2010-08-31

# Structural and Functional Studies of Proteins Involved in Antigen Processing: A Dissertation

Tina T. Nguyen
*University of Massachusetts Medical School*

# STRUCTURAL AND FUNCTIONAL STUDIES OF PROTEINS INVOLVED IN ANTIGEN PROCESSING

A Dissertation Presented

By

Tina T. Nguyen

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 31, 2010

Biochemistry and Molecular Pharmacology

STRUCTURAL AND FUNCTIONAL STUDIES OF PROTEINS INVOLVED IN ANTIGEN PROCESSING

A Dissertation Presented By

Tina T. Nguyen

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

_____

Lawrence J. Stern, Ph.D., Thesis Advisor

_____

Reid Gilmore, Ph.D., Member of Committee

_____

Kenneth Rock, M.D., Member of Committee

_____

William Royer, Ph.D., Member of Committee

_____

Ian York, Ph.D., External Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____

Celia Schiffer, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduatio requirements of the school.

_____

Anthony Carruthers, Ph.D.
Dean of the Graduate School of Biomedical Sciences

Program in Biochemistry & Molecular Pharmacology

August 31, 2010

**Acknowledgements**

First, I would like to thank Larry Stern for advising me throughout graduate school, allowing me to conduct experiments in the lab and teaching me not only about crystallography, structural biology or immunology but also about New England such as fried clams and ice cream. Who would have known that those two would go together? I would also like to thank my thesis advisory committee, Celia, Reid, Bill and Ken, for hearing me out and giving suggestions throughout the years. I can't say that I've done everything that was suggested but there were still helpful discussions.

I would also like to thank the Stern Lab members past and present for their support, companionship, helpful discussions, randomness, entertainment and listening to me throughout my time in graduate school. Though members have come and gone, the lab has always been a good supportive non-hostile learning environment.

Last but not least, I would like to thank my family (including my boyfriend, fiancé, and now husband, Will, who have held all those titles at one time or another during my time in graduate school). Thanks for keeping me grounded and reminding me of my roots. I will always remember the pep talks and laughter shared during this stage of my life.

## THESIS ABSTRACT

This thesis is comprised of studies of proteins involved in class I and class II major histocompatibility complex (MHC) antigen procressing. In class I MHC processing, structural and functional studies were conducted of an aminopeptidase, ERAP1, that mediates the final step in antigen processing to understand how it is particularly suitable for cleavage of antigenic peptides for class I MHC presentation. In the class II MHC antigen presentation pathway, structural studies were conducted to characterize a fluorogenic peptide that can be used to understand peptide loading events *in vivo* and in real time. Also structural studies of class II MHC and peptide complexes were conducted to understand the nature of an unique C-terminal secondary structure element exhibited by an HIV derived peptide in the peptide binding groove of class II MHC. The studies discussed in this thesis provide insights into the proteins involved in the class I and class II MHC antigen presentation pathway.

The endoplasmic reticulum (ER) aminopeptidase, ERAP1, is a 941 amino acid member of the M1 family of zinc metalloaminopeptidases. Unlike other aminopeptidases, ERAP1 has a length and C-terminal preference for its substrates. Interestingly, ERAP1 has been shown to trim antigenic peptides to lengths of 8 or 9 amino acids long. This length matches the length required to bind into the peptide binding groove of class I MHC molecules. In addition, ERAP1 is upregulated in the ER of cells treated with interferon gamma (IFN-γ).

Knock-down of ERAP1 by siRNA results in less overall antigenic presentation during IFN-γ treatment, although the knock-down does not affect all class I MHC epitopes equally. Knock-out studies show that ERAP1 effects the antigen repertoire at the cell surface. These and other data implicate ERAP1 as an important player in class I MHC antigen presentation. A chapter of this thesis will describe the crystallographic work describing the structures of ERAP1 with an aminopeptidase inhibitor, bestatin, and ERAP1 without an inhibitor that suggest possible peptide binding site in ERAP1 that will allow it to generate suitable substrates for a subset of class I MHC alleles.

Class II MHC plays a key role in the immune response by presenting antigenic peptides on CD4+ cytotoxic cell surfaces for T-cell response. The binding of peptides onto the MHC is an important step in creating an immune response. Structures of peptide bound MHC class II show conserved side chain binding pockets within the overall peptide-binding groove. In HLA-DR1, a common human class II MHC, the P1 pocket shows a preference for large hydrophobic side chains. Development of environmentally sensitive peptide analogs, that can bind into the class II MHC the same way as native peptides, can assist in visualizing the antigen binding process. A chapter in this thesis describes the crystallographic work showing that (4-DAPA)-HA can be used to study antigen-presenting processes in a cell by visualizing the changes in fluorescence of the synthesized peptide upon antigen loading.

Crystallographic analysis of MHC class II, HLA-DR1, in complex with HIV gag-derived peptide, GagP16(PEVIPMFSALSEGATP), and superantigen, SEC3-3B2, reveals the conventional polyproline conformation up to MHC binding pocket residue, P9, while the C-terminus of GagP16 adopts an unusual β- hairpin loop structure. Additionally, interactions between the leucine at P8 (LeuP8) and other residues on the loop such as ThrP16 and AlaP14 of the hairpin loop, was observed. Importantly, GagP16 requires the last 4 amino acids (P13-P16), which is part of the hairpin loop, for T-cell recognition. Understanding what dictates the C-terminal hairpin loop and the interaction motif of HLA-DR1/GagP16 complex with its TCR will provide insights on why it is important for T cell activation. A chapter in this thesis discusses the structural investigation conducted to understand the determinants of the loop at the C-terminus of GagP16 using designed peptides. It will also discuss work involving HLA-DR1 with the T cell receptor, AC25, that was cloned from T cells that are specific to HLA-DR1 in complex with the GagP16 peptide.

## TABLE OF CONTENTS

## List of Tables

## List of Figures

## Abbreviations

MHC - major histocompatibility complex

HLA- human leukocyte antigen

CLIP- Class II-associated invariant chain

TCR- T cell receptor

T cells- T lymphocytes

B cells- B lymphocytes

IRAP- Insulin regulated aminopeptidase

ERAP- Endoplasmic reticulum aminopeptidase

DAPA- 4-*N*,*N*,-dimethylaminophthalimidoalanine

HA- Influenza-derived peptide residues 306-PKYVKQNTLKLAT-318

GagP16- Gag caspid derived peptide residues 34-PEVIPMFSALEGATP-48

APC- Antigen presenting cell

IFN- Interferon

HC- Heavy chain

IgG- Immunogloblin type G

TAP- transporter associated with antigen processing

β2M- β2 microglobulin

## Preface

The work presented in chapter II on structural characterization of ERAP1 was done in collaboration with the laboratories of Efstratios Stratikos and Alfred Goldberg, who provided biochemical data.

The worked presented in chapter III on the characterizing the fluorogenic peptide, (4-DAPA)-HA was done in collaboration with a former Stern lab member, Prassana Venkatraman, Mauricio Calle-Calvo and the laboratory of Barbara Imperiali. The fluorophore (4-DAPA) was provided by the Imperiali lab and Prassana Venkatraman incorporated (4-DAPA) into the HA peptide and characterized the its spectroscopic activity when in complex with HLA-DR1 (Figure 3. 1). Mauricio Calle-Calvo executed the Il-2 bioassay with me (Figure 3.2).

### *References*

Nguyen TT, Chang SC, Evnouchidou I, York IA, Zikos C, Rock KL, Goldberg AL, Stratikos E, Stern LJ., Structural Basis For Antigenic Peptide Precursor Processing by the Endoplasmic Reticulum Aminopeptidase ERAP1, Manuscript Submitted.

Venkatraman P, Nguyen TT, Sainlos M, Bilsel O, Chitta S, Imperiali B, Stern LJ. Fluorogenic probes for monitoring peptide binding to class II MHC proteins in living cells. Nat Chem Biol. 2007 Apr;3(4):222-8.

