger H, 2006. Tricine-SDS-PAGE. *Nat Protoc* **1**, 16-22.

Shemella P, Pereira B, Zhang Y*, et al.*, 2007. Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations. *Biophys J* **92**, 847-53.

Shen R, Drelichman ER, Bimczok D*, et al.*, 2010. GP41-specific antibody blocks cell-free HIV type 1 transcytosis through human rectal mucosa and model colonic epithelium. *J Immunol* **184**, 3648-55.

Shi W, Bohon J, Han DP, *et al.*, 2010. Structural characterization of HIV gp41 with the membrane-proximal external region. *J Biol Chem* **285**, 24290-8.

Song L, Sun ZY, Coleman KE, *et al.*, 2009. Broadly neutralizing anti-HIV-1 antibodies disrupt a hinge-related function of gp41 at the membrane interface. *Proc Natl Acad Sci U S A* **106**, 9057-62.

Soudeyns H, Paolucci S, Chappey C, *et al.*, 1999. Selective pressure exerted by immunodominant HIV-1-specific cytotoxic T lymphocyte responses during primary infection drives genetic variation restricted to the cognate epitope. *Eur J Immunol* **29**, 3629-35.

Spence JC, Doak RB, 2004. Single molecule diffraction. *Phys Rev Lett* **92**, 198102.

Spence JC, Weierstall U, Chapman HN, 2012. X-ray lasers for structural and dynamic biology. *Rep Prog Phys* **75**, 102601.

Steckbeck JD, Craigo JK, Barnes CO, Montelaro RC, 2011. Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. *J Biol Chem* **286**, 27156-66.

Sun JB, Holmgren J, Czerkinsky C, 1994. Cholera toxin B subunit: an efficient transmucosal carrier-delivery system for induction of peripheral immunological tolerance. *Proc Natl Acad Sci U S A* **91**, 10795-9.

Tartoff K, Hobb C, 1987. Improved Media for Growing Plasmid and Cosmid Clones, B. *Bethesda Res. Lab. Focus* **9**, 12.

Teixeira C, Gomes JR, Gomes P, Maurel F, Barbault F, 2011. Viral surface glycoproteins, gp120 and gp41, as potential drug targets against HIV-1: brief overview one quarter of a century past the approval of zidovudine, the first anti-retroviral drug. *Eur J Med Chem* **46**, 979-92.

Travers SA, O'connell MJ, Mccormack GP, Mcinerney JO, 2005. Evidence for heterogeneous selective pressures in the evolution of the env gene in different human immunodeficiency virus type 1 subtypes. *J Virol* **79**, 1836-41.

Tudor D, Derrien M, Diomede L, *et al.*, 2009. HIV-1 gp41-specific monoclonal mucosal IgAs derived from highly exposed but IgG-seronegative individuals block HIV-1 epithelial transcytosis and neutralize CD4(+) cell infection: an IgA gene and functional analysis. *Mucosal Immunol* **2**, 412-26.

Unaids JUNPOHA, 2013. Global Report: UNAIDS report on the global AIDS epidemic 2013. In. http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf.

Valdiserri RO, Ogden LL, Mccray E, 2003. Accomplishments in HIV prevention science: implications for stemming the epidemic. *Nat Med* **9**, 881-6.

Van Den Akker F, Feil IK, Roach C, Platas AA, Merritt EA, Hol WG, 1997. Crystal structure of heat-labile enterotoxin from Escherichia coli with increased thermostability introduced by an engineered disulfide bond in the A subunit. *Protein Sci* **6**, 2644-9.

Wang W, Nie J, Prochnow C, *et al.*, 2013. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10**, 14.

Weierstall U, Doak RB, Spence JCH, *et al.*, 2008. Droplet streams for serial crystallography of proteins. *Experiments in Fluids* **44**, 675-89.

Weierstall U, Spence JC, Doak RB, 2012. Injector for scattering measurements on fully solvated biospecies. *Rev Sci Instrum* **83**, 035108.

White TA, Kirian RA, Martin AV, *et al.*, 2012. CrystFEL: a software suite for snapshot serial crystallography. *J Appl Crystallogr* **45**, 335-41.

Woolfson MM, 1997. *An introduction to X-ray crystallography*. Cambridge ; New York, NY, USA: Cambridge University Press.

Wyatt R, Sodroski J, 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**, 1884-8.

Yasuda Y, Matano K, Asai T, Tochikubo K, 1998. Affinity purification of recombinant cholera toxin B subunit oligomer expressed in Bacillus brevis for potential human use as a mucosal adjuvant. *FEMS Immunol Med Microbiol* **20**, 311-8.

Zolla-Pazner S, 2004. Identifying epitopes of HIV-1 that induce protective antibodies. *Nature Reviews Immunology* **4**, 199-210.

Zolla-Pazner S, Kong XP, Jiang X, *et al.*, 2011. Cross-clade HIV-1 neutralizing antibodies induced with V3-scaffold protein immunogens following priming with gp120 DNA. *J Virol* **85**, 9887-98.

Zrimi J, Ng Ling A, Giri-Rachman Arifin E, Feverati G, Lesieur C, 2010. Cholera toxin B subunits assemble into pentamers--proposition of a fly-casting mechanism. *PLoS One* **5**, e15347.

Zwick MB, 2005. The membrane-proximal external region of HIV-1 gp41: a vaccine target worth exploring. *AIDS* **19**, 1725-37.

Zwick MB, Labrijn AF, Wang M, *et al.*, 2001. Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *J Virol* **75**, 10892-905.