Structural Studies of the Transmembrane and Membrane Proximal Domains of HIV-1 gp41 by X-Ray Crystallography

by

Zhen Gong

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

Approved November 2014 by the
Graduate Supervisory Committee

Petra Fromme, Co-Chair
Tsafrir Mor, Co-Chair
Alexandra Ros
Kevin Redding

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

The transmembrane subunit (gp41) of the envelope glycoprotein of HIV-1 associates noncovalently with the surface subunit (gp120) and together they play essential roles in viral mucosal transmission and infection of target cells. The membrane proximal region (MPR, residues 649-683) of gp41 is highly conserved and contains epitopes of broadly neutralizing antibodies. The transmembrane (TM) domain (residues 684-705) of gp41 not only anchors the envelope glycoprotein complex in the viral membrane but also dynamically affects the interactions of the MPR with the membrane. While high-resolution X-ray structures of some segments of the MPR were solved in the past, they represent the pre-fusion and post-fusion conformations, most of which could not react with the broadly neutralizing antibodies 2F5 and 4E10. Structural information on the TM domain of gp41 is scant and at low resolution.

This thesis describes the structural studies of MPR-TM (residues 649-705) of HIV-1 gp41 by X-ray crystallography. MPR-TM was fused with different fusion proteins to improve the membrane protein overexpression. The expression level of MPR-TM was improved by fusion to the C-terminus of the Mistic protein, yielding ~1 mg of pure MPR-TM protein per liter cell culture. The fusion partner Mistic was removed for final crystallization. The isolated MPR-TM protein was biophysically characterized and is a monodisperse candidate for crystallization. However, no crystal with diffraction quality was obtained even after extensive crystallization screens. A novel construct was designed to overexpress MPR-TM as a maltose binding protein (MBP) fusion. About 60 mg of MBP/MPR-TM recombinant protein was obtained from 1 liter of cell culture. Crystals of

MBP/MPR-TM recombinant protein could not be obtained when MBP and MPR-TM were separated by a 42 amino acid (aa)-long linker but were obtained after changing the linker to three alanine residues. The crystals diffracted to 2.5 Å after crystallization optimization. Further analysis of the diffraction data indicated that the crystals are twinned. The final structure demonstrated that MBP crystallized as a dimer of trimers, but the electron density did not extend beyond the linker region. We determined by SDS-PAGE and MALDI-TOF MS that the crystals contained MBP only. The MPR-TM of gp41 might be cleaved during or after the process of crystallization. Comparison of the MBP trimer reported here with published trimeric MBP fusion structures indicated that MBP might form such a trimeric conformation under the effect of MPR-TM.

DEDICATION

I dedicate this dissertation to my parents, Wu Gong and Jingxiu Wang, my husband Chi Xu, my son Vincent Xu and my brother Wuwei Gong. I would like to specially thank my mom for her great contribution to our family. She has offered me tremendous love, support and encouragement since I was born. It would be impossible for me to finish this dissertation without her help of taking care of my son. I want to thank my father, who encourages me to continue higher education and pursue a doctoral degree. Special appreciation goes to my husband Chi Xu, for his enormous love, understanding, support and encouragement. I also want to thank my little sweetheart Vincent Xu, for all the unparalleled happiness that he brings to us. Finally, I want to thank my brother Wuwei Gong, for his understanding and support.

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge my advisor Dr. Petra Fromme for her expertise, mentoring, support and encouragement through the past five years. Her enthusiasm for daily life especially science greatly affected me. Her broad knowledge and earnest attitude towards science guided me through my research.

I would like to thank my advisor Dr. Tsafrir Mor for his guidance and support. I started to learn the basic experimental techniques in his lab. His profound knowledge and competence on scientific paper writing greatly improved my scientific writing skills. I am also grateful to my committee members, Dr. Alexandra Ros and Dr. Kevin Redding for their time and advice.

I want to thank all the past and present graduate students, postdoctoral researchers, and lab members. I give my special thanks to Dr. Raimund Fromme. He brought me to the wonderful world of crystallography. He taught me all the basic knowledge and experimental techniques about crystallography and helped me collect the X-ray diffraction data. I would like to thank Ho-Hsien Lee, who is always happy to share his valuable experience with me. I am very grateful for Dr. Rob Lawrence and Jay-How Yang's tremendous effort to make our lab clean and tidy. I also want to thank Shibom Basu, who helped me collect the X-ray diffraction data. I would like to thank Dr. Ingo Grotjohann, for his countless help in the lab and also being a good friend.

This work has been realized with the help of many fellow researchers from MPID (center for membrane proteins in infectious Diseases). Dr. Sarah Kessans worked closely with me on this project at the beginning and taught me all the basic experimental techniques. I give my special thanks to Dr. Katerina Dörner and Dr. Jose Martin-Garcia.

They not only taught me useful techniques, but also had countless discussions with me, which finally lead to promising achievements. I also want to thank them for their encouragement and being my good friends. I would like to thank Dr. Sashka Daskalova and Felicia Craciunescu for designing and cloning the plasmid; Dr. Liqing Chen for setting up crystallization trays with robot; Dr. Brenda Hogue and Dr. Debra Hansen for thoughtful discussions. I want to thank Tina Esquerra. As the manager of MPID, she is extremely nice and helpful with many daily affairs. I sincerely appreciate Dr. Lusheng Song's help for surface plasmon resonance measurements.

I truly acknowledge APS data collection workshop and CCP4 school. I want to specially thank Dr. Ruslan Sanishvili, Dr. Ronan Keegan, Dr. Andrey Lebedev and Dr. Tim Grüne. I am grateful for the huge effort that Dr. Ruslan Sanishvili has put on organizing the workshop and also his competence on X-ray data collection. I am greatly impressed with Dr. Andrey Lebedev and Dr. Ronan Keegan's expertise on data analysis. The structure will not be solved without their professional and generous help.

Additionally, I want to thank the support from Arizona State University and Department of Chemistry and Biochemistry. I would like to thank all the staff in Chemistry Department, who heartily care us and make our life and work much easier.

The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: MAb to HIV-1 gp41 2F5 and 4E10 from Dr. Hermann Katinger. This work was supported by the Center for Membrane Proteins in Infectious Diseases (MPID), which is funded by the Protein Structure Initiative of NIH PSI:Biology of the National Institute of Health (grant number 1U54GM094625).

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABBREVIATIONS

Ab……Antibody

AIDS……....Acquired Immune Deficiency Syndrome

ASU…….....Asymmetric Unit

BLV-1……..bovine leukemia virus type 1

BSA……Bovine Serum Albumin

CBD……Chitin Binding Domain

CC…….coiled coil

CD……Circular Dichroism

CHR……...C-terminal Heptad Repeat Region of Gp41

CTD……..Cytoplasmic Domain of Gp41

CTT……...C-terminal Tail of Gp41

CV…….Column Volume

DLS…….Dynamic light Scattering

EM…….Electron Microscopy

Env…….Envelope Glycoprotein

FP……...Fusion Peptide of Gp41

FPRR……...Fusion Peptide Proximal Region of Gp41

GalCer……Galactosyl Ceramide

HIV-1……...Human Immunodeficiency Virus Type 1

HTLV-1…… human T cell leukemia virus type 1

IPTG……Isopropyl β-D-1-Thiogalactopyranoside

KE……..Kennedy Epitope

LB……Luria-Bertani

LDAO……Lauryldimethylamine-Oxide

mAb……..Monoclonal Antibody

MALDI-TOF……...Matrix-Assisted Laser Desorption/Ionization-Time of Flight

MBP……Maltose Binding Protein

Mcm10……minichromosome maintenance protein 10

MPER……..Membrane Proximal External Region of Gp41

MPR…….……Membrane Proximal Region of Gp41

MS……Mass Spectrometry

MSD……...Membrane-Spanning Domain

MW……...…Molecular Weight

NHR…….....N-terminal Heptad Repeat Region of Gp41

Ni-NTA……..Nickel-Nitrilotriacetic Acid

PBS……...Phosphate Buffered Saline

PBST……...Phosphate Buffered Saline supplemented with Tween 20

PBST-M…....Phosphate Buffered Saline supplemented with Tween 20 and Non-Fat Milk

PDB……..Protein Data Bank

RT…….……..Reverse Transcriptase

RMSD…….....Root-Mean-Square Deviation

SEC…….……...Size Exclusion Chromatography

SPRi……..Surface Plasmon Resonance Imaging

SRP……Signal Recognition Particle

TCR……T-cell Receptor

TEV…….....Tobacco Etch Virus

TM…….....Transmembrane Domain of Gp41

TMD……Transmembrane Domain of Gp41

WHO……..World Health Organization

βDM……n-Decyl-β-D-Maltoside

βDDM…….....n-dodecyl-β-D-maltoside

CHAPTER 1

INTRODUCTION

1.1 Current Status Of HIV Infections Worldwide

The human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS), has become one of the world's most serious health problems. AIDS was first reported in 1981 (Patrick et al., 2013). According to the World Health Organization (WHO), almost 75 million people have been infected with HIV virus and about 36 million people have died of HIV since the beginning of the epidemic. An estimation of 35 million people are living with HIV globally (The Gap report - unaids, 2013). However, there is no effective vaccine against HIV-1. The HIV-1 virus has evolved multiple strategies to escape the antibody binding (Johnson and Desrosiers, 2002). The envelope glycoprotein (Env) glycoprotein of HIV-1 is heavily glycosylated, and the poorly immunogenic glycans inhibit antibody access to the underlying peptide (Burton et al., 2004). Trimerization of Env makes epitopes in Env less accessible to antibodies compared with monomeric subunit (Burton et al., 2004). Conformational change of Env during virus fusion with target cell makes epitopes in Env only accessible in the intermediate stage. High mutation rate of HIV due to low fidelity of the reverse transcriptase (RT) is another important reason why no vaccine exists so far (Preston et al., 1988). These strategies make the development of vaccine against HIV-1 more difficult.

1.2 The Env of HIV-1

The Env of HIV-1 plays an essential role for virus attachment and fusion with target cells and is a primary target for vaccine development (Hughson, 1997). The HIV-1 glycoprotein is initially expressed as a precursor form gp160, which is then

proteolytically cleaved into two non-covalently associated subunits: gp120 and gp41 (Decroly et al., 1994). Gp120, a surface subunit, is responsible for recognizing and binding to cell surface receptors CD4 and a coreceptor, usually CCR5 or CXCR4 (Broder and Dimitrov, 1996; Moore et al., 1997), while gp41, a transmembrane subunit, mediates fusion between viral and cellular membranes through its own conformational change (Chan and Kim, 1998). At the surface of the virus, HIV-1 Env oligomerize into a trimer of heterodimers consisting of gp120 and gp41, in which gp120 forms a cap covering most of the gp41(Chan et al., 1997; Weissenhorn et al., 1997).

The Env of HIV-1 gp41 consists of an ectodomain (residues 512–683), a transmembrane domain (residues 684–705), and a cytoplasmic domain (residues 706–856) (Weiss, 2003; Checkley et al., 2011). Biophysical and structural studies delineate further distinct structural and functional features within the ectodomain of gp41 (Fig. 1.1). The very N-terminus of gp41 is a hydrophobic glycine-rich fusion peptide (FP, residues 512-527), which is exposed and inserted into the target cell membrane at an early step of the fusion process. The N-terminal heptad repeat (NHR or HR1, residues 540-590) is adjacent to the FP, while the C-terminal heptad repeat (CHR or HR2, residues 628-661) precedes the membrane proximal external region (MPER, residues 662-682) and transmembrane (TM, residues 684-705) domain. The membrane proximal region (MPR) with residues 649 to 683 is composed of part of CHR (residues 647 to 661) and the whole MPER (residues 662 to 683). MPR-TM is among the most highly conserved regions in the HIV-1 Env and contains epitopes for the neutralizing monoclonal antibodies (mAbs) 2F5 (Purtscher et al., 1994), 4E10 (Zwick et al., 2001) and 10E8 (Huang et al., 2012).

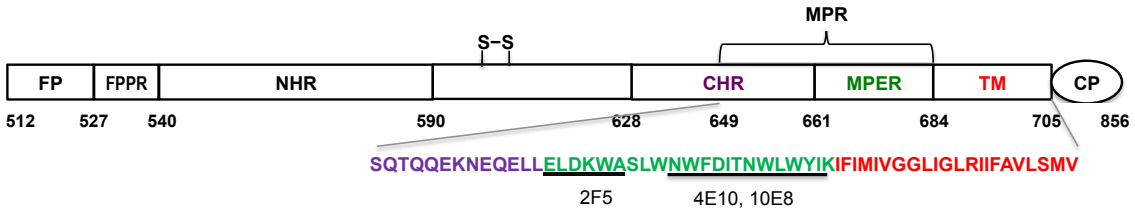


Fig. 1.1 Schematic diagram of the sequence of HIV-1 gp41. FP (residues 512-527): fusion peptide; FPRR (residues 528-539): fusion peptide proximal region; NHR (residues 540-590): N-terminal heptad repeat region; S-S: a disulfide linkage; CHR (residues 628-661): C-terminal heptad repeat region; MPER (residues 662-683): membrane proximal external region; MPR (residues 649-683): membrane proximal region; TM (residues 684-705): transmembrane domain; CP (residues 706-856): cytoplasmic domain. 2F5, 4E10 and 10E8 are epitopes for three broadly neutralizing mAbs.

1.3 Current Model of HIV-1 Membrane Fusion

One current model for HIV-1 viral membrane fusion is shown in Fig. 1.2A (Eckert and Kim, 2001). The conformation of gp41 undergoes at least three different stages during this process: native trimer, pre-hairpin intermediate and postfusion six-helix bundle. In the native stage, Env is a trimer of heterodimers consisting of gp120 and gp41. The FP and NHR (or HR1) are buried in the interior of gp120. Binding of gp120 to the target cell receptor CD4 and a coreceptor CCR5 or CXCR4 triggers the conformational change of the Env. FP and NHR (or HR1) are exposed and extended into the target cell membrane, which leads to the prehairpin intermediate of gp41. The postfusion conformation of gp41 is a trimer of hairpin formed by antiparallel association of NHR (or HR1) and CHR (or HR2). Three NHRs form an interior, parallel coiled-coil trimer, while three CHRs pack in an antiparallel manner on the surface of NHRs. This rearrangement results in membrane apposition.

The fusion process has provided a target for anti-HIV-1 drug therapy. Some HIV-1 fusion inhibitors, C-peptides (Wild et al., 1992; Jiang et al., 1993; Wild et al., 1994; Lu et al., 1995; Wild et al., 1995) for example, are synthetic peptides derived from the CHR,

which can bind the NHR of gp41 during the intermediate stage and therefore interrupt the fusion process by preventing the post-fusion stage (Eckert and Kim, 2001). Cyclic D peptides, which target a hydrophobic picket on the NHR of gp41, inhibit HIV-1 entry in the same manner (Eckert et al., 1999; Cole and Garsky, 2001). In addition, N-peptides, which are synthetic peptides derived from the NHR of gp41, can also inhibit HIV-1 entry (Fig. 1.2B) (Wild et al., 1992; Lu et al., 1995). There are two possible mechanisms. First, synthetic N-peptides may target an exposed region in the CHR of gp41. Second, synthetic N-peptides may intercalate with NHR of gp41, forming a heterotrimeric coiled coil and inhibiting the formation of NHR coiled coil (Wild et al., 1992).

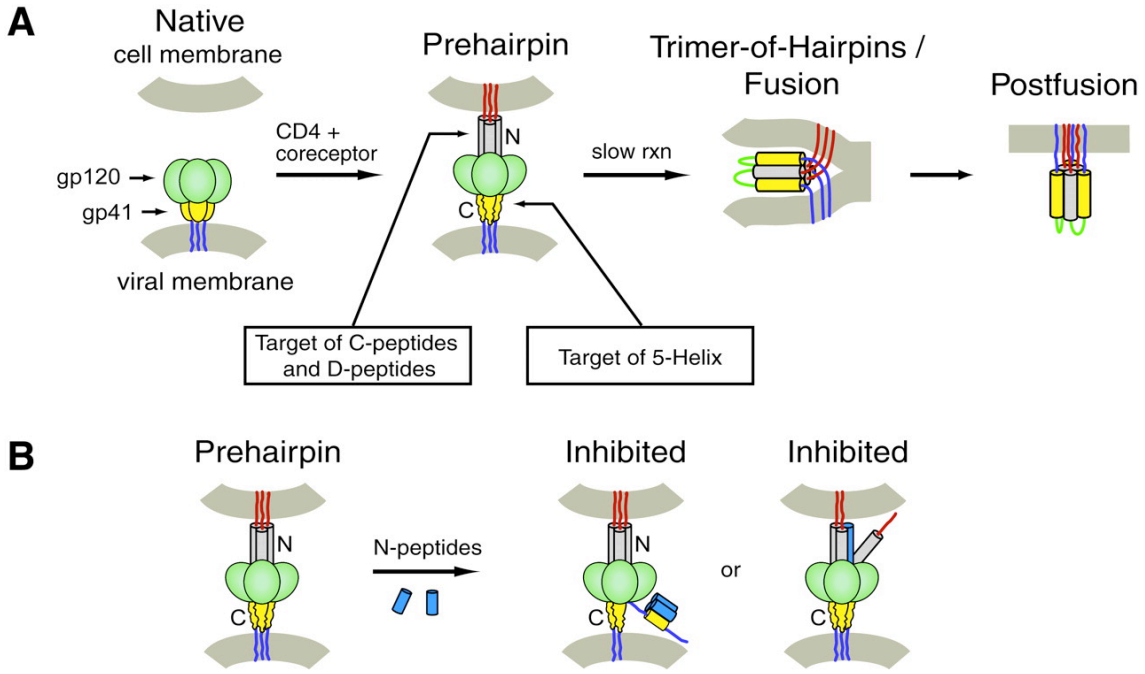


Fig. 1.2 (A) One of the current models of HIV-1 membrane fusion and its inhibition. (B) Representation of two possible mechanisms for synthetic N-peptides inhibitory activity as proposed by Eckert and Kim (Eckert and Kim, 2001). (©2001 by Proceeding of the National Academy of Sciences of the USA)

1.4 Accessibility of Epitopes on MPR During Viral Membrane Fusion

Although there are three broadly neutralizing mAbs (2F5, 4E10 and 10E8) recognizing the MPR of HIV-1 gp41, the epitopes are not always accessible for Ab binding. Frey et al. reported that mAbs 2F5 and 4E10 were not able to bind the native trimer of gp140 (the ectodomain of the precursor gp160), which indicated that epitopes of 2F5 and 4E10 are either shielded or in a nonantigenic conformation on the native gp140 trimer (Frey et al., 2008). They also detected that postfusion gp41 binds the 2F5 Fab very weekly ($K_D \approx 1.4$ μM). Therefore, the 2F5 epitope in postfusion gp41 does not have the optimal binding conformation. Frey and co-workers have designed an intermediate gp41 by adding an extra HR2 at the N-terminus of gp41 ectodomain (see gp41-inter in Fig. 1.3). The extra HR2 at the N-terminus interacted with HR1 and formed a six-helix bundle with HR1, which leaves the real HR2 at the C-terminus and MPER in the coiled coil conformation, mimicking the prehairpin intermediate during viral membrane fusion. The Fab fragment of mAb 2F5 and the single-chain Fv fragment of mAb 4E10 bind gp41-inter proteins (schematic representation shown in Fig. 1.3) very tightly ($K_D < 10$ nM). It is therefore concluded by Frey et al. that the epitopes of 2F5 and 4E10 are only exposed on a form designed to mimic a prehairpin intermediate state during viral membrane fusion.

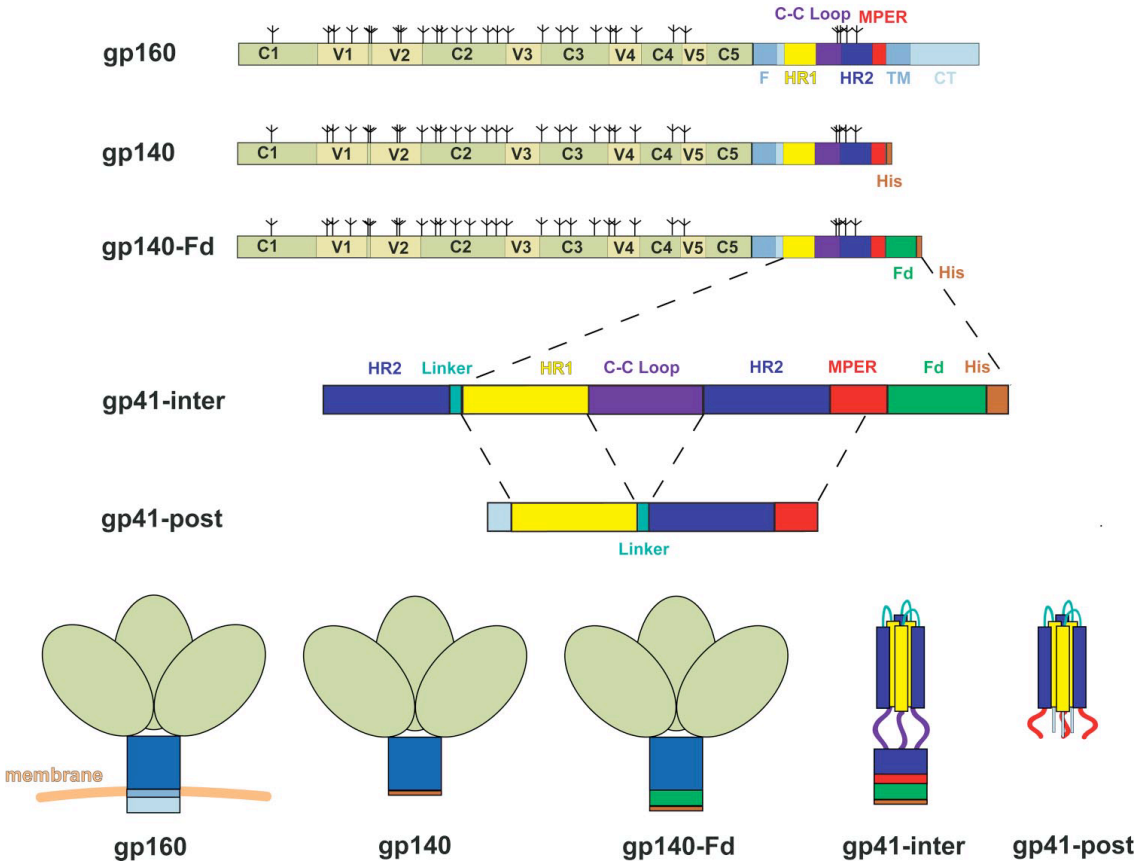


Fig. 1.3 Schematic representation of HIV-1 Env. Gp160, the full length precursor; gp140, uncleaved ectodomain of gp160 with a C-terminal His tag; gp140-Fd, uncleaved ectodomain of gp160 with a trimerization tag and a C-terminal His tag; gp41-inter, gp41 in the prehairpin intermediate conformation trapped by an N-terminal HR2 peptide and a C-terminal foldon tag; gp41-post, gp41 in the six-helix conformation with partial MPER (Frey et al., 2008). (©2008 by The National Academy of Sciences of the USA)

1.5 Current Structural Models of HIV-1 Gp41 Solved in Different Conformations

Structures of trimeric MPER have been solved in the pre-fusion conformation (Pancera et al., 2014) and post-fusion "six-helix bundle" state (Buzon et al., 2010), respectively. However, both of them could not be recognized by the broadly neutralizing antibodies 2F5 or 4E10, which is consistent to Frey et al's conclusion discussed previously. Another structure of trimeric MPER has been solved in the postfusion state (a six-helix bundle) containing a shortened NHR (HR1) region, which left MPER accessible to the 2F5 antibodies (Shi et al., 2010). More recently, Reardon et al. reported a trimeric

MPER structure solved by NMR spectroscopy in a putative pre-fusion intermediate state (a three-helix bundle) (Reardon et al., 2014). In their structure, the N termini of the MPER helices are closely associated with each other while the C termini gradually separate, which leaves space for antibody binding. However, in their construct, MPER was fused to the C terminus of a 27-residue trimerization domain from bacteriophage T4 fibritin (the foldon domain). Although Reardon et al. reported that MPER was linked to the foldon motif through a flexible linker GSSG, which is intended to minimize the effect of the structured trimerization motif on the conformation and dynamics of MPER, it is still not experimentally confirmed that MPER forms a trimer in the pre-fusion intermediate form without the effect of a trimerization motif.

Mao et al. reported a 6-Å structure of the membrane-bound HIV-1 envelope glycoprotein trimer in its uncleaved state by using cryo-electron microscopy (EM), which included the TM domain of gp41 (Mao et al., 2013). Their low-resolution structural model proposed that the TM domain of gp41 might form a left-handed three α-helical coiled coil. This is the only structural information of the TM domain of HIV-1 gp41 reported so far. The atomic structure of the gp41 TM domain is still unknown.

1.6 Crystallization of Membrane Proteins

Crystallization of membrane proteins is very challenging. With more than 100,000 protein structures in Protein Data Bank (PDB), there are only 504 unique structures of membrane proteins deposited so far (Berman et al., 2000; Berman et al., 2002). Each step in the process of crystallization of membrane proteins can be more difficult than that of soluble proteins. First, it is difficult to produce large amounts of correctly folded membrane proteins. The limited membrane surface area in the expression cell line may

not only limit the total amount of properly-folded recombinant proteins, but also may have cytotoxic consequences by competitively reducing the production of vital host membrane proteins or by negatively affecting membrane physiology (Roosild et al., 2005). Even if the target membrane proteins are expressed in large amounts, it cannot be guaranteed that the over-expressed proteins are correctly folded in the membrane. When expressed in *E. coli*, the overexpressed membrane proteins form aggregates within the cell in form of inclusion bodies (Ni da et al., 2011). As proteins in inclusion bodies are normally mistakenly folded or unfolded, they cannot be directly used for crystallization. Refolding of membrane proteins from inclusion bodies to reconstitute properties appropriate to their native counterparts is complicated (Rogl et al., 1998; Ni da et al., 2011) but has been achieved in some cases (Kiefer et al., 1999; Janovick et al., 2007; Dockter et al., 2009). Second, purification of membrane proteins is particularly delicate. The initial step in membrane protein purification is to extract and solubilize membrane proteins in a correctly folded conformation from the membrane. Detergents need to be screened to extract the target membrane protein in their native state. Special care needs to be taken during the concentration step to avoid the concentration of detergents, which may denature the protein (Wiener, 2004). In addition, the detergent/protein complex needs to be monodisperse prior to its entry into the crystallization pipeline (Prive, 2007; Prince and Jia, 2013). However, fulfillment of all these critical criteria did not necessarily yield diffracting crystals. No wonder some scientists often say crystallization is more of an art than a science (Desiraju, 2001).

1.7 Background on Recombinant Protein Expression of the Transmembrane and Membrane Proximal Domains of HIV-1 Gp41

In this dissertation, MPR-TM is expressed with different fusion proteins to help with the expression, purification and structural studies by X-ray crystallography.

The use of P8CBD (P8 is the coat protein of M13 bacteriophage and CBD is a chitin binding domain from *Bacillus circulans*) was initially tested as a protein fusion partner, which can promote the over-expression of membrane proteins by utilizing the co-translational Signal Recognition Particle (SRP) pathway to target heterogonous proteins to the *E. coli* inner membrane (Luo et al., 2009). The P8CBD fusion partner begins with the M13 bacteriophage coat protein P8, which was designed for optimal membrane targeting. P8 is followed by a chitin binding domain (CBD) from *Bacillus circulans*, which serves as an affinity tag for future purification. However, use of the P8CBD expression vector resulted in extremely poor accumulation of properly targeted MPR-TM.

Mistic, a *Bacillus subtilis* integral membrane protein that autonomously folds into the membrane, was fused to the N-terminus of MPR-TM to improve the expression yield (Gong et al., 2014). About 1 mg of monodisperse MPR-TM protein was obtained from 1 liter of cell culture, but no crystals were obtained even after extensive crystallization screens. This work is discussed in Chapter 2.

A novel construct was designed to overexpress MPR-TM as a maltose binding protein (MBP) fusion. MBP is a commonly used fusion partner, capable of improving the solubility and expression level of the target protein (Kapust and Waugh, 1999; Kobe et al., 1999; Do et al., 2014; Nguyen et al., 2014; Raran-Kurussi and Waugh, 2014). The

expression, purification, biophysical characterization and crystallization of MPR-TM as MBP fusion protein are discussed in Chapter 3 and 4.

CHAPTER 2

RECOMBINANT EXPRESSION, PURIFICATION, AND BIOPHYSICAL CHARACTERIZATION OF THE TRANSMEMBRANE AND MEMBRANE PROXIMAL DOMAINS OF HIV-1 GP41 AS MISTIC FUSION PROTEIN

Text and Figures in this chapter were reprinted with permission from Gong, Z., Kessans, S. A., Song, L., Dorner, K., Lee, H. H., Meador, L. R., LaBaer, J., Hogue, B. G., Mor, T. S., and Fromme, P. (2014) Recombinant expression, purification and biophysical characterization of the transmembrane and membrane proximal domains of HIV-1 gp41, *Protein Sci*. It has been included in the dissertation with permission from the journal.

2.1 Abstract

The transmembrane subunit (gp41) of the envelope glycoprotein of HIV-1 associates noncovalently with the surface subunit (gp120) and together they play essential roles in viral mucosal transmission and infection of target cells. The membrane proximal region (MPR) of gp41 is highly conserved and contains epitopes of broadly neutralizing antibodies. The transmembrane (TM) domain of gp41 not only anchors the envelope glycoprotein complex in the viral membrane but also dynamically affects the interactions of the MPR with the membrane. While high-resolution X-ray structures of some segments of the MPR were solved in the past, they represent the post-fusion forms. Structural information on the TM domain of gp41 is scant and at low resolution. Here we describe the design, expression and purification of a protein construct that includes MPR and the transmembrane domain of gp41 (MPR-$TM_{TEV\text{-}6His}$), which reacts with the broadly neutralizing antibodies 2F5 and 4E10 and thereby may represent an immunologically relevant conformation mimicking a prehairpin intermediate of gp41. The expression level

of MPR-$TM_{TEV\text{-}6His}$ was improved by fusion to the C-terminus of Mistic protein, yielding ~1 mg of pure protein per liter. The isolated MPR-$TM_{TEV\text{-}6His}$ protein was biophysically characterized and is a monodisperse candidate for crystallization. This work will enable further investigation into the structure of MPR-$TM_{TEV\text{-}6His}$, which will be important for the structure-based design of a mucosal vaccine against HIV-1.

2.2 Introduction

The envelope glycoprotein of the human immunodeficiency virus type 1 (HIV-1) plays essential roles in virus attachment and fusion with target cells (Hughson, 1997) and is also a primary target for vaccine design (Burton et al., 2004). It is a complex consisting of two non-covalently associated subunits that are cleaved off their precursor polyprotein to form the surface (gp120) and the transmembrane subunit (gp41) (Checkley et al., 2011). The transmembrane anchor, gp41, consists of an ectodomain (Residues 512-683, Fig. 2.1), a transmembrane domain (Residues 684-705), and a cytoplasmic domain (Residues 706-856) (Weiss, 2003; Checkley et al., 2011). Biophysical and structural studies delineate further distinct structural and functional features within the ectodomain of gp41 (Fig. 2.1) including the N-terminal and C-terminal heptad repeat regions (NHR and CHR) (Chan et al., 1997; Weissenhorn et al., 1997; Checkley et al., 2011) that are flanked by the fusion peptide (FP) and the fusion peptide proximal region (FPPR) on one side, and the membrane proximal external region (MPER) on the other side.

Target-cell infection by HIV-1 is initiated when gp120 binds to its primary receptor CD4 and coreceptor, usually CCR5 or CXCR4 (Broder and Dimitrov, 1996; Moore et al., 1997). The exact mechanism leading to virus entry is still not known. A current model proposes that following binding of gp120 to its receptors, the gp41 subunit is exposed,

triggering drastic sequential changes in its conformation culminating in fusion between the viral envelope and the target cell's plasma membrane (Gallo et al., 2003; Melikyan, 2008; Ashkenazi and Shai, 2011). According to this model, NHR and CHR of gp41 are partially shielded by gp120, and switch to an extended conformation upon the latter's removal to allow insertion of the FP into the target cell's membrane (Liu et al., 2009; Buzon et al., 2010; Shi et al., 2010). After the fusion of the juxtaposed viral and cellular membranes, gp41's core regains the 6-helical bundle conformation of post-fusion complex (Chan and Kim, 1998).

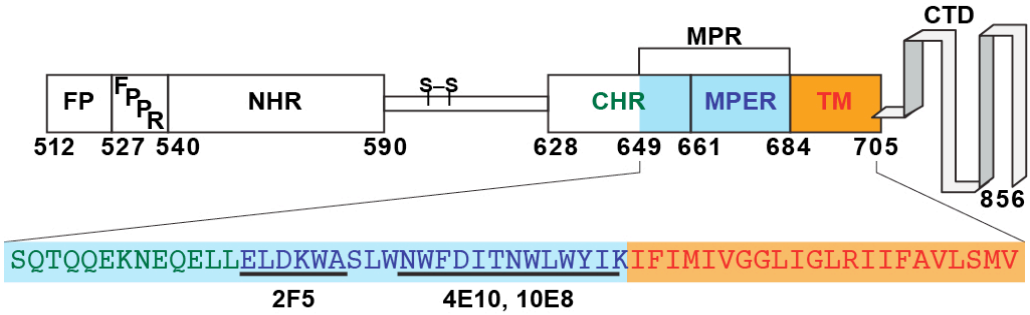


Figure 2.1. Schematic diagram of HIV-1 gp41. FP (Residues 512-527): fusion peptide; FPRR (Residues 528-539): fusion peptide proximal region; NHR (Residues 540-590): N-terminal heptad repeat region; S-S: a disulfide linkage; CHR (Residues 628-661): C-terminal heptad repeat region; MPER (Residues 662-683): membrane proximal external region; MPR (Residues 649-683): membrane proximal region; TM (Residues 684-705): transmembrane domain; CTD (Residues 706-856): cytoplasmic domain. 2F5, 4E10 and 10E8 are epitopes for three broadly neutralizing mAbs.