# CHAPTER 1

# INTRODUCTION

## *Overview of Studies in Thesis*

The focus of this thesis is on the molecules involved in the peptide antigen processing for loading onto the major histocompatibility complex (MHC) that is important for activation of T- lymphocytes (T cells). It will include structural characterization of an endoplasmic reticulum aminopeptidase, ERAP1, a fluorogenic peptide, (4-DAPA)-HA, that can be used for visualizing peptide loading only MHC molecules, and peptide MHC complexes that was designed to understand an unique C-terminal hairpin structure exhibited by an HIV derived peptide in the context of MHC molecules. These studies provide further structural understanding of the molecules involved in antigen presentation.

## *Overview of the Immune System*

The immune system defends vertebrates from foreign agents by differentiating between self and non-self, and then responding to eliminate non-self. It plays a major role in many health conditions such as tumor recognition, autoimmunity, grafted organ acceptance, allergies and protection against infectious agents. Immune responses can be categorized as innate or adaptive responses. The innate immune system recognizes and responds to conserved microbial pathogen-associated molecular patterns such as lipopolysaccharide,

and results in immediate effects such as inflammation. The adaptive immune system, also referred to as the specific immune system, is slower to respond relative to the innate immune system, but is more selective, and can “learn” to recognize extremely varied antigens with great specificity. B- lymphocytes (B cells) recognize epitopes directly via their clonotypic B cell receptors that can be secreted in soluble form as soluble antibodies, as in the case of neutralizing antibodies. T cells, however, recognize foreign pathogens that are presented on the surface of other cells through presenter molecules such as MHC proteins (Janeway, 2001).

In the adaptive immune system, antigen presenting cells (APC) present antigens such as peptides, DNA or carbohydrates on the cell surface for T cell recognition and elimination. T cells are categorized based on their expression of one of two glycoprotein surface markers, CD8 or CD4. T cells expressing the CD4 surface marker (“Helper” T cells) generally release cytokines and chemokines upon recognition of specific antigen, resulting in recruitment and activation of macrophages for phagocytosis, and stimulation B cells for antibody production against the antigen recognized. T cells expressing the CD8 surface marker generally play a cytotoxic role and recognize cells presenting non-self antigens and directly eliminate those cells. However, some CD4 expressing T cells were shown to have cytotoxic roles and some CD8 expressing T cells secrete cytokines (Janeway, 2001).

T cell antigen recognition is mediated by antigen presenting molecules, MHC proteins, that present self and non-self peptides on the surface of APCs. CD4+ T cells recognize class II MHC molecules and CD8+ T cells recognize class I MHC molecules. While both classes of MHC molecules present peptides on the cell surface, they differ structurally, in peptide binding specificity, in peptide antigen processing pathways and in their source of peptide antigens (Braciale et al., 1987a; Madden, 1995; Stern et al., 1994). Both classes of MHC are loaded with peptide epitopes through a multi-step process that includes folding in the endoplasmic reticulum (ER), trafficking through the Golgi apparatus and presentation on the cell surface for T cell recognition. In general, for both class I and class II MHC complexes, antigen must be processed to be suitable for loading onto the MHC molecules, and the nascent MHC molecules must be matured into a receptive state for loading the antigens (Figure 1.1) (Braciale et al., 1987b).

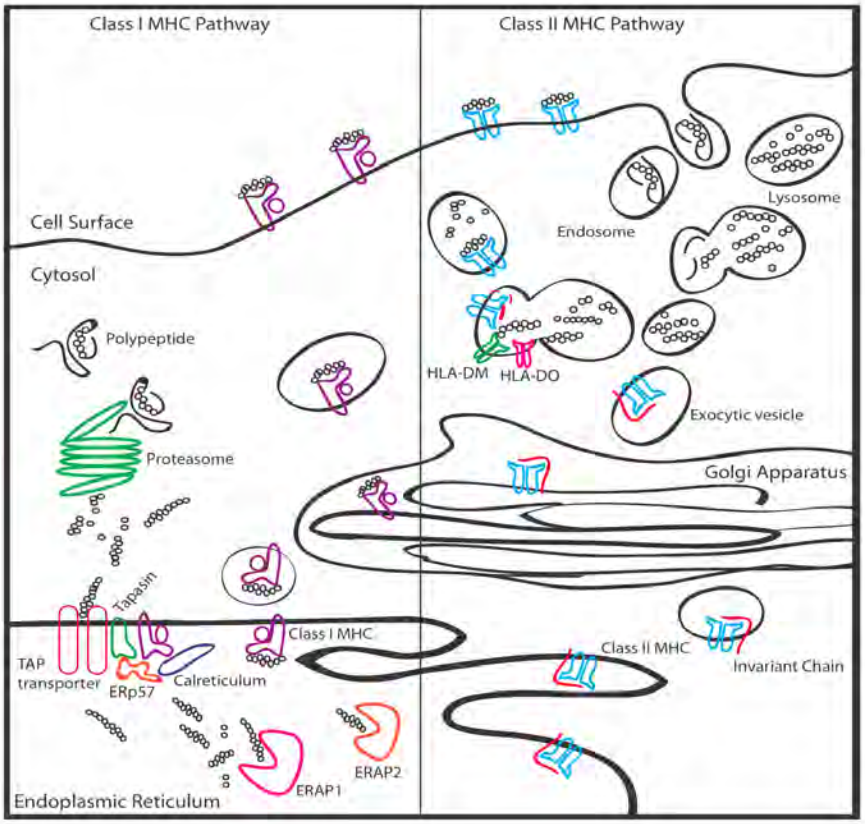

Class I MHC Pathway
Class II MHC Pathway
Cell Surface
Cytosol
Polypeptide
Proteasome
Tapasin
TAP
transporter
ERp57
Calreticulum
Class I MHC
ERAP1
ERAP2
Endoplasmic Reticulum
Endosome
Lysosome
HLA-DM
HLA-DO
Exocytic vesicle
Golgi Apparatus
Class II MHC
Invariant Chain

**Figure 1.1. Illustration of MHC Antigen Loading Pathways.** Left side shows class I MHC peptide loading pathway where a polypeptide in the cytosol is cleaved by the proteasome into shorter peptides. The peptides are transported into the endoplasmic reticulum by TAP transporter molecules where peptides can bind into class II MHC molecules. Empty class I MHC molecules interact with members of the antigen loading complex, tapasin, ERp57 and calreticulum in the ER. Longer peptides can be further trimmed by ER aminopeptidases, ERAP1 and/or ERAP2 before MHC loading. Class I MHC bound to a peptide dissociates from the loading complex are transported through the Golgi apparatus to the cell surface for CD8+ T cell recognition. Right side shows the class II MHC peptide loading pathway where foreign proteins are endocytosed into the cell through endosomes which fuses with the lysosome where polypeptides can be cleaved into peptides by low pH activated proteases. Next the endosome merges with exocytic vesicles that carry class II MHC associated with the invariant chain. The invariant chain gets cleaved to a shorter CLIP peptide, which occupies the peptide binding groove of class II MHC until HLA-DM that is regulated by HLA-DO assists peptide loading onto class II MHC in the merged endosomes. Peptide bound class II MHC then gets transported to the cell surface for CD4+ T cell recognition.

***Structural Features of Class I MHC***

Class I MHC molecules present self and foreign peptide antigens for CD8+ cytotoxic T cell recognition. Class I MHC molecules are heterodimers consisting of a polymorphic heavy chain (HC) and a monomorphic immunoglobulin (IgG) like light chain (β2M). HC is a glycoprotein that consists of 4 domains: α1 and α2 make up the peptide binding groove for antigens, α3 is a membrane-proximal IgG-like domain and the fourth domain consist of a C-terminal transmembrane domain and a short cytoplasmic tail (Korman et al., 1982; Orr et al., 1979; Townsend et al., 1990). The light chain β2M is not membrane-bound; non-covalent interactions between HC and β2M stabilize upon peptide binding (Figure 1.2) (Townsend et al., 1990). Peptides that bind into the binding groove are typically 8-10 amino acids in length, usually with a positively charged or hydrophobic C-terminal residue (Rammensee et al., 1995). Pockets along the length of the peptide binding groove provide further peptide sequence specificity (Collins et al., 1995; Mattson et al., 1989). Since the MHC gene is highly polymorphic, different class I MHC alleles may vary in peptide binding specificity because a majority of the polymorphic amino acids are in the peptide binding groove (Parham et al., 1989). For example, two mouse class I MHC alleles, H2-$K^d$ and H2-$K^b$, exhibit different peptide specificities. H2-$K^d$ binds 9-mers and prefers the amino acids, TYR or PHE, at the $8^{th}$ position from the C-terminus which like H2-$K^b$ prefers positively charged or hydrophobic amino acids (for example, Y/FXXXXXX-positive/hydrophobic) (Falk et al., 1991). H2-$K^b$ binds

8-mers and prefers TYR or PHE at the 4$^{th}$ position from the C-terminus (for example, XXXXY/FXX-positive/hydrophobic), but shows no preference for a particular amino acid at the 8$^{th}$ position like that observed in H2-K$^{d}$ (Falk et al., 1991; Rammensee, 1995). Unlike class II MHC, class I MHC binds to its peptide substrate in a way that the both N- and C- terminus of the peptide are in the pockets of the class I MHC (Bjorkman et al., 1987; Stern et al., 1994).

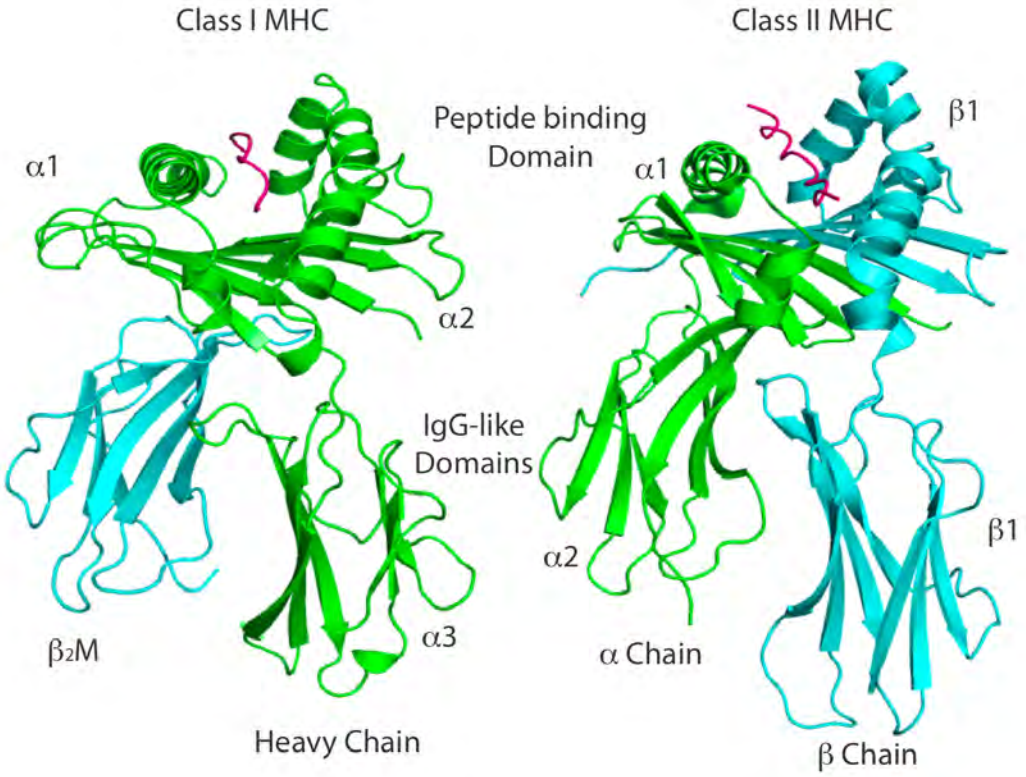

Class I MHC
Class II MHC
Peptide binding
Domain
α1
α2
IgG-like
Domains
β2M
α3
Heavy Chain
α1
β1
α2
β1
α Chain
β Chain

**Figure 1.2. Class I and Class II MHC are Structurally Homologous.** MHC molecules are represented in ribbon diagram. Left side shows a class I MHC molecule heavy chain is green and β2M is cyan. The class I molecule, H2-Kb, is bound to the ovalbumin derived peptide, SIINFEKL (PDB 1VAC). Right side panel shows a class II MHC molecule. The α chain is green and β chain is cyan. The class II molecule, HLA-DR1, is bound to the influenza hemagglutinin derived peptide, HA (PDB 1DLH). Peptides in both molecules are shown in magenta cartoon.

### *Class I MHC Antigen Processing and Presentation Pathway*

Class I MHC loading requires two elements to come together: first, the nascent class I MHC molecule is synthesized in the ER stabilized by the antigen loading complex; and second, cytosolic proteins must be proteolized by the proteasome and transported into the ER for further trimming and presentation (Figure 1.1). Many of the proteins involved in this pathway are upregulated upon cytokine stimulation (Aki et al., 1994; Klar and Hammerling, 1989; Saric et al., 2002). Modulation of proteins in this pathway can result in changes in the cell surface antigen repertoire and can influence T cell response (Boulanger et al., 2010; Garbi et al., 2000; York et al., 2006). For example, herpes simplex virus, HSV, can inhibit a crucial transporter protein, TAP, involved in transporting antigen precursors into the ER. This reduces the ability of the infected cell to present HSV protein epitopes bound to class I MHC on their surface for T cells to detect (Galocha et al., 1997).

Class I MHC antigen processing begins in the cytosol where the proteasome subunits β1, β2, β5 are replaced by β1i, β2i and β5i, resulting in a proteasome termed the “immunoproteasome” because of its different enzymatic activity (Aki et al., 1994; Frentzel et al., 1994). The immunoproteasome was shown to produce more N-extended (9-15 residues) peptides compared to the proteasome (Cascio et al., 2001). Additionally, in studies using the ovalbumin derived peptide, SIINFEKL, the C-termini of peptide antigens are formed in the cytosol by the proteasome (Cascio et al., 2001; Craiu et al., 1997). When the

immunoproteasome peptide products are not degraded in the cytosol by other peptidases such as leucine aminopeptidase into single amino acids, the peptides can be transported into the ER by an ATP dependent transporter, transporter associated with antigen processing (TAP), which is also upregulated upon cytokine stimulation (Bacik et al., 1994; Beninga et al., 1998; Reits et al., 2004). TAP can transport peptides that are at 8-16 residues long but prefers 9-12 residues peptides (Schumacher et al., 1994; van Endert et al., 1994). Similar to class I MHC molecules and the aminopeptidase ERAP1, TAP prefers to transport peptides with a hydrophobic or arginine C- terminal residues (F, L, R, Y, V) (Uebel and Tampe, 1999). N-extended peptides in the ER require further processing before loading onto class I MHC. Inside the ER, peptide antigen can be further trimmed from the amino terminus to the suitable size of 8-10 amino acids by the M1 family of aminopeptidase members, ERAP1 or ERAP2 (Saric et al., 2002; Tanioka et al., 2003; York et al., 2002). Even though not all antigens require further processing in the ER, a majority of peptides entering the ER through TAP will need further processing because of the specificity of TAP for peptides of at least 8 residues and the production of longer peptides by the immunoproteasome (Cascio et al., 2001; Schumacher et al., 1994). The peptide bound class I MHC complex is released from the antigen loading complex and processed through the Golgi apparatus for presentation on the cell surface for CD8+ T cell recognition.