In addition to its well-recognized role in infection of target $CD4^+$ cells, gp41 is instrumental during early steps in various processes leading to mucosal transmission of the virus (Wu, 2008; Bomsel et al., 2011; Shen et al., 2011; Puryear and Gummuluru, 2013). The virus is utilizing several routes to cross epithelial surfaces including capture by dendritic and Langerhans cells prevailing in pluristratified epithelia and transcytosis that is particularly important in simple epithelia (Ganor and Bomsel, 2011).

A key player in the transcytosis process is a region of the gp41 corresponding to residues 649-683, which includes the MPER and part of the CHR and will be referred to here in correspondence with Matoba *et al.* (Matoba et al., 2006) as "MPR". Transcytosis is initiated when the MPR binds to the glycosphingolipid galactosyl ceramide (GalCer) and the co-receptor CCR5 (Alfsen and Bomsel, 2002; Meng et al., 2002; Puryear and Gummuluru, 2013). GalCer is enriched at the apical membrane of epithelial cells (Simons and van Meer, 1988) and is involved in the establishment of lipid rafts (Brown and London, 1998; Rietveld and Simons, 1998), which are proposed to act as platforms for HIV-1 entry (Popik et al., 2002), transcytosis (Alfsen and Bomsel, 2002), virion assembly and budding (Hollmann et al., 2013).

The minimal region of gp41 that can bind GalCer is the MPR. This region, together with the adjacent transmembrane domain, is the most highly conserved element of the envelope protein (Alfsen et al., 2001; Alfsen and Bomsel, 2002; Miyauchi et al., 2005; Checkley et al., 2011). The MPR is the target of secretory IgAs that can be found in mucosal secretions of highly exposed, persistently seronegative individuals and may constitute one of very few potential correlates of protection against HIV-1 infection (Mayr et al., 1978; Devito et al., 2000; Devito et al., 2000; Miyazawa et al., 2009; Tudor et al., 2009; Trabattoni et al., 2012). These mucosal antibodies (Abs) were shown to possess anti-HIV responses including neutralization and blocking of transcytosis (Devito et al., 2000; Tudor et al., 2009). Significantly, epitopes within the MPER are recognized by three of only a handful of broadly neutralizing monoclonal Abs (mAbs) characterized thus far. Among them are 2F5 (Purtscher et al., 1994), 4E10 (Zwick et al., 2001) and more recently 10E8 (Huang et al., 2012). These mAbs also have other anti-HIV-1

activities including transcytosis-blocking (Shen et al., 2010) and Fc-mediated cytotoxicity (Tudor and Bomsel, 2011) and were shown to provide full protection against mucosal challenge when delivered intravenously (Baba et al., 2000; Mascola et al., 2000; Hessell et al., 2010). These attributes make the MPR a particularly interesting target for the development of a prophylactic vaccine against HIV-1 (Matoba et al., 2004; Matoba et al., 2008; Matoba et al., 2009; Bomsel et al., 2011; Burton et al., 2012; Leroux-Roels et al., 2013).

Consequently, elucidation of the structure of the MPR is of interest and importance as it will instruct mucosal vaccine design against HIV-1. Liu *et al*. have reported that a portion of the MPER (Residues 662-683) forms a parallel three-stranded coiled coil stabilized by the addition of a C-terminal isoleucine zipper motif (Liu et al., 2009). However, it is difficult to determine whether MPER forms the three-stranded coiled coil by itself or if this MPER is forced into this conformation by its attachment to the three helix bundles with the isoleucine zipper, which forms a classic coiled coil (Harbury et al., 1994). Instead of using an artificial motif to stabilize MPR, we want to study the structure of MPR together with its native transmembrane domain, which, similarly to the MPR, is highly conserved (Checkley et al., 2011).

Studies on the function of the transmembrane domain of HIV-1 gp41 are limited. The transmembrane domain of HIV-1 gp41 plays an important role in anchoring the glycoprotein envelope complex into the viral membrane and is also crucial for its biological function in fusion and virus entry (Shang et al., 2008; Yue et al., 2009; Shang and Hunter, 2010; Montero et al., 2012). Mao *et al.* recently obtained a 6-Å structure of the membrane-bound HIV-1 envelope glycoprotein trimer in its uncleaved state by using

cryo-electron microscopy (EM), which included the transmembrane domain of gp41 (Mao et al., 2013). Their low-resolution structural model proposed that the transmembrane domain of gp41 might form a left-handed three α-helical coiled coil, with a crossing angle of about 35°. This is the only structural information of the transmembrane domain of gp41 reported so far. The structure of MPR-TM at higher resolution is still needed for the structure-based design of a vaccine against HIV-1.

There are two bottlenecks in membrane protein structure determination: high-yield membrane protein production and crystallization. There are two major reasons that account for the difficulty in producing large amounts of correctly-folded membrane proteins in bacteria. Most eukaryotic membrane proteins are inserted into the membrane in a process, which combines translation, targeting, folding and post-translational modifications. Despite some similarities and homologous elements, the membrane-targeting pathways in bacteria are different enough to require engineering of eukaryotic genes to optimize their expression and accumulation. In addition, the limited membrane surface area in *Escherichia coli* may not only limit the total amount of properly-folded recombinant proteins made in this system, but also may have cytotoxic consequences by competitively reducing the production of vital host membrane proteins or by negatively affecting membrane physiology (Roosild et al., 2005). Mistic is a *Bacillus subtilis* integral membrane protein that autonomously folds into the membrane. It acts as a targeting signal and can be used for over-expression of other membrane proteins in their native conformations (Roosild et al., 2005). In our study, we developed an expression and purification strategy of MPR-$TM_{TEV\text{-}6His}$ fused to the C-terminus of Mistic. Surface plasmon resonance (SPR) measurements and ELISA experiments were carried out to test

if the epitope on the purified MPR-$TM_{TEV-6His}$ was exposed and could be recognized by the broadly neutralizing mAb 2F5 and 4E10. The purified MPR-$TM_{TEV-6His}$ was also biophysically characterized by size exclusion chromatography (SEC), MALDI-TOF MS, CD spectroscopy and dynamic light scattering (DLS).

2.3 Materials and Methods

2.3.1 Cloning, Bacterial Strains and Growth Conditions

The MPR-$TM_{649-705}$ construct is based on a deconstructed HIV-1 gp41 (dgp41) gene (GenBank Accession number JX534518) (Kessans et al., 2013), a chimera comprising the gp41 MPR derived from the B-clade MN isolate (GenBank accession number AF075722) and the transmembrane domain and cytoplasmic tail region of the C clade 1084i isolate (GenBank accession number AY805330). A more detailed description can be found in the Supporting information. Briefly, two tobacco etch virus (TEV) protease recognition sites were added to flank the coding sequence of MPR-$TM_{649-705}$ and the construct was introduced into Gateway entry vector pCR8/GW/TOPO (Invitrogen). The construct was then fused to the *B, subtilis* Mistic protein by recombination into the Gateway destination vector pMistic (DNASU: pMIS2.1mv), which was a kind gift of Dr. Mark Vega, Salk Institute. For expression, the recombinant plasmid pMistic-MPR-$TM_{649-705}$ was transformed into *E. coli* C41 (DE3). Cell culture growth conditions and recombinant protein induction are described in the Supporting information.

2.3.2 Purification of MPR-$TM_{TEV-6His}$

The purification protocol is described in details in the Supporting information and will be only outlined here. Harvested cells were lysed with a microfluidizer (Microfluidics). Water-soluble proteins were separated from membrane proteins (and

other water-insoluble material) by centrifugation and discarded. The Mistic-MPR-$TM_{TEV-6His}$ fusion protein was extracted by resuspending the pellet in ice-cold extraction buffer (PBS, 1% βDDM and protease inhibitor cocktail) and incubation with gentle shaking for 3 h at 4 °C. Following centrifugation, the supernatant was collected and the protein was purified by TALON metal affinity chromatography (Clontech Laboratories, see Supporting information for details). The eluate was dialyzed (2000 Da cut-off dialysis tube, Sigma) overnight at 4 °C against 20 mM NaCl, 20 mM HEPES, pH 7.5. After dialysis, Tris-HCl pH 8.0, EDTA and DTT were added to the dialyzed protein (final concentrations: 50 mM, 0.5 mM and 1 mM, respectively). Protein preparation was proteolytically digested with TEV (Invitrogen, protease:substrate ratio of 20 U/182 μg, 2 h at room temperature) resulting in >95% cleavage of Mistic-MPR-$TM_{TEV-6His}$.

The cleaved MPR-$TM_{TEV-6His}$ protein preparation was further purified by anion exchange chromatography using ÄKTApurifier 10 (GE Healthcare) and a Mono Q 5/50 GL column (see Supporting information for details). Fractions containing cleaved MPR-$TM_{TEV-6His}$ were pooled together for further biochemical and biophysical analysis.

2.3.3 Size Exclusion Chromatography

The purified MPR-$TM_{TEV-6His}$ preparations were characterized by SEC using ÄKTApurifier 10 (GE Healthcare) and a Superdex 200 10/300 GL column (24 mL bed volume, GE Healthcare). The CD spectroscopy-compatible mobile phase was 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5, 0.02% βDDM as detailed in the Supporting information.

2.3.4 Protein Determination, SDS-PAGE, Immunoblotting and ELISA

Protein determination in crude and enriched preparations was carried out by the modified Lowry assay (Markwell et al., 1978). Protein concentration of pure preparations

of MPR-$TM_{TEV\text{-}6His}$ was determined by measuring A_{280} (ε = 32,290 $cm^{-1}M^{-1}$, obtained using Peptide Property Calculator at http://www.basic.northwestern.edu/ biotools/proteincalc.html).

Proteins were separated by SDS-PAGE (Schägger, 2006) and were subjected to silver staining (Lawrence et al., 2011) or to immunoblotting (see Supporting information for details) (Kessans et al., 2013). The authors thank the NIH AIDS Research and Reference Reagent Program (Divisions of AIDS, NIAID, NIH) for donation of the mAbs 2F5 (catalog number 1475) and 4E10 (catalog number 10091). Chemiluminescence was detected using the BioSpectrum 500 C Imaging System (Ultra-Violet Products Ltd). ELISA was performed essentially as previously described (Matoba et al., 2008) and as detailed in the Supporting information.

2.3.5 MALDI-TOF MS, CD Spectroscopy and DLS

We used MALDI-TOF MS (Applied Biosystems) to accurately measure the molecular weight of the purified MPR-$TM_{TEV\text{-}6His}$ protein as detailed in the Supporting information.

A JASCO J-710 CD spectropolarimeter was used for measuring the CD spectra of purified sample and the procedure, detailed in the Supporting information, was essentially according to Greenfield (Greenfield, 2006). Data analysis was performed using the CONTINLL program in CDPro software package by comparing the measured data with reference set option 10, which included 13 membrane proteins along with 43 soluble proteins (Sreerama and Woody, 2000).

DynaPro NanoStar M3300 from Wyatt Technology was used to carry out DLS measurements in the same buffer used for CD spectroscopy (see the Supporting information for details).

2.3.6 Surface Plasmon Resonance (SPR)

All experiments were performed on a K_X5 Surface Plasmon Resonance Imaging (SPRi) System (Plexera). The K_X5 SPRi procedure was previously described (Song et al., 2013) and is detailed in the Supporting information. Preparation of custom SPRi chips is described in the Supporting information. We used gold chips coated with covalently linked Protein A/G that allowed immobilization of the test Abs through their Fc region, ensuring unimpeded interactions with antigens. To prevent nonspecific adsorption, the chip was blocked with BSA (5 mg/mL) before further analysis. The running buffer and dilution buffer of the analyte was 1xPBS containing 0.02% βDDM. In sequential runs, CTB (the negative control) at 850 n*M*, CTB-MPR at 600 n*M* and MPR-$TM_{TEV\text{-}6His}$ at 840 n*M* were passed over the ligand surface at a flow rate of 1 μL/s, with a 300-s association and a 600-s dissociation. The chip was regenerated between runs with H_3PO_4 (1 : 200 of 85% w/w) for 100 s followed by recoating with the desired antibody. Identical injections over blank protein A/G surfaces were subtracted from the data for kinetic analysis. SPRi data consisting of video images at 1-s resolution were analyzed with Data Analysis Module software from Plexera. The binding curve was analyzed and fitted with 1:1 interaction model with Scrubber 2 software (Biologic Software).

2.4 Results and Discussion

2.4.1 Cloning and Expression of MPR-TM$_{649-705}$

The membrane proximal region (MPR) of HIV-1 gp41 is important for the design of a mucosal vaccine against HIV-1. The transmembrane (TM) domain of HIV-1 gp41 plays an essential role in anchoring the envelope complex into the viral membrane and is also crucial for its biological function in fusion and virus entry (Shang et al., 2008; Yue et al., 2009; Shang and Hunter, 2010). Bacterial expression of these two hydrophobic domains of HIV-1 has proved to be difficult and previous experiments in our laboratories making use of the P8CBD expression vector (Luo et al., 2009) resulted in extremely poor accumulation of properly-targeted MPR-TM (Gong, Kessans, Fromme and Mor, unpublished). In our study, the portion of the HIV-1 Env gene encoding for MPR-TM was cloned into the expression vector pMIS2.1mv to obtain pMistic-MPR-TM$_{TEV-6His}$ (Fig. 2.2A). This vector directs the tightly regulated expression of a C-terminal translation fusion between the *B. subtilis* integral membrane protein Mistic and MPR-TM$_{TEV-6His}$ in *E. coli* (Fig. 2.2B, 2.2C). Mistic was previously shown to improve as a translational-fusion partner the expression and accumulation levels of several membrane proteins in their native conformations (Roosild et al., 2005). To allow the removal of the Mistic fusion partner prior to future crystallization experiments, two tobacco etch virus (TEV) protease recognition sites (Parks et al., 1994) were introduced by PCR primers. One TEV protease recognition site was introduced at the N-terminus of MPR-TM$_{649-705}$ and the other was located at the C-terminus (Fig. 2.2B, 2.2C). The recombinant plasmid pMistic-MPR-TM$_{TEV-6His}$ was transformed into *E. coli* C41 (DE3) cells for expression.

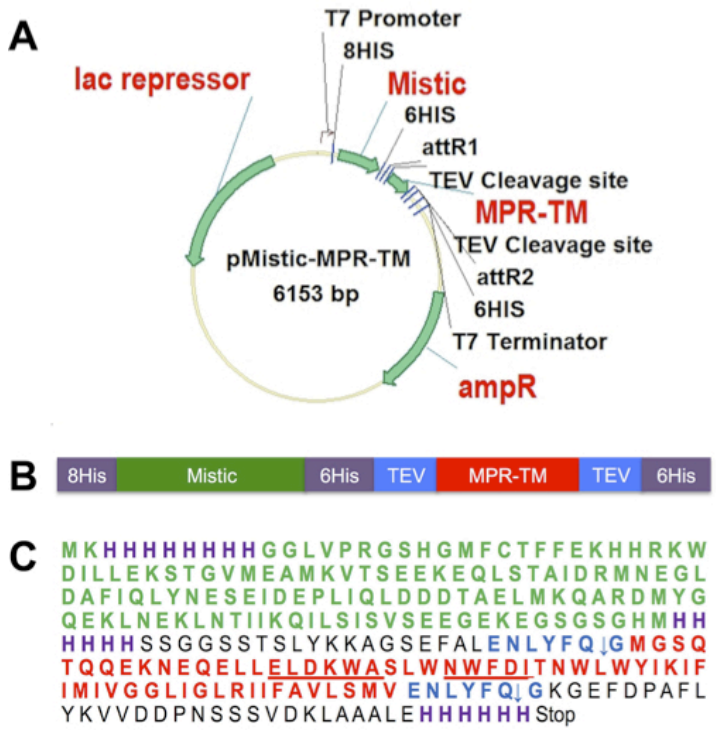


Figure 2.2 (A) Construction of the expression vector for pMistic-MPR-TM$_{TEV-6His}$. See text and Supporting information for details. (B) Scheme of the Mistic-MPR-TM$_{TEV-6His}$ fusion protein. (C) Amino acid sequence of Mistic-MPR-TM$_{TEV-6His}$. Purple: His-tag; Green: Mistic; red: MPR-TM; underlined ELDKWA: the mAb 2F5 core epitope; underlined NWFDI: the 4E10 mAb core epitope; blue: TEV recognition sites. Note that the cleavage occurred between Q and G residues of the TEV recognition sequence ENLYFQG.

2.4.2 Purification of MPR-TM$_{TEV-6His}$

After cells were lysed by microfluidization, the majority of the fusion protein Mistic-MPR-TM$_{TEV-6His}$ was found in the membrane fraction, and a purification protocol was developed to allow efficient purification without compromising the structural integrity of the protein (Fig. 2.3). Following extensive screening of various detergents (data not shown), βDDM at 1% was used to extract the fusion protein from the membrane. The βDDM extract was subjected to TALON metal affinity chromatography to separate His-tagged Mistic-MPR-TM$_{TEV-6His}$ from other proteins (Porath et al., 1975; Hochuli et al., 1987). The second elution fraction contained the majority of His-tagged

proteins, which showed a complex pattern of banding upon SDS-PAGE fractionation (Fig. 2.4). The very top band (marked with blue arrows) corresponded to the fusion protein Mistic-MPR-$TM_{TEV\text{-}6His}$, with a molecular mass of 31 kDa, in good concordance with its calculated expected size. Four contaminant bands were visible. The second band from the top on the silver-stained SDS-PAGE in Figure 4 corresponded to a His-tagged fragment of Mistic and MPR-$TM_{TEV\text{-}6His}$ as it could be detected in immunoblots by the MPER-specific mAb 2F5 (Fig. 2.4). In contrast, the three lower bands, clearly visible on the silver-stained gels, did not react with the 2F5 Ab (Fig. 2.4) and could be His-tagged fragments of Mistic without MPR or unrelated *E. coli* proteins. We observed that degradation was a common problem for Mistic fusion constructs (data not shown).

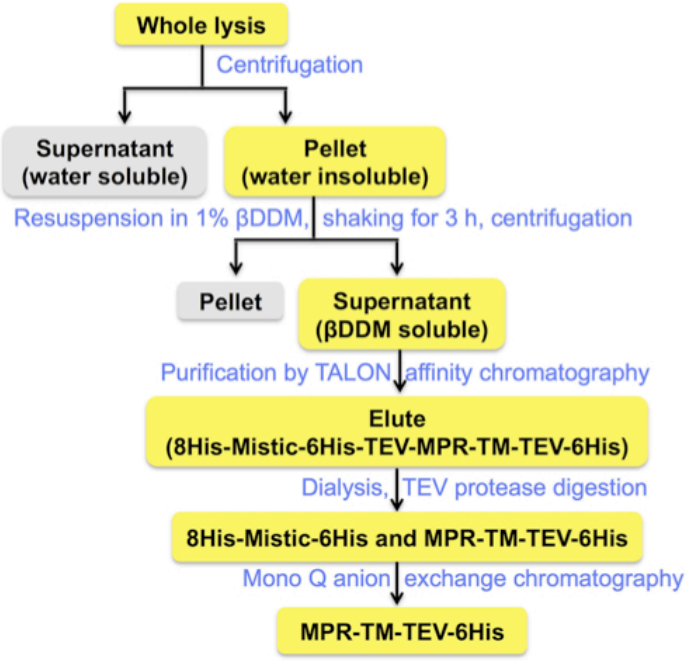


Figure 2.3. MPR-$TM_{TEV\text{-}6His}$ purification scheme.

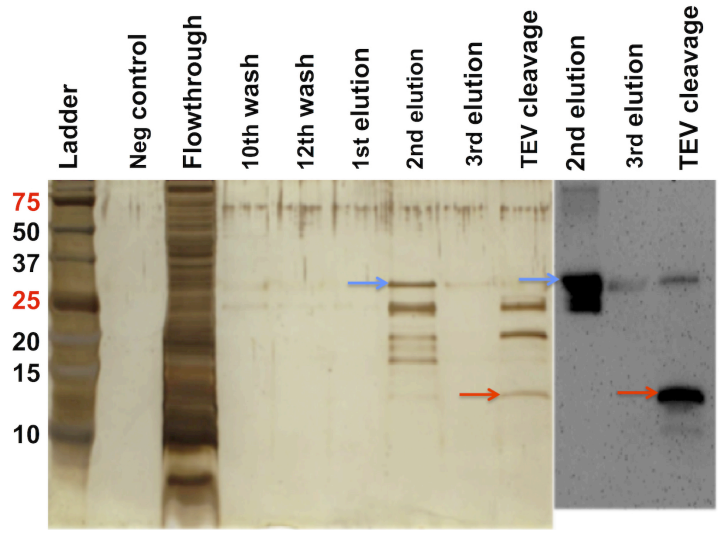


Figure 2.4. Purification of Mistic-MPR-TM$_{TEV\text{-}6His}$ by metal affinity chromatography and its cleavage by TEV protease. Fractions were resolved by SDS-PAGE and visualized by silver staining (left) or immunoblotting with the mAb 2F5 (right). Blue arrows: Mistic-MPR-TM$_{TEV\text{-}6His}$. Red arrows: cleaved MPR-TM$_{TEV\text{-}6His}$.

The next step in our purification scheme (Fig. 2.3) was the specific cleavage of the fusion protein followed by anion exchange chromatography to separate the MPR-TM$_{TEV\text{-}6His}$ protein from its Mistic fusion partner. The TALON column eluates were dialyzed to remove the imidazole and the ionic conditions of the buffer were adjusted to ensure efficient cleavage by the TEV protease. We expected the fully-processed MPR-TM protein to have a molecular weight of 7.8 kDa, if both the N-terminal and C-terminal TEV recognition sites were to be cleaved. However, extensive digestion by TEV yielded a protein band (indicated by red arrows in Fig. 2.4) with apparent molecular mass greater than 10 kDa that cross-reacted with the MPR-specific mAb 2F5. Moreover, subjecting the cleavage products to a second TALON purification step, demonstrated that this >10 kDa cleavage protein retained a functional His-tag that allowed its efficient binding to the column and required high concentration of imidazole (250 mM) for elution (data not shown). Our results are therefore compatible with the lack of TEV cleavage at its C-

terminal site of the protein. The final product therefore consists of the MPR-TM with its C-terminal His-tag still attached (called hereafter "MPR-$TM_{TEV\text{-}6His}$"). The expected molecular mass of this polypeptide was 11.9 kDa. Lack of cleavage at the C-terminal TEV recognition site could be explained by its proximity to TM domain, which is likely to be fully embedded in the detergent micelle, thereby the hydrophilic sugar heads of the βDDM molecules may obscure the TEV cleavage site or may otherwise impede the enzyme's proteolytic activity.

Retention of the C-terminal His-tags on the MPR-$TM_{TEV\text{-}6His}$ protein interfered with our original plan to separate it from the other cleavage products containing the Mistic protein by the second TALON metal-affinity chromatography step. Instead, we turned to size-exclusion chromatography (SEC) and experimented with two types of SEC columns, Superdex 75 10/300 GL and Superdex 200 10/300 GL, to separate MPR-$TM_{TEV\text{-}6His}$ from $_{8His}Mistic_{6His}$ and other degradation products, but MPR-$TM_{TEV\text{-}6His}$ could not be purified by either of the two columns (data not shown).

Finally, Mono Q anion exchange chromatography was used to further purify the MPR-$TM_{TEV\text{-}6His}$ protein. The optimal conditions for Mono Q ion exchange chromatography were chosen based on several prior small-scale tests. MPR-$TM_{TEV\text{-}6His}$ eluted as the main elution peak (A9 in Fig. 2.5A) while $_{8His}Mistic_{6His}$ and other degradation proteins were in the flowthrough (A1 in Fig. 2.5A) and a shoulder peak (A 10 and A11 in Fig. 2.5A) of the main elution peak. The protein elution was monitored at 280 nm and the fractions were analyzed by SDS-PAGE (Fig. 2.5B). The band labeled by red arrowheads corresponds to MPR-$TM_{TEV\text{-}6His}$ while the band pointed out by green arrowheads corresponds to the cleaved $_{8His}Mistic_{6His}$. Fraction A9 was overloaded so it

was very difficult to determine if the strong band was MPR-$TM_{TEV\text{-}6His}$ only or the mixture of $_{8His}Mistic_{6His}$ and MPR-$TM_{TEV\text{-}6His}$. Fraction A9 was diluted 10 times and reanalyzed by SDS-PAGE (Fig. 2.5C). The result shown in Fig. 2.5C confirms that fraction A9 only contains MPR-$TM_{TEV\text{-}6His}$. The protein preparations corresponding to fraction A9 were used for future analysis.

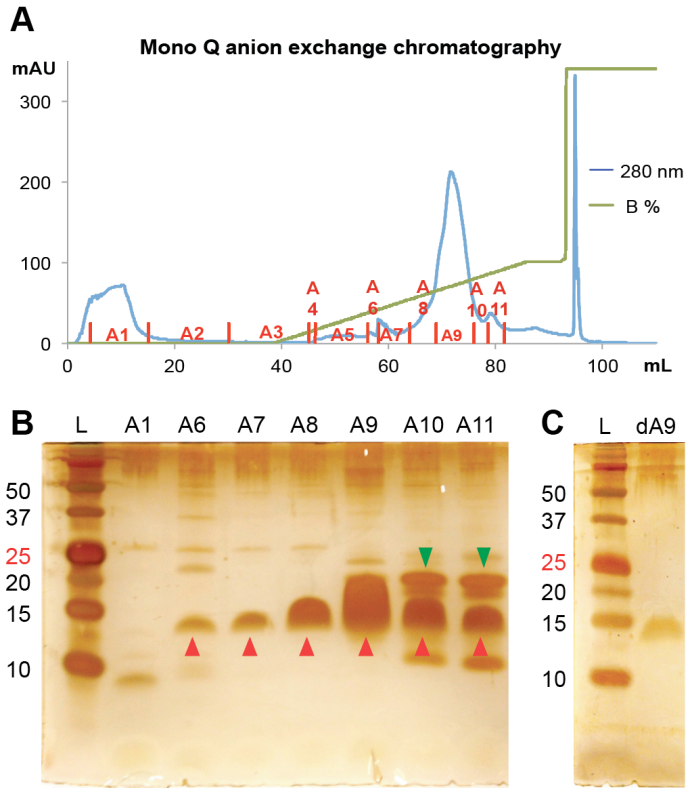


Figure 2.5. Separation of Mistic-MPR-$TM_{TEV\text{-}6His}$ cleavage products by Mono Q anion exchange chromatography. (A) Chromatogram. (B) Fractions were resolved by SDS-PAGE and visualized by silver staining. Red arrowheads: MPR-$TM_{TEV\text{-}6His}$; green arrowheads: $_{8His}Mistic_{6His}$. (C) SDS-PAGE analysis of 10-fold diluted fraction A9 (dA9). L: ladder.

The presence of the uncleaved C-terminal tail containing the C-terminal TEV recognition site, an attR2 site and a 6His-tag (Fig. 2.2C) may raise the concern that it might affect future crystallization of MPR-$TM_{TEV\text{-}6His}$. However, this seems unlikely as the protein data bank (PDB) contains several examples of high-resolution structures

containing such an artificial protein domain including a 1.55 Å crystal structure of the thioredoxin domain of human thioredoxin-like protein 2 (PDB: 2WZ9).

2.4.3 Purified MPR-$TM_{TEV\text{-}6His}$ is Folded and Stable

Circular dichroism (CD) spectroscopy was used to estimate the secondary structural content of MPR-$TM_{TEV\text{-}6His}$. The CD spectra of MPR-$TM_{TEV\text{-}6His}$ displayed one positive band at 195 nm and two negative bands at 208 nm and 222 nm (Fig. 2.6), which is characteristic of α-helical proteins (Greenfield, 2006). Data analysis with CONTINLL in the CDPro software package (Sreerama and Woody, 2000) produced an estimation of 52.2% α helices, 6.3% β sheets, 14.9% turns and 26.5% random coils and the root-mean-square deviation (RMSD) value was 0.055. Estimation with CONTINLL was similar to the secondary structure prediction with server APSSP2 (Raghava, 2002), which predicted 59.6% α helices. The 2-Å crystal structure of the $gp41_{528\text{-}683}$ indicated that the MPR (Residues 649 to 683) might form an α-helix (Buzon et al., 2010). As reported by Mao et al. (Mao et al., 2013), The 6-Å cryo-EM structure of the uncleaved HIV-1 envelope glycoprotein trimer suggests that the transmembrane domain (Residues 684- 705) of gp41 might be an α-helix as is generally assumed (but see Steckbeck and co-workers for a different view (Steckbeck et al., 2013)). Please note that the Env structure proposed by Mao et al. (Mao et al., 2013) differs from other recent structures based on crystallographic and EM studies, and some aspects of that model remain controversial (Buzon et al., 2010; Shi et al., 2010; Lyumkis et al., 2013). In particular, the organization of secondary structure elements in the gp41 ectodomain of the structure proposed by Mao et al. (Mao et al., 2013) is dramatically different from that of the six α-helix bundles seen in the post-fusion form. Specifically, according to Mao et al. (Mao et al., 2013), the NHR

and CHR domains are broken into eight short α-helices in striking opposition to six helical bundle. In any event, the CD spectra indicate that MPR-$TM_{TEV\text{-}6His}$ is folded after purification.

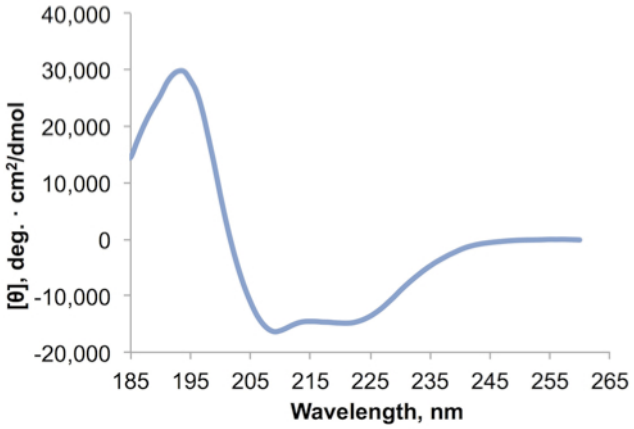


Figure 2.6. CD spectrometry demonstrated that purified MPR-$TM_{TEV\text{-}6His}$ was α-helical. With one positive band at 195 nm and two negative bands at 208 nm and 222 nm, the CD spectra of the protein were typical for α-helical proteins

We used dynamic light scattering (DLS) to further demonstrate that purified *E. coli*-derived MPR-$TM_{TEV\text{-}6His}$ is stably folded and monodisperse as these are critical factors affecting crystallization (Ferré-D'Amaré and Burley, 1997). DLS is a technique that is very sensitive for the detection of aggregates. Our DLS analysis demonstrated that purified MPR-$TM_{TEV\text{-}6His}$ was monodisperse, with polydispersity of the protein-detergent complex estimated to be only 12.3% (Fig. 2.7 and Supporting Table 2.1). Moreover, we have used DLS to monitor the stability of purified MPR-$TM_{TEV\text{-}6His}$ protein at 4 °C over time. Our result shows that MPR-$TM_{TEV\text{-}6His}$ remains monodisperse for at least 10 days at 4 °C (Fig. 2.7 insert, Supporting Fig. 2.1, page 48). The DLS results indicate that MPR-$TM_{TEV\text{-}6His}$ is a monodisperse and stable candidate for crystallization.

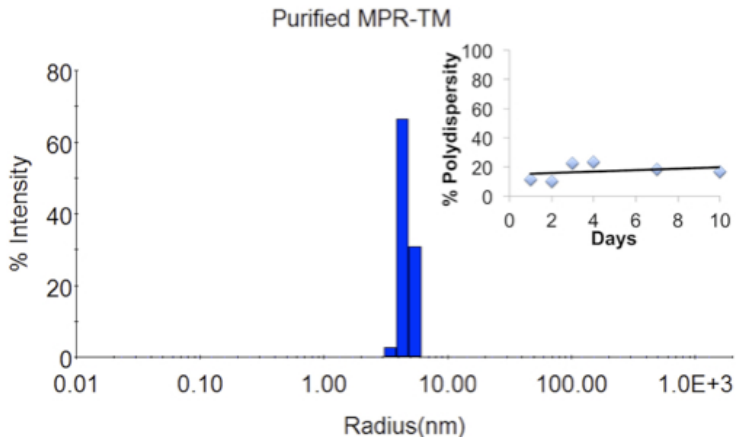


Figure 2.7. DLS demonstrated that purified MPR-$TM_{TEV\text{-}6His}$ was highly monodisperse. Sample contained MPR-$TM_{TEV\text{-}6His}$ (0.22 mg/mL) in 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5, 0.02% βDDM. Insert: the purified MPR-$TM_{TEV\text{-}6His}$ was stored at 4 °C and measured by DLS at day 1, 2, 3, 4, 7 and 10 respectively. The polydispersity remained below 25% (please refer to Supporting Table 2.2 and Supporting Figure 2.1 for details), which indicated that purified MPR-$TM_{TEV\text{-}6His}$ was monodisperse for at least 10 days at 4 °C.