### *Class I MHC Antigen Loading Complex*

Newly synthesized class I MHC in the ER associates with the loading complex before a peptide is loaded, and the peptide-MHC complex is transported to the cell surface. It was shown that before peptide loading onto class I MHC molecules, HC and β2M interact with ER chaperones, calreticulin or calnexin, as well as with the protein sulfide isomerase family member, ERp57 (Degen and Williams, 1991; Morrice and Powis, 1998). Calnexin is a membrane bound chaperone that interacts with nascent HC before the closely related soluble chaperone binds (Kang et al., 2009). The chaperones have been shown to interact with HC through their lectin like domains and stabilize HC with β2-m (Degen and Williams, 1991; Leach et al., 2002). ERp57 is a member of the PDI family and is composed of four thioredoxin domains, of which only two contain an active site motif and was shown to affect the rate of disulfide bond formation in HC (Dick et al., 2002). It was suggested that upon antigen binding, conserved Cys 101 and 164 in the peptide binding site of HC are shielded from ERp57 reduction (Dick et al., 2002). However, it is unclear what redox activities are involved in the loading of antigen onto class I MHC molecules. HC was, also shown to be, tethered to TAP by a membrane bound protein, tapasin (Suh et al., 1999). Tapasin was proposed to be a peptide editor for class I MHC and to assist in peptide loading (Garbi et al., 2000; Momburg and Tan, 2002). HC, β2M, TAP, tapasin, ERp57 and calreticulin collectively make up what is termed the class I MHC antigen loading complex. The antigen loading complex members

have not been shown to interact with HC after antigen is loaded into the class I MHC molecule. Since ERAP1 is in the ER and mediates the final processing step of the antigen before loading onto class I MHC, it is reasonable to predict that ERAP1 retention in the ER can be through interactions with a component of the antigen loading complex though it has not been shown.

### *The ER Aminopeptidase, ERAP1*

The ER aminopeptidase, ERAP1, is a 941 residue long aminopeptidase in the M1 family of zinc aminopeptidases (Hattori et al., 1999). ERAP1 has also been referred to as type 1 tumor necrosis factor receptor shedding aminopeptidase regulator (ARTS-1), adipocyte-derived leucine aminopeptidase (ALAP), puromycin-insensitive leucine aminopeptidase (PILS-AP) and ER aminopeptidase associated with antigen processing (ERAAP) because of ERAP1's characterization by different research groups (Cui et al., 2002; Hattori et al., 1999; Schomburg et al., 2000; Serwold et al., 2002). Knockout studies have shown that ERAP1 to be the major antigen processing enzyme in the ER and knock down studies show that ERAP1 depletion effects the presentation of some epitopes (Chang et al., 2005; York et al., 2006). ERAP1 was shown to effect the overall cell surface antigen repertoire but does not affect all antigens, or even all class I MHC, in CMV infected cells and in non infected cells (Blanchard et al., 2010). Additionally, ERAP1 expression was altered in human tumors (Fruci et al., 2008; Kazeto et al., 2003; Mehta et al., 2009; Watanabe et

al., 2003). Studies comparing tumor and non tumor cell lines show that ERAP1 expression is decreased in melanoma, carcinoma and lymphoma cells (Fruci et al., 2008). ERAP1 polymorphisms have been observed to correlate with increased risk of ankylosing spondylitis, a chronic inflammatory autoimmune disease that highly correlates with the class I MHC allele HLA-B27 (Reveille et al., 2010; Schlosstein et al., 1973).

Pull-down studies have shown that ERAP1 interacts with a homologous M1 aminopeptidase, ERAP2 (Saveanu et al., 2005; Tanioka et al., 2003). Additionally, ERAP1 and ERAP2 have been shown to cleave the same peptide (HIV derived peptide, RSLYNTVATL), and have been shown to have different peptide cleavage specificity. ERAP1 prefers hydrophobic residues at the N-terminus and ERAP2 prefers basic residues at the N-terminus (Hattori et al., 1999; Tanioka et al., 2003). Aside from being upregulated in the ER by cytokines where nascent class I MHC awaits peptide loading, ERAP1 has enzymatic characteristics that make it specialized for generating peptide antigen (Saric et al., 2002). ERAP1 shows substrate size specificity by cleaving peptides from 9-16 amino acids long more rapidly than shorter peptides (Chang et al., 2005; Saric et al., 2002; York et al., 2002). ERAP1 also has an N- and C- terminal residue preference for small hydrophobic residues in addition to an internal preference for a positively charged residue around 3 residues from the C-terminus (Chang et al., 2005; Evnouchidou et al., 2008). These characteristics make it unique compared to other M1 aminopeptidase family members, and particularly suitable

for class I MHC antigen processing. Since neither ERAP1 nor ERAP2 has an ER retention sequence, it remains a mystery as to how these ER resident aminopeptidases stay in the ER. It is possible but not shown that ERAP1 and ERAP2 interact with members of the antigen loading complex that stabilize the nascent class I MHC before peptide loading. Many questions about ERAP1's mechanisms and activities still remain elusive. How does ERAP1 differ from other aminopeptidases that allows it to bind N-extended peptides in the ER and cleave the peptide to the appropriate length for MHC binding? Chapter 2 reports the crystal structure of ERAP1 in complex with the inhibitor, bestatin, and suggests a structural basis for peptide trimming.

### *M1 Family of Aminopeptidases*

The M1 family of aminopeptidases, of which ERAP1 is a member, are zinc aminopeptidases with a conserved zinc binding motif HExxH$x_{18}$E and a conserved GXMEN motif. Members of this family play a variety of biological roles, including insulin regulation, angiotensin III degradation in blood pressure regulation, maternal serum regulation in child birth, as well as peptide cleavage for antigen processing (Cui et al., 2002; Hattori et al., 1999; Saric et al., 2002; Schomburg et al., 2000; Tsujimoto and Hattori, 2005; York et al., 2002). In some organisms, such as *A. thaliana* and *P. falciparum*, M1 aminopeptidases are targeted as a receptor for pesticides (Azimzadeh et al., 2010; Peer et al., 2009). Currently, there are structures of eight M1 family members including ERAP1.

The structure of ERAP1 will be presented and discussed in detail in Chapter 2. The other M1 aminopeptidases for which structures have been solved are human leukotriene hydrolase (LTA4H), cold active aminopeptidase (CoAP) from *C. psychrerythrea,* the tricorn interacting factor f3 (TIFF3) from *T. acidophilium*, aminopeptidase N (ePepN) from *E. coli,* aminopeptidase M1 (PfaM1) from *P. falciparum* and aminopeptidase M1 from *N. meningitidis* (Bauvois et al., 2008; Ito et al., 2006; Kyrieleis et al., 2005; Nocek et al., 2008; Tholander et al., 2008). All the above M1 aminopeptidases exhibit primary sequence homology mostly around the catalytic region, which is a thermolysin-like domain, and very little homology in the C-terminal region. However, the structures exhibit a homologous domain arrangement consisting of an $\alpha/\beta$ N-terminal domain, a thermolysin-like catalytic domain, a helical C-terminal domain and in some cases a small (inserted) domain between the catalytic and C-terminal domain (Figure 1.3) (Bauvois et al., 2008; Ito et al., 2006; Kyrieleis et al., 2005; Nocek et al., 2008; Tholander et al., 2008). Substrate specificity and hints on mechanism have been suggested for LTA4H and PepN through mutagenesis, inhibition and crystallographic studies (Addlagatta et al., 2006; Jiang et al., 2010; Minami et al., 1992; Tholander et al., 2008). Homology modeling was used to predict that the active site of ERAP1 is located in the thermolysin like domain; however, the environment of the rest of the peptide binding domain remained elusive. The structure of ERAP1 presented in Chapter 2 suggests a common domain arrangement among the M1 family of aminopeptidases and a possible

mechanism by which ERAP1 can interact with the peptide substrate via the C-terminal domain that does not have sequence homology to other M1 family members.

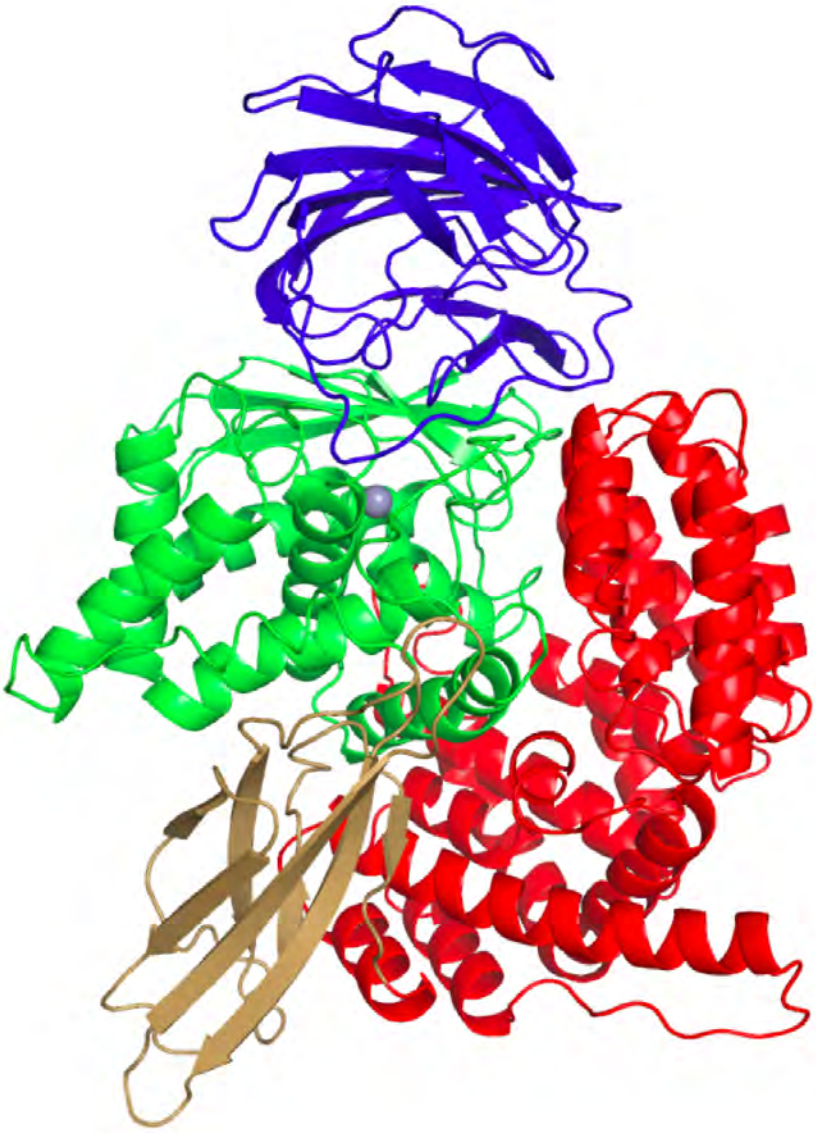


**Figure 1.3. Structure of Representative M1 Aminopeptidase Family Member, M1 Alanylaminopeptidase from Malaria.** The aminopeptidase is shown in ribbon representation and the zinc atom is a grey sphere. The domains are colored as follows: N-terminal domain I in blue, catalytic domain II in green, inserted domain III in brown, C-terminal domain IV in red. The PDB ID for this structure is 3EBG.

### *The Oxytocinase Subfamily M1 Family of Aminopeptidases*

ERAP1, ERAP2, and insulin regulated aminopeptidase (IRAP) make up the oxytocinase subfamily of M1 aminopeptidase because of their activity on peptide hormones and their homology (Tanioka et al., 2003). Primary sequence alignment of the three aminopeptidases is shown in Figure 1.4. ERAP1 and ERAP2 are ER residents that have been shown to trim antigenic precursors to the appropriate length for class I MHC loading and have different N-terminal substrate specificities (Saric et al., 2002; Tanioka et al., 2003; York et al., 2002). The third member, IRAP, is a membrane protein that co-localizes with the glucose transporter 4 in intracellular vesicles targeted to the cell surface (Ross et al., 1996). IRAP have been shown to be involved in class I MHC cross presentation pathways in endosomes (Saveanu et al., 2009). IRAP unlike ERAP1 cleaves SIINFEKL to INFEKL, which is not suitable for class I MHC binding (Georgiadou et al., 2010). Since the three oxytocinase subfamily members share 40-49% identity, a structural comparison among the family members may lead to insight into how these aminopeptidases can control their product sizes.


**Figure 1.4. Sequence Alignment of Human ERAP1, ERAP2, and IRAP**

### *Structural Features of Class II MHC*

Class II MHC molecules present self and foreign peptide antigens to CD4+ T cells. Class II MHC molecules are heterodimers consisting of two subunits, the $\alpha$ chain and the $\beta$ chain. Each chain has an upper domain that contributes to the peptide binding groove, a membrane-proximal IgG-like domain, a transmembrane domain, and a cytoplasmic tail (Bjorkman et al., 1987; Stern et al., 1994). Peptide elution studies suggest that the average-sized peptide bound to a class II MHC is about 13-20 residues in length, which is longer than peptides that bind to class I MHC (Chicz et al., 1992; Rudensky et al., 1991). The class II MHC peptide binding domain is made of both the $\alpha$ and $\beta$ chain, each contributing one helix lining the groove and half of the beta sheet that makes up the "floor" of the binding groove (Figure 1.4). Crystal structures show that conserved interactions between the main chain of the bound peptide and main chain and side chains of the MHC hold the peptide in a polyproline type II conformation independent of the peptide sequence (Figure 1.5) (Stern et al., 1994). The amino and carboxyl-termini of peptides bound to class II MHC are free to extend beyond the boundaries of the peptide binding groove (Stern et al., 1994). Although main chain interactions hold the peptide in the groove, peptide side chain sequence specificity still exists, and is dictated by pockets lining the peptide binding groove (Figure 1.6) (Stern et al., 1994; Zavala-Ruiz et al., 2004a; Zavala-Ruiz et al., 2003). The binding specificity of the class II MHC, HLA-DR1, has been well studied, and is usually discussed as a nine amino acid binding

“register” (P1-P9). HLA-DR1 has a P1 binding pocket that can accommodate large hydrophobic residue. Binding studies have shown that the P1 position plays a major role in dictating whether a peptide will bind to the class II MHC molecule (Jardetzky et al., 1990). Other pockets on the MHC also contribute to peptide binding specificity. For HLA-DR1, the P4 pocket prefers aliphatic amino acids; the P6 pocket is more of a shelf that prefers small amino acids; P9 prefers aliphatic residues; and P10 (outside the canonical “register”) shows a preference for phenylalanine or tyrosine (Jardetzky et al., 1990; Stern et al., 1994; Zavala-Ruiz et al., 2004a; Zavala-Ruiz et al., 2003). The crystal structure of HLA-DR1 with the T cell receptor, HA1.7, specific for HLA-DR1 bound to the flu hemagglutinin-derived peptide, HA, (PKYVKQNTLKLAT) shows that the italicized residues to be anchored into pockets in the MHC and the underlined residues protrude out of the peptide binding groove and act as T cell contacts that were shown to be important in T cell activation studies (Hennecke and Wiley, 2002). Prediction programs such as RANKPEP and SYFPEITHI can take a long peptide sequence and suggest a binding motif in the peptide binding register of the MHC and suggest residues that may interact with the T cell receptor (DiBrino et al., 1994; Rammensee et al., 1999; Reche et al., 2004). However, the information about flanking residues outside the peptide binding pockets is limited and Chapter 4 will discuss studies of an HIV derived peptide that exhibits a C-terminal hairpin loop that is important for T cell activation.