2.4.4 Molecular Mass of MPR-$TM_{TEV\text{-}6His}$ and Its Oligomeric State

We used MALDI-TOF MS to determine the accurate molecular mass of MPR-$TM_{TEV\text{-}6His}$ and the resulting spectrum revealed a protein peak of 11,874 ± 4 Da (Fig. 2.8). This agreed very well with the theoretical molecular weight of MPR-$TM_{TEV\text{-}6His}$, which was predicted to be 11,872 Da based on the sequence (Fig. 2.2) and the SDS-PAGE analysis (Fig. 2.4). The shoulder peak at 12,138 Da could be assigned to the complex of MPR-$TM_{TEV\text{-}6His}$ and the matrix sinapinic acid whose molecular weight was 224 Da.

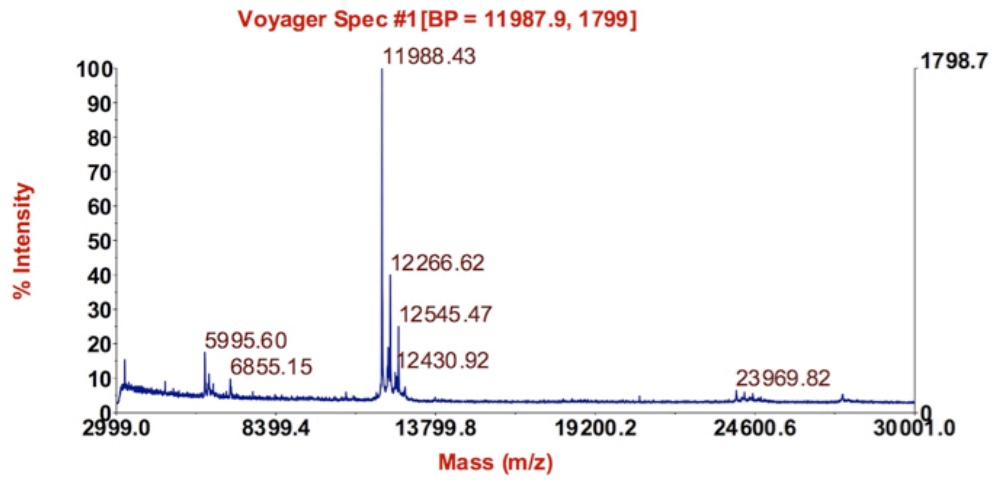


Figure 2.8. MALDI-TOF spectra of MPR-$TM_{TEV\text{-}6His}$ were in perfect agreement with the calculated molecular mass of the protein at 11,872 Da.

The MS results confirm the predicted molecular mass of the MPR-$TM_{TEV\text{-}6His}$. Proteins are denatured by MALDI using sinapinic acid and do not provide information about the oligomeric information. It was, therefore, of interest to check the oligomerization state of the purified protein. Our DLS results showed that the Stokes radius of the detergent-protein complex was 4.68 nm, which corresponded to a molecular mass of 124 kDa (Supporting Table 2.1). This indicates that MPR-$TM_{TEV\text{-}6His}$ polypeptides form a larger complex consisting of several monomeric subunits. However, the oligomeric state of the complex is difficult to determine, as it exists in the form of protein-detergent complex.

We have used analytical SEC to verify the quaternary structure of purified MPR-$TM_{TEV\text{-}6His}$ and to provide additional estimate as to its molecular mass and its oligomeric state in its detergent-solubilized state (Fig. 2.9). The SEC chromatogram revealed a single symmetric peak that eluted at 13.70 mL (Fig. 2.9A). The molecular weight of this peak was ~123 kDa, calculated based on the standard curve (Fig. 2.9B). This size corresponds to the MPR-$TM_{TEV\text{-}6His}$ oligomer embedded into a βDDM micelle and was very similar to

the molecular weight estimation obtained by DLS (124 kDa, Supporting Table 2.1 and Fig. 2.7).

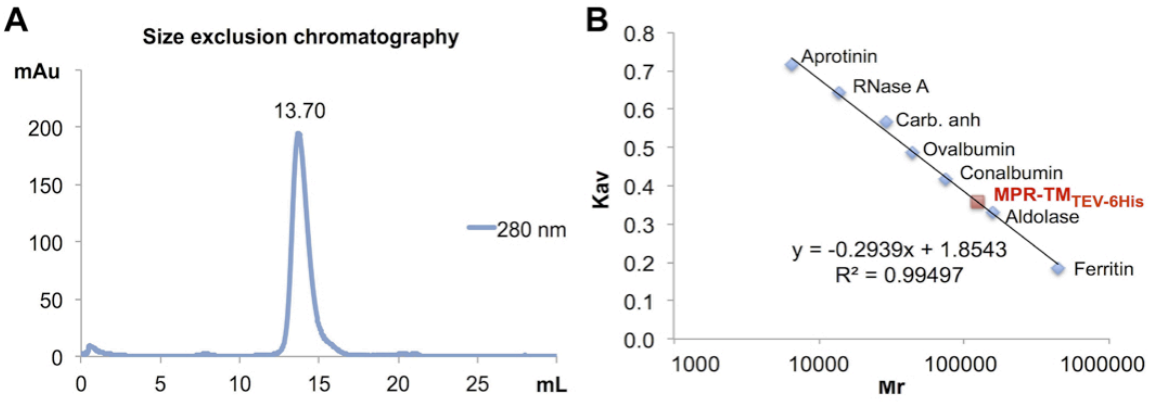


Figure 2.9. Estimation of the molecular mass of purified native MPR-$TM_{TEV\text{-}6His}$. (A) SEC of MPR-$TM_{TEV\text{-}6His}$ revealed a single symmetric peak eluted at 13.7 mL corresponding to ~123 kDa based on the standard curve. (B) The standard curve of Superdex 200 10/300 GL column using the following standard proteins: aprotinin (6.5 kDa), RNase A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), aldolase (158 kDa) and ferritin (440 kDa). Kav is the partition coefficient, which can be calculated as following: (elution volume – void volume) / (total volume – void volume).

The excellent correspondence of the DLS and SEC data supported our aim to use the information to assess the oligomeric structure of MPR-$TM_{TEV\text{-}6His}$. The average molecular weight of "empty" βDDM micelles measured by DLS was 68 kDa (data not shown). However, detergent shell in the protein-detergent complex is often larger than the empty micelle because the detergent must cover all the hydrophobic surface of the protein. Therefore, the estimation of the oligomeric state of the protein inside the micelle is complicated. It is likely that the presence of embedded proteins should change the expected size of the micelle. Other estimates published in the past indicated that detergent to protein ratio values rang 2.4-3.5 (g/g) (Butler et al., 2004; Bamber et al., 2006). A trimer embedded in βDDM micelles could therefore show an apparent molecular mass in the range of 121-160 kDa. In either case, the values for a dimer (81 – 107 kDa) would be

lower than the observed value. While the definitive subunit composition of the MPR-$TM_{TEV\text{-}6His}$ is hard to resolve at this stage, it is clear that the protein is oligomeric, an important structure-function attribute of the native gp41 molecule.

2.4.5 Purified MPR-$TM_{TEV\text{-}6His}$ is Recognized by the Broadly Neutralizing mAbs 2F5 and 4E10

An important structure-function attribute of gp41 is its ability to bind to broadly neutralizing mAbs. It was therefore of great importance to test if the deconstructed version of the transmembrane subunit of the envelope protein could be recognized by the broadly neutralizing mAbs 2F5 and 4E10. The 2F5 and 4E10 mAbs interacts with a highly conserved sequence of MPER. This part is "hidden" inside a tight helix bundle in most of the structural models reported from MPER (Liu et al., 2009; Buzon et al., 2010). Only one structure has been solved containing a shortened NHR (HR1) region, which left MPER accessible to 2F5 Abs (Shi et al., 2010). We used ELISA to determine if the 2F5 antibody was able to bind to purified MPR-$TM_{TEV\text{-}6His}$ in its non-denatured state (Fig. 2.10). The results (Fig. 2.10A, rows C and D) clearly indicate that the MPR-$TM_{TEV\text{-}6His}$ protein is indeed recognized by the 2F5 Abs. As expected, a positive control consisting of a fusion protein of a fusion protein of the cholera toxin subunit B with MPR also react with the 2F5 Abs (CTB-MPR, Fig. 2.10A, rows E and F). CTB-MPR has previously been shown to react with 2F5 Abs (Matoba et al., 2004) and was able to elicit Abs that could block the transcytosis progression of HIV through the tight epithelia models (Matoba et al., 2004). In CTB-MPR, the pentameric nature of CTB is thought to hinder MPR from assuming the trimeric post-fusion conformation that does not allow for antibody access to the 2F5-binding site (Lee et al., 2014).

To affirm the ELISA results and to quantitatively assess the affinity of MPR-$TM_{TEV\text{-}6His}$ to the two broadly neutralizing Abs 2F5 and 4E10, we employed surface plasmon resonance (SPR) measurements (Fig. 2.10B, C and Table 2.1). The results demonstrate very high affinity (nanomolar and subnanomolar range) of the MPR in the context of its transmembrane domain (MPR-$TM_{TEV\text{-}6His}$, Fig. 2.10B) and as a fusion protein with CTB (Fig. 2.10C), in good agreement to our previously published results concerning CTB-MPR (Matoba et al., 2008). The calculated dissociation constant (K_D) of 2F5 from MPR-$TM_{TEV\text{-}6His}$ and CTB-MPR was 2.2 ± 0.2 nM and 0.8 ± 0.2 nM, respectively. The calculated dissociation constant (K_D) of 4E10 from MPR-$TM_{TEV\text{-}6His}$ and CTB-MPR was 2.1 ± 0.0 nM and 0.5 ± 0.2 nM, respectively. As a negative control, we tested CTB by itself, which showed no appreciable binding to either 2F5 or 4E10 (data not shown).

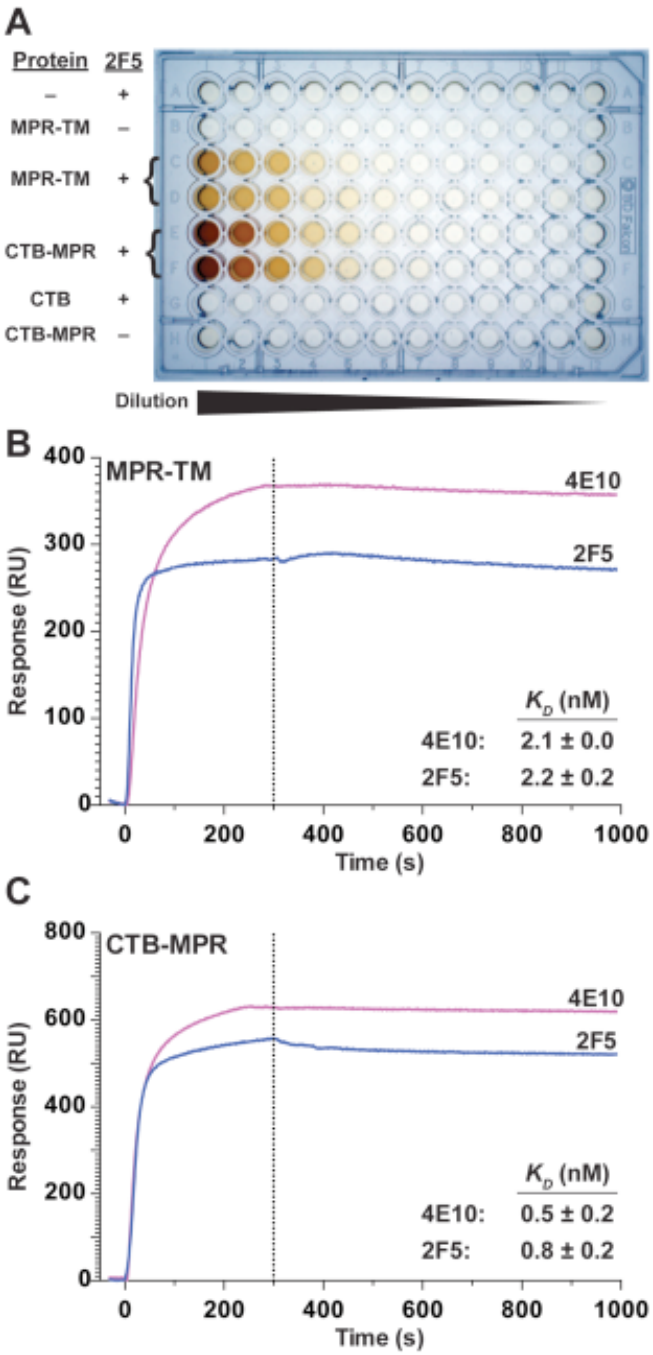


Figure 2.10. Purified native MPR-$TM_{TEV\text{-}6His}$ can bind to 2F5 and 4E10 mAbs. (A) Image of the developed ELISA plate. Row A: control uncoated wells. Rows B-D: wells coated with serially diluted MPR-$TM_{TEV\text{-}6His}$. Rows E, F and H: wells coated with serially diluted CTB-MPR. Row G: wells coated with serially diluted CTB. All rows except B and H were overlaid with the mAb 2F5. (B, C) SPR analysis. Association/dissociation traces of MPR-$TM_{TEV\text{-}6His}$ (B) and CTB-MPR (C) with either 2F5 or 4E10 mAbs. Traces are average of four independent measurements and the dissociation constants (K_D) are listed in the inserts (mean ± SD).

Immobilized ligand	Flowing analyte	k_a, ms^{-1}	k_d, s^{-1}	K_D, M
2F5	MPR-TM	4.7±0.3E4	9.9±0.1E-5	2.2±0.2E-9
4E10	MPR-TM	2.3±0.1E4	5.0±0.4E-5	2.1±0.0E-9
2F5	CTB-MPR	5.5±0.7E4	4.2±1.0E-5	0.8±0.2E-9
4E10	CTB-MPR	4.7±0.3E4	2.5±1.3E-5	0.5±0.2E-9

Table 2.1. Association and dissociation rate constants derived from SPR analysis.

The ELISA and SPR results indicated that the 2F5 and 4E10 epitopes in MPR-$TM_{TEV\text{-}6His}$ was exposed and accessible for strong 2F5 (and 4E10) binding. These results are in excellent agreement with experiments of other groups aimed at recreating an early fusion intermediate in which the epitopes of 2F5 and 4E10 are exposed to allow interactions with the two neutralizing monoclonal mAbs (Qiao et al., 2005; Kim et al., 2007; Frey et al., 2008; Wang et al., 2011; Lutje Hulsik et al., 2013). For example, Frey *et al.* (Frey et al., 2008) created a prehairpin intermediate consisting of a trimerization tag fused to the MPER, and to the NHR sandwiched between a duplicated CHR. The NHR was apparently prevented from masking the MPER, thus allowing its interactions with 2F5. More recently, Lutje-Hulsik et al. (Lutje Hulsik et al., 2013) have used an even simpler construct that contained the TM domain of gp41, the MPER and the CHR to demonstrate similar high affinities to several neutralizing monoclonal Abs. Our construct contains no artificial trimerizing fusion partners, and contains only the C-terminal half of the CHR.

In contrast, both Frey et al. and Lutje-Hulsik et al. (Frey et al., 2008; Lutje Hulsik et al., 2013), as well as others, demonstrated that gp41 in its prefusion conformation (present on the virions, for example see (Cavacini et al., 2002)) cannot interact with 2F5 or 4E10 prior to Env's interactions with the CD4 receptor (Cavacini et al., 2002; Leaman et al., 2010). Similarly, a post-fusion conformation precludes 2F5 and 4E10 binding (Frey

et al., 2008; Lutje Hulsik et al., 2013). For example, Liu *et al*. (Liu et al., 2009) reported tight helix bundle structure for MPER. In their study, MPER of gp41 was fused to a trimeric C-terminal isoleucine zipper motif and formed a parallel three-stranded coiled coil. The 2F5 epitopes were buried within the interface between the MPER helices and could not be recognized by 2F5 (Liu et al., 2009). Our results provide further experimental evidence of the importance of the transmembrane domain of gp41 to preserve the immunological signature of the membrane proximal region of gp41 (Montero et al., 2012), probably mimicking in a pre-hairpin intermediate. It also demonstrates the need to remove the heptad repeat regions.

2.5 Conclusion

In summary, we describe here the design, expression and purification of a protein construct that includes MPR and the transmembrane domain of gp41 (MPR-$TM_{TEV\text{-}6His}$), which reacts with the broadly neutralizing Abs 2F5 and 4E10 and thereby may represent an immunologically relevant conformation mimicking a pre-hairpin intermediate of gp41. The quantity and quality of purified MPR-$TM_{TEV\text{-}6His}$ reported here make the protein suitable for crystallization experiments and NMR studies as a prerequisite for structural studies, which may guide the structure-based design of vaccines against HIV-1 in the future.

2.6 Supporting Information

2.6.1 Supporting Material and Methods

2.6.1.1 Cloning, Bacterial Strains and Growth Conditions

The MPR-$TM_{649\text{-}705}$ construct is based on a deconstructed HIV-1 gp41 (dgp41) gene (GenBank Accession number JX534518) (Kessans et al., 2013), a chimera

comprising the gp41 MPR derived from the B-clade MN isolate (GenBank accession number AF075722) and the transmembrane domain and cytoplasmic tail region of the C clade 1084i isolate (GenBank accession number AY805330). The coding sequence of MPR-$TM_{649-705}$ was amplified by PCR from pTM601 (Kessans et al., 2013) using the forward primer (5' GAGAATCTTTATTTTCCAGGGCATGGGATCTCAAACTCAAC 3') and the reverse primer (5' GCCCTGAAAATAAAGATTCTCTTACACCATAGAC AACACAG 3') and then ligated into the pCR8/GW/TOPO vector (pCR8/GW/TOPO TA Cloning Kit, Invitrogen) to yield a Gateway entry clone. The entry clone was screened by restriction analysis and sequencing to confirm the presence and correct orientation of MPR-$TM_{649-705}$. Two recognition sites for the tobacco etch virus (TEV) protease were introduced by PCR primers. One was on the N-terminus of MPR-$TM_{649-705}$ and the other was on the C-terminus. MPR-$TM_{649-705}$ was further cloned into a Gateway destination vector pMistic (Vector name: pMIS2.1mv, DNASU) from the entry clone pCR8/GW/TOPO using a LR recombinant reaction (Gateway LR Clonase II Enzyme Mix, Invitrogen) to generate the expression clone pMistic-MPR-$TM_{649-705}$. The expression vector pMistic (Vector name: pMIS2.1mv, DNASU) was a kind gift of Dr. Mark Vega (Center for Structures of Membrane Proteins, PSI, Salk Institute) to the Center for Membrane Protein in Infectious Diseases (MPID) of Arizona State University. The recombinant plasmid pMistic-MPR-$TM_{649-705}$ was transformed into *E. coli* C41 (DE3) (Lucigen) for expression. Recombinant *E. coli* C41 (DE3) cells were grown in Terrific Broth medium containing 50 μg/mL of ampicillin in a shaker at 37 °C and 200 rpm. Recombinant expression of our protein in *E. coli* C41 (DE3) cells was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 200 μM

when the OD_{600} reached 0.6-0.8. The cells were then incubated for another 24 hours at 25 °C and 200 rpm. The OD_{600} after 24-hours induction was 2.0. The cells were harvested by centrifugation at 5000 x*g* for 15 min at 4 °C. Cell pellets were stored at -80 °C for future use.

2.6.1.2 Purification of MPR-TM

Cells stored at -80 °C were thawed and resuspended in ice-cold phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) with Sigma EDTA-free protease inhibitor cocktail tablet. Cell pellets (15 g) were resuspended in 100 mL of PBS containing one Sigma EDTA-free protease inhibitor cocktail tablet. Cells were then lysed by passing through a microfluidizer (Microfluidics) twice at 90 psi. The cell lysate was centrifuged at 20,000 x*g* for 20 min at 4 °C. The supernatant was discarded and the pellet, containing the membrane fraction, peptidoglycan cell wall and other aqueous-insoluble material was stored at -80 °C overnight. We found that freezing and thawing the sample makes resuspension easier, probably because high molecular weight DNA in the sample becomes sheared in the process. If processed without the freeze-thaw treatment, the detergent extract is highly viscous and too sticky to pass through the metal affinity column in the next step.

To extract the MPR-TM from the membrane, the membrane pellet (obtained from 15 g of cells) was thawed and resuspended in 100 mL of ice-cold PBS with 1% n-β-dodecyl-D-maltoside (βDDM) and a Sigma EDTA-free protease inhibitor cocktail tablet. The βDDM-containing suspension was incubated with gentle shaking for 3 h at 4 °C and then centrifuged down at 20,000 *xg* for 20 min at 4 °C. The supernatant was collected and the protein was purified by TALON metal affinity chromatography (Clontech

Laboratories) using a hybrid batch/gravity-flow procedure. In this hybrid procedure, the binding step was performed in a batch format in a column (Bio-Rad Econo-Column, 5 cm × 20 cm, maximum volume: 393 mL) that accommodated ~20 times the resin bed volume (typically 20 mL) for homogeneous binding. The washing and elution steps were then performed by gravity-flow. Prior to use, the column was washed with 3 column volumes (CVs) of H_2O and equilibrated with 3 CVs of binding buffer (20 mM bicine pH 8.0, 500 mM NaCl, 0.05% βDDM). The βDDM extract was loaded onto the column and incubated with TALON resin while gently shaking for 1 hour at 4 °C. The resin was let to settle in the column and the flowthrough was collected for SDS-PAGE analysis. The resin was washed with 6 CVs of binding buffer followed by 12 CVs of wash buffer (20 mM bicine, 500 mM NaCl, pH 8.0, 10 mM imidazole, 0.05% βDDM). The 10th and 12th washes were collected for SDS-PAGE analysis. After the washing steps, elution buffer (20 mM bicine, 500 mM NaCl, pH 8.0, 250 mM imidazole, 0.05% βDDM, total volume 2.5 CVs) was applied onto the column to elute His-tagged proteins. The eluate was collected in 10 mL (first) or 20 mL (all subsequent) fractions. Eluate samples were kept for SDS-PAGE analysis.

The eluted proteins were dialyzed against 20 mM NaCl, 20 mM HEPES, pH 7.5 using a 2000 Da cut-off dialysis tube (Sigma) overnight at 4 °C. After dialysis, Tris-HCl pH 8.0, EDTA and dithiothreitol (DTT) were added to the dialyzed protein to final concentrations of 50 mM, 0.5 mM and 1 mM, respectively. TEV protease (Invitrogen) was then added to a final protease:substrate ratio of 20 U/ 182 μg, and the reaction mixture was incubated for 2 h at room temperature. Under these conditions >95% of the

protein was cleaved. The protein concentration was determined by modified Lowry assay (Markwell et al., 1978).

The cleaved MPR-TM protein sample was further purified by Mono Q anion exchange chromatography using an ÄKTApurifier 10 (GE Healthcare) and a Mono Q 5/50 GL column (1 mL bed volume, GE Healthcare). Optimal conditions for Mono Q anion exchange chromatography purification of MPR-TM were chosen based on the results of extensive small scale tests. All buffers used for FPLC were filtered through a 0.2 μm membrane (Millipore) and degassed. After the column was equilibrated with 10 ml of buffer A (20 mM HEPES, pH 7.5, 0.02% βDDM), the dialyzed MPR-TM sample was injected onto the column. The column was then washed with 30 ml of buffer A at 2 mL/min, followed by an increasing linear gradient of buffer B (20 mM HEPES, pH 7.5, 1 M NaCl, 0.02% βDDM) from 0% to 25% at 2 ml/min in 20 min, leading to the elution of cleaved MPR-TM. After elution, the column was washed with 100% buffer B to wash the column at the end of the run. Protein elution was monitored by absorbance at 280 nm and the fractions were analyzed by 14% SDS-PAGE and immunoblot analysis. Fractions containing cleaved MPR-TM were pooled together for further biophysical analysis.

2.6.1.3 SDS-PAGE and Western Blot Detection

Proteins were separated by SDS-PAGE carried out as described by Schägger on 14% polyacrylamide gels(Schägger, 2006) and either stained by the silver staining method(Lawrence et al., 2011) or subjected to immunoblotting. Proteins in gels were transferred onto a PVDF membrane (Bio-Rad) in the presence of transfer buffer (192 mM glycine, 24.9 mM Tris base, 20% v/v methanol) using a Bio-Rad Mini Trans-Blot Module at 220 mA for 60 min. The PVDF membrane was incubated with 5% PBST-M

(PBS supplemented with 0.05% Tween 20 and 5% non-fat milk) for 60 min at room temperature. The HIV-1 gp41 mAb 2F5 (cat# 1475), obtained from Hermann Katinger through the NIH AIDS Research and Reference Reagent Program, Divisions of AIDS, NIAID, NIH, was used as the primary detecting Ab. Following another 60-min incubation with the primary Ab (1:10000 dilution), the membrane was rapidly washed three times with PBST (PBS supplemented with 0.05% Tween 20) and followed by a 30-min incubation with PBST. The membrane was then incubated for 60 min with a rabbit anti-human IgG-HRP conjugate (Santa Cruz Biotechnology) as the secondary antibody and StrepTactin-HRP conjugate (Bio-Rad, to detect the MW standards) that were diluted, respectively, 1:20,000 or 1: 5,000 in 5% PBST-M. The membrane was washed in PBST as described above. Following the fourth PBST wash, the wet membrane was developed using Immun-Star HRP substrate kit (Bio-Rad) per manufacturer's instructions. Chemiluminescence was detected using the BioSpectrum 500 C Imaging System (Ultra-Violet Products Ltd).

2.6.1.4 Size Exclusion Chromatography

The purified MPR-TM from Mono Q anion exchange chromatography was characterized by size exclusion chromatography (SEC) using an ÄKTApurifier 10 (GE Healthcare) and a Superdex 200 10/300 GL column (24 ml bed volume, GE Healthcare). The mobile phase was 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5, 0.02% βDDM. This buffer was used because it has little absorption in the UV for the subsequent circular dichroism (CD) measurements. The column was calibrated using the Gel Filtration LMW Calibration Kit and Gel Filtration HMW Calibration Kit from GE Healthcare. The kit contained blue dextran 2000 (2000 kDa) and eight standard proteins: aprotinin (6.5 kDa),

RNase A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), aldolase (158 kDa), ferritin (440 kDa) and thyroglobulin (669 kDa). Blue dextran 2000 was used to determine the void volume of the column. Seven standard proteins except thyroglobulin (669 kDa) were used to obtain the standard curve for molecular mass estimation.

2.6.1.5 Mass Spectrometry

We used matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) Mass spectrometry (MS) to accurately measure the molecular weight of the purified MPR-TM protein. Purified MPR-TM (1 μL of 1 mg/mL) was added to 4 μl of sinapinic acid matrix, which was prepared daily as a saturated solution in 50% acetonitrile/H_2O and 0.1% trifluoroacetic acid (TFA). The protein/ matrix mixture (1 μL) was added onto a steel target plate (Applied Biosystems) and allowed to dry in air. The plate was then placed into Applied Biosystems DE-STR MALDI-TOF mass spectrometer, and spectra were collected in a positive linear mode over a mass range from 3 to 30 kDa. Final results represented the average of 10 separate spectra with each spectrum in turn the average of 100 laser shots.

2.6.1.6 ELISA

ELISA plates (96-wells, Becton Dickinson) were coated by the test proteins, 5-fold serially diluted (starting concentration of 200 μg/mL) in coating buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, 3 mM NaN_3, pH 9.6) at 37 °C for 1 h. Test proteins included purified MPR-TM (this work), CTB-MPR (Lee *et al*, in preparation and reference (Matoba et al., 2008)), a fusion protein between the MPR and the B subunit of cholera toxin (CTB), and

CTB serving as a negative control (List Biological Laboratories). Blank (uncoated) wells and omission of the primary Ab served as additional negative controls.

After protein coating, the wells were washed two times with PBST buffer, followed by incubation with 5% PBST-M buffer at room temperature for 1 h. Subsequently, the wells were rinsed two times with deionized H_2O. The plates were then incubated at 37 °C for 1 h with 1% PBST-M (PBS supplemented with 0.05% Tween 20 and 1% non-fat milk) with the mAb 2F5 added (50 ng) to the indicated wells. Subsequently, all wells were washed three times with PBST buffer, followed by incubation with goat anti-human IgG (Sigma) at 1:1000 dilution in 1% PBST-M at 37 °C for 1 h. The wells were then washed three times with deionized H_2O and developed with Sigma *FAST* OPD (*o*-phenylenediamine dihydrochloride) substrate (Sigma). The plates were imaged and the absorbance at 490 nm was measured using a microplate reader (Spectra Max 340PC, Molecular Devices). Absorbance data plotted against protein concentrations were fitted by nonlinear regression using GraphPad Prism 4.0 to obtain approximate dissociation constant (K_d) values of antigen-Ab interactions.

2.6.1.7 Circular Dichroism (CD) Spectroscopy and Dynamic Light Scattering (DLS)

A JASCO J-710 CD spectropolarimeter was used for measuring the CD spectra of purified sample. The SEC-purified MPR-TM (eluted in 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5, 0.02% βDDM) was concentrated to 0.22 mg/ml by a 50-kDa concentrator (polyethersulfone membrane, Sartorius) and used for CD measurement. Buffer-only samples were measured as blank and the blank values were subtracted from the CD measurement of MPR-TM. CD spectra were recorded from 185 to 260 nm at 25 °C using a 0.1 cm quartz cuvette. Parameters were set at 1 nm data pitch, continuous scanning

mode, a scanning speed of 50 nm/min, a response of 4 s, and a spectral bandwidth of 1 nm. Output spectra were generated based on an accumulation of five scans. The molar ellipticity in deg.·cm^2/dmol was calculated as described by Greenfield's *et al.* (Greenfield, 2006). Data analysis was performed using the CONTINLL program in CDPro software package by comparing the measured data with reference set option 10, which included 13 membrane proteins along with 43 soluble proteins(Sreerama and Woody, 2000). The secondary structure content of purified MPR-TM was estimated based on the CONTINLL analysis.

DynaPro NanoStar M3300 from Wyatt Technology was used to carry out DLS measurements in the same buffer used for CD spectroscopy. In addition, DLS measurements were conducted with solutions containing 1%, 0.1% and 0.02% βDDM in 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5 to estimate the molecular mass of βDDM micelles. A 120 mW laser of 660 nm was used as the light source. For each measurement, the number of acquisitions was 10 and each acquisition time was 20 s. All measurements were carried out at 20 °C.

2.6.1.8 Determination of the Protein Concentration

Protein determination was carried out by the modified Lowry assay(Markwell et al., 1978). Protein concentration of pure preparations of MPR-TM was determined by measuring A_{280} (ε= 32,290 cm^{-1}M^{-1}, obtained using Peptide Property Calculator at http://www.basic.northwestern.edu/biotools/proteincalc.html).

2.6.1.9 Surface Plasmon Resonance (SPR)

All experiments were performed on a K_X5 Surface Plasmon Resonance Imaging (SPRi) System (Plexera). The Kx5 SPRi procedure were previously described (Song et

al., 2013). The SPRi chip was a 25 mm x 75 mm BK7 optical glass slide coated with a 50 nm-thick gold layer and a 1.5 nm-thick chromium adhesive layer (Plexera). The gold slide was cleaned with oxygen plasma for 2 min at 29.6 W, and was immediately incubated for 16 h at 4 ºC with 20-(11-mercaptoundecanoyl)-3,6,9,12,18-hexaoxaeicosanoicacid (1 mM in 100% ethanol) to form a carboxylate-terminated, self-assembled monolayer (SAM) on the slide. After rinsing it with 100% ethanol and water in turn, the gold chip was dried with compressed air and incubated with a mixture of 0.2 M ethyl(dimethylaminopropyl) carbodiimide (EDC) and 0.05 M *N*-Hydroxysuccinimide (NHS) for 15 min to activate the carboxylate groups need for the subsequent immobilization of Protein A/G (200 μg/ml, Thermo Scientific) through amide bond formation with primary amine groups of Protein A/G. Parallel micro-channels (300 mm x 20 mm x 10 mm, W x H x L) were formed by sealing the protein A/G-coated gold chip with a polydimethylsiloxane (PDMS) slab that was embedded with the channels' pattern. Immobilization of the test Abs was achieved by injecting 2F5 and 4E10 (80 mg/mL) to individual channels and incubated for 1 h. Protein A/G captures Abs by their Fc region, allowing their unimpeded interactions with antigens. After immobilization of the Abs, the PDMS slab was peeled off and the gold chip was rinsed with running buffer and then assembled with a mono-channel flow cell over the whole detection region (10 mm x 10 mm). To prevent non-specific adsorption, the chip was blocked with BSA (5 mg/mL) before further analysis. The running buffer and dilution buffer of the analyte was 1xPBS containing 0.02% βDDM. In sequential runs, CTB (the negative control) at 850 nM, CTB-MPR at 600 nM and MPR-$TM_{TEV\text{-}6His}$ at 840 nM were passed over the ligand surface at a flow rate of 1 μl/s, with a 300-sec association and a 600-sec dissociation. The

chip was regenerated between runs with H_3PO_4 (1:200 of 85% w/w) for 100 s followed by recoating with the desired antibody. Identical injections over blank protein A/G surfaces were subtracted from the data for kinetic analysis. SPRi data consisting of video images at 1 s resolution were analyzed with Data Analysis Module software from Plexera. The binding curve was analyzed and fitted with 1:1 interaction model with Scrubber 2 software (Biologic Software).