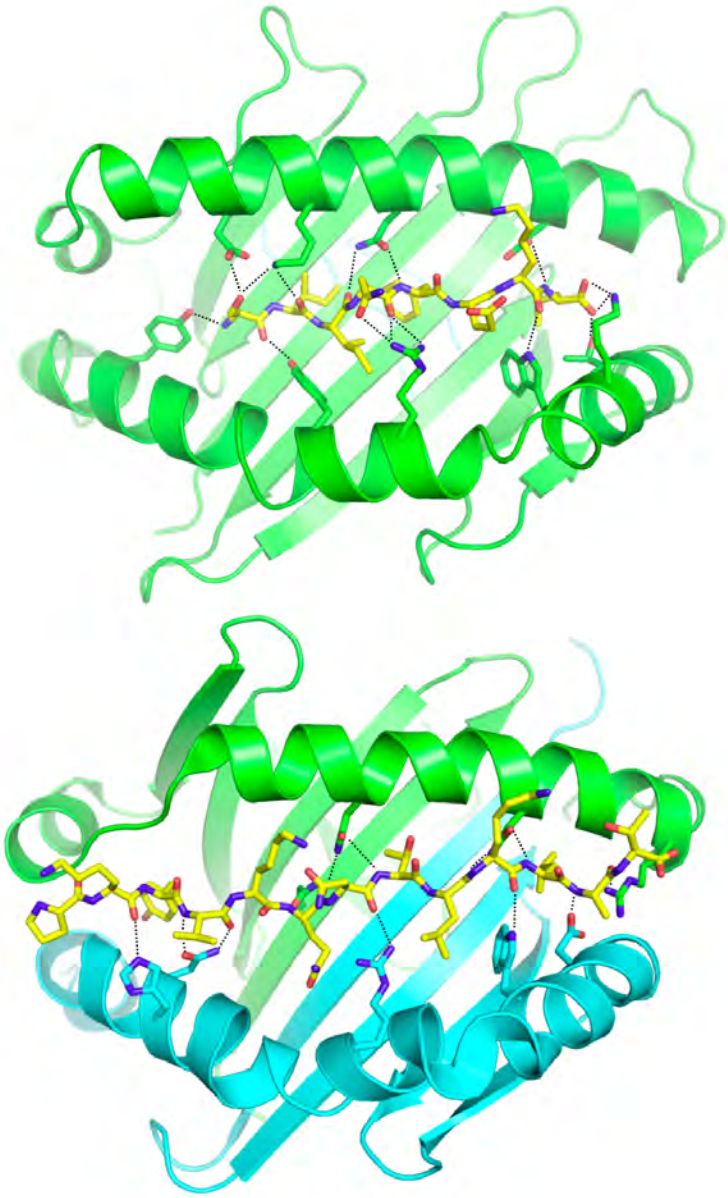


**Figure 1.5. Class I and Class II MHC Main Chain with Peptide Interactions.** MHC molecules are represented in ribbon diagram. Top panel shows a class I MHC molecule heavy chain is green. Both termini of the peptide interact with the MHC molecule. Bottom panel shows a class II MHC molecule. The $\alpha$ chain is green and the $\beta$ chain is cyan. The termini of the peptide extends outside the groove. The hydrogen bonding network holds the peptide in a polyproline type II conformation. Peptides in both molecules are represented in yellow sticks representation. Hydrogen bonding interactions are represented in dotted lines. The same PDB ID is in Figure 1.2.

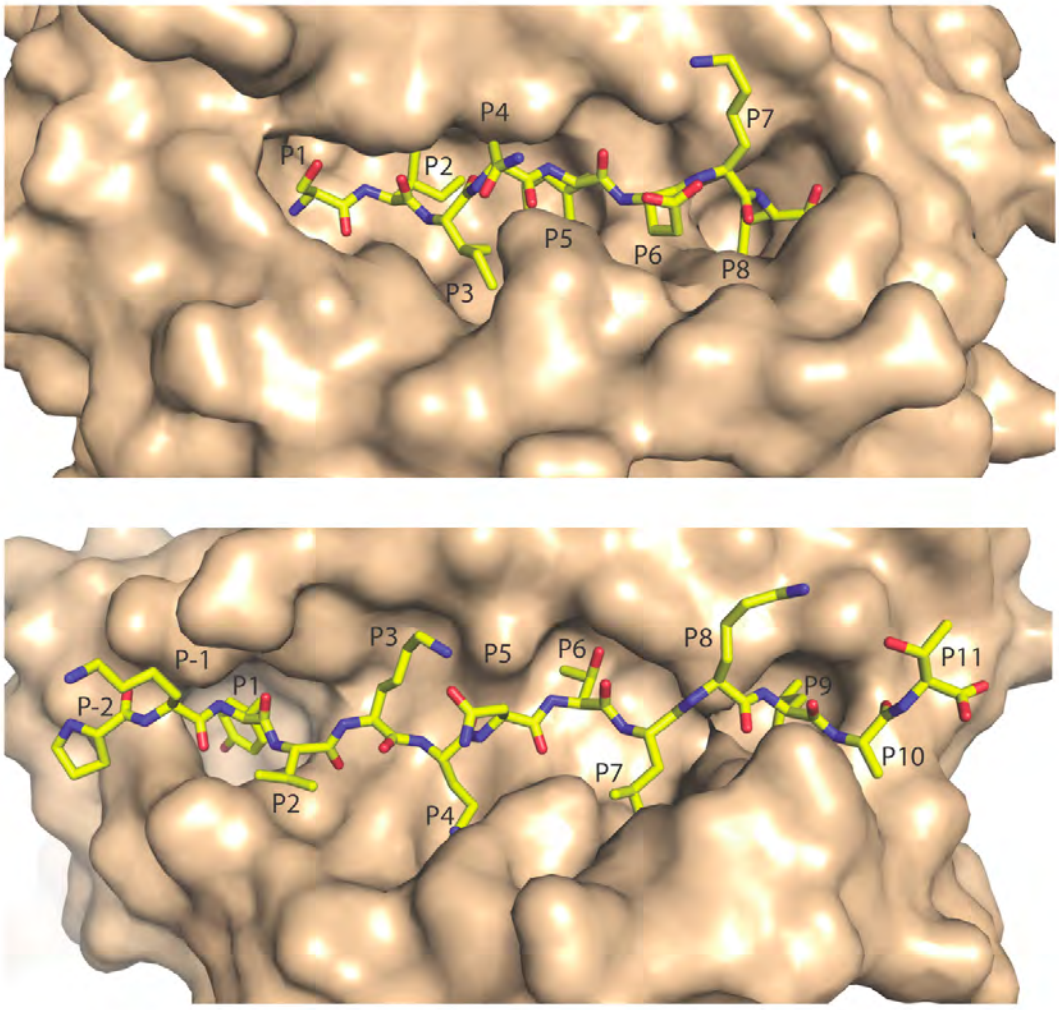

P4
P7
P1
P2
P5
P6
P8
P3
P3
P5
P6
P8
P11
P-1
P1
P9
P-2
P10
P7
P2
P4

**Figure 1.6. Peptide Binding Pockets in Class I and Class II MHC.** The surface of MHC molecules is shown here is brown and peptides are represented as yellow sticks. Peptide labeling is coordinated with its binding pocket in the MHC. Top panel shows class I MHC, H2-Kb, having large pockets at P5 and P8 that explains the preference for binding to large hydrophobic groups in those positions. Bottom panel show the class II MHC, HLA-DR1, having pockets at P1, P4, P6, and P9 that can allow for binding specificities. Same PDB ID is in Figure 1.2.

### *Class II MHC Antigen Processing Pathway*

Peptide sources for class II MHC typically are proteins that are endocytosed into endosomes, where they are degraded into small peptide fragments and loaded onto class II MHC molecules. The $\alpha$ and $\beta$ chains of class II MHC molecules are synthesized in the ER where the peptide binding groove is occupied by a chaperone called the invariant chain (Lamb and Cresswell, 1992; Roche and Cresswell, 1990). The class II MHC/invariant chain complex then traffics into endosomes where cleaved peptide fragments await. The lowered pH in the endosomal compartments activates proteases that cleave long versions of the invariant chain, leaving a shorter peptide fragment called “CLIP” in the peptide-binding groove (Lamb and Cresswell, 1992; Riberdy et al., 1992). Between the early endosome and late endosome, the peptide antigen replaces CLIP in the peptide-binding groove of the MHC. Class II MHC-like molecules, HLA-DM and HLA-DO, play a role in loading of peptides into MHC (Kropshofer et al., 1998; van Lith et al., 2001). HLA-DM has been termed the “peptide editor” because it facilitates the exchange of peptides in the groove including release of CLIP, and may affect the suitability of peptides that are eventually bound and presented at the cell surface (Denzin and Cresswell, 1995; Kropshofer et al., 1998; Sherman et al., 1995). HLA-DO can modulate HLA-DM and has inhibitory functions in peptide loading. HLA-DM is active between neutral and slightly acidic pH but optimal at pH 5.5. However, HLA-DO does not affect HLA-DM activity at acidic pH (Alfonso et al., 1999; Liljedahl et al., 1998). The molecular mechanisms

of peptide loading/exchange and the interactions between HLA-DM and HLA-DO with class II MHC remain unclear. The balance of the activities of HLA-DM, HLA-DO, and other facets of the endosomal environment result in the final array of class II MHC peptide complexes presented on the cell surface for CD4+ T cell recognition. Chapter 3 describes characterization of an environmentally sensitive fluorogenic peptide that can be utilized for real time visualization of the peptide loading process.

### *MHC/Peptide and TCR Modes of Interaction*

While there is much diversity in alleles of class I and class II MHC, there is significantly more diversity in T cell receptors (TCRs). TCRs exhibit similar structures to antibodies including two IgG-like constant domains (C$\alpha$ and C$\beta$) and two IgG-like variable domains (V$\alpha$ and V$\beta$) (Reinherz et al., 1999). The variable domains consist of complementarity-determining regions are made of loops (CDR loops) with extreme variability, and these loops that interact with the MHC-peptide complex (Hennecke and Wiley, 2002). While TCRs are selected during development are “restricted” by an MHC allele, the receptor will still not be activated unless an appropriate antigenic peptide is presented. Binding studies suggested that the interactions between the TCR and MHC follow a two step mechanism, where first the TCR interacts with regions of the MHC, which then interact with the peptide in the groove (Wu et al., 2002). Currently there are twelve structures of $\alpha/\beta$ TCRs bound to MHC peptide complexes. The

interactions between the TCR and MHC peptide complex can be described as having a diagonal or non-diagonal mode of interaction. In the diagonal mode of interaction is when the TCR interact with the MHC $\alpha$1 helix near the peptide N-terminus and diagonally across the peptide binding groove towards the peptide C-terminus and interact the MHC $\beta$1 helix. TCR interaction is over the center of the peptide (Figure 1.5) (Hennecke and Wiley, 2002). The non-diagonal mode of interaction is exhibited when the TCR is tilted and causes a biased towards the N-terminus of the peptide as observed in TCR Ob.1A12-HLA-DR2-mylin basic protein derived peptide (Maynard et al., 2005). The observed non-diagonal TCR/ MHC structures were of complexes consisting autoreactive TCR and self peptide MHC peptide complexes (Deng et al., 2007; Hahn et al., 2005; Maynard et al., 2005). These structures suggest that the MHC and TCR interactions are weak and allow for escape of TCR negative selection where the stronger self TCR are selected out (Rudolph et al., 2006). These different modes of interactions show that the MHC/TCR binding orientation is not strictly conserved (Figure 1.7).

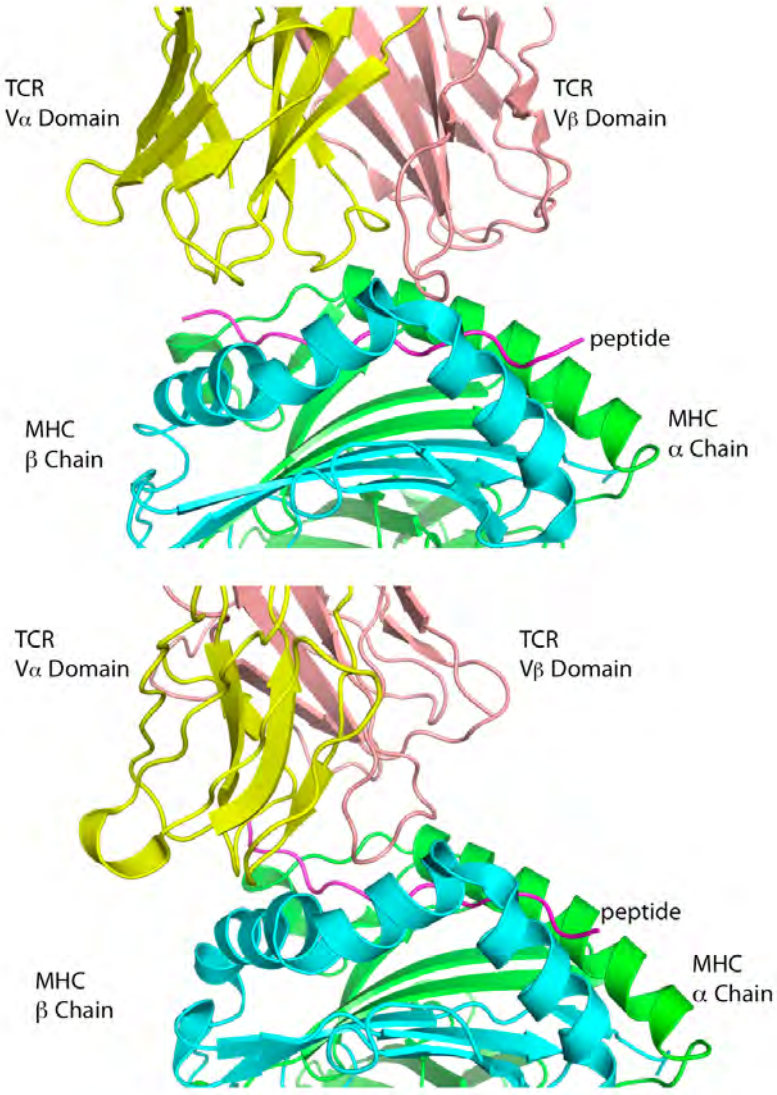

TCR
Vα Domain
TCR
Vβ Domain
peptide
MHC
β Chain
MHC
α Chain
TCR
Vα Domain
TCR
Vβ Domain
peptide
MHC
β Chain
MHC
α Chain

**Figure 1.7. Modes of Interaction Between α/β TCR and Class II MHC.** The molecules are represented in ribbon where the MHC α chain is green and β chain is cyan. The TCR α chain is yellow and β chain is pink. The peptide is magenta. The top panel shows the HA.17 TCR with the class II MHC, HLA-DR1, bound to HA peptide (PDB 1FYT). This complex structure exhibits a diagonal interaction between the MHC and TCR. The TCR interactions start from the $\alpha$1 helix and diagonally across the peptide binding groove to $\beta$1 helix of the MHC and the TCR interaction is over the center of the peptide. The bottom panel shows the Ob.1A12 TCR with the class II MHC, HLA-DR2, bound a self peptide derived from myelin basic protein, MBP peptide (PDB 1YMM). This exhibits a non-diagonal interaction between the MHC and TCR. The TCR interacts less with the center of the peptide but instead the N-terminal part of the peptide and MHC β chain.