2.6.2 Supporting Tables

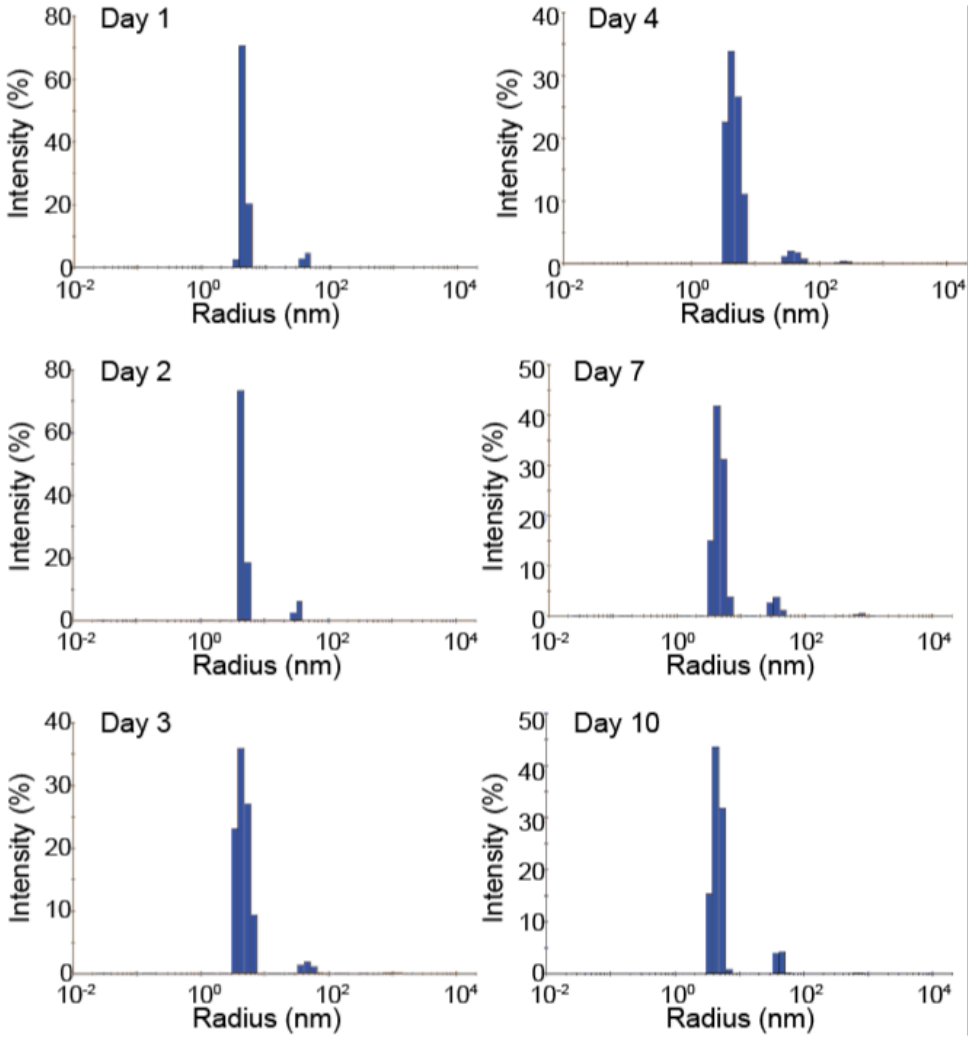
Intensity Distribution	Radius (nm)	Polydispersity (%)	Mw-R (kDa)	Intensity (%)	Mass (%)
Peak 1	4.7±0.1	11.8±0.6	121.0±3.0	963.5±3.5	100.0±0.0

Supporting Table 2.1. Result from DLS measurement of purified MPR-TM.

Intensity Distribution	Radius (nm)	Polydispersity (%)	Mw-R (kDa)	Intensity (%)	Mass (%)
Day 1					
Peak 1	4.6	11.2	118	93.0	100.0
Peak 2	43.1	11.1	22441	7.0	0.0
Day 2					
Peak 1	4.6	10.2	118	91.5	100.0
Peak 2	34.7	10.0	13543	8.5	0.0
Day 3					
Peak 1	4.7	22.6	127	95.3	100.0
Peak 2	47.4	19.6	28120	4.3	0.0
Peak 3	1062.0	21.9	40520700	0.4	0.0
Day 4					
Peak 1	4.8	23.3	130	93.9	100.0
Peak 2	42.2	22.9	20264	5.5	0.0
Peak 3	259.5	16.7	1498970	0.7	0.0
Day 7					
Peak 1	4.7	18.5	126	91.9	100.0
Peak 2	35.6	16.5	14334	7.3	0.0
Peak 3	743.5	13.8	17596700	0.8	0.0
Day 10					
Peak 1	4.6	16.8	121	91.4	100.0
Peak 2	42.1	12.6	21277	8.3	0.0
Peak 3	690.2	11.4	14782400	0.3	0.0

Supporting Table 2.2. DLS measurements of purified MPR-TM subjected to prolonged incubation at 4 °C. Samples were taken at the indicated time points

2.6.3 Supporting Figures



Supporting Fig. 2.1. DLS measurement of the stability of purified MPR-TM. Purified MPR-TM was kept at 4 °C, and DLS measurments were carried out at day 1, 2, 3, 4, 7 and 10. The result shows that MPR-TM remains monodisperse for at least 10 days.

CHAPTER 3

BIOPHYSICAL CHARACTERIZATION OF A VACCINE CANDIDATE AGAINST HIV-1: THE TRANSMEMBRANE AND MEMBRANE PROXIMAL DOMAINS OF HIV-1 GP41 AS A MALTOSE BINDING PROTEIN FUSION

3.1 Abstract

The membrane proximal region (MPR, residues 649-683) and transmembrane domain (TMD, residues 684-705) of the gp41 subunit of HIV-1's envelope protein are highly conserved and are important in viral mucosal transmission, virus attachment and membrane fusion with target cells. Several structures of the trimeric membrane proximal external region (residues 662-683) of MPR have been reported at the atomic level; however, the atomic structure of TMD still remains unknown. To elucidate the structure of both MPR and TMD, we expressed the region spanning both domains MPR-TM (residues 649-705) in *Escherichia coli* as a fusion protein with *E. coli*'s maltose binding protein (MBP). MPR-TM was initially fused to the C-terminus of MBP via a 42 aa-long linker containing a TEV protease recognition site (MBP-linker-MPR-TM). The purified MBP-linker-MPR-TM protein was biophysically characterized by size exclusion chromatography, CD spectroscopy and dynamic light scattering. Our results indicated that the purified MBP-linker-MPR-TM protein was a monodisperse and stable candidate for crystallization. However, crystals of the MBP-linker-MPR-TM protein could not be obtained in extensive crystallization screens. It is possible that the 42 residue-long linker between MBP and MPR-TM was interfering with crystal formation. To test this hypothesis the 42 residue-long linker was replaced with three alanine residues. This fusion protein, MBP-AAA-MPR-TM, was similarly purified and characterized. The

crystallization experiments of MBP-AAA-MPR-TM are currently undergoing. Significantly, both MBP-linker-MPR-TM and MBP-AAA-MPR-TM protein strongly interacted with broadly neutralizing monoclonal antibodies 2F5 and 4E10. With epitopes accessible to the broadly neutralizing antibodies, MBP/MPR-TM recombinant proteins may be in an immunologically relevant conformation mimicking a pre-hairpin intermediate of gp41.

3.2 Introduction

The transmembrane (TM) domain of HIV-1 gp41 is one of the most highly conserved regions of the envelope glycoprotein (Env) of HIV-1 (Shang et al., 2008; Checkley et al., 2011). This region is involved in many essential biological functions (recently reviewed by Steckbeck and co-workers (Steckbeck et al., 2013)). The primary role of gp41 TM domain is to anchor Env in both viral and cellular membranes (Gabuzda et al., 1991). It has been recently reported that the TM domain induces lipid mixing and associates with the fusion peptide (FP) of HIV-1 gp41 during the viral fusion process (Reuven et al., 2012). The gp41 TM peptide was able to inhibit virus-cell fusion because it associates strongly with FP and thus may interfere with FP's inserting into the target cell membrane, which makes gp41 TM peptide a new fascinating HIV-1 entry inhibitor (Reuven et al., 2012). Additionally, the gp41 TM domain shares a motif with the α subunit of the T-cell receptor (TCR) TM domain (Checkley et al., 2011). The gp41 TM peptide co-localizes with CD3 in the TCR complex and inhibits T cell proliferation *in vitro*, which could be one of the strategies that the HIV-1 virus has developed to evade the immune response (Cohen et al., 2010).

Mao et al. reported a 6-Å structure of the membrane-bound HIV-1 envelope glycoprotein trimer in its uncleaved state by using cryo-electron microscopy (EM), which included the TM domain of gp41 (Mao et al., 2013). This is the only structural information of the TM domain of gp41 reported so far. The atomic structure of the gp41 TM domain remains unknown perhaps due to its high hydrophobicity making its expression, purification and crystallization difficult. The traditional model of the gp41 TM domain is a single membrane-spanning α-helix (residues 684-705) followed by a intracytoplasmic C-terminal tail (CTT) (Checkley et al., 2011). An alternative model was proposed by Hollier and Dimmock (Hollier and Dimmock, 2005) to explain the observation that an epitope in the C-terminal tail of HIV-1 gp41 (the so-called Kennedy epitope (Hilf and Dutzler, 2008)) is extracellulary exposed under some conditions. To account for the drastic change in the membrane topology necessary for the exposure of the Kennedy epitope, Hollier and Dimmock speculatively suggested that the transmembrane region assumes a different secondary structure consisting of three membrane-spanning β-sheets under certain conditions (Hollier and Dimmock, 2005).

Another highly conserved segment of HIV-1 gp41 is the membrane proximal region (MPR, residues 650-685) (Miyauchi et al., 2005; Checkley et al., 2011), which contains the membrane proximal external region (MPER, residues 662-683) and part of the C-terminal heptad repeat region (CHR, residues 650-661) of HIV-1 gp41. HIV-1 gp41 is essential in transcytosis, a process leading to mucosal transmission of the virus (Bomsel et al., 2011; Shen et al., 2011; Puryear and Gummuluru, 2013). Transcytosis is initiated when gp41 binds to the epithelial glycosphingolipid galactosyl ceramide (Gal Cer), the epithelial cell receptor for HIV (Alfsen et al., 2001). The minimal region

required for gp41 to bind Gal Cer is the membrane proximal region (MPR, residues 650-685) (Alfsen and Bomsel, 2002). Furthermore, MPR is the target of secretory IgAs, which block HIV-1 transcytosis by neutralizing gp41 binding to Gal Cer (Devito et al., 2000; Tudor et al., 2009). Significantly, epitopes in the MPER are recognized by three broadly neutralizing antibodies 2F5 (Purtscher et al., 1994), 4E10 (Zwick et al., 2001) and 10E8 (Huang et al., 2012).

HIV-1 gp41 mediates the membrane fusion between target cell and virus through its own conformational change: native trimer prior to the interaction between gp120 and CD4, pre-hairpin intermediate and postfusion trimer of hairpins (or a six-helix bundle) (Chan and Kim, 1998). Structures of trimeric MPER have been solved in the pre-fusion (Pancera et al., 2014) and postfusion conformations (a six-helix bundle) (Buzon et al., 2010; Shi et al., 2010). However, both of them could not be recognized by the broadly neutralizing antibodies 2F5 or 4E10. Another structure of trimeric MPER has been solved in the postfusion state (a six-helix bundle) containing a shortened NHR (HR1) region, which left MPER accessible to 2F5 antibody (Shi et al., 2010). More recently, Reardon et al. reported a trimeric MPER structure solved by NMR spectroscopy in a putative pre-fusion intermediate state (a three-helix bundle) (Reardon et al., 2014). In their structure, the N termini of the MPER helices are closely associated with each other while the C termini gradually separate, which leaves space for antibody binding. However, in their construct, MPER was fused to the C terminus of a 27-residue trimerization domain from bacteriophage T4 fibritin (the foldon domain). Although Reardon et al. reported that MPER was linked to the foldon motif through a flexible linker GSSG, which is intended to minimize the effect of the structured trimerization motif on the conformation and

dynamics of MPER, it is still not experimentally confirmed that MPER forms a trimer in the pre-fusion intermediate form without the effect of the trimerization motif. Moreover, the tight association of the MPER trimer at N terminus could be due to its proximity to the tight foldon motif while C terminus of the MPER trimer separated from each other because it has less effect from the trimerization motif. To better understand HIV-1 gp41 and instruct design of vaccines and therapeutics against HIV-1, elucidating the structure of the gp41 MPR and TM domains is required.

Mistic, a *Bacillus subtilis* integral membrane protein (Roosild et al., 2005), was previously used as a fusion partner in our laboratory to overexpress MPR-TM of HIV-1 gp41, where the fusion partner Mistic was removed for crystallization (Gong et al., 2014). However, no crystals were obtained even after extensive crystallization screens. It may be due to the highly hydrophobic property of MPR-TM and/or lack of crystal contacts between MPR-TM molecules. In this work reported here, a novel construct was designed to overexpress MPR-TM as a maltose binding protein (MBP) fusion. MBP is a commonly used fusion partner, capable of improving the solubility and expression level of the target protein (Kapust and Waugh, 1999; Kobe et al., 1999; Do et al., 2014; Nguyen et al., 2014; Raran-Kurussi and Waugh, 2014). MBP may function as a chaperon to assist the correct folding of the target proteins in active form (Kapust and Waugh, 1999). Furthermore, MBP may provide a large hydrophilic interaction surface for formation of crystal lattice contacts thereby facilitating crystal formation (Moon et al., 2010). Structures of dozens of MBP recombinant proteins have been solved at atomic resolution by X-ray structure analysis in the past decades (reviewed in Refs. (Smyth et

al., 2003; Moon et al., 2010)), which indicated that MBP could be used as a fusion partner for structural studies.

In the present study, we reported the expression, purification and biophysical characterization of MPR-TM (residues 649 to 705) as a fusion to the C terminus of 8xHis-tagged MBP for structural determination by X-ray crystallography. Surface plasmon resonance (SPR) was used to test if the epitope on the purified protein was exposed and could be recognized by the broadly neutralizing mAbs 2F5 and 4E10. Crystals of the MBP/MPR-TM recombinant protein could not be obtained when MPR-TM was fused to the C-terminus of MBP via a 42-residue linker. The linker was changed to three alanine residues (MBP-AAA-MPR-TM) and crystallization experiments are undergoing.

3.3 Materials and Methods

3.3.1 Cloning and Expression of MBP-linker-MPR-TM and MBP-AAA-MPR-TM

The coding region of gp41 MPR-TM flanked by a TEV cleavage site was subcloned from pMistic-MPR-TM (Gong et al., 2014) into the pCR8/GM/Topo vector (Invitrogen, Carlsbad, CA), sequence-verified and shuttled into the pVP16 destination vector (Thao et al., 2004) by Gateway® recombination cloning (Invitrogen, Carlsbad, CA). An aliquot of the LR reaction was used to transform One Shot® TOP10 chemically competent *E. coli* cells following the protocol provided by the manufacturer (Invitrogen, Carlsbad, CA). Recombinant clones were selected on LB agar plates supplemented with 100 μg/mL of ampicillin. Plasmid DNA was isolated from several colonies using QIAGEN Plasmid Mini Kit (QIAGEN, Valencia, CA) and digested with *EcoR*I and *Kpn*I (Promega, Madison, WI) to confirm the presence of the insert into pVP16. The verified

recombinant expression vector of MBP-linker-MPR-TM was further transformed into NEB Express competent cells (New England Biolabs, Ipswich, MA), an enhanced BL21 derivative that allows inducible recombinant protein expression from T5 and other non-T7 promoters. Recombinant clones of MBP-linker-MPR-TM were selected on LB plates supplemented with 100 μg/mL of ampicillin.

To construct the gene encoding MBP-AAA-MPR-TM, a 153 bp *Bgl*II and *Sal*I restriction fragment that encompasses a region of the MBP-linker-MPR-TM coding sequence that encodes the long linker and a TEV protease recognition site (spanning EALKDAQ…ENLYFQG, Supporting Fig. 3.1) was replaced with a DNA sequence that encodes the protein sequence AALAAAQTNAAA. This cloning was accomplished with the following PCR reactions using Takara PrimeSTAR Max DNA Polymerase (Clontech, Mountain View, CA; see Supporting Table 3.1 (page 79) for the primer sequences). PCR1 used primers MBP-MPR-fuseF1 and insert1-PCR1R, and template MBP-linker-MPR-TM; PCR2 used primers MBP-MPR-fuseF1 and insert1-PCR2R, and template PCR1; PCR3 used primers insert2-PCR1F and MBP-MPR-fuseR1, and template MBP-linker-MPR-TM; PCR4 used primers MBP-MPR-fuseF2 and MBP-MPR-fuseR1, and template PCR2. The products of PCR3 and PCR4 were simultaneously inserted into *Bgl*II- and *Sal*I-digested MBP-linker-MPR-TM using the ligation-independent InFusion HD Cloning Plus system (Clontech) (Hamilton et al., 2007). DNA sequence confirmation was performed at the School of Life Sciences DNA Laboratory at Arizona State University. Vectors and their sequences have been deposited and are available from the PSI:Biology-Materials Repository at DNASU (http://dnasu.org) (Seiler et al., 2014). The verified recombinant expression vector of MBP-AAA-MPR-TM was further transformed

into BL21(DE3) competent cells (Invitrogen). Recombinant clones of MBP-AAA-MPR-TM were selected on LB plates supplemented with 100 μg/mL of ampicillin.

Bacterial overexpression of both fusion proteins was done as follows. A single bacterial colony was used to inoculate 100 mL pre-culture in LB containing 100 μg/mL of ampicillin in a shaker-incubator (37 °C, 200 rpm) overnight. This culture was used to inoculate 3 L of LB broth containing 100 μg/mL of ampicillin at a dilution ration of 1:30. Cells were grown until the culture density reached OD_{600} between 0.6 and 0.8. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 500 μM and continued incubation for additional 4 h. Bacteria were harvested by centrifugation (5000 ×g, 10 min, 4 °C) and cell pellets (12-14 g fresh weight per 3 L culture) were then stored at -80 °C until further use.

3.3.2 Preparation of the Crude Membrane Fractions of MBP-linker-MPR-TM and MBP-AAA-MPR-TM

Preparation procedures of the crude membrane fractions of MBP-linker-MPR-TM and MBP-AAA-MPR-TM were same. Cell pellets stored at -80 °C were thawed and re-suspended in ice-cold phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO4, pH 7.4 with 1× EDTA-free protease inhibitor (Sigma). 100 mL of PBS buffer containing 1× EDTA-free protease inhibitor (Sigma) was used to re-suspend per 15 grams of wet cell pellet. The suspension was disrupted by ultrasonic homogenizer (Model 300 V/T, Biologics) at 150 W and 10 kHz for 1 min on and 1 min off while kept on ice. This step was repeated three times in total. The cell lysate was centrifuged at 20,000 × *g* for 20 min at 4 °C. The supernatant was discarded and the

pellet, containing the membrane fraction, peptidoglycan cell wall and other aqueous-insoluble material was stored at -80 °C overnight.

3.3.3 Detergent Solublization of MBP-linker-MPR-TM and MBP-AAA-MPR-TM

The frozen crude membrane fraction of MBP-linker-MPR-TM was thawed and resuspended in ice-cold PBS buffer with 1× EDTA-free protease inhibitor (Sigma, S8830). 100 mL of PBS buffer containing 1× EDTA-free protease inhibitor (Sigma) was used to re-suspend per 15 grams of crude membrane fraction. Detergent screens were conducted by adding the following detergents to 1% (w/v) final concentration: lauryldimethylamine-oxide (LDAO), n-decyl-β-D-maltoside (β-DM), and n-dodecyl-β-D-maltoside (β-DDM). These suspensions were incubated at 4 °C for 2 h. Screening for the optimal length of time required for efficient detergent extractions was done essentially as described above except that the incubation time was varied among 0.5 h, 1 h, 2 h and 3 h respectively. Following their incubation, the suspensions were centrifuged at 20,000 ×g for 20 min and the supernatant liquid, which contained detergent-soluble proteins, was then used for further analyses. Large scale detergent extractions of both MBP-linker-MPR-TM and MBP-AAA-MPR-TM were performed with 1% βDDM at 4 °C for 1 h.

3.3.4 Purification of MBP-linker-MPR-TM and MBP-AAA-MPR-TM

Purification procedures of MBP-linker-MPR-TM and MBP-AAA-MPR-TM were same. The detergent-solubilized proteins were purified by metal-affinity FPLC using nickel-nitrilotriacetic acid (Ni-NTA) Superflow column (QIAGEN). The Ni-NTA Superflow resin was manually packed into the Tricon Empty 10/100 column (GE Healthcare) with a bed volume of about 8 ml. The columns were equilibrated with buffer A (500 mM NaCl, 20 mM bicine pH 8.0 and 0.05% β-DDM). The detergent soluble

fraction was injected onto the Ni-NTA Superflow column by a 10 mL superloop (GE Healthcare) and washed with buffer A until the A280 was stable below 10 mAu. To remove non-specifically-bound proteins, the column was then washed with 2% buffer B (Buffer A with 500 mM imidazole; final concentration of imidazole was 10 mM) until the A280 was stable below 20 mAu. Specifically-bound proteins were eluted off the column by application of a linear gradient of buffer B from 2% to 50% in 25 min at 1 mL/min. The eluted proteins were concentrated to 10-20 mg/mL using concentrators (100-kDa cutoff polyethersulfone membrane, Sartorius). Concentrated proteins were then subjected to size exclusion chromatography (SEC) using a Superdex-200 HR 10/300 gel filtration column (GE Healthcare) in SEC buffer (20 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.015% β-DDM). The Superdex-200 column was calibrated using standards aprotinin (6.5 kDa), RNase A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), aldolase (158 kDa) and ferritin (440 kDa). Protein purity was analyzed by SDS-PAGE.

The protein concentration of MBP-linker-MPR-TM and MBP-AAA-MPR-TM were determined spectrophotometrically at 280 nm using extinction coefficients of 98,290 $cm^{-1}M^{-1}$ and 94,450 $cm^{-1}M^{-1}$, respectively (extinction coefficients were obtained using Peptide Property Calculator at http://www.basic.northwestern.edu/biotools/-proteincalc.html).

3.3.5 SDS-PAGE, Silver Stain, Immunoblot Detection and Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS)

Proteins were resolved by SDS-PAGE and gels were either silver stained or subjected to immunoblotting as previously described (Gong et al., 2014). Briefly, SDS-

PAGE was performed with Tricine-SDS gels with 4% stacking and 8% separating gels respectively. The primary antibody (1:2000 dilution) used in immunoblots was BSA-free Penta·His antibody (QIAGEN) and the secondary antibody (1:2000 dilution) was goat anti-mouse IgG-HRP (Invitrogen). MALDI-TOF-MS was performed as previously described (Gong et al., 2014).

3.3.6 Circular Dichroism (CD) and Dynamic Light Scattering (DLS) Measurements

The CD spectra of purified sample were measured by A JASCO J-710 CD spectropolarimeter and DLS measurements were carried out using DynaPro NanoStar M3300 from Wyatt as previously described (Gong et al., 2014).

In order to study the effect of ionic strength on the secondary structure, the protein sample was concentrated to 5 mg/mL as described above and then diluted by 20 times into buffers containing 20 mM Tris-HCl, pH 7.5, 0.02% β-DDM and 0, 20, 50, 100, 150, 200, 250 or 300 mM NaCl. In order to identify the effect of pH on the secondary structure, the 5 mg/mL of protein sample was diluted by 20 times into buffers containing 150 mM NaCl, 0.02% β-DDM and at pH 6.5, 7.0, 7.5, 8.0 or 8.5. Protein samples were incubated in the buffer overnight before CD measurements.

Analysis of the secondary structure content based on CD data was carried out using the CDPro software package (Sreerama and Woody, 2000). Three methods CDSSTR (Compton and Johnson, 1986; Manavalan and Johnson, 1987; Sreerama and Woody, 2000), CONTIN/LL (Provencher and Glockner, 1981; Vanstokkum et al., 1990), and SELCON3 (Sreerama and Woody, 1993, 2000) were run using a protein reference set with 43 soluble proteins and 13 membrane proteins (SMP56) (Sreerama and Woody, 2000). The average and standard deviation of the results from these three methods were

calculated.

For thermal denaturation measurements monitored by CD spectroscopy, the CD signal at 220 nm was recorded from 15 to 90 °C with a temperature change of slope at a rate of 2 °C/min. The apparent midpoint temperature of the transition was calculated from the first derivative of the thermal melting curve.

3.3.7 Surface Plasma Resonance (SPR)

All experiments were performed in quadruplicate on a KX5 Surface Plasmon Resonance Imaging (SPRi) System (Plexera) as previously described (Gong et al., 2014). The mAbs 2F5 and 4E10 were immobilized onto the surface of the gold chip (Plexera). MBP (fractions A2 from SEC purification, Fig. 3.1C) at 240 nM was used as a negative control. The purified MBP-linker-MPR-TM protein at 187 nM and MBP-AAA-MPR-TM protein at 205 nM were measured respectively.

3.4 Results and Discussion

3.4.1 Cloning and Expression of MBP-linker-MPR-TM

In order to improve the expression level and solubility of the membrane proximal region and transmembrane (MPR-TM) domains of HIV-1 gp41, MPR-TM was fused to the C terminus of the maltose binding protein (MBP) by a 42 aa-long linker containing a tobacco etch virus (TEV) protease cleavage site (Supporting Fig. 3.1). An 8xHis-tag was cloned onto the N-terminus of MBP for purification purpose (Supporting Fig. 3.1). The MBP-linker-MPR-TM recombinant protein was expressed in NEB Express E. coli, a BL21 derivative-strain as described in Materials and Methods.

3.4.2 Detergent Solublization of MBP-linker-MPR-TM

The expressed cells were lysed by ultrasonic homogenizer as described in

Materials and Methods and then pelleted by centrifugation. The homogenized cell lysate prior to centrifugation and the supernatant containing water-soluble proteins after centrifugation were analyzed by anti-His immunoblot. A small portion of MBP-linker-MPR-TM protein was found to be in the supernatant (Supporting Fig. 3.2A (page 79)). Most of the fusion protein, however, was found in the pellet, which contained the membrane fraction, peptidoglycan cell wall and other water-insoluble material, and had to be extracted by detergents. To establish the optimal solubilization conditions, we have tested three different detergents: 1% LDAO, 1% β-DM and 1% β-DDM. The three detergents were able to extract the majority of the MBP-linker-MPR-TM protein from the pellet and their extraction efficiencies were similar as judged by immunoblot analysis (Supporting Fig. 3.2B). β-DDM was chosen for future purification because it has been widely and successfully used in crystallization of membrane proteins (Newstead et al., 2008).

In order to determine the time needed for efficient detergent extraction, crude membrane suspension solution was incubated in 1% βDDM at 4 °C for 0.5 h, 1 h, 2 h and 3 h respectively. Our results showed that 1 h incubation was sufficient to extract the majority of the MBP-linker-MPR-TM protein. Incubation for 2 h and 3 h did not significantly increase the extraction efficiency (Supporting Fig. 3.2C). Therefore, crude membrane suspension solution was incubated in 1% β-DDM at 4 °C for 1 h for future purification.

3.4.3 Purification of MBP-linker-MPR-TM

Purification of βDDM-solubilized MBP-linker-MPR-TM protein was accomplished by FPLC-connected Ni-affinity chromatography followed by size exclusion

chromate-graphy (SEC) in presence of βDDM. A single elution peak was eluted from the Ni-affinity chromatogram by a linear increasing gradient of 10 to 250 mM imidazole (Fig. 3.1A). The flowthrough and elution peaks were collected and analyzed by SDS-PAGE (Fig. 3.1B). Two protein bands were identified: one above the 50 kDa marker and the other below the 50 kDa marker (Fig. 3.1B). These two proteins were further successfully separated by SEC (Superdex 200) (Fig. 3.1C) and analyzed by SDS-PAGE analysis (Fig. 3.1D). Both bands were recognized by anti-His antibodies (Fig. 3.1D), indicative of an intact 8xHis tag. MALDI-TOF MS analysis demonstrated that SEC peak A1 contains a protein of 53,319 Da (Fig. 3.2A), possibly representing the MBP-linker-MPR-TM fusion protein (theoretical molecular weight: 53,389 Da). The signal at *m/z* 26,641 is likely to correspond to MBP-linker-MPR-TM charged with two protons. Peak A2 from SEC purification contained proteins with masses ranging from 43,818 Da to 45,814 Da (Fig. 3.2B). The theoretical molecular weight of the MBP itself is 41,694 Da (Fig. 3.2C). MBP including the linker would have a molecular weight of 46,376 Da (Fig. 3.2C). Therefore, proteins eluted in SEC peak A2 represent a number of C-terminally truncated versions of the recombinant fusion protein, which includes intact MBP (Fig. 3.1C).

The protein eluted in SEC peak A1 (Fig. 3.1C) was used for further analyses. About 60 mg of pure MBP-linker-MPR-TM protein was obtained from 1 liter of LB cell culture.

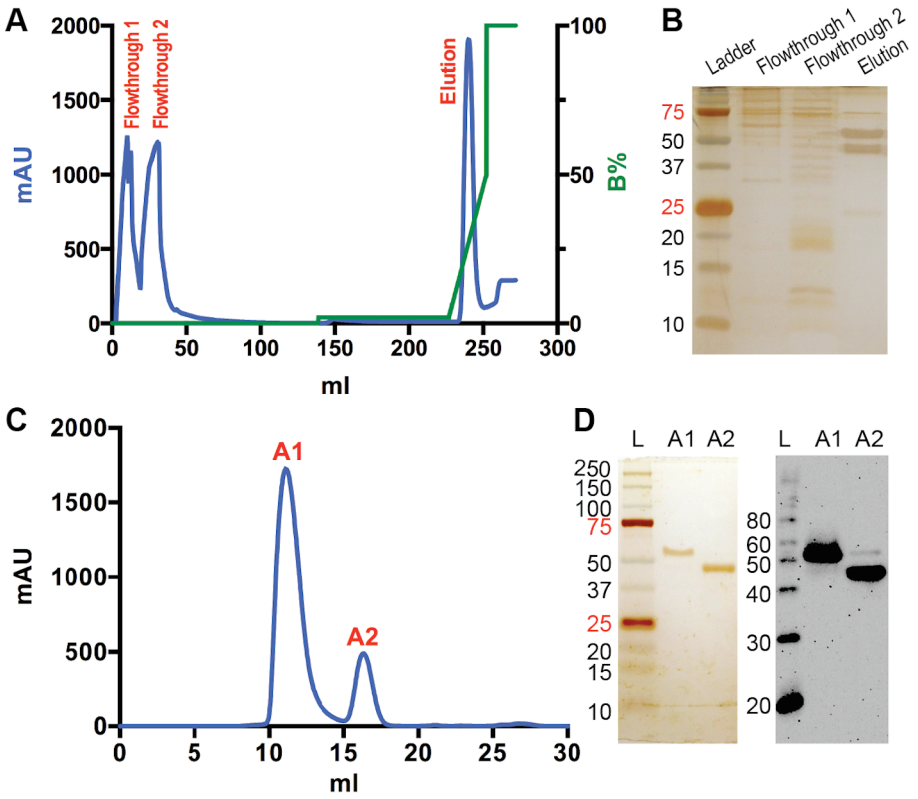


Fig. 3.1. Purification of MBP-linker-MPR-TM. (A) Ni-affinity chromatogram of the βDDM extraction. Blue curve: UV absorbance at 280 nm; green curve: percentage of buffer B. (B) SDS-PAGE analysis of the flowthrough peaks and elution peak of the Ni-affinity chromatography. (C) SEC of the elution from Ni-affinity chromatography. Blue curve: UV absorbance at 280 nm. Peak A1 eluted at 11.09 mL and peak A2 eluted at 16.33 mL. (D) Silver stained SDS-PAGE (left) and anti-His Western blot (right) analysis of the peak A1 and A2 from SEC.

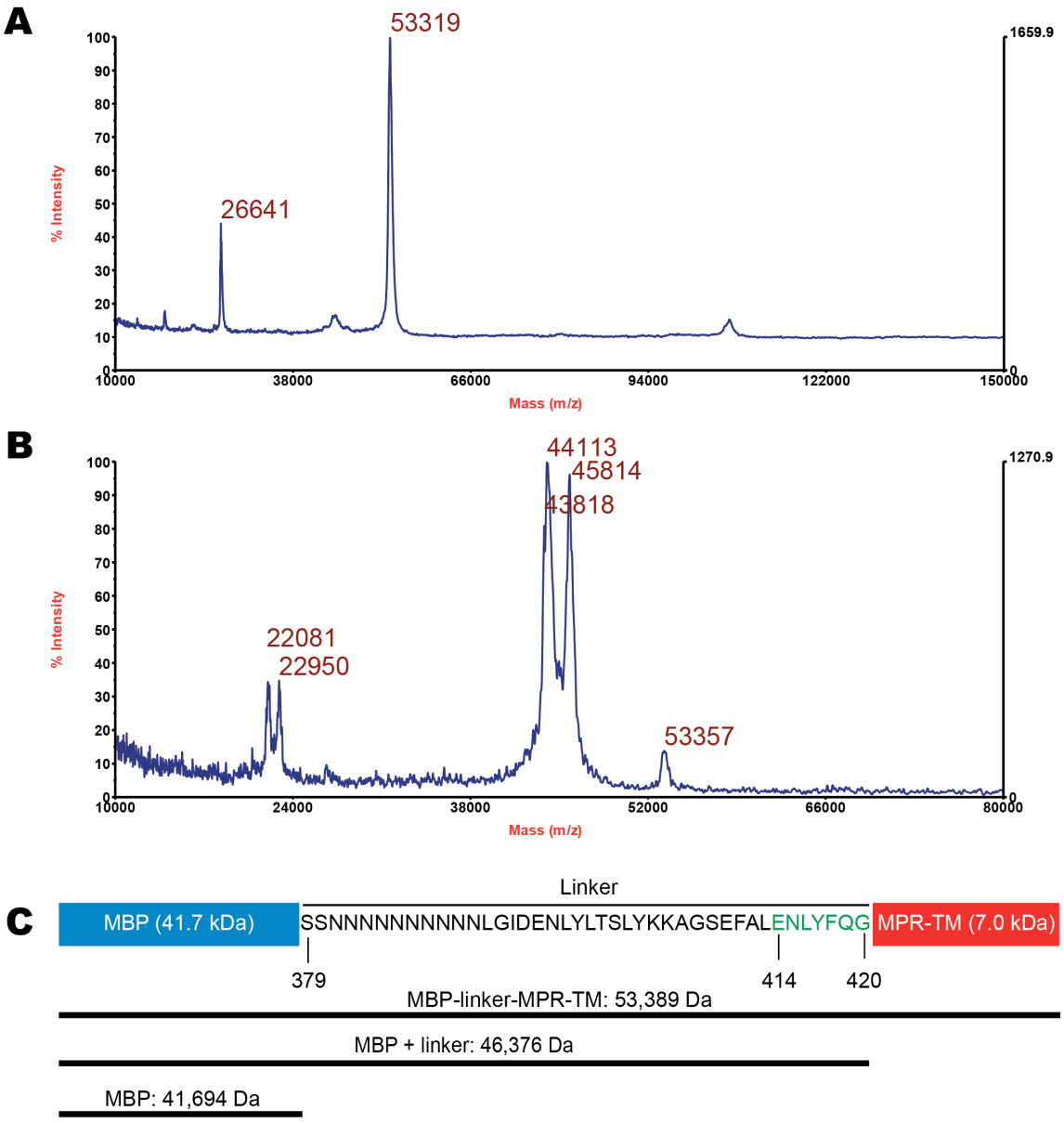


Fig. 3.2. Elutions from SEC were analyzed by MALDI-TOF MS. (A) MALDI-TOF MS analysis of the peak A1 from SEC (Fig. 3.1C). (B) MALDI-TOF MS analysis of the peak A2 from SEC (Fig. 3.1C). (C) Schematic representation of the MBP-linker-MPR-TM protein and molecular weight of different protein fragments. Residues 414-420 in green: TEV protease recognition site.

3.4.4 MBP-linker-MPR-TM Forms Oligomers

SDS-PAGE and MALDI-TOF MS analysis can only be used to determine the molecular weight of the monomeric MBP-linker-MPR-TM because the proteins were

denatured and lost their structural integrity and oligomeric state in these two analysis methods. The molecular weight of MBP-linker-MPR-TM in its detergent solubilized state was analyzed by analytical SEC and dynamic light scattering (DLS).

The elution fraction from Ni-affinity chromatography purification were analyzed by analytical SEC (Fig. 3.1C). Based on the standard curve calibration, the A2 SEC peak (Fig 3.1C, centering at 16.33 mL) corresponded to proteins with MW of <50 kDa, which indicated that the various partially truncated fusion proteins containing intact MBP remain monomeric as previously reported for MBP (Spurlino et al., 1991). The A1 SEC peak (Fig. 1B, centering at 11.09 mL) corresponded to proteins with MW of ~470 kDa based on our calibration. Considering the proteins are embedded in large β-DDM micelles, the exact oligomeric state is difficult to assess, but it is likely at least a hexamer. Previous studies demonstrated that MBP by itself is a monomer (Spurlino et al., 1991) but recombinant MBP fusion proteins may form oligomers depending on the nature of the fusion partner (Kobe et al., 1999; Ke et al., 2012; Tan et al., 2013). Our SEC results therefore provide an indication that it is the MPR-TM that is responsible for the oligomerization of MBP-linker-MPR-TM.

The size of a protein-detergent micelle estimated by SEC provides just a rough estimation of its size (Andrew, 2006). DLS was utilized in addition to estimate the molecular weight of MBP-linker-MPR-TM in form of the protein-detergent complex (Fig. 3.3). The Stokes radius of the detergent-protein complex was determined to be 7.7 ± 1.0 nm, which corresponded to a molecular mass of about 400 kDa (Supporting table 3.2). This result is slightly different from the estimation by SEC (~470 kDa). Please note that both SEC and DLS can be used to roughly estimate the size of molecules in solution,

but the estimation by neither of them is accurate for membrane proteins because both DLS and SEC assume a spherical particle that is very likely not the case for our fusion protein. The most important function of DLS in our study is to measure the sample's polydispersity distribution, which is the standard deviation of the histogram that refers to the width of the peak (Chu, 1970). The percent polydispersity (%Polydispersity) is the polydispersity divided by the estimated hydrodynamic radius multiplied by 100. The level of homogeneity is considered high when the percent polydispersity is less than 15% (Proteau et al., 2010). As shown in Figure 3.3 and Supporting Table 3.2, the percent polydispersity of the peak is 13.4% (Supporting Table 3.2), which indicated that the purified MBP-linker-MPR-TM is a monodisperse candidate for crystallization. It is concluded from the SEC and DLS analysis that MBP-linker-MPR-TM formed an oligomer in the presence of detergent.

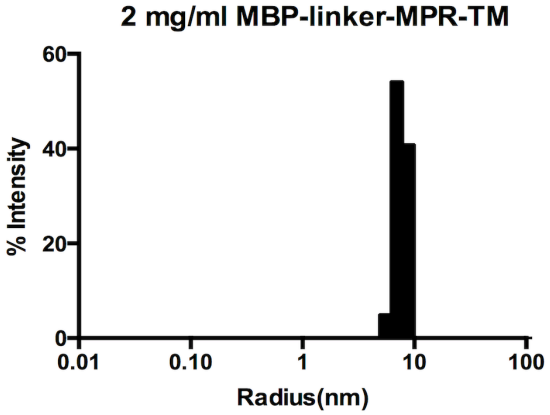


Fig. 3.3. DLS measurement of the purified MBP-linker-MPR-TM protein showed one monodisperse peak at 7.7 ± 1.0 nm.

3.4.5 Secondary Structure Estimation of MBP-linker-MPR-TM by Circular Dichroism (CD) Spectroscopy

It is essential to test if the purified MBP-linker-MPR-TM is folded well by CD before carrying out crystallization screens. The result of CD measurement (Fig. 3.4A)

displayed one positive band at 193 nm and two negative bands at 208 and 222 nm, which is characteristic of a protein with a large fraction of α-helices (Greenfield, 2006). Analysis of the secondary structure content of MBP-linker-MPR-TM by CDPro provided the following values: 39.3 ± 2.3% α-helix and 13.5 ± 1.8% β-sheet. Comparison of the CD spectrum of MBP (fraction A2 in Fig. 3.1C) with that of MBP-linker-MPR-TM (fraction A1 in Fig. 3.1C) demonstrated that the two minima of MBP-linker-MPR-TM were slightly lower than that of MBP (Supporting Fig. 3.3, page 79). Analysis of the secondary structure content of MBP by CDPro showed that MBP contained 33.0 ± 2.6% α-helix and 15.2 ± 2.4% β-sheet. The higher α-helix content in MBP-linker-MPR-TM might be due to the presence of MPR-TM. The crystal structure of the pre-fusion and post-fusion HIV-1 gp41 indicate that the MPR region in these status may form an α-helix (Buzon et al., 2010; Shi et al., 2010; Pancera et al., 2014) and the cryo-EM structure of the uncleaved HIV-1 envelope glycoprotein indicate that the TM region may also be helical in the complex (Mao et al., 2013). CD measurements here function as a quality control of the purified protein sample to make sure that the protein is well folded, which is a prerequisite condition prior to crystallization screens.

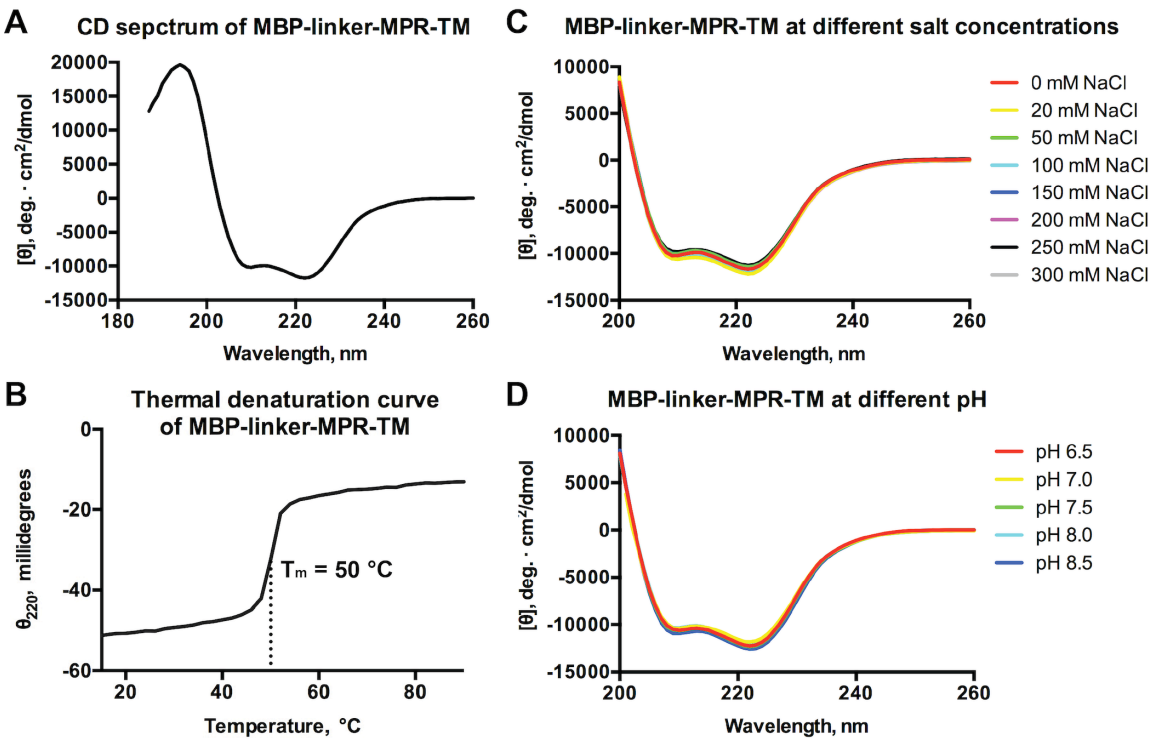


Fig. 3.4. CD spectra of MBP-linker-MPR-TM. (A) CD spectrum of MBP-linker-MPR-TM displayed one positive band at 193 nm and two negative bands at 208 and 222 nm, which is characteristic for α-helical proteins. Buffer: 100 mM NaF, 20 mM NaH_2PO_4, pH 7.5 and 0.02% βDDM. Protein concentration was 0.26 mg/mL. (B) Thermal denaturation curve of MBP-linker-MPR-TM measured at 220 nm and at a rate of temperature change of 2 °C/min. The apparent midpoint "denaturation" temperature of the protein was 50 °C. Protein concentration was 0.5 mg/mL. (C) Effect of ionic strength on the secondary structure of MBP-linker-MPR-TM. CD measurements were performed in 20 mM Tris, pH 7.5, 0.02% βDDM containing 0 to 300 mM NaCl. (D) Effect of pH on the secondary structure of MBP-linker-MPR-TM. CD spectra were recorded in 150 mM NaCl, 0.02% βDDM at pH from 6.5 to 8.5.

3.4.6 Thermal Stability of MBP-linker-MPR-TM

In order to estimate the thermal stability of the MBP-linker-MPR-TM protein and to guide decisions of the optimal temperature at which crystallization screens should be performed, CD spectroscopy was used to measure the apparent midpoint "denaturation" temperature of MBP-linker-MPR-TM. The CD spectra were recorded at 220 nm from 15 to 90 °C with a temperature change slope at 2 °C/min. The result showed that the apparent midpoint "denaturation" temperature of the sample was 50 °C (Fig. 3.4B),

which indicated good thermal stability. According to the melting curve obtained (Fig. 3.4B), the MBP-linker-MPR-TM protein would not denature below 45 °C. Therefore, crystallization screens of MBP-linker-MPR-TM could be performed at room temperature or lower temperatures.

3.4.7 Effect of Ionic Strength and pH on the Secondary Structure of MBP-Linker-MPR-TM

Ionic strength and pH are two important parameters which could be adjusted in crystallization screens to promote the crystal formation. We aimed to investigate the effect of ionic strength and pH on the secondary structure of the MBP-linker-MPR-TM protein to determine the range in which these two parameters could be adjusted to ensure that the protein would be correctly folded under crystallization conditions. We used CD measurements under various conditions of pH and ionic composition of the medium to determine how they would affect the secondary structure of MBP-linker-MPR-TM. The purified MBP-linker-MPR-TM protein was concentrated to 5 mg/mL, diluted into buffers containing 0 to 300 mM NaCl at pH 7.5 to test the effects of ionic strength (Fig. 3.4C), or diluted into buffers containing 150 mM NaCl at different pH values ranging from 6.5 to 8.5 to test the effect of pH (Fig. 3.4D). The samples were equilibrated overnight at 4 °C and CD spectra were measured the next day.

The CD spectra of MBP-linker-MPR-TM in buffers containing different ionic strength (Fig. 3.4C) and pH (Fig. 3.4D) were very similar: they all showed negative bands at 208 and 222 nm of almost the same molar ellipticity (deg. · cm^2/dmol). Our results indicate that the fusion protein MBP-linker-MPR-TM is stable under all tested conditions and crystallization screens of MBP-linker-MPR-TM could potentially be

carried out at a large ionic strength and pH range, at which the protein will not be denatured.

3.4.8 MBP-linker-MPR-TM is Stable at High Protein Concentration

Crystallization screens often start with a protein concentration at 10 mg/mL although the optimal protein concentration depends on the molecular weight of the protein, the stability of the protein at high concentration and some other properties. Here SEC and DLS were used to test the stability of MBP-linker-MPR-TM at high protein concentration. The protein was concentrated to 10 mg/mL by a 100-kDa cut-off concentrator to avoid concentration of the detergent βDDM, whose micelle molecular weight is 68 kDa (Gong et al., 2014). The concentrated sample was stored at 4 °C and analyzed by SEC and DLS at day 1, 3 and 7 respectively.

The size exclusion chromatogram (Fig. 3.5A) revealed a main peak eluting at about 10 mL and a very small minor peak eluting at 14.5 mL from day 1 to day 7. The main peak contained oligomeric MBP-linker-MPR-TM/detergent complex while the small minor peak at 14.5 mL might represent the MBP protein. The small minor peak increased slightly from day 1 to day 7, which represented 0.4%, 0.7% and 1.5% of the total area respectively (insert in Fig. 3.5A), but was still very small compared with the main peak after one week storage at 4 °C.

The DLS measurements (Fig. 3.5BCD and Supporting Table 3.3) revealed a narrow peak at 7.1 to 7.6 nm with the percent polydispersity below 12% from day 1 to day 7, which indicates that the protein sample is in a monodisperse condition for seven days at 4 °C (Fig. 3.5BCD and Supporting Table 3.3). Although a small aggregation peak at 90.2 nm was detected on day 7, the aggregation peak represented only 0.1% of the total

amount (mass) of the protein because the increase in scattered intensity is proportional to r^6 (r is the particle radius) (Hoo et al., 2008). In conclusion, 10 mg/mL MBP-linker-MPR-TM was homogeneous and monodisperse for 7 days and crystallization screens could be performed at this concentration.

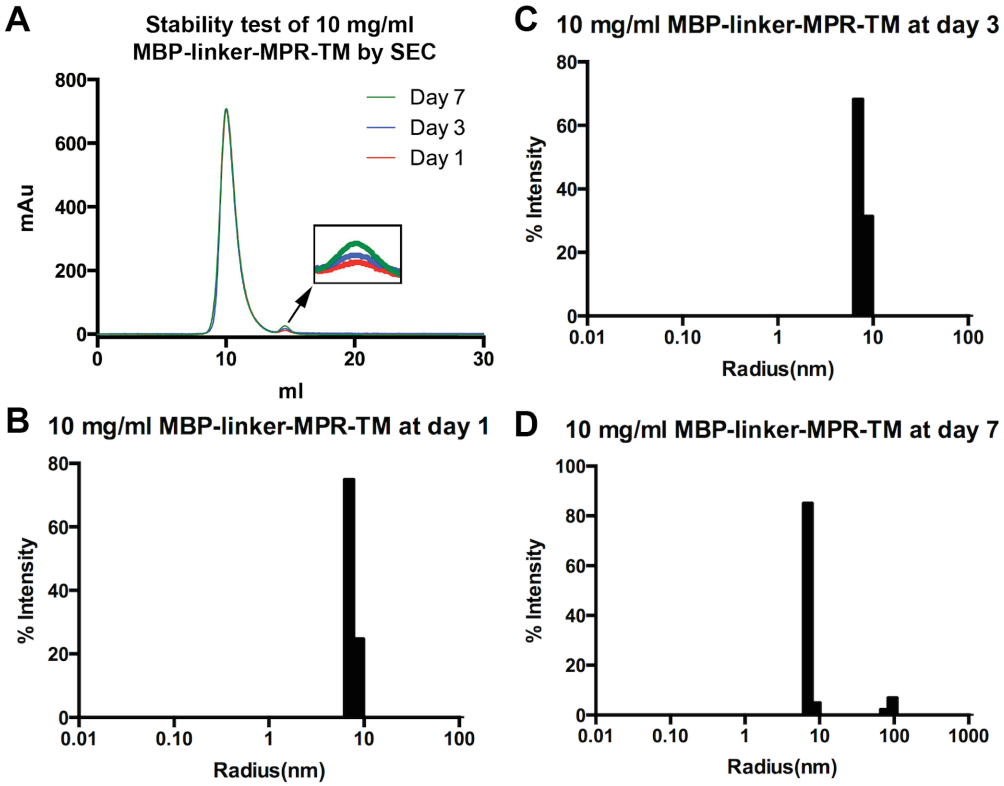


Fig. 3.5. Stability test of 10 mg/mL MBP-linker-MPR-TM. 10 mg/mL MBP-linker-MPR-TM was stored at 4 °C and measured by SEC and DLS on day 1, 3 and 7. (A) Stability test of 10 mg/mL MBP-linker-MPR-TM by SEC. Insert: magnification of the degradation peak, which indicated slight increase of the MBP-linker-MPR-TM degradation from day 1 to day 7. (B-D) Stability test of 10 mg/mL MBP-linker-MPR-TM by DLS. The protein sample was homogeneous and monodisperse for 7 days at 4 °C; only a little protein aggregation was detected on day 7.

3.4.9 MBP-linker-MPR-TM is Recognized by the Broadly Neutralizing mAbs 2F5 and 4E10

An important feature of gp41 is that it contains the epitopes for broadly neutralizing antibodies 2F5 and 4E10, which makes it an attractive target for vaccine

design. However, the epitopes for 2F5 and 4E10 cannot bind the antibodies in the pre-fusion and post-fusion conformations of gp41. HIV-1 gp41 mediates the membrane fusion between target cell and virus through its own conformational change: native trimer prior to the interaction between gp120 and CD4, pre-hairpin intermediate and postfusion trimer of hairpins (or a six-helix bundle) (Chan and Kim, 1998). It was reported by Frey at al that gp41 in its prefusion conformation could not interact with 2F5 or 4E10; gp41 in its postfusion conformation binds 2F5 very weakly ($K_d \approx 1.4$ μM), while gp41 in its pre-hairpin intermediate state binds 2F5 and 4E10 very strongly ($K_d < 10$ nM) (Frey et al., 2010). Therefore, it is of great interest to measure the binding affinities of MBP-linker-MPR-TM to 2F5 and 4E10 by surface plasma resonance (SPR), from which the conformational information of MPR-TM could be estimated.

In SPR measurements, mAbs 2F5 and 4E10 were immobilized onto the surface of the gold chip. The negative control MBP (fractions A2 from SEC purification, Fig. 3.1B) and the purified MBP-linker-MPR-TM protein were passed over the ligand surface to measure their binding affinities to 2F5 and 4E10. Identical injections over blank surfaces were subtracted from the data for kinetic analysis. The result showed that mAbs 2F5 and 4E10 bound strongly to MBP-linker-MPR-TM with a K_D of 0.5 ± 0.2 nM for the 2F5 Ab and a K_D of 0.3 ± 0.1 nM for the 4E10 Ab (Fig. 3.6 and Table 3.1). This result strongly indicates that MBP-linker-MPR-TM may be in the pre-hairpin intermediate conformation, at which stage the epitopes for 2F5 and 4E10 are exposed and are available for antibody binding. The weak unspecific binding of 2F5 and 4E10 to the negative control MBP (fractions A2 from SEC purification, Fig. 3.1B) may be due to the existence of high concentration of maltose (10% maltose) in the antibody solution. However, the

binding affinities of MBP-linker-MPR-TM to mAbs 2F5 and 4E10 are much stronger than that of MBP with a K_D of 674.3 ± 306.9 nM for the 2F5 Ab and a K_D of 543.8 ± 337.7 nM for the 4E10 Ab. Therefore, the strong binding of MBP-linker-MPR-TM to mAbs 2F5 and 4E10 is due to the presence of MPR-TM instead of MBP.

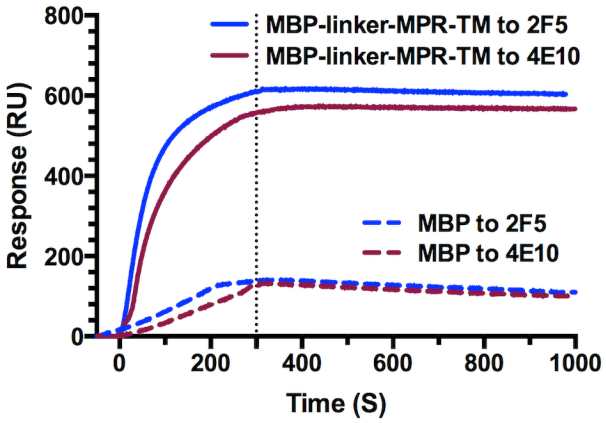


Fig. 3.6. Surface plasma resonance (SPR) analysis.

Immobilized ligand	Flowing analyte	k_a, s^{-1}	k_d, s^{-1}	K_D, nM
2F5	MBP-linker-MPR-TM	7.0 ± 1.2E4	3.5 ± 1.2E-5	0.5 ± 0.2
4E10	MBP-linker-MPR-TM	4.8 ± 0.1E4	1.8 ± 0.5E-5	0.3 ± 0.1

Table 3.1. Association and dissociation rate constants of MBP-linker-MPR-TM derived from SPR analysis (Results are average of four independent measurements and are listed as in mean ± SD).

3.4.10 Change the Linker of MBP-linker-MPR-TM to Three Alanine Residues

Extensive crystallization trials with screens of thousands of conditions using both vapor diffusion method and liquid cubic phase did not yield crystals of MBP-linker-MPR-TM with diffraction quality. Similar problems were reported by Center et al., who described that crystals could not be obtained when the ectodomain of gp21 (human T cell leukemia virus type I transmembrane protein) fused to the C-terminus of MBP via a flexible linker containing 25 residues, but were obtained after changing the linker to three alanine residues (Center et al., 1998). Their strategy was applied to our study. The 42 aa-

long linker containing a TEV protease recognition site was replaced by a short three-alanine linker (Fig. 3.7). In addition, three charged residues at the C-terminus of MBP (Glu-370, Lys-373 and Asp-374) were replaced by alanines to avoid potential electrostatic repulsion between MBP monomers if MPR-TM forms a trimer (Center et al., 1998). The new fusion protein was dubbed MBP-AAA-MPR-TM (Fig. 3.7).

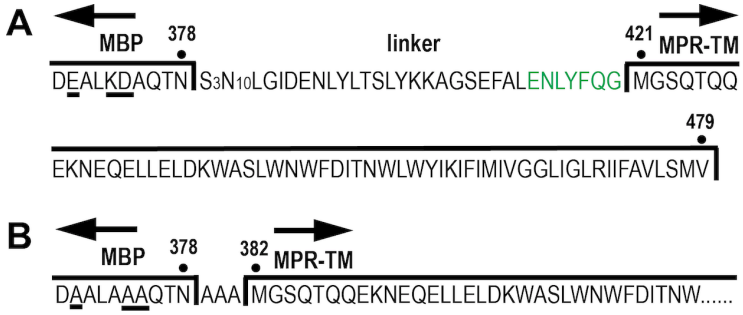


Fig. 3.7. Schematic representation of the sequence of MBP-linker-MPR-TM (A) and MBP-AAA-MPR-TM (B). Amino acid changes in the C-terminus of MBP are underlined.

We employed the purification procedure that was devised for MBP-linker-MPR-TM to purify MBP-AAA-MPR-TM. Ni-affinity chromatography of MBP-AAA-MPR-TM resulted in a single band at the expected apparent MW (49 kDa, Fig. 3.8B), unlike the case of preparations of the longer-linker fusion protein, which contained prominent contaminating degradation products (Fig. 3.1A and B). The purity of the affinity chromatography eluate was further demonstrated by a subsequent SEC, which exhibited a single peak (Fig. 3.8A). About 60 mg of pure MBP-AAA-MPR-TM protein was obtained from 1 liter of bacterial culture.

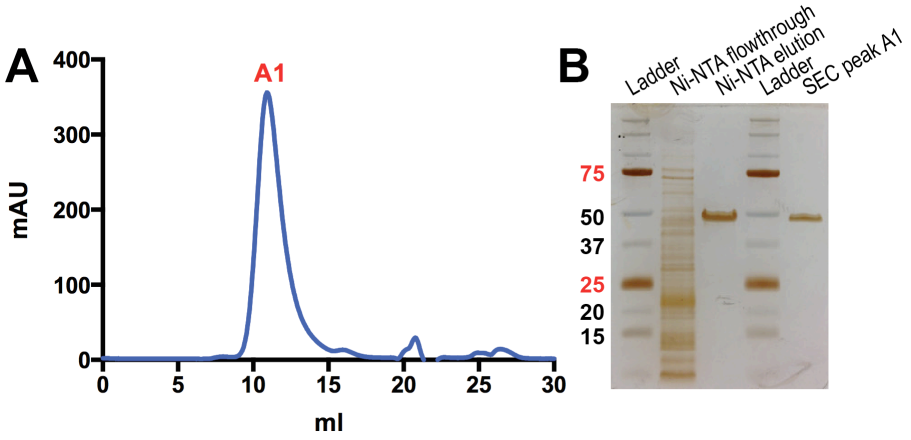


Fig. 3.8. Purification of MBP-AAA-MPR-TM. (A) SEC of the Ni-affinity elution of MBP-AAA-MPR-TM. (B) Silver stained SDS-PAGE analysis of Ni-NTA flowthrough, Ni-NTA elution and SEC peak A1.

3.4.11 CD, DLS and SPR measurements of MBP-AAA-MPR-TM

CD spectroscopy measurements showed that the CD spectrum of MBP-AAA-MPR-TM was very similar to that of MBP-linker-MPR-TM (Fig. 3.9A). Analysis of the secondary structure content of MBP-AAA-MPR-TM by CDPro provided the following values: 38.8 ± 2.3% α-helix and 13.9 ± 2.6% β-sheet, which is in good agreement with the secondary structure content of MBP-linker-MPR-TM (39.3 ± 2.3% α-helix and 13.5 ± 1.8% β-sheet). These results indicate that changing of the linker to three alanine residues did not affect the secondary structure of the recombinant protein.

DLS analysis displayed one monomeric peak at 7.4 ± 0.8 nm, which corresponded to a protein/detergent complex of 360 kDa (Fig. 3.9B and Supporting Table 3.2). The size estimation of MBP-AAA-MPR-TM is slightly smaller than that of MBP-linker-MPR-TM (7.7 ± 1.0 nm and ~400 kDa, Fig. 3.3 and Supporting Table 3.2), which reflects the slightly smaller mass of MBP-AAA-MPR-TM.

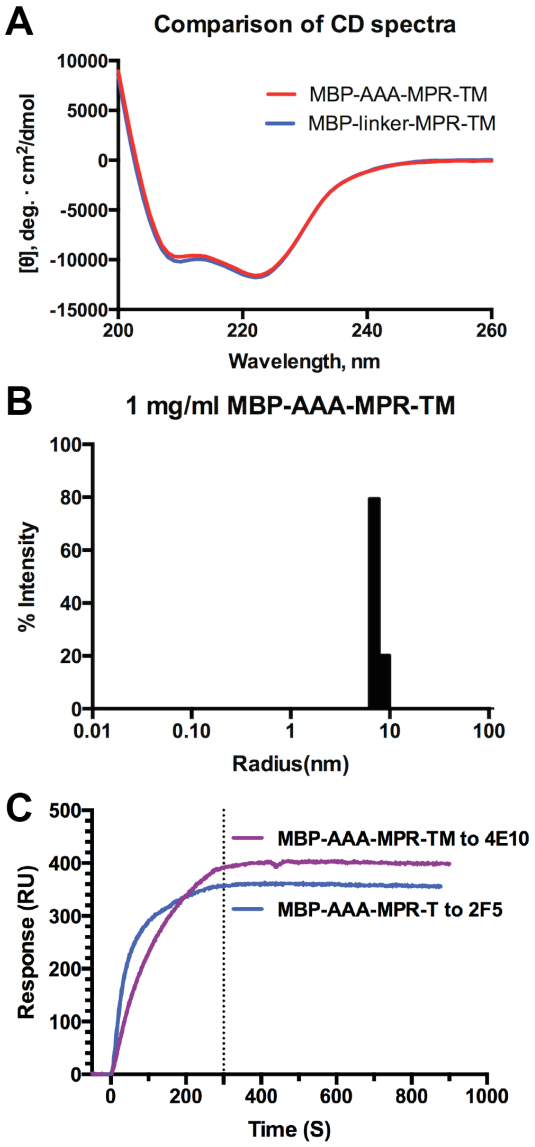


Fig. 3.9. CD, DLS and SPR measurements of MBP-AAA-MPR-TM. (A) Comparison of CD spectra of MBP-linker-MPR-TM and MBP-AAA-MPR-TM. (B) DLS measurement of MBP-AAA-MPR-TM showed one monodisperse peak at 7.4 ± 0.8 nm. (C) SPR analysis.

Immobilized ligand	Flowing analyte	k_a, s^{-1}	k_d, s^{-1}	K_D, nM
2F5	MBP-AAA-MPR-TM	2.1 ± 0.3E4	2.1 ± 0.8E-5	1.0 ± 0.5
4E10	MBP-AAA-MPR-TM	7.3 ± 1.3E3	3.7 ± 1.8E-6	0.5 ± 0.1

Table 3.2. Association and dissociation rate constants of MBP-AAA-MPR-TM derived from SPR analysis (Results are average of four independent measurements and are listed as in mean±SD).

As the longer linker could provide more flexibility of the MPR region for binding the broadly neutralizing Abs, we conducted SPR measurements to determine if the MBP-AAA-MPR-TM protein still binds the Abs with high affinity. The SPR result shown in Table 3.2 indicated that both 2F5 and 4E10 bind to MBP-AAA-MPR-TM with nanomolar to subnanomolar affinities: K_D values of 1.0 ± 0.5 nM were determined for 2F5 Ab and K_D values of 0.5 ± 0.1 nM were determined for 4E10 Ab (Fig. 3.9C and Table 3.2). It is therefore concluded that MBP-AAA-MPR-TM protein binds 2F5 and 4E10 Abs with nanomolar to subnanomolar affinities which makes it an excellent candidate for vaccine developments against HIV-1.

3.5 Conclusion

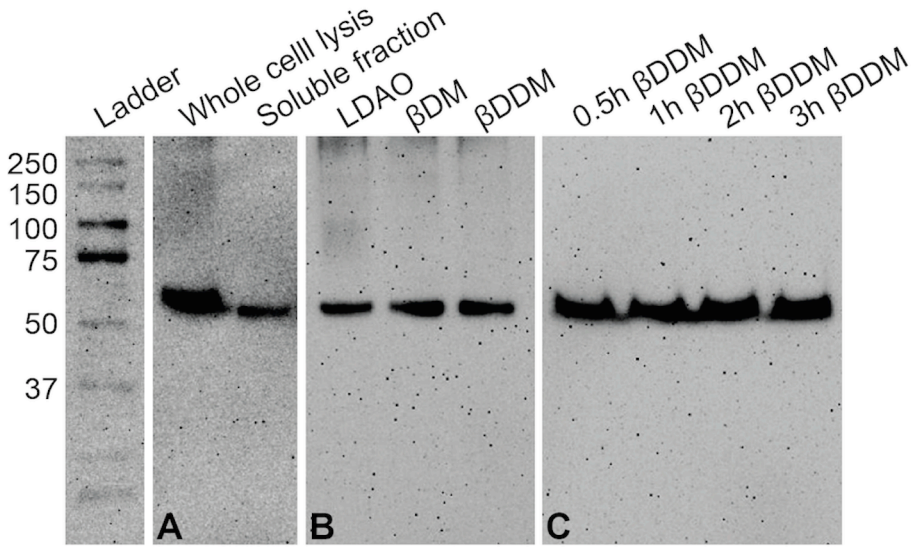
In summary, we describe here the expression and purification of two recombinant protein variants consisting of a fusion between MBP and MPR-TM of HIV-1 gp41. In one of these variants, MPR-TM was fused to the C-terminus of MBP via a 42 aa-long linker containing a TEV protease recognition site. In the second variant, the long linker was replaced by a short and structured peptide consisting of three alanine residues. Both proteins were purified to homogeneity and were shown to be stable in solution under various conditions and remained monodisperse over time. Both proteins were able to strongly interact in solution with the broadly neutralizing mAbs 2F5 and 4E10 with nanomolar to subnanomolar affinities, in good agreement with our previously published results concerning MPR-TM (Gong et al., 2014). The longer linker variant MBP-linker-MPR-TM was not amenable to crystallization under exhaustive screening. The crystallization experiments of MBP-AAA-MPR-TM are currently undergoing.

3.6 Supporting Information

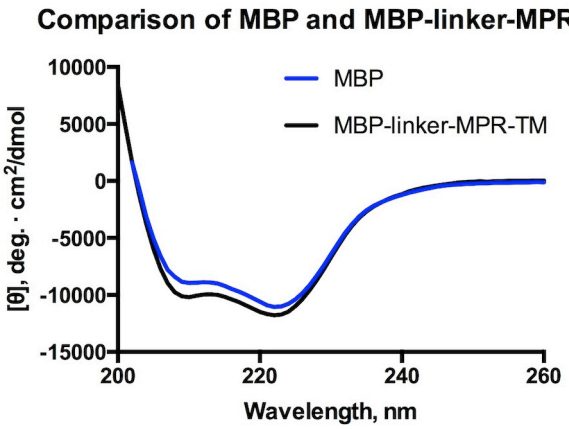
3.6.1 Supporting Figures

1 80
MGHHHHHHHHHSKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGG
160
YAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
240
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP
320
WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEE
379 400
ELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIDENLYL
414 421 479
TSLYKKAGSEFALENLYFQGMGSQTQQEKNEQELLELDKWASLWNWFDITNWLWYIKIFIMIVGGLIGLRIIFAVLSMV

Supporting Fig. 3.1. Amino acid sequence of MBP-linker-MPR-TM. The color scheme is as followings: blue, 8His-MBP (residues 1-368); black and green: 42 aa-long linker containing a TEV protease recognition site (residues 379-420); green: TEV protease recognition site (residues 414-420); red: MPR-TM (residues 421-479).



Supporting Fig. 3.2. Anti-His Western blot analysis of the detergent solublization of MBP-linker-MPR-TM. (A) A small portion of MBP-linker-MPR-TM protein was found to be in the supernatant. (B) Comparison the extract efficiency of different detergents. (C) Determination of the time needed for efficient detergent extraction.



Supporting Fig. 3.3. Comparison of the CD spectra of MBP and MBP-linker-MPR-TM.

3.6.2 Supporting Tables

Primer	Sequence (5' to 3')
MBP-MPR-fuseF1	CTGATTTATAACAAAGATCTGCTGCCGAACCCG
insert1-PCR1R	CAGGGCTGCATCGACAGTC
insert1-PCR2R	ATTAGTCTGCGCGGCTGCCAGGGCTGCATCGACAGTC
insert2-PCR1F	GCAGCTGCCATGGGATCTCAAACTCAACAAGAGAAG
MBP-MPR-fuseR1	CGATGGTACCGTCGACGTCCTACAGGCGCGCC
MBP-MPR-fuseF2	GCCGCGCAGACTAATGCAGCTGCCATGGGATCTC

Supporting Table 3.1. Primer sequences

Intensity Distribution	Radius (nm)	% Polydispersity	Mw-R (kDa)	% Intensity	% Mass
Peak 1	7.7 ± 0.5	13.4	397	100.0	100.0

Supporting Table 3.2. DLS measurement of purified MBP-linker-MPR-TM

Intensity Distribution	Radius (nm)	Polydispersity (%)	Mw-R (kDa)	Intensity (%)	Mass (%)
Day 1					
Peak 1	7.5 ± 0.4	10.9	370	100.0	100.0
Day 3					
Peak 1	7.6 ± 0.4	11.5	385	100.0	100.0
Day 7					
Peak 1	7.1 ± 0.2	6.1	330	90.5	99.9
Peak 3	90.2 ± 4.4	9.7	126571	9.5	0.1

Supporting Table 3.3. DLS measurements of 10 mg/ml MBP-linker-MPR-TM subjected to prolonged incubation at 4 °C. Samples were taken at the indicated time points.

Intensity Distribution	Radius (nm)	Polydispersity (%)	Mw-R (kDa)	Intensity (%)	Mass (%)
Peak 1	7.4 ± 0.4	10.2	360	100.0	100.0

Supporting Table 3.4. DLS measurement of purified MBP-AAA-MPR-TM

CHAPTER 4

CRYSTALLIZATION AND CRYSTALLOGRAPHIC ANALYSIS OF CRYSTALS OBTAINED FROM A SOLUTION OF MBP-AAA-MPR-TM

4.1 Abstract

The envelope glycoprotein of HIV-1 (gp120/gp41) plays an essential role in the initial process of viral infection. Three broadly neutralizing antibodies (2F5, 4E10 and 10E8) bind to the membrane proximal region (MPR, residues 649-683) of HIV-1 gp41, which makes the MPR a fascinating template for developing vaccines that can elicit antibodies with similar properties. The transmembrane (TM) domain (residues 684-795) of gp41 is involved in anchoring the envelope glycoprotein complex in the viral membrane but its atomic resolution structure is still unknown. In order to provide a structural basis for rational vaccine design, we attempted to crystallize a fusion protein of maltose binding protein (MBP) with MPR-TM of gp41 connected by a three-alanine linker (MBP-AAA-MPR-TM). The crystal diffracted to 2.5 Å after crystallization optimization. Further analysis of the diffraction data indicated that the crystals were twinned. The final structure demonstrated that MBP oligomerized as dimers of trimers, but the electron density did not extend beyond the linker region. SDS-PAGE and MALDI-TOF MS analysis of dissolved crystals confirmed that the crystal contained MBP only. As the MBP-AAA-MPR-TM protein is highly purified and is very stable, the cleavage may be induced by the process of crystallization. Based on comparison of the MBP trimer reported here with published trimeric MBP fusion structures and the fact that MBP on its own crystalizes as a monomer, it is indicated that MBP might form such a dimer of trimers induced by MPR-TM.

4.2 Introduction

Human immunodeficiency virus type 1 (HIV-1) is a retrovirus that causes AIDS. The envelope glycoprotein (Env) of HIV-1 mediates virion attachment to cells and subsequent fusion between viral and cellular membranes (Chan et al., 1997). The HIV-1 Env consists of a surface subunit, gp120 and a transmembrane subunit, gp41, formed by the proteolytic cleavage of a polyprotein precursor, gp160 (Chan and Kim, 1998). Gp120 and gp41 oligomerize as trimers and are noncovalently associated on the surface of the virion (Mao et al., 2012).

Despite considerable sequence variety, the membrane proximal region (MPR, residues 649-683) and transmembrane (TM) domain (residues 684-705) of HIV-1 gp41 are highly conserved among different subtypes of HIV-1 (Miyauchi et al., 2005; Shang et al., 2008; Checkley et al., 2011) and have many important biological functions. Transcytosis is one of the initial mucosal transmission pathways of HIV-1, which translocate the virus across epithelial cells. Transcytosis is initiated when HIV-1 Env binds to the epithelial cell galactosyl ceramide (GalCer) and the minimal region required for gp41 to bind GalCer is MPR (Alfsen and Bomsel, 2002). Furthermore, MPR contains epitopes for three broadly neutralizing antibodies 2F5 (Purtscher et al., 1994), 4E10 (Zwick et al., 2001) and 10E8 (Huang et al., 2012). The primary role of gp41 TM domain is to anchor the envelope glycoprotein complex in the viral membrane (Yue et al., 2009).

The HIV-1 Env undergoes dramatic conformational change during membrane fusion process. The first state is the pre-fusion state that exists before host-cell encounter and receptor binding. The second state is the pre-fusion intermediate state, where gp41 is extended and epitopes for broadly neutralizing antibodies are available for antibody

binding. The final state is the post-fusion "six-helix bundle" conformation, whose structure has been well studied by both X-ray crystallography and NMR (reviewed by Merk and Subramaniam (Merk and Subramaniam, 2013)). The structure of a trimeric pre-fusion form of HIV-1 Env has recently been solved at 3.5 Å by X-ray crystallography (Pancera et al., 2014). The HIV-1 Env trimer was captured in a potential pre-fusion mature close state by neutralizing antibodies PGT122 (Walker et al., 2011) and 35O22 (Huang et al., 2014). In the prefusion conformation, the ectodomain of gp41 contains four α-helices and one β-strand, which is very different from the post-fusion "six-helix bundle" conformation. This pre-fusion conformation of gp41 indicated that structural arrangements are needed for fusion activation. However, gp41 in neither pre-fusion nor post-fusion conformation could react with the broadly neutralizing antibody 2F5 (Frey et al., 2008), whose epitope is on MPR. Accessibility of antibodies to the epitope is of great significance for vaccine design. Here, maltose binding protein (MBP) was used as a fusion partner to overexpress MPR-TM of gp41. The purified MBP/MPR-TM recombinant protein strongly interacts with monoclonal antibodies 2F5 and 4E10 and thereby may represent an immunologically relevant conformation mimicking a pre-hairpin intermediate of gp41. Structural studies of the MBP/MPR-TM recombinant protein might be important for the structural-based design of vaccine against HIV-1.

Crystals from a highly purified and monodisperse solution of the MBP/MPR-TM recombinant protein were obtained when MPR-TM was fused to the C-terminus of HIV-1 gp41 through a three-alanine linker. Crystals of MBP-AAA-MPR-TM diffracted to 2.5 Å after crystallization optimization. However, attempts of molecular replacement using MBP as the search model failed due to the presence of four-domain twinning.

Crystallization and crystallographic analysis of crystals from MBP-AAA-MPR-TM are reported here. The final structure demonstrated that MBP oligomerized as dimers of trimers, but the electron density did not extend beyond the linker region. Analysis of the crystals revealed that they contained MBP only. Comparison of the MBP trimer reported here with published trimeric MBP fusion structures indicated that MBP might form such a trimer under the influence of MPR-TM.

4.3 Materials and Methods

4.3.1 Crystallization and Data Collection

Expression and purification of MBP-AAA-MPR-TM were carried out as described in Chapter 3. Initial crystallization condition of MBP-AAA-MPR-TM was found using commercial screens from Hampton Research, Qiagen and Molecular Dimensions. Very thin needle crystals (Fig. 4.1A) were obtained using hanging drop vapor diffusion method with a 1:1 drop ratio of 2 µl of protein (~15 mg/ml in 20 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.015% βDDM) plus 2 µl of reservoir solution (0.1 M Tris-HCl, pH 8.5, 25% PEG 8000) equilibrated against 800 µl of reservoir solution. However, these needles, which diffracted to ~7 Å, were so thin that it was nearly impossible to separate and shoot one of them. The diffracted image was overlaps of diffractions from several needle crystals and could not be indexed. The initial crystallization condition was further optimized by varying the concentration of PEG 8000 and/or adding the CALIXAR^{TM} additives (Molecular Dimensions). The optimal crystallization condition without additive was as follows: 2 µl of protein (~15 mg/ml in 20 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.015% βDDM) was mixed with 2 µl of reservoir solution (0.1 M Tris-HCl, pH 8.5, 18 - 25% PEG 8000) and equilibrated over 800 µl of reservoir solution. The final conditions with

CALIXAR™ additives were as follows: 2 µl of protein (~15 mg/ml in 20 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.015% βDDM) was mixed with 0.4 µl of CALIXAR™ additive and 1.6 µl of reservoir solution (0.1 M Tris-HCl, pH 8.5, 18 - 25% PEG 8000). The mixture was equilibrated over 800 µl of reservoir solution, which did not contain any additive. Crystals were cryoprotected in 0.1 M Tris-HCl, pH 8.5, 30% PEG 8000 for a few seconds before being flash frozen in liquid nitrogen.

Diffraction data were collected at Advanced Light Source (ALS), beamline 821 and Advanced Photon Source (APS), Sector 19BM, 23B and 23D. The optimal data set was collected using an ADSC 315 CCD detector at ALS beamline 821 at a wavelength of 0.9999 Å. The distance between the crystal and the detector is 320 mm. Three data sets were collected from one crystal. The oscillation per frame is 0.3333° and 360 frames in total were collected, which corresponding to 120°.

4.3.2 Structure Determination and Refinement

Diffraction images were processed using XDS (Kabsch, 2010) and the integrated intensities were scaled using AIMLESS from the CCP4 program suite (Collaborative Computational Project, 1994). The phenix.xtriage module of the PHENIX program (Adams et al., 2002; Adams et al., 2010) was used to perform the twinning tests of collected data. Molecular replacement was carried out with the program Phaser (McCoy et al., 2007) or Molrep (Vagin and Teplyakov, 1997) using the structure of an open-form MBP (PDB code IFQA) as the search model. The structure was refined using REFMAC (Murshudov et al., 1997). All structural figures were produced using PyMol (Schrodinger, 2010).

4.3.3 SDS-PAGE and MALDI-TOF MS

Tricine-SDS-PAGE was performed with stacking and resolving gels of 4 % and 8 %, respectively (Schagger, 2006) and stained by Instant*Blue* (Expedeon).

The protocol of MALDI-TOF MS analysis is described in detail previously (Gong et al., 2014) and will be only outlined here. Applied Biosystems DE-STR MALDI-TOF mass spectrometer was used to accurately measure the molecular weight of the protein. A saturated solution in 50% acetonitrile/H_2O and 0.1% trifluoroacetic acid (TFA) was used as matrix. Spectra were collected in a positive linear mode over a mass range from 15000 to 90000 Da. Final results represented the average of 10 separate spectra.

4.4 Results

4.4.1 Crystallization

The CALIXAR^{TM} Additive Kit was used to optimize the initial crystallization condition. It is reported that the CALIXAR^{TM} additives promote the crystallization of membrane proteins by intercalating between membrane protein molecules and forming salt-bridges with positively charged residues on the surface of membrane proteins (Falson et al., 2010; Matar-Merheb et al., 2011). With a hydrophilic charged head and a hydrophobic alkyl tail of different lengths (Supporting Fig. 4.1), the CALIXAR^{TM} additives (C4Cn, where n represents the number of carbons in the alkyl tail) have detergent-like properties. They form micelles of 5-24 nm with negatively charged head outside and hydrophobic tail inside. It is the additive micelles that intercalate between membrane protein/detergent complexes and therefore facilitate the crystal formation via ionic interactions.

Crystallization optimization trials were carried out with the CALIXAR™ additives or without any additive. Big needle crystals (Fig. 4.1B) grew after three weeks in conditions with every individual CALIXAR™ additive from the kit (1 to 10 mM C4C1, 1 to 10 mM C4C3, 5 to 35 mM C4C5, 5 to 15 C4C7, 0.3 to 3 mM C4C9 and 0.3 to 3 mM C4C10, respectively) or without any additive. X-ray diffraction data were collected as discussed in Materials and Methods. Crystals with 20% PEG 8000 and 5 mM C4C1 diffracted to the highest resolution (2.5 Å).

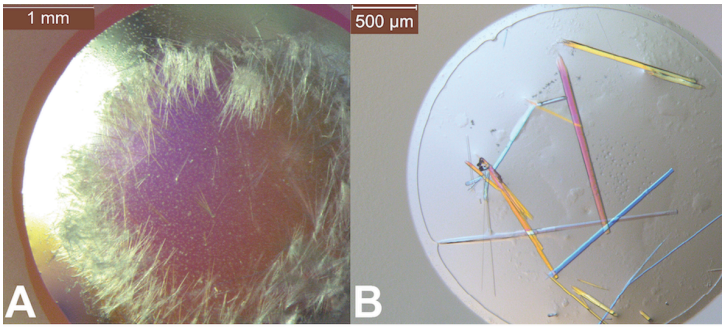


Figure. 4.1. Crystals of MBP-AAA-MPR-TM. (A) Very thin needle crystals of MBP-AAA-MPR-TM were obtained from initial crystallization screens. (B) Big need crystals of MBP-AAA-MPR-TM were obtained after crystallization optimization.

4.4.2 Space group Determination and Structure Solution

The crystal of MBP-AAA-MPR-TM with additive C4C1 diffracted to 2.5 Å. The most likely space group was $P6_2$, with unit cell parameters a = 171.93, b = 171.93, c = 70.08, $\alpha = \beta = 90°$ and $\gamma = 120°$. The structure of MBP was used as the search model for molecular replacement. MBP consists of two distinct domains connected by a hinge region that move from an open to a closed conformation when the ligand maltose binds to the hinge region (Bertz and Rief, 2009). Since no maltose was used during protein purification and crystallization, it was assumed that MBP was in the open conformation. Therefore, the structure of an open-form MBP (PDB code IFQA) was initially used as the search model for molecular replacement. However, only partial solution was obtained in

$P6_2$ from molecular replacement. A considerable part of the structure could not be built and refinement stalled at high R factors. The structure of a closed-form MBP (PDB code 1FQC) was then used as the search model. However, no improvement was achieved compared with using the open-form MBP as the search model. In addition, attempts to use the separate domain as search models for molecular replacement also failed.

By using the phenix.xtriage module of the PHENIX program (Adams et al., 2002; Adams et al., 2010), it was determined that the diffraction data were twinned. Strong twinning adds extra symmetry to the diffraction data, which can mislead the space-group assignment (Barends et al., 2005) and lead to the failure of molecular replacement in the wrong space group. Therefore, molecular replacement with an open-form MBP (PDB code IFQA) and refinement were repeated using the lower symmetry space groups P3, $P3_1$ and $P3_2$. Molecular replacement in space group $P3_2$ using Phaser yielded a partial solution of four MBP monomers in one asymmetric unit, which had reasonable geometry and no clashes. However, there was still an appreciable part of the structure that could not be built.

The most probable content of the asymmetric unit, estimated using Matthews from the CCP4 program suite (Collaborative Computational Project, 1994), contains 5-6 molecules with a solvent content between 50.67 and 40.80 under the assumption that the true space group is $P3_2$. The $P3_2$-model with four MBP monomers obtained using Phaser previously was refined using Refmac (Murshudov et al., 1997). The fifth MBP monomer was positioned using Molrep (Vagin and Teplyakov, 1997) and refined. This step was repeated to position the sixth MBP monomer. Please note that the amplitude based twin refinement was essential to be used during refinement; otherwise the position of the fifth

MBP monomer was not correctly assigned. The final $P3_2$-model with six MBP monomers (dimer of trimers) underwent several rounds of refinement with global NCS, no NCS and local NCS successively. The final crystallographic data statistics are summarized in Table 4.1.

4.4.3 Space Group Validation with Zanuda

The space-group assignment was further validated by Zanuda, a program that was developed by Lebedev and Isupov to automate the validation of space group assignment, especially in presence of pseudosymmetry or twinning (Lebedev and Isupov, 2014). A $P3_1$-model with six MBP monomers was prepared using similar strategies of preparing the $P3_2$-model. The $P3_1$-model was finally refined to an R factor of 0.283 and an R_{free} of 0.317 and submitted to Zanuda. As a result, the input $P3_1$-model is converted into $P3_2$ model, which conforms the correct space group assignment of $P3_2$.

4.5 Discussion

4.5.1 Overview Structure

The structure was refined to 2.49 Å in space group $P3_2$ with good refinement statistics (Table 4.1). The overall final R factor and R_{free} were 0.159 and 0.194 respectively. Each asymmetric unit (ASU) contains 6 monomers of MBP in the open conformation, which form two layers of trimers (or dimer of trimers, top view in Fig. 4.2A). These two trimers are oriented in different direction (front and back trimer in Fig. 4.2A) forming a "sandwich" with neighbor trimers. The C-termini of MBP are at the center of the trimer, which are indicated by red asterisks (Fig. 4.2A). However, there was no electron density visible beyond Ala381, which is the third residue of the three-alanine linker between MBP and MPR-TM (Fig. 4.3). There might be two reasons: First, the

target protein MPR-TM could be unstructured in the crystal, because the MPR-TM region could be flexible. Second, the target protein could be cleaved or degraded during or after the process of crystallization.

Table 4.1. Data collection and refinement statistics

Data collection	
Resolution (Å)	43.88 - 2.49
Wavelength used (Å)	0.9999
Space group	$P3_2$
Unit cell parameters	
a, b, c (Å)	171.93, 171.93, 70.08
α, β, γ (deg)	90, 90, 120
No. of reflections	252285
No. of unique reflections	68170
Rmerge (%)	12.1 (52.8)
I/σ(I)	7.7 (1.7)
Redundancy	3.7 (2.2)
Completeness (%)	84.4 (68.8)
CC (1/2)	0.986 (0.375)
No. of molecules per ASU	6
Solvent content (%)	39.72
Refinement	
Resolution range (Å)	43.88 - 2.49
R factor	0.159 (0.194)
R_{free}	0.194 (0.285)
Rmsd for bond length (Å)	0.014
Rmsd for bond angles (deg)	1.55

Crystals of MBP-AAA-MPR-TM were dissolved and analyzed by SDS-PAGE and MALDI-TOF MS to determine if MPR-TM is still present in the crystal. The protein from dissolved crystals migrates faster than the purified fusion protein MBP-AAA-MPR-TM prior to crystallization (Fig. 4.4A). In addition, the purified fusion protein MBP-AAA-MPR-TM prior to crystallization showed a peak of 48750 Da in the Mass spectrum (Fig. 4.4B), which agreed well with its theoretical molecular weight (48761 Da). The protein from dissolved crystals showed a peak of 41097 – 41675 Da in Mass spectrum

(Fig. 4.4B), which indicated that dissolved crystals contained MBP only, whose molecular weight is 41694 Da.

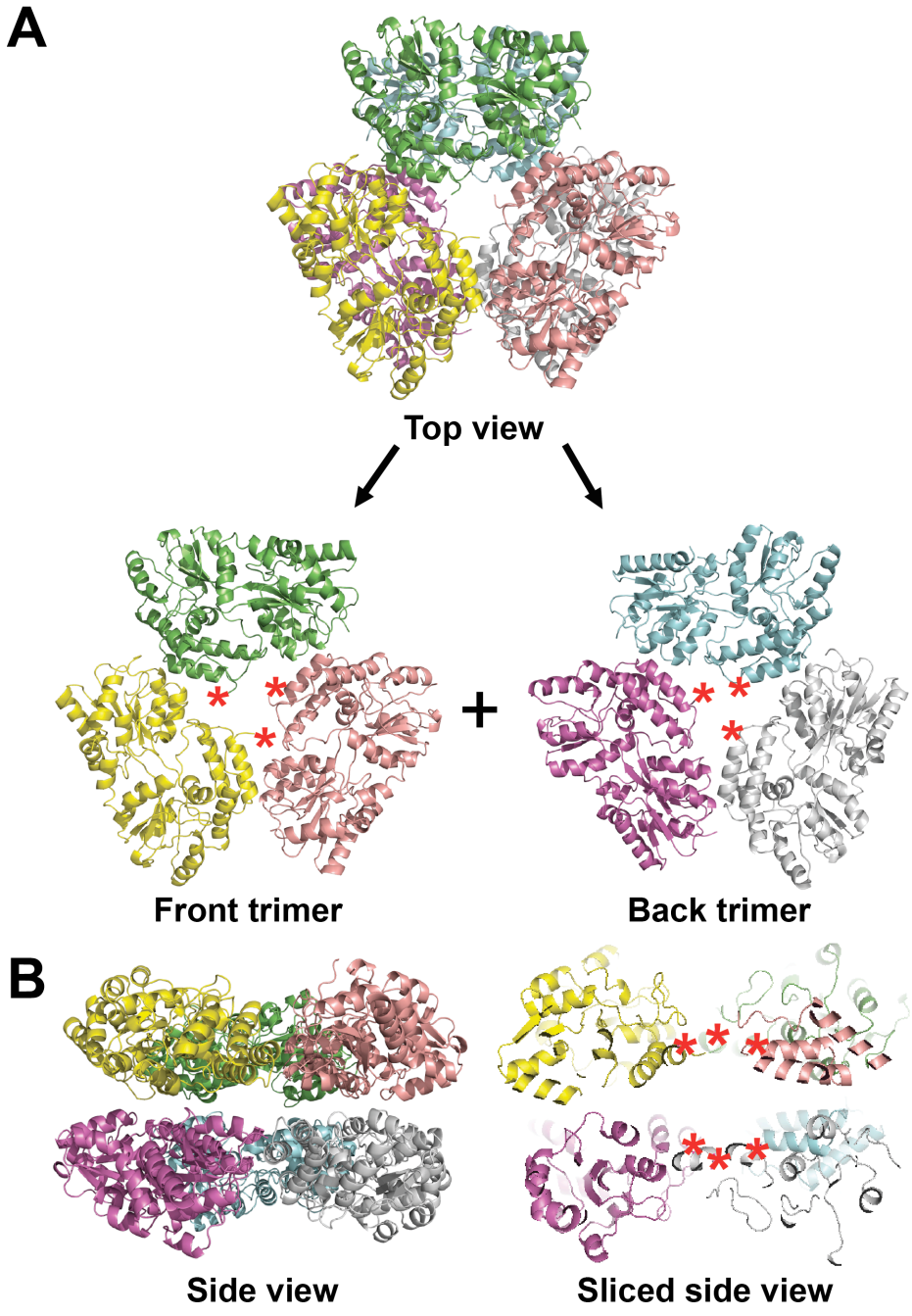


Figure 4.2. Cartoon diagrams of the six monomers of MBP in one asymmetric unit. (A) Top view. The six MBP molecules form two layers of trimmers, which are oriented in different directions. The C-termini of MBP are at the center of each trimer, which are indicated by red asterisks. (B) Side view of the six MBP molecules (left). A sliced side view (right) is presented in order to show the C-termini of MBP.

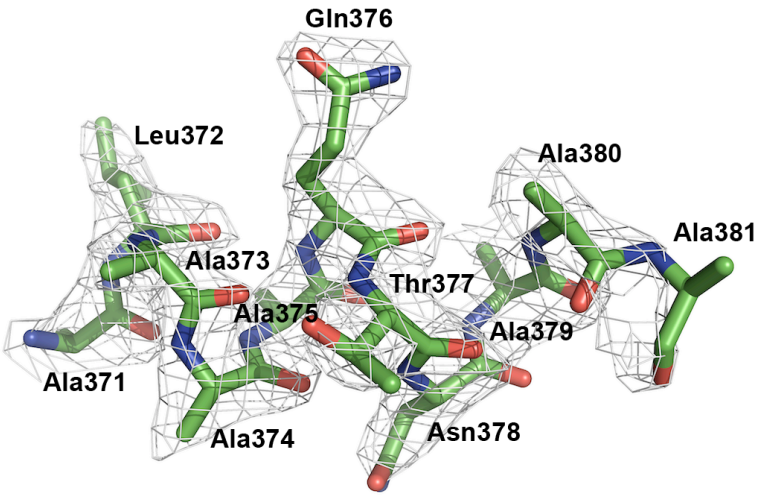


Figure 4.3. The C-terminus of the model with a 2Fo – Fc electron density map. Residues 371 – 381 of one of the six molecules are shown in stick representation. There is no electron density beyond Ala381, which is the third residue of the three-alanine linker between MBP and MPR-TM.

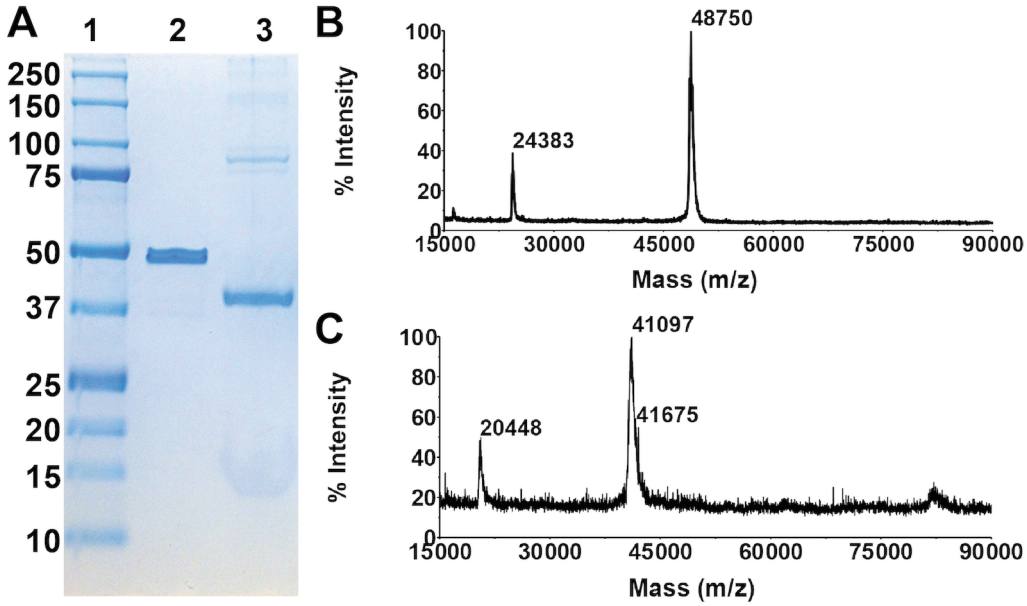


Figure 4.4. Analysis of dissolved crystals. (A) SDS-PAGE analysis. Lane 1: ladder; lane 2: purified MBP-AAA-MPR-TM fusion protein before crystallization; lane 3: protein from dissolved crystal. (B) MALDI MS analysis of MBP-AAA-MPR-TM before crystallization. (C) MALDI-TOF MS analysis of the protein from dissolved crystals.

4.5.2 The Molecular Interface and Packing Interactions

The six MBP molecules in the ASU are actually identical, with an RMSD less than 0.199 Å between corresponding C^{α} positions. Protein contacts are observed and highlighted in Fig 4.5. The interface area between each monomer inside the trimer of the same layer is 566.4 $Å^2$ (Fig. 4.5A). The interface area between each monomer from the two layers of trimers is 289.4 $Å^2$ (Fig. 4.5B). The interfacing residues are listed in Supporting Table 4.1 and Supporting Table 4.2.

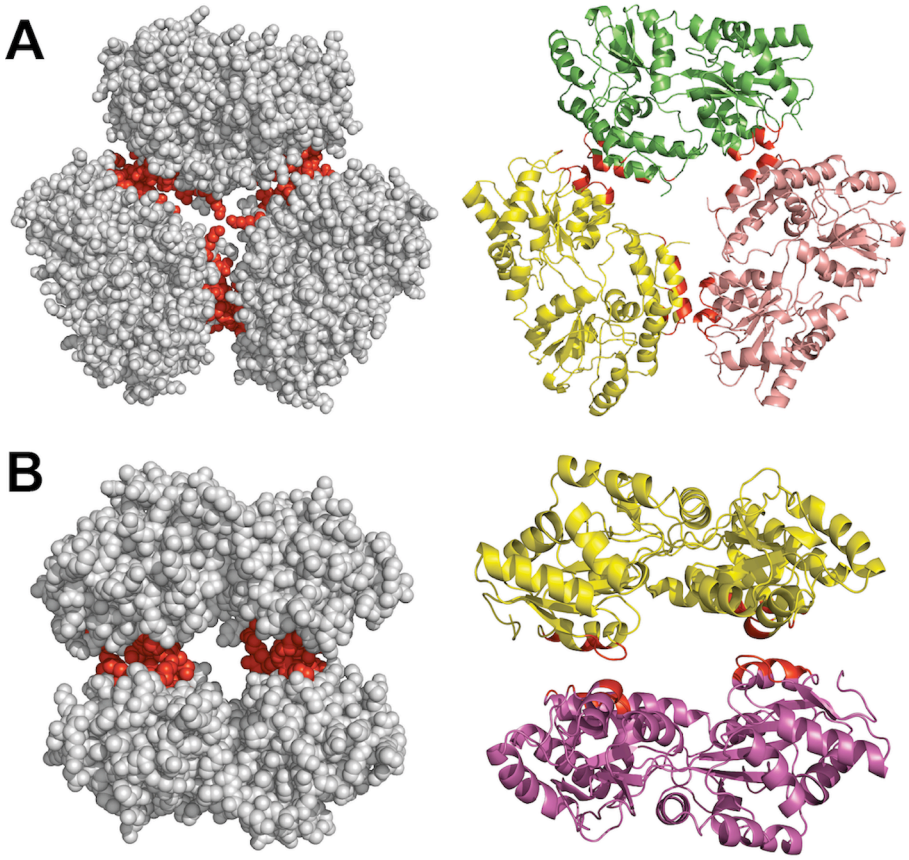


Figure 4.5. Molecular interface. (A) Molecular interface between each monomer inside the trimer of the same layer in sphere representation (left) and ribbon representation (right). Interface residues are highlighted in red. (B) Molecular interface between each monomer from the two layers of trimers in sphere representation (left) and ribbon representation (right). Interface residues are highlighted in red.

MBP is a monomer by itself (Spurlino et al., 1991). It forms different oligomers depending on the target proteins it fused with. It is reported that MBP exists as a trimer (Kobe et al., 1999; Song et al., 2003; Bethea et al., 2008; Kim et al., 2009; Lamb et al., 2011; Du et al., 2013), a dimer (Liu et al., 2001; Zhang et al., 2011) or a monomer (Ke et al., 2012; Patrick et al., 2013) in the crystal structure when it fused with different target proteins.

The MPR-TM (residues 649 – 705) is part of the envelope glycoprotein of HIV-1 gp41 (residues 512 - 856), which forms a trimeric spike on the surface of the HIV-1 virion (Buzon et al., 2010; Shi et al., 2010). The sequence of HIV-1 gp41 contains a 4-3 hydrophobic repeat, with hydrophobic amino acids spaced every four and then every three residues apart, which is characteristic of forming coiled coil (Fass and Kim, 1995; Lu et al., 1995). Trimerization of gp41 is one of the mechanisms that HIV-1 has evolved to escape antibody binding. Vulnerable epitopes are less exposed on gp41-trimer than on the individual monomeric subunit (Burton et al., 2004).

Interestingly, the arrangement of the MBP monomers inside the MBP-trimer is very similar to some of the published MBP fusion protein trimers (Fig. 4.6). Further analysis of the trimeric MBP fusion proteins indicated that the target proteins contain a 4-3 hydrophobic repeat (coiled coil motif) (Kobe et al., 1999; Lamb et al., 2011; Du et al., 2013), which is characteristic for HIV-1 gp41 as well (Lu et al., 1995). Figure 4.6B shows the superposition of the structure of our MBP (green) on the structure of MBP-gp21 chimera (shown in blue, which actually contains MBP and the ectodomain of gp21) (Kobe et al., 1999). Gp21 is the transmembrane protein of human T cell leukemia virus type 1 (HTLV-1), which is a retrovirus. Figure 4.6C shows the superposition of the

structure of MBP (green) on the structure of MBP-gp30 chimera (represented in pink, which is actually composed of MBP and the ectodomain of gp30) (Lamb et al., 2011). Gp30 is the transmembrane protein of bovine leukemia virus (BLV-1). BLV-1 is also a retrovirus. Figure 4.6D shows the superposition of the structure of MBP (green) on the structure of MBP-CC chimera (yellow) (Du et al., 2013). CC stands for the coiled coil motif of the minichromosome maintenance protein 10 (Mcm10). The MBP-gp21, MBP-gp30 and MBP-CC chimera trimers all have a mushroom shape, with the coiled coil motif (ectodomain of gp21, ectodomain of gp30 and coiled coil motif of Mcm10 respectively) forming the stalk and the MBP trimer constituting the head. The MBP-trimer represented here (Fig. 4.6A) only contains the head of the mushroom. Although SDS-PAGE and MALDI-TOF MS analysis indicated that dissolved crystals contained MBP only (Fig. 4.4), it is possible that MBP forms a trimer under the effect of MPR-TM, which is part of the trimeric envelope glycoprotein of HIV-1 gp41.

However, the crystal packings of our trimeric MBP and other trimeric MBP fusion proteins are different (Fig. 4.7). The front trimers and back trimers of our MBP are alternatively superimposed on each other (Fig. 4.7A). The crystal packings of MBP-gp21 ectodomain chimera (Fig. 4.7B) and MBP-gp30 ectodomain chimera (Fig. 4.7C) are similar: the “stalks” of the “mushrooms” are pointed at one direction and the “heads” of the “mushrooms” are shifted among different layers. In the case of MBP-CC chimera, the “heads” of the “mushrooms” are shifted among different layers but the “stalks” of the “mushrooms” are pointed at opposite directions.

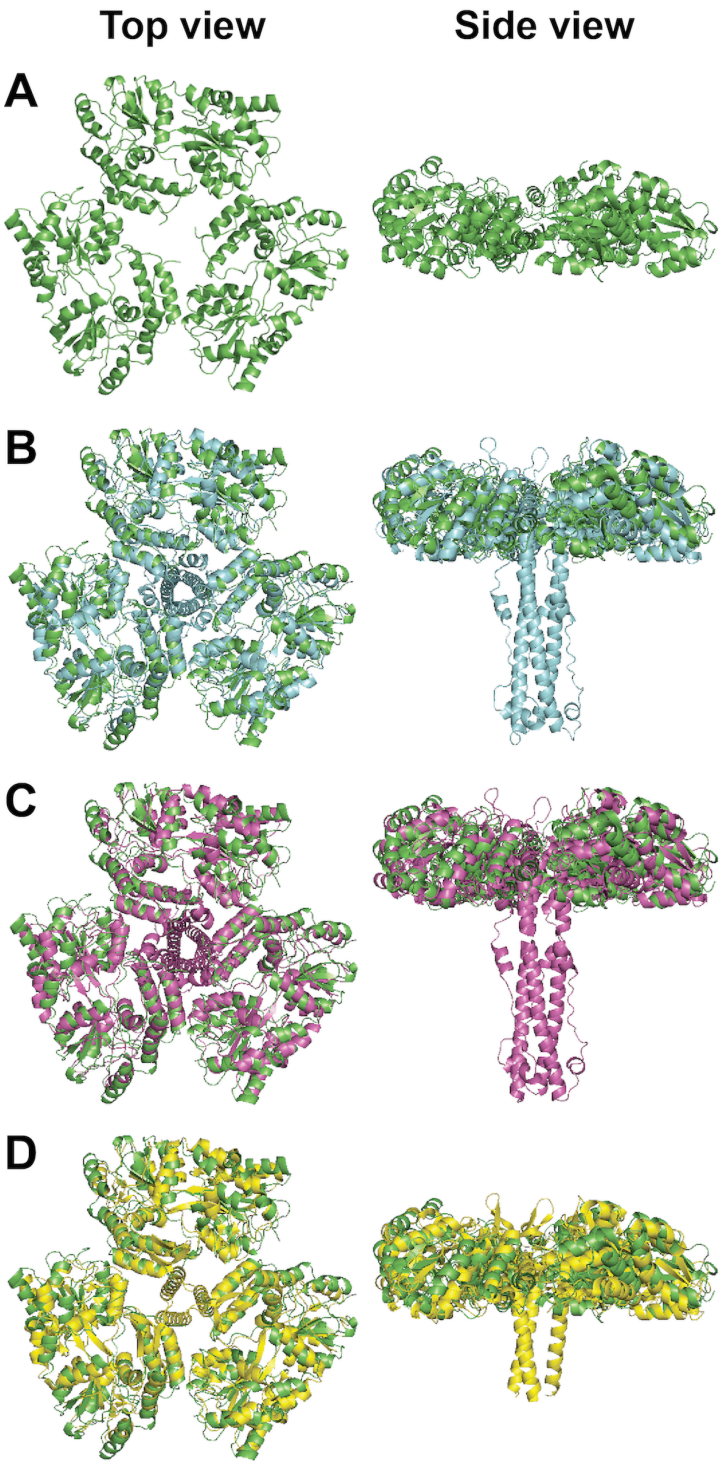


Figure 4.6. Comparison of trimeric MBP with other trimeric MBP fusion proteins. (A) MBP (only the front trimer is shown for clear representation). (B-D) Superposition of the structure of MBP on the structure of (B) MBP-gp21 ectodomain chimera (PDB ID 1MG1) (Kobe et al., 1999), (C) MBP-gp30 ectodomain chimera (PDB ID 2XZ3) (Lamb et al., 2011) and (D) MBP-CC chimera (PDB ID 4JBZ) (Du et al., 2013), respectively.

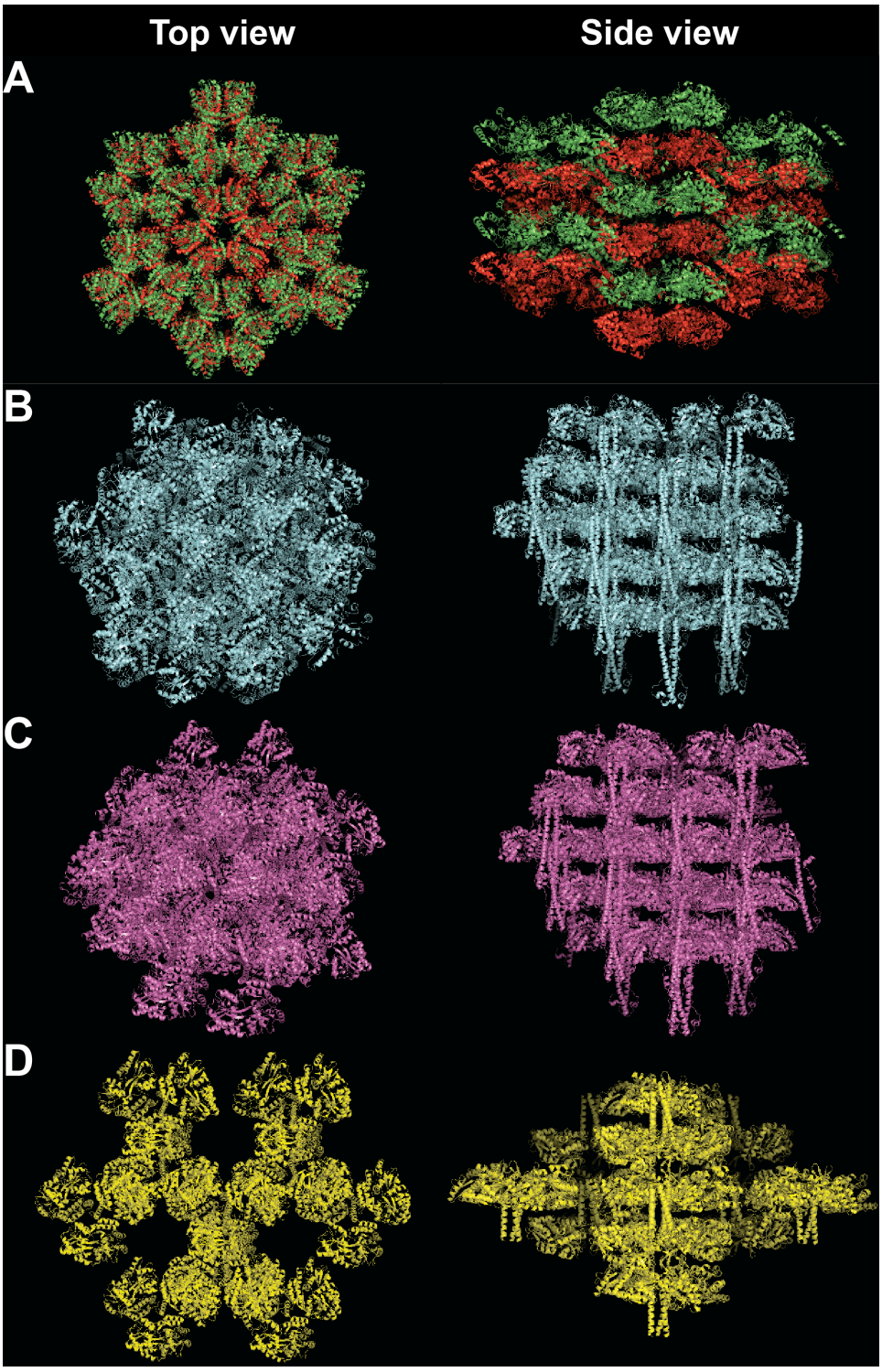


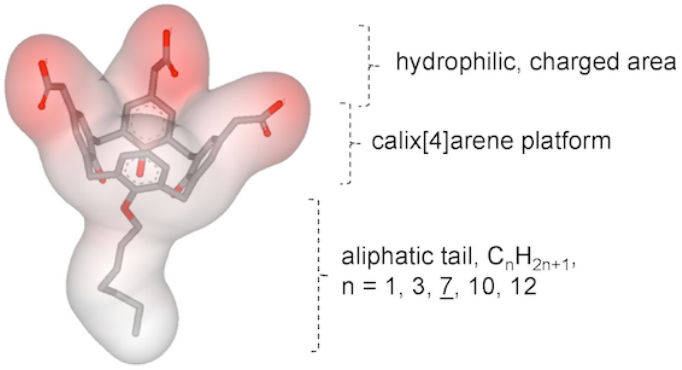
Figure 4.7. The crystal packing of trimeric MBP and other MBP fusion protein trimers. (A) MBP. The front trimers are demonstrated in green and back trimers are shown in red. (B) MBP-gp21 ectodomain chimera (PDB ID 1MG1) (Kobe et al., 1999). (C) MBP-gp30 ectodomain chimera (PDB ID 2XZ3) (Lamb et al., 2011). (D) MBP-CC chimera (4JBZ) (Du et al., 2013).

4.6 Conclusion

Crystallization of MBP-AAA-MPR-TM was optimized by adding the CALIXAR^{TM} additive. Crystals with additive C4C1 diffracted to the highest resolution. The final structure was refined to 2.49 Å in space group $P3_2$ with R factor of 0.159 and R_{free} of 0.194, respectively. Each asymmetric unit (ASU) contains 6 monomers of MBP in the open conformation, which form two layers of trimers rotated in opposite orientations. However, there is no electron density beyond Ala381, which is the third residue of the three-alanine linker between MBP and MPR-TM. The crystals were dissolved and analyzed by SDS-PAGE and MALDI-TOF MS. The result demonstrated that the crystals contained MBP only. MPR-TM might be cleaved during or after the process of crystallization. Comparison of the structures of published trimeric MBP fusion proteins indicated that they have a mushroom shape, with the coiled coil motif (ectodomain of gp21, ectodomain of gp30 and coiled coil motif of Mcm10 respectively) forming the stalk and the trimeric MBP moieties constituting the head. The MBP trimer represented here has a very similar “mushroom head” but does not contain the “mushroom stalk”. It is possible that MBP forms a trimer under the effect of MPR-TM, which is part of the trimeric envelope glycoprotein of HIV-1 gp41.

4.7 Supporting Information

4.7.1 Supporting Figures



Supporting Fig. 4.1. Chemical structure of the CALIXAR™ additives, C4Cn (Matar-Merheb et al., 2011).

4.7.2 Supporting Tables

Number	Source atoms	Target atoms	Distance (Å)
1	D/ 366(GLN). / CA [C]	E/ 63(ALA). / O [O]	3.46
2	D/ 367(THR). / N [N]	E/ 63(ALA). / O [O]	2.85
3	D/ 367(THR). / OG1[O]	E/ 63(ALA). / O [O]	3.36
4	D/ 367(THR). / CG2 [C]	E/ 65(GLY). / CA [C]	3.39
5	D/ 373(ALA). / CB [C]	E/ 84(SER). / C [C]	3.46
6	D/ 373(ALA). / CB [C]	E/ 84(SER). / O [O]	3.35
7	D/ 369(ASP). / CB [C]	E/ 85(GLY). / O [O]	3.37

Supporting Table 4.1. Interfacing residues between each monomer inside the trimer of the same layer.

Number	Source atoms	Target atoms	Distance (Å)
1	A/ 220(ASP). / OD2[O]	B/ 52(ASP). / N [N]	3.12
2	A/ 222(SER). / OG [O]	B/ 52(ASP). / CG [C]	3.11
3	A/ 222(SER). / OG [O]	B/ 52(ASP). / OD1[O]	2.48
4	A/ 220(ASP). / CB [C]	B/ 52(ASP). / OD2[O]	3.42
5	A/ 220(ASP). / CG [C]	B/ 52(ASP). / OD2[O]	3.32
6	A/ 221(TYR). / N [N]	B/ 52(ASP). / OD2[O]	3.23
7	A/ 222(SER). / N [N]	B/ 52(ASP). / OD2[O]	3.02
8	A/ 222(SER). / OG [O]	B/ 52(ASP). / OD2[O]	2.97
9	A/ 218(ASP). / OD1[O]	B/ 57(LYS). / NZ [N]	3.36
10	A/ 57(LYS). / NZ [N]	B/ 218(ASP). / OD1[O]	3.41
11	A/ 52(ASP). / OD2[O]	B/ 220(ASP). / CG [C]	3.50
12	A/ 52(ASP). / N [N]	B/ 220(ASP). / OD2[O]	3.24
13	52(ASP). / OD2[O]	B/ 222(SER). / N [N]	3.08
14	52(ASP). / OD2[O]	B/ 222(SER). / CB [C]	3.42
15	A/ 52(ASP). / CG [C]	B/ 222(SER). / OG [O]	3.06
16	A/ 52(ASP). / OD2[O]	B/ 222(SER). / OG [O]	2.68
17	A/ 52(ASP). / OD1[O]	B/ 222(SER). / OG [O]	2.59

Supporting Table 4.2. Interfacing residues between each monomer from the two layers of trimers.

CHAPTER 5

SUMMARY AND OUTLOOK

5.1 Summary

Mistic, a *Bacillus subtilis* integral membrane protein, could be used as a fusion partner to overexpress the MPR-TM of HIV-1 gp41. The expression level of MPR-TM was improved by fusion to the C-terminus of Mistic, yielding ~1 mg of pure MPR-TM protein per liter. The fusion partner Mistic was removed for final crystallization. The isolated MPR-TM protein was biophysically characterized and is a monodisperse candidate for crystallization. However, no crystal with diffraction quality was obtained after extensive crystallization screens. It may be due to the highly hydrophobic property of MPR-TM and/or lack of crystal contacts between MPR-TM molecules. Significantly, the purified MPR-TM protein reacts strongly with the broadly neutralizing Abs 2F5 and 4E10 with nanomolar to subnanomolar affinities. Our result indicated that the isolated MPR-TM protein might be in a pre-hairpin intermediate state, in which epitopes for mAbs are available for antibody binding (Frey et al., 2008). Although no crystal with diffraction quality was obtained, the protocol of purifying monodisperse MPR-TM was established. The purified MPR-TM protein could be used for other purposes instead of crystallization, for instance, structural studies of MPR-TM by NMR.

A novel construct was designed to overexpress MPR-TM as a maltose binding protein (MBP) fusion. MBP not only improves the solubility and expression level of MPR-TM, but may also provide a large hydrophilic interaction surface for formation of crystal lattice contacts thereby facilitating crystal formation. About 60 mg of MBP/MPR-TM recombinant protein was obtained from 1 liter of cell culture. Crystals of MBP/MPR-

TM recombinant protein could not be obtained when MBP and MPR-TM were separated by a 42 aa-long linker but were obtained after changing the linker to three alanine residues.

The crystal of MBP-AAA-MPR-TM diffracted to 2.5 Å after crystallization optimization by adding the CALIXAR^{TM} additive. Crystals with additive C4C1 diffracted to the highest resolution. Further analysis of the diffraction data indicated that the crystal is twinned. The final structure was refined to 2.49 Å in space group $P3_2$ with R factor of 0.164 and R_{free} of 0.196, respectively. Each ASU contains 6 monomers of MBP in the open conformation, which form two layers of trimers rotated in opposite orientations. However, the electron density did not extend beyond the linker region. It is later figured out that the crystal contained MBP only. The MPR-TM of gp41 might be cleaved by protease impurities from protein purification or degrade because of instability during or after the process of crystallization. Comparison of the MBP trimer reported here with published trimeric MBP fusion structures indicated that MBP might form such a trimeric conformation under the effect of MPR-TM.

In general, MBP could be used as a potential fusion partner to improve the expression level of membrane proteins. The linker of MBP fusion protein plays a crucial role in crystallization. MBP together with the three-alanine linker is a useful tool to study the structure of membrane proteins.

5.2 Outlook

5.2.1 The Structure of Gp41 in Prehairpin Intermediate State is Highly Demanded

The HIV-1 gp41 undergoes at least three distinct conformational states during membrane fusion process: the pre-fusion conformation, an extended prehairpin

intermediate and the post-fusion "six-helix bundle" conformation. The structure of gp41 in post-fusion conformation has been well studied since 1997 (Chan et al., 1997; Weissenhorn et al., 1997) by both X-ray crystallography and NMR (reviewed in ref. (Merk and Subramaniam, 2013)). The structure of HIV-1 Env in pre-fusion conformation has recently been solved at 3.5 Å by X-ray crystallography (Pancera et al., 2014). However, gp41 in neither pre-fusion nor post-fusion conformation could react with the broadly neutralizing antibody 2F5, whose epitope is on MPR. It is very important to study the structure of gp41 in the prehairpin intermediate state, in which epitopes for mAbs are available for antibody binding (Frey et al., 2008). Structural studies of gp41 in pre-fusion, prehairpin intermediate and post-fusion conformations will help us understand the viral and cellular membrane fusion process and design structural-based vaccines against HIV-1.

Interestingly, none of the structures of gp41 solved by X-ray crystallography or NMR included the TM domain of gp41. In this thesis, the isolated MPR-TM protein removed from Mistic-MPR-TM fusion and the MBP/MPR-TM recombinant protein react strongly with the broadly neutralizing Abs 2F5 and 4E10. Our results provide further experimental evidence of the importance of the transmembrane domain of gp41 to preserve the immunological signature of the membrane proximal region of gp41 (Montero et al., 2012), probably mimicking in a pre-hairpin intermediate. It also demonstrates the need to remove the heptad repeat regions. Our study will give some hints for future construct design to study the intermediate state of gp41.

5.2.2 Crystallization with Antibodies

The final structure demonstrated that MBP oligomerized as trimers, but the electron density did not extend beyond the linker region. MPR-TM might degrade because of instability during or after the process of crystallization. Many proteins have been crystallized with in complex with high affinity antibody fragments (Acharya et al., 2014; Pancera et al., 2014; Robin et al., 2014; Rudicell et al., 2014; Wright et al., 2014). Co-crystallization with antibody is believed to increase the probability of crystallization by providing a large hydrophilic contact for initial crystallization formation and/or limiting the conformational flexibility of solvent exposed loop regions (Hunte and Michel, 2002; Lam et al., 2009). The broadly neutralizing mAbs 2F5 and 4E10 displayed high binding affinity to purified MPR-TM protein and MBP/MPR-TM recombinant protein. These antibodies could be used for co-crystallization with MPR-TM or MBP-AAA-MPR-TM.

5.2.3 Structural Studies of the C-terminal Tail of Gp41

As the only viral protein exposed on the virion surface, the Env of gp41 has been the primary target for structural studies and vaccine development. Much research has been focused on gp120 and the ectodomain of gp41. In contrast, the C-terminal tail (CTT, residues 706-856) of gp41, which is approximately 150 aa long, has received relatively little attention. The CCT of gp41 is involved in many important functions, such as Env trafficking and virion incorporation (Freed and Martin, 1996), virion maturation (Jiang and Aiken, 2007), Env endocytosis (Byland et al., 2007), modulation of the overall Env structure (Edwards et al., 2002) and so on. In spite of functional significance, the atomic level structure of CTT is still unknown.

Early topogenesis studies classified gp41 as a type I membrane protein, including an N-terminal extracellular domain, a single α-helical membrane-spanning domain (MSD) and a long cytoplasmic CTT (Haffar et al., 1988). An alternative model has been proposed by Hollier and Dimmock (Hollier and Dimmock, 2005) with three membrane-spanning β-sheets based on the fact that the Kennedy epitope (KE) (Kennedy et al., 1986), which is located in the cytoplasmic CTT according to the single MSD model, is extracellularly exposed under some circumstances. KE (residues 735-752) is a strongly reactive peptide located in the CTT identified in earlier studies attempting to generate antibodies recognizing the precursor Env (gp160) (Kennedy et al., 1986). Interestingly, the sera against the KE strongly neutralized virus, indicating the extracellular exposure of the epitope, because antibodies cannot cross intact lipid membranes. The two proposed MSD models are shown in Fig. 5.1 (Steckbeck et al., 2010).

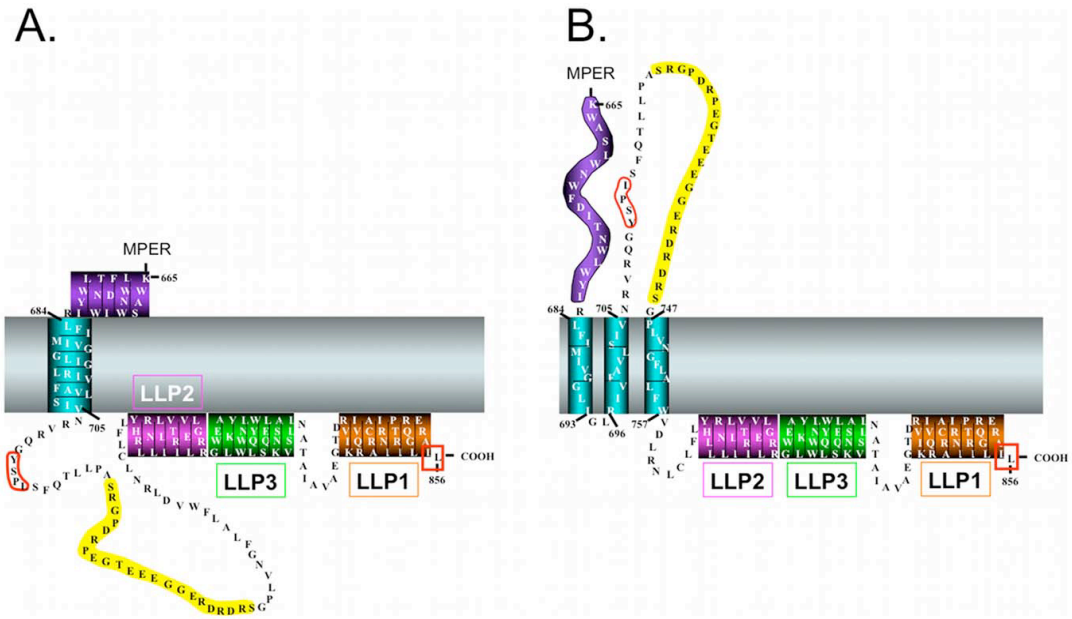


Fig. 5.1. Proposed topology models for HIV-1 gp41 CTT (Steckbeck et al., 2010). (A) Traditional CTT model with one membrane-spanning α-helix. (B) Alternative CTT model with multiple MSD segments proposed by Holllier and Dimmonck (Hollier and Dimmock, 2005). This model proposes three membrane-spanning β-sheets and an extracellular localization of the KE, which is highlighted in yellow. LLP: lentivirus lytic peptide.

Additionally, the MBP with three-alanine linker studied in this thesis could be used as a helpful tool to study the structure of gp41 CCT.

5.2.4 Utilize Eukaryotic Expression Systems

HIV-1 is a human virus and it replicates itself in the infected human cells. Although *E. coli* is the most frequently used expression system for prokaryotic and eukaryotic membrane proteins (Gordon et al., 2008), E. coli's ability to perform posttranslational modifications is poor and the posttranslational modifications are usually essential for proper folding, targeting and function (Blum et al., 2000). Therefore, it is worth trying to express the membrane protein of HIV-1 in eukaryotic expression systems.

REFERENCES

Acharya P, Tolbert WD, Gohain N, Wu X, Yu L, Liu T, Huang W, Huang CC, Kwon YD, Louder RK, Luongo TS, McLellan JS, Pancera M, Yang Y, Zhang B, Flinko R, Foulke JS, Jr., Sajadi MM, Kamin-Lewis R, Robinson JE, Martin L, Kwong PD, Guan Y, DeVico AL, Lewis GK, Pazgier M (2014) Structural Definition of an Antibody-Dependent Cellular Cytotoxicity Response Implicated in Reduced Risk for HIV-1 Infection. J Virol **88:** 12895-12906

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography **66:** 213-221

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallographica Section D-Biological Crystallography **58:** 1948-1954

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. Journal of Biological Chemistry **277:** 25649-25659

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. The Journal of biological chemistry **277:** 25649-25659

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:** 6257-6265

Andrew P (2006) Estimation of Molecular Size and Molecular Weights of Biological Compounds by Gel Filtration. *In* Methods of Biochemical Analysis. John Wiley & Sons, Inc., pp 1-53

Ashkenazi A, Shai Y (2011) Insights into the mechanism of HIV-1 envelope induced membrane fusion as revealed by its inhibitory peptides. European biophysics journal : EBJ **40:** 349-357

Baba TW, Liska V, Hofmann-Lehmann R, Vlasak J, Xu W, Ayehunie S, Cavacini LA, Posner MR, Katinger H, Stiegler G, Bernacky BJ, Rizvi TA, Schmidt R, Hill LR, Keeling ME, Lu Y, Wright JE, Chou TC, Ruprecht RM (2000) Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. Nat Med **6:** 200-206

Bamber L, Harding M, Butler PJ, Kunji ER (2006) Yeast mitochondrial ADP/ATP carriers are monomeric in detergents. Proceedings of the National Academy of Sciences of the United States of America **103:** 16224-16229

Barends TR, de Jong RM, van Straaten KE, Thunnissen AM, Dijkstra BW (2005) Escherichia coli MltA: MAD phasing and refinement of a tetartohedrally twinned protein crystal structure. Acta Crystallogr D Biol Crystallogr **61:** 613-621

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) The Protein Data Bank. Acta Crystallogr D Biol Crystallogr **58:** 899-907

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res **28:** 235-242

Bertz M, Rief M (2009) Ligand binding mechanics of maltose binding protein. J Mol Biol **393:** 1097-1105

Bethea HN, Xu D, Liu J, Pedersen LC (2008) Redirecting the substrate specificity of heparan sulfate 2-O-sulfotransferase by structurally guided mutagenesis. Proceedings of the National Academy of Sciences of the United States of America **105:** 18724-18729

Blum A, Martin HJ, Maser E (2000) Human 11beta-hydroxysteroid dehydrogenase 1/carbonyl reductase: recombinant expression in the yeast Pichia pastoris and Escherichia coli. Toxicology **144:** 113-120

Bomsel M, Tudor D, Drillet AS, Alfsen A, Ganor Y, Roger MG, Mouz N, Amacker M, Chalifour A, Diomede L, Devillier G, Cong Z, Wei Q, Gao H, Qin C, Yang GB, Zurbriggen R, Lopalco L, Fleury S (2011) Immunization with HIV-1 gp41 Subunit Virosomes Induces Mucosal Antibodies Protecting Nonhuman Primates against Vaginal SHIV Challenges. Immunity **34:** 269-280

Bomsel M, Tudor D, Drillet AS, Alfsen A, Ganor Y, Roger MG, Mouz N, Amacker M, Chalifour A, Diomede L, Devillier G, Cong Z, Wei QA, Gao H, Qin CA, Yang GB, Zurbriggen R, Lopalco L, Fleury S (2011) Immunization with HIV-1 gp41 Subunit Virosomes Induces Mucosal Antibodies Protecting Nonhuman Primates against Vaginal SHIV Challenges. Immunity **34:** 269-280

Broder CC, Dimitrov DS (1996) HIV and the 7-transmembrane domain receptors. Pathobiology **64:** 171-179

Brown DA, London E (1998) Functions of lipid rafts in biological membranes. Annu Rev Cell Dev Biol **14:** 111-136

Burton DR, Ahmed R, Barouch DH, Butera ST, Crotty S, Godzik A, Kaufmann DE, McElrath MJ, Nussenzweig MC, Pulendran B, Scanlan CN, Schief WR, Silvestri G, Streeck H, Walker BD, Walker LM, Ward AB, Wilson IA, Wyatt R (2012) A Blueprint for HIV Vaccine Discovery. Cell host & microbe **12:** 396-407

Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, Moore JP, Nabel GJ, Sodroski J, Wilson IA, Wyatt RT (2004) HIV vaccine design and the neutralizing antibody problem. Nat Immunol **5:** 233-236

Butler PJ, Ubarretxena-Belandia I, Warne T, Tate CG (2004) The Escherichia coli multidrug transporter EmrE is a dimer in the detergent-solubilised state. Journal of molecular biology **340:** 797-808

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

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 Pathogens **6**

Byland R, Vance PJ, Hoxie JA, Marsh M (2007) A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. Mol Biol Cell **18:** 414-425

Cavacini LA, Duval M, Robinson J, Posner MR (2002) Interactions of human antibodies, epitope exposure, antibody binding and neutralization of primary isolate HIV-1 virions. AIDS **16:** 2409-2417

Center RJ, Kobe B, Wilson KA, Teh T, Howlett GJ, Kemp BE, Poumbourios P (1998) Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. Protein Science **7:** 1612-1619

Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell **89:** 263-273

Chan DC, Kim PS (1998) HIV entry and its inhibition. Cell **93:** 681-684

Checkley MA, Luttge BG, Freed EO (2011) HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. Journal of molecular biology **410:** 582-608

Checkley MA, Luttge BG, Freed EO (2011) HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol **410:** 582-608

Chu B (1970) Laser Light Scattering. Annual Review of Physical Chemistry **21:** 145-+

Cohen T, Cohen SJ, Antonovsky N, Cohen IR, Shai Y (2010) HIV-1 gp41 and TCRalpha trans-membrane domains share a motif exploited by the HIV virus to modulate T-cell proliferation. PLoS Pathog **6:** e1001085

Cole JL, Garsky VM (2001) Thermodynamics of peptide inhibitor binding to HIV-1 gp41. Biochemistry **40:** 5633-5641

Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr **50:** 760-763

Compton LA, Johnson WC (1986) Analysis of Protein Circular-Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication. Analytical Biochemistry **155:** 155-167

Decroly E, Vandenbranden M, Ruysschaert JM, Cogniaux J, Jacob GS, Howard SC, Marshall G, Kompelli A, Basak A, Jean F, et al. (1994) The convertases

furin and PC1 can both cleave the human immunodeficiency virus (HIV)-1 envelope glycoprotein gp160 into gp120 (HIV-1 SU) and gp41 (HIV-I TM). J Biol Chem **269:** 12240-12247

Desiraju GR (2001) Chemistry beyond the molecule. Nature **412:** 397-400

Devito C, Broliden K, Kaul R, Svensson L, Johansen K, Kiama P, Kimani J, Lopalco L, Piconi S, Bwayo JJ, Plummer F, Clerici M, Hinkula J (2000) Mucosal and plasma IgA from HIV-1-exposed uninfected individuals inhibit HIV-1 transcytosis across human epithelial cells. J Immunol **165:** 5170-5176.

Devito C, Broliden K, Kaul R, Svensson L, Johansen K, Kiama P, Kimani J, Lopalco L, Piconi S, Bwayo JJ, Plummer F, Clerici M, Hinkula J (2000) Mucosal and plasma IgA from HIV-1-exposed uninfected individuals inhibit HIV-1 transcytosis across human epithelial cells. Journal of Immunology **165:** 5170-5176

Devito C, Hinkula J, Kaul R, Lopalco L, Bwayo JJ, Plummer F, Clerici M, Broliden K (2000) Mucosal and plasma IgA from HIV-exposed seronegative individuals neutralize a primary HIV-1 isolate. AIDS **14:** 1917-1920.