Whether other and more different TCR and peptide MHC interaction orientations occur remains to be seen. T cell lines grown from HIV patients that respond specifically to a C-extended gag envelope-derived peptide, GagP16 (PEVIPMFSALSEGATP), but not the shorter version of the same peptide, GagP13 (PEVIPMFSALSEG), suggest a possibility of another mode of TCR and peptide MHC interaction biased toward residues extending beyond the C-terminus of the peptide-binding register of class II MHC (Figure 1.8) (Norris et al., 2006; Zavala-Ruiz et al., 2004b). Structural studies of HLA-DR1 bound to the GagP16 and HLA-DR1 bound to GagP13 shows that both peptides bind into the HLA-DR1 peptide binding groove in the same register with an exception of a hairpin-like loop in the C-terminus of the GagP16 peptide composed of the last five residues (EGATP) where the proline loops back towards the glutamate (Zavala-Ruiz et al., 2004b). Predicted T cell and peptide interactions are present in both gag envelope-derived peptides (Zavala-Ruiz et al., 2004b). Future studies of co-crystal structures of this TCR with the two peptide-MHC complexes will show whether the docking orientation in this system is different in a way that is biased towards the C-terminus of the peptide. In Chapter 4, structural studies will be discussed about the determinants of the C-terminal hairpin loop that are important for T cell activation.

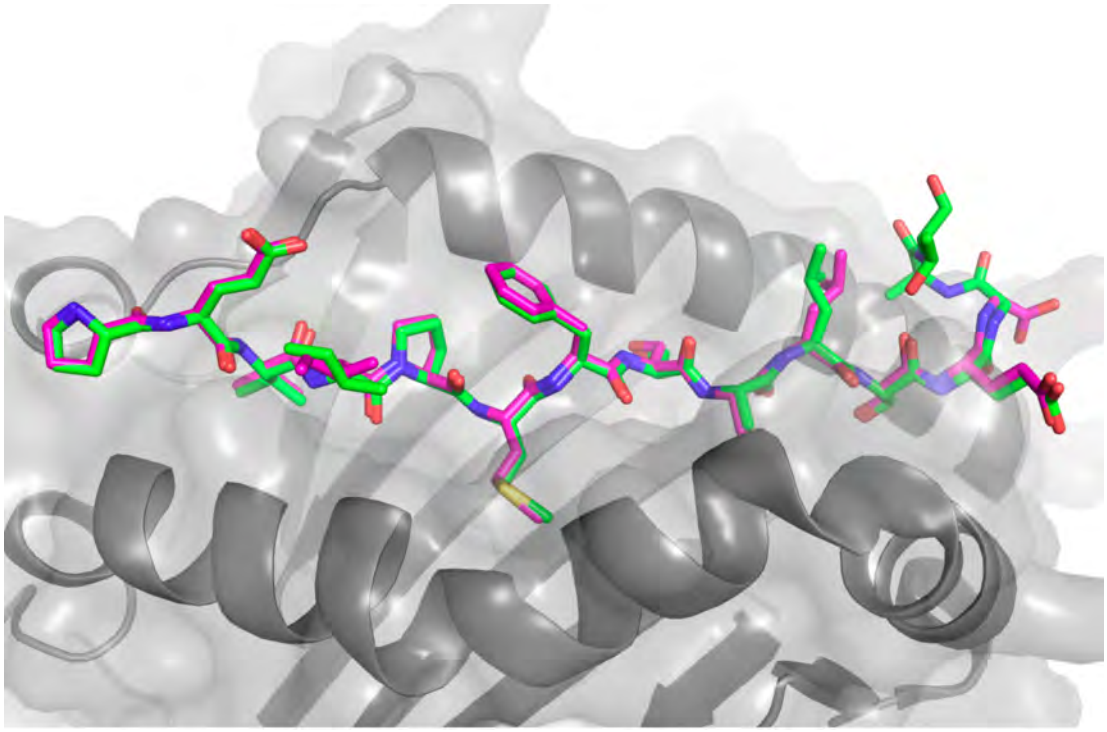


**Figure 1.8. HIV Gag Derived Peptides in Class II MHC.** The structures of the two gag derived peptides, GagP13 and GagP16, were overlay in the peptide binding groove of HLA-DR1 shown in grey ribbon and surface. The surface of the MHC molecule is transparent gray to allow visualization of peptide side chains inside the binding groove. This illustrates that GagP13 and GagP16 binds in the MHC molecule identically and the major difference between the two peptides in the context of HLA-DR1 is that the C-terminal hairpin loop exhibited only by GagP16 can activate AC25 TCR expressing T cells. The PDB IDs used in this figure are 1SJE and 1SJH.

***Studies in This Thesis***

The work in this thesis will contribute to understanding class I and class II antigen processing and class II MHC interaction with αβ TCR. In class I antigen presentation, the structural studies of ERAP1 will allow an understanding of how ERAP1 can exhibit unusual aminopeptidase specificity compared to other members of the M1 family of aminopeptidases, allowing ERAP1 to perform antigen processing activity. In class II antigen processing, structural characterization of an environmentally sensitive peptide probe bound to HLA-DR1 allowed for peptide binding studies *in vivo* to further understand endosomal antigen processing. To study class II MHC interaction with αβ TCR, HLA-DR1 bond to a series of gag-derived peptides were structurally characterized to investigate the role of the hairpin loop in class II MHC interaction for αβ TCR. These studies are meant to increase our understanding of the peptidases and presenting molecules that are involved in antigen peptide presentation for T cell recognition.

# CHAPTER 2

## Structural Basis For Antigenic Peptide Precursor Trimming By ERAP1, The Endoplasmic Reticulum Aminopeptidase Involved In Antigen Processing

The work presented in this chapter is in collaboration with the laboratories of Alfred Goldberg and Efstratios Stratikos. I did the crystallographic work, the Goldberg laboratory did the biochemistry experiments showing that short peptides are activators of small fluorogenic substrates and the Stratikos laboratory executed the normal mode analysis and enzymatic experiments with peptide substrates.

## INTRODUCTION

Most major histocompatibility complex (MHC) class I proteins have strict requirements on the lengths of peptide ligands that can be bound with high affinity, typically 8-10 residues, and therefore peptides destined for presentation by class I MHC molecules must be precisely cleaved before loading (Rock et al., 2004). Proteasomal degradation and cytosolic aminopeptidase processing are sufficient for the generation of some class I MHC antigens, but a majority of epitopes require an additional processing step in the endoplasmic reticulum (ER) after TAP-mediated transport (Fruci et al., 2001; Serwold et al., 2001). A search for the peptidase(s) responsible for that activity identified an ER-resident, interferon-$\gamma$ inducible, metalloaminopeptidase that was termed ERAP1 (ER-

associated aminopeptidase 1) or ERAAP (ER aminopeptidase associated with antigen processing) (Saric et al., 2002; Serwold et al., 2002).

ERAP1 processing impacts a large number of self and foreign epitopes. Studies of ERAP1-deficient cells have shown that ERAP1 generates many presented peptide epitopes from larger precursors, but it can destroy other epitopes by trimming them below the minimal size needed for MHC class I binding (York et al., 2002). The peptide repertoire carried by MHC molecules of wild-type and ERAP1–deficient mice differ substantially (Blanchard et al., 2010; Hammer et al., 2007). Epitope immunodominance hierarchies differ in wild-type and ERAP1–deficient mice infected with LCMV (York et al., 2006), and T cell responses against many model antigens were reduced in ERAP1 deficient mice (Yan et al., 2006). Furthermore, cross-presentation of cell-associated epitopes was lower in ERAP1-deficient mice (Firat et al., 2007). ERAP1-dependent epitopes were shown to be important in resistance to *Toxoplasma gondii* (Blanchard et al., 2008).

Allelic variants of ERAP1 have been linked to a number of human diseases, including the autoimmune disease ankylosing spondylitis (AS) (reviewed in (Brown, 2010)), diabetes (Fung et al., 2009), some forms of cervical cancer (Mehta et al., 2009), and high blood pressure (Yamamoto