Do BH, Ryu HB, Hoang P, Koo BK, Choe H (2014) Soluble Prokaryotic Overexpression and Purification of Bioactive Human Granulocyte Colony-Stimulating Factor by Maltose Binding Protein and Protein Disulfide Isomerase. Plos One **9**

Dockter C, Volkov A, Bauer C, Polyhach Y, Joly-Lopez Z, Jeschke G, Paulsen H (2009) Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR. Proc Natl Acad Sci U S A **106:** 18485-18490

Du WY, Josephrajan A, Adhikary S, Bowles T, Bielinsky AK, Eichman BF (2013) Mcm10 Self-Association Is Mediated by an N-Terminal Coiled-Coil Domain. Plos One **8**

Eckert DM, Kim PS (2001) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. Proc Natl Acad Sci U S A **98:** 11187-11192

Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS (1999) Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell **99:** 103-115

Edwards TG, Wyss S, Reeves JD, Zolla-Pazner S, Hoxie JA, Doms RW, Baribaud F (2002) Truncation of the cytoplasmic domain induces exposure of conserved regions in the ectodomain of human immunodeficiency virus type 1 envelope protein. J Virol **76:** 2683-2691

Falson PG, Coleman AW, Matar MR, Leydier A, HUCHÉ F (2010) Additive for the crystallization of proteins, use and process. *In*. Google Patents

Fass D, Kim PS (1995) Dissection of a retrovirus envelope protein reveals structural similarity to influenza hemagglutinin. Curr Biol **5:** 1377-1383

Ferré-D'Amaré AR, Burley SK (1997) Dynamic light scattering in evaluating crystallizability of macromolecules. Methods Enzymol **276:** 157-166

Freed EO, Martin MA (1996) Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. Journal of Virology **70:** 341-351

Frey G, Chen J, Rits-Volloch S, Freeman MM, Zolla-Pazner S, Chen B (2010) Distinct conformational states of HIV-1 gp41 are recognized by neutralizing and non-neutralizing antibodies. Nat Struct Mol Biol **17:** 1486-1491

Frey G, Peng H, Rits-Volloch S, Morelli M, Cheng Y, Chen B (2008) A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. Proceedings of the National Academy of Sciences of the United States of America **105:** 3739-3744

Frey G, Peng H, Rits-Volloch S, Morelli M, Cheng Y, Chen B (2008) A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. Proc Natl Acad Sci U S A **105:** 3739-3744

Gabuzda D, Olshevsky U, Bertani P, Haseltine WA, Sodroski J (1991) Identification of membrane anchorage domains of the HIV-1 gp160 envelope glycoprotein precursor. J Acquir Immune Defic Syndr **4:** 34-40

Gallo SA, Finnegan CM, Viard M, Raviv Y, Dimitrov A, Rawat SS, Puri A, Durell S, Blumenthal R (2003) The HIV Env-mediated fusion reaction. Biochim Biophys Acta **1614:** 36-50

Ganor Y, Bomsel M (2011) HIV-1 Transmission in the Male Genital Tract. Am J Reprod Immunol **65:** 284-291

The Gap report - unaids (2013). *In*, p 120

Gong Z, Kessans SA, Song L, Dorner K, Lee HH, Meador LR, LaBaer J, Hogue BG, Mor TS, Fromme P (2014) Recombinant expression, purification and biophysical characterization of the transmembrane and membrane proximal domains of HIV-1 gp41. Protein Sci

Gordon E, Horsefield R, Swarts HG, de Pont JJ, Neutze R, Snijder A (2008) Effective high-throughput overproduction of membrane proteins in Escherichia coli. Protein Expr Purif **62:** 1-8

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc **1:** 2876-2890

Haffar OK, Dowbenko DJ, Berman PW (1988) Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. J Cell Biol **107:** 1677-1687

Hamilton MD, Nuara AA, Gammon DB, Buller RM, Evans DH (2007) Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. Nucleic Acids Res **35:** 143-151

Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. Nature **371:** 80-83

Hessell AJ, Rakasz EG, Tehrani DM, Huber M, Weisgrau KL, Landucci G, Forthal DN, Koff WC, Poignard P, Watkins DI, Burton DR (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. Journal of virology **84:** 1302-1313

Hilf RJ, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature **452:** 375-379

Hochuli E, Dobeli H, Schacher A (1987) New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighboring Histidine-Residues. J Chromatogr **411:** 177-184

Hollier MJ, Dimmock NJ (2005) The C-terminal tail of the gp41 transmembrane envelope glycoprotein of HIV-1 clades A, B, C, and D may exist in two conformations: An analysis of sequence, structure, and function. Virology **337:** 284-296

Hollmann A, Matos PM, Augusto MT, Castanho MA, Santos NC (2013) Conjugation of cholesterol to HIV-1 fusion inhibitor C34 increases peptide-membrane interactions potentiating its action. PloS one **8:** e60302

Hoo CM, Starostin N, West P, Mecartney ML (2008) A comparison of atomic force microscopy (AFM) and dynamic light scattering (DLS) methods to characterize nanoparticle size distributions. Journal of Nanoparticle Research **10:** 89-96

Huang J, Kang BH, Pancera M, Lee JH, Tong T, Feng Y, Georgiev IS, Chuang GY, Druz A, Doria-Rose NA, Laub L, Sliepen K, van Gils MJ, de la Pena AT, Derking R, Klasse PJ, Migueles SA, Bailer RT, Alam M, Pugach P, Haynes BF, Wyatt RT, Sanders RW, Binley JM, Ward AB, Mascola JR, Kwong PD, Connors M (2014) Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. Nature

Huang J, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, Alam SM, Wang T, Yang Y, Zhang B, Migueles SA, Wyatt R, Haynes BF, Kwong PD, Mascola JR, Connors M (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature **491:** 406-412

Huang JH, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, Alam SM, Wang T, Yang YP, Zhang BS, Migueles SA, Wyatt R, Haynes BF, Kwong PD, Mascola JR, Connors M (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature **491:** 406-+

Hughson FM (1997) Enveloped viruses: A common mode of membrane fusion? Current Biology **7:** R565-R569

Hunte C, Michel H (2002) Crystallisation of membrane proteins mediated by antibody fragments. Curr Opin Struct Biol **12:** 503-508

Janovick JA, Brothers SP, Cornea A, Bush E, Goulet MT, Ashton WT, Sauer DR, Haviv F, Greer J, Conn PM (2007) Refolding of misfolded mutant GPCR: post-translational pharmacoperone action in vitro. Mol Cell Endocrinol **272:** 77-85

Jiang J, Aiken C (2007) Maturation-dependent human immunodeficiency virus type 1 particle fusion requires a carboxyl-terminal region of the gp41 cytoplasmic tail. J Virol **81:** 9999-10008

Jiang S, Lin K, Strick N, Neurath AR (1993) HIV-1 inhibition by a peptide. Nature **365:** 113

Johnson WE, Desrosiers RC (2002) Viral persistence: HIV's strategies of immune system evasion. Annu Rev Med **53:** 499-518

Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr **66:** 125-132

Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science **8:** 1668-1674

Ke W, Laurent AH, Armstrong MD, Chen YC, Smith WE, Liang J, Wright CM, Ostermeier M, van den Akker F (2012) Structure of an Engineered beta-Lactamase Maltose Binding Protein Fusion Protein: Insights into Heterotropic Allosteric Regulation. Plos One **7**

Kennedy RC, Henkel RD, Pauletti D, Allan JS, Lee TH, Essex M, Dreesman GR (1986) Antiserum to a synthetic peptide recognizes the HTLV-III envelope glycoprotein. Science **231:** 1556-1559

Kessans SA, Linhart MD, Matoba N, Mor T (2013) Biological and biochemical characterization of HIV-1 Gag/dgp41 virus-like particles expressed in Nicotiana benthamiana. Plant Biotechnology Journal **11:** 681–690

Kessans SA, Linhart MD, Matoba N, Mor T (2013) Biological and biochemical characterization of HIV-1 Gag/dgp41 virus-like particles expressed in Nicotiana benthamiana. Plant Biotechnol J **11:** 681-690

Kiefer H, Maier K, Vogel R (1999) Refolding of G-protein-coupled receptors from inclusion bodies produced in Escherichia coli. Biochem Soc Trans **27:** 908-912

Kim M, Qiao Z, Yu J, Montefiori D, Reinherz EL (2007) Immunogenicity of recombinant human immunodeficiency virus type 1-like particles expressing gp41 derivatives in a pre-fusion state. Vaccine **25:** 5102-5114

Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I (2009) Secondary Structure of Huntingtin Amino-Terminal Region. Structure **17:** 1205-1212

Kobe B, Center RJ, Kemp BE, Poumbourios P (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. Proceedings of the National Academy of Sciences of the United States of America **96:** 4319-4324

Lam AY, Pardon E, Korotkov KV, Hol WG, Steyaert J (2009) Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus. J Struct Biol **166:** 8-15

Lamb D, Schuttelkopf AW, van Aalten DMF, Brighty DW (2011) Charge-Surrounded Pockets and Electrostatic Interactions with Small Ions Modulate the Activity of Retroviral Fusion Proteins. Plos Pathogens **7**

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 expression and purification **76:** 15-24

Leaman DP, Kinkead H, Zwick MB (2010) In-solution virus capture assay helps deconstruct heterogeneous antibody recognition of human immunodeficiency virus type 1. Journal of virology **84:** 3382-3395

Lebedev AA, Isupov MN (2014) Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. Acta Crystallogr D Biol Crystallogr **70:** 2430-2443

Lee H-H, Cherni I, Yu H, Fromme R, Doran JD, Grotjohann I, Mittman M, Basu S, Deb A, Dörner K, Aquila A, Barty A, Boutet S, Chapman HN, Doak RB, Hunter MS, James D, Kirian RA, Kupitz C, Lawrence RM, Liu H, Nass K, Schlichting I, Schmidt KE, Seibert MM, Shoeman RL, Spence JCH, Stellato F, Weierstall U, Williams GJ, Yoon C, Wang D, Zatsepin NA, Hogue BG, Matoba N, Fromme P, Mor TS, (2014) Expression, purification and crystallization of CTB-MPR, a candidate mucosal vaccine component against HIV-1. IUCrJ **1:** In press

Leroux-Roels G, Maes C, Clement F, van Engelenburg F, van den Dobbelsteen M, Adler M, Amacker M, Lopalco L, Bomsel M, Chalifour A, Fleury S (2013) Randomized Phase I: Safety, Immunogenicity and Mucosal Antiviral Activity in Young Healthy Women Vaccinated with HIV-1 Gp41 P1 Peptide on Virosomes. PloS one **8:** e55438

Liu J, Deng Y, Dey AK, Moore JP, Lu M (2009) Structure of the HIV-1 gp41 membrane-proximal ectodomain region in a putative prefusion conformation. Biochemistry **48:** 2915-2923

Liu YF, Manna A, Li RG, Martin WE, Murphy RC, Cheung AL, Zhang GY (2001) Crystal structure of the SarR protein from Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America **98:** 6877-6882

Lu M, Blacklow SC, Kim PS (1995) A Trimeric Structural Domain of the Hiv-1 Transmembrane Glycoprotein. Nature Structural Biology **2:** 1075-1082

Luo J, Choulet J, Samuelson JC (2009) Rational design of a fusion partner for membrane protein expression in E. coli. Protein Sci **18:** 1735-1744

Lutje Hulsik D, Liu YY, Strokappe NM, Battella S, El Khattabi M, McCoy LE, Sabin C, Hinz A, Hock M, Macheboeuf P, Bonvin AM, Langedijk JP, Davis D, Forsman Quigley A, Aasa-Chapman MM, Seaman MS, Ramos A, Poignard P, Favier A, Simorre JP, Weiss RA, Verrips CT, Weissenhorn W, Rutten L (2013) A gp41 MPER-specific llama VHH requires a hydrophobic CDR3 for neutralization but not for antigen recognition. PLoS pathogens **9:** e1003202

Lyumkis D, Julien JP, de Val N, Cupo A, Potter CS, Klasse PJ, Burton DR, Sanders RW, Moore JP, Carragher B, Wilson IA, Ward AB (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science **342:** 1484-1490

Manavalan P, Johnson WC (1987) Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular-Dichroism Spectra. Analytical Biochemistry **167:** 76-85

Mao Y, Wang L, Gu C, Herschhorn A, Desormeaux A, Finzi A, Xiang SH, Sodroski JG (2013) Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. Proc Natl Acad Sci U S A **110:** 12438-12443

Mao YD, Wang LP, Gu C, Herschhorn A, Xiang SH, Haim H, Yang XZ, Sodroski J (2012) Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. Nature Structural & Molecular Biology **19:** 893-899

Markwell MA, Haas SM, Bieber LL, Tolbert NE (1978) A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. Anal Biochem **87:** 206-210

Mascola JR, Stiegler G, VanCott TC, Katinger H, Carpenter CB, Hanson CE, Beary H, Hayes D, Frankel SS, Birx DL, Lewis MG (2000) Protection of macaques against vaginal transmission of a pathogenic HIV- 1/SIV chimeric virus by passive infusion of neutralizing antibodies. Nat Med **6:** 207-210.

Matar-Merheb R, Rhimi M, Leydier A, Huche F, Galian C, Desuzinges-Mandon E, Ficheux D, Flot D, Aghajari N, Kahn R, Di Pietro A, Jault JM, Coleman AW, Falson P (2011) Structuring detergents for extracting and stabilizing functional membrane proteins. PLoS One **6:** e18036

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:** 5047-5055

Matoba N, Griffin TA, Mittman M, Doran JD, Alfsen A, Montefiori DC, Hanson CV, Bomsel M, Mor TS (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:** 218-229

Matoba N, Kajiura H, Cherni I, Doran JD, Alfsen A, Bomsel M, Fujiyama K, Mor TS (2009) Biochemical and immunological characterization of the plant-derived candidate HIV-1 mucosal vaccine CTB $MPR_{649\ 684}$. Plant Biotechnol J **7:** 129-145

Matoba N, Magerus A, Geyer BC, Zhang Y, Muralidharan M, Alfsen A, Arntzen CJ, Bomsel M, Mor TS (2004) A mucosally targeted subunit vaccine candidate eliciting HIV-1 transcytosis-blocking Abs. Proc Natl Acad Sci U S A **101:** 13584-13589

Mayr A, Stickl H, Muller HK, Danner K, Singer H (1978) [The smallpox vaccination strain MVA: marker, genetic structure, experience gained with the parenteral vaccination and behavior in organisms with a debilitated defence mechanism (author's transl)]. Zentralblatt fur Bakteriologie, Parasitenkunde,

Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe B: Hygiene, Betriebshygiene, praventive Medizin **167:** 375-390

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr **40:** 658-674

Melikyan GB (2008) Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology **5:** 111

Meng G, Wei X, Wu X, Sellers MT, Decker JM, Moldoveanu Z, Orenstein JM, Graham MF, Kappes JC, Mestecky J, Shaw GM, Smith PD (2002) Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells. Nat Med **8:** 150-156

Merk A, Subramaniam S (2013) HIV-1 envelope glycoprotein structure. Current Opinion in Structural Biology **23:** 268-276

Miyauchi K, Komano J, Yokomaku Y, Sugiura W, Yamamoto N, Matsuda Z (2005) Role of the specific amino acid sequence of the membrane-spanning domain of human immunodeficiency virus type 1 in membrane fusion. J Virol **79:** 4720-4729

Miyazawa M, Lopalco L, Mazzotta F, Lo Caputo S, Veas F, Clerici M (2009) The 'immunologic advantage' of HIV-exposed seronegative individuals. AIDS **23:** 161-175

Montero M, Gulzar N, Klaric KA, Donald JE, Lepik C, Wu S, Tsai S, Julien JP, Hessell AJ, Wang S, Lu S, Burton DR, Pai EF, Degrado WF, Scott JK (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:** 2930-2941

Moon AF, Mueller GA, Zhong XJ, Pedersen LC (2010) A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. Protein Science **19:** 901-913

Moore JP, Trkola A, Dragic T (1997) Co-receptors for HIV-1 entry. Curr Opin Immunol **9:** 551-562

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr **53:** 240-255

Newstead S, Ferrandon S, Iwata S (2008) Rationalizing alpha-helical membrane protein crystallization. Protein Sci **17:** 466-472

Nguyen MT, Koo BK, Thi Vu TT, Song JA, Chong SH, Jeong B, Ryu HB, Moh SH, Choe H (2014) Prokaryotic soluble overexpression and purification of bioactive human growth hormone by fusion to thioredoxin, maltose binding protein, and protein disulfide isomerase. PLoS One **9:** e89038

Ni da Q, 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:** 157-168

Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, Stewart-Jones GB, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L, Munro JB, Blanchard SC, Mothes W, Connors M, Kwong PD (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature

Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. Analytical biochemistry **216:** 413-417

Patrick AN, Cabrera JH, Smith AL, Chen XS, Ford HL, Zhao R (2013) Structure-function analyses of the human SIX1-EYA2 complex reveal insights into metastasis and BOR syndrome. Nat Struct Mol Biol **20:** 447-453

Popik W, Alce TM, Au WC (2002) Human immunodeficiency virus type 1 uses lipid raft-colocalized CD4 and chemokine receptors for productive entry into CD4(+) T cells. Journal of virology **76:** 4709-4722

Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionation. Nature **258:** 598-599

Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. Science **242:** 1168-1171

Prince CC, Jia Z (2013) Detergent quantification in membrane protein samples and its application to crystallization experiments. Amino Acids

Prive GG (2007) Detergents for the stabilization and crystallization of membrane proteins. Methods **41:** 388-397

Proteau A, Shi R, Cygler M (2010) Application of dynamic light scattering in protein crystallization. Curr Protoc Protein Sci **Chapter 17:** Unit 17 10

Provencher SW, Glockner J (1981) Estimation of Globular Protein Secondary Structure from Circular-Dichroism. Biochemistry **20:** 33-37

Purtscher M, Trkola A, Gruber G, Buchacher A, Predl R, Steindl F, Tauer C, Berger R, Barrett N, Jungbauer A, et al. (1994) A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. AIDS research and human retroviruses **10:** 1651-1658

Purtscher M, Trkola A, Gruber G, Buchacher A, Predl R, Steindl F, Tauer C, Berger R, Barrett N, Jungbauer A, et al. (1994) A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. AIDS Res Hum Retroviruses **10:** 1651-1658

Purtscher M, Trkola A, Gruber G, Buchacher A, Predl R, Steindl F, Tauer C, Berger R, Barrett N, Jungbauer A, Katinger H (1994) A Broadly Neutralizing Human Monoclonal-Antibody against Gp41 of Human-Immunodeficiency-Virus Type-1. Aids Research and Human Retroviruses **10:** 1651-1658

Puryear WB, Gummuluru S (2013) Role of glycosphingolipids in dendritic cell-mediated HIV-1 trans-infection. Advances in experimental medicine and biology **762:** 131-153

Puryear WB, Gummuluru S (2013) Role of Glycosphingolipids in Dendritic Cell-Mediated HIV-1 Trans-infection. Hiv Interactions with Dendritic Cells: Infection and Immunity **762:** 131-153

Qiao ZS, Kim M, Reinhold B, Montefiori D, Wang JH, Reinherz EL (2005) Design, expression, and immunogenicity of a soluble HIV trimeric envelope fragment adopting a prefusion gp41 configuration. The Journal of biological chemistry **280:** 23138-23146

Raghava GPS (2002) APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. CASP5**:** A-132

Raran-Kurussi S, Waugh DS (2014) Unrelated solubility-enhancing fusion partners MBP and NusA utilize a similar mode of action. Biotechnol Bioeng

Reardon PN, Sage H, Dennison SM, Martin JW, Donald BR, Alam SM, Haynes BF, Spicer LD (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:** 1391-1396

Reuven EM, Dadon Y, Viard M, Manukovsky N, Blumenthal R, Shai Y (2012) HIV-1 gp41 transmembrane domain interacts with the fusion peptide: implication in lipid mixing and inhibition of virus-cell fusion. Biochemistry **51:** 2867-2878

Rietveld A, Simons K (1998) The differential miscibility of lipids as the basis for the formation of functional membrane rafts. Biochimica Et Biophysica Acta-Reviews on Biomembranes **1376:** 467-479

Robin G, Sato Y, Desplancq D, Rochel N, Weiss E, Martineau P (2014) Restricted Diversity of Antigen Binding Residues of Antibodies Revealed by Computational Alanine Scanning of 227 Antibody-Antigen Complexes. J Mol Biol

Rogl H, Kosemund K, Kuhlbrandt W, Collinson I (1998) Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. FEBS Lett **432:** 21-26

Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science **307:** 1317-1321

Rudicell RS, Kwon YD, Ko SY, Pegu A, Louder MK, Georgiev IS, Wu X, Zhu J, Boyington JC, Chen X, Shi W, Yang ZY, Doria-Rose NA, McKee K, O'Dell S, Schmidt SD, Chuang GY, Druz A, Soto C, Yang Y, Zhang B, Zhou T, Todd JP, Lloyd KE, Eudailey J, Roberts KE, Donald BR, Bailer RT, Ledgerwood J, Program NCS, Mullikin JC, Shapiro L, Koup RA, Graham BS, Nason MC, Connors M, Haynes BF, Rao SS, Roederer M, Kwong PD, Mascola JR, Nabel GJ (2014) Enhanced Potency of a Broadly Neutralizing HIV-1 Antibody In Vitro Improves Protection against Lentiviral Infection In Vivo. J Virol **88:** 12669-12682

Schagger H (2006) Tricine-SDS-PAGE. Nat Protoc **1:** 16-22

Schägger H (2006) Tricine-SDS-PAGE. Nature protocols **1:** 16-22

Schrodinger, LLC (2010) The PyMOL Molecular Graphics Systemm, Version 1.3,.

Seiler CY, Park JG, Sharma A, Hunter P, Surapaneni P, Sedillo C, Field J, Algar R, Price A, Steel J, Throop A, Fiacco M, LaBaer J (2014) DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. Nucleic Acids Res **42:** D1253-1260

Shang L, Hunter E (2010) Residues in the membrane-spanning domain core modulate conformation and fusogenicity of the HIV-1 envelope glycoprotein. Virology **404:** 158-167

Shang L, Yue L, Hunter E (2008) Role of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein in cell-cell fusion and virus infection. J Virol **82:** 5417-5428

Shen R, Drelichman ER, Bimczok D, Ochsenbauer C, Kappes JC, Cannon JA, Tudor D, Bomsel M, Smythies LE, Smith PD (2010) GP41-specific antibody blocks cell-free HIV type 1 transcytosis through human rectal mucosa and model colonic epithelium. J Immunol **184:** 3648-3655

Shen R, Richter HE, Smith PD (2011) Early HIV-1 target cells in human vaginal and ectocervical mucosa. American journal of reproductive immunology **65:** 261-267

Shen RZ, Richter HE, Smith PD (2011) Early HIV-1 Target Cells in Human Vaginal and Ectocervical Mucosa. American Journal of Reproductive Immunology **65:** 261-267

Shi W, Bohon J, Han DP, Habte H, Qin Y, Cho MW, Chance MR (2010) Structural characterization of HIV gp41 with the membrane-proximal external region. J Biol Chem **285:** 24290-24298

Simons K, van Meer G (1988) Lipid sorting in epithelial cells. Biochemistry **27:** 6197-6202

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. Protein Science **12:** 1313-1322

Song JJ, Liu JD, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nature Structural Biology **10:** 1026-1032

Song L, Wang Z, Zhou D, Nand A, Li S, Guo B, Wang Y, Cheng Z, Zhou W, Zheng Z, Zhu J (2013) Waveguide coupled surface plasmon resonance imaging measurement and high-throughput analysis of bio-interaction. Sensors and Actuators B: Chemical **181:** 652-660

Spurlino JC, Lu GY, Quiocho FA (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem **266:** 5202-5219

Sreerama N, Woody RW (1993) A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular-Dichroism. Analytical Biochemistry **209:** 32-44

Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem **287:** 252-260

Steckbeck JD, Kuhlmann AS, Montelaro RC (2013) C-terminal tail of human immunodeficiency virus gp41: functionally rich and structurally enigmatic. J Gen Virol **94:** 1-19

Steckbeck JD, Sun C, Sturgeon TJ, Montelaro RC (2010) Topology of the C-terminal tail of HIV-1 gp41: differential exposure of the Kennedy epitope on cell and viral membranes. PLoS One **5:** e15261

Steckbeck JD, Sun C, Sturgeon TJ, Montelaro RC (2013) Detailed topology mapping reveals substantial exposure of the "cytoplasmic" C-terminal tail (CTT) sequences in HIV-1 Env proteins at the cell surface. PLoS One **8:** e65220

Tan MH, Zhou XE, Soon FF, Li X, Li J, Yong EL, Melcher K, Xu HE (2013) The crystal structure of the orphan nuclear receptor NR2E3/PNR ligand binding domain reveals a dimeric auto-repressed conformation. PLoS One **8:** e74359

Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Riters M, Newman CS, Fox BG, Wrobel RL (2004) Results from high-throughput DNA cloning of Arabidopsis thaliana target genes using site-specific recombination. J Struct Funct Genomics **5:** 267-276

Trabattoni D, Biasin M, Clerici M (2012) Mucosal immunoglobulin A in HIV-exposed seronegative individuals. AIDS **26:** 2247-2250

Tudor D, Bomsel M (2011) The broadly neutralizing HIV-1 IgG 2F5 elicits gp41-specific antibody-dependent cell cytotoxicity in a FcgammaRI-dependent manner. AIDS **25:** 751-759

Tudor D, Derrien M, Diomede L, Drillet AS, Houimel M, Moog C, Reynes JM, Lopalco L, Bomsel M (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 Immunology **2:** 412-426

Tudor D, Derrien M, Diomede L, Drillet AS, Houimel M, Moog C, Reynes JM, Lopalco L, Bomsel M (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:** 412-426

Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography **30:** 1022-1025

Vanstokkum IHM, Spoelder HJW, Bloemendal M, Vangrondelle R, Groen FCA (1990) Estimation of Protein Secondary Structure and Error Analysis from Circular-Dichroism Spectra. Analytical Biochemistry **191:** 110-118

Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P, Investigators PGP (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature **477:** 466-U117

Wang J, Tong P, Lu L, Zhou L, Xu L, Jiang S, Chen YH (2011) HIV-1 gp41 core with exposed membrane-proximal external region inducing broad HIV-1 neutralizing antibodies. PloS one **6:** e18233

Weiss CD (2003) HIV-1 gp41: mediator of fusion and target for inhibition. AIDS Rev **5:** 214-221

Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC (1997) Atomic structure of the ectodomain from HIV-1 gp41. Nature **387:** 426-430

Wiener MC (2004) A pedestrian guide to membrane protein crystallization. Methods **34:** 364-372

Wild C, Greenwell T, Shugars D, Rimskyclarke L, Matthews T (1995) The Inhibitory Activity of an Hiv Type-1 Peptide Correlates with Its Ability to Interact with a Leucine-Zipper Structure. Aids Research and Human Retroviruses **11:** 323-325

Wild C, Oas T, Mcdanal C, Bolognesi D, Matthews T (1992) A Synthetic Peptide Inhibitor of Human-Immunodeficiency-Virus Replication - Correlation between Solution Structure and Viral Inhibition. Proceedings of the National Academy of Sciences of the United States of America **89:** 10537-10541

Wild CT, Shugars DC, Greenwell TK, Mcdanal CB, Matthews TJ (1994) Peptides Corresponding to a Predictive Alpha-Helical Domain of Human-Immunodeficiency-Virus Type-1 Gp41 Are Potent Inhibitors of Virus-Infection. Proceedings of the National Academy of Sciences of the United States of America **91:** 9770-9774

Wright KE, Hjerrild KA, Bartlett J, Douglas AD, Jin J, Brown RE, Illingworth JJ, Ashfield R, Clemmensen SB, de Jongh WA, Draper SJ, Higgins MK (2014) Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. Nature

Wu L (2008) Biology of HIV mucosal transmission. Current opinion in HIV and AIDS **3:** 534-540

Yue L, Shang L, Hunter E (2009) Truncation of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein defines elements required for fusion, incorporation, and infectivity. J Virol **83:** 11588-11598

Yue L, Shang L, Hunter E (2009) Truncation of the Membrane-Spanning Domain of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Defines Elements Required for Fusion, Incorporation, and Infectivity. Journal of Virology **83:** 11588-11598

Zhang Y, Gao X, Michael Garavito R (2011) Structural analysis of the intracellular domain of (pro)renin receptor fused to maltose-binding protein. Biochem Biophys Res Commun **407:** 674-679

Zwick MB, Labrijn AF, Wang M, Spenlehauer C, Saphire EO, Binley JM, Moore JP, Stiegler G, Katinger H, Burton DR, Parren PW (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. J Virol **75:** 10892-10905.

Zwick MB, Labrijn AF, Wang M, Spenlehauer C, Saphire EO, Binley JM, Moore JP, Stiegler G, Katinger H, Burton DR, Parren PWHI (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. Journal of Virology **75:** 10892-10905

APPENDIX A

PERMISSIONS TO USE COPYRIGHTED MATERIALS

Copyright Permission

3 messages

Zhen Gong <gzhen@asu.edu> Fri, Sep 19, 2014 at 4:50 PM
To: PNASPermissions@nas.edu

Dear Editor,

I am a Ph.D candidate in Arizona State University and I'd like to use one of the figures published in PNAS in my dissertation.

Here is the information about the original material:

1. My name: Zhen Gong;
   my affiliation**:** Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona;
   my title: research assistant
2. My mailing address: 1718 S Jentilly Ln, Apt 219, Tempe, AZ, 85281;
   phone number: 602-628-1216
   email: gzhen@asu.edu
3. PNAS volume number: 98; issue number: 20; issue date: Sep 25, 2001
4. PNAS article title: Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region
5. PNAS authors' names: Debra M. Eckert and Peter S. Kim
6. Page numbers of ties to be reprinted: 11188
7. Figure number to be reprinted: Fig. 1

Here is the information about the intended use of the material:

1. Title of Dissertation: Structural Studies of the Transmembrane and Membrane Proximal Domains of HIV-1 gp41 by

X-Ray Crystallography

2. Authors of work: Zhen Gong

Please let me know if you need any other information. Thank you.

Zhen

PNAS Permissions <PNASPermissions@nas.edu> Mon, Sep 22, 2014 at 7:15 AM
To: Zhen Gong <gzhen@asu.edu>

Permission is granted for your use of the figure as described in your message. Please cite the PNAS article in full, and include "Copyright (2001) National Academy of Sciences, U.S.A." as a copyright note. Because this material published between 1993 and 2008, a copyright note is needed. Let us know if you have any questions.

Best regards,

Kay McLaughlin for

Diane Sullenberger

Copyright Permission

Zhen Gong <gzhen@asu.edu> Mon, Sep 22, 2014 at 11:44 AM
To: PNAS Permissions <PNASPermissions@nas.edu>

Dear Editor,

I am a Ph.D candidate in Arizona State University and I'd like to use one of the figures published in PNAS in my dissertation.

Here is the information about the original material:

1. My name: Zhen Gong;
   my affiliation**:** Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona;
   my title: research assistant
2. My mailing address: 1718 S Jentilly Ln, Apt 219, Tempe, AZ, 85281;
   phone number: 602-628-1216
   email: gzhen@asu.edu
3. PNAS volume number: 105; issue number: 10; issue date: Mar 11, 2008
4. PNAS article title: A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies
5. PNAS authors' names: Gary Frey, Hanqin Peng, Sophia Rits-Volloch, Marco Morelli, Yifan Cheng, and Bing Chen.
6. Page numbers of ties to be reprinted: 3740
7. Figure number to be reprinted: Fig. 1

Here is the information about the intended use of the material:

1. Title of Dissertation: Structural Studies of the Transmembrane and Membrane Proximal Domains of HIV-1 gp41 by

X-Ray Crystallography

2. Authors of work: Zhen Gong

Please let me know if you need any other information. Thank you.

Best regards,
Zhen Gong
Ph.D candidate
Department of Chemistry and Biochemistry
Arizona State University

PNAS Permissions <PNASPermissions@nas.edu> Mon, Sep 22, 2014 at 12:59 PM
To: Zhen Gong <gzhen@asu.edu>

Permission is granted for your use of the figure as described in your message. Please cite the PNAS article in full, and include "Copyright (2008) National Academy of Sciences, U.S.A." as a copyright note. Because this material published between 1993 and 2008, a copyright note is needed. Let us know if you have any questions.

Best regards,

Kay McLaughlin for

Diane Sullenberger

Executive Editor

PNAS

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Sep 16, 2014

This is a License Agreement between Zhen Gong ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3471031042021
License date	Sep 16, 2014
Licensed content publisher	John Wiley and Sons
Licensed content publication	Protein Science
Licensed content title	Recombinant expression, purification, and biophysical characterization of the transmembrane and membrane proximal domains of HIV-1 gp41
Licensed copyright line	© 2014 The Protein Society
Licensed content author	Zhen Gong,Sarah A. Kessans,Lusheng Song,Katerina Dörner,Ho-Hsien Lee,Lydia R Meador,Joshua LaBaer,Brenda G. Hogue,Tsafrir S. Mor,Petra Fromme
Licensed content date	Sep 3, 2014
Start page	n/a
End page	n/a
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Structural studies of the transmembrane and membrane proximal domains of HIV-1 gp41 by X-ray crystallography
Expected completion date	Nov 2014
Expected size (number of pages)	200
Total	0.00 USD
Terms and Conditions	

APPENDIX B

CO-AUTHOR APPROVAL

I verify that the following co-authors have approved of my use of our publications in my dissertation.

Petra Fromme (Arizona State University)

Tsafrir S. Mor (Arizona State University)

Brenda G Hogue (Arizona State University)

Joshua LaBaer (Arizona State University)

Sarah A. Kessans (Arizona State University[1])

Lusheng Song (Arizona State University)

Katerina Dörner (Arizona State University)

Ho-Hsien Lee (Arizona State University)

Lydia R Meador (Arizona State University)

[1]Note: The author's address is listed as when the research was performed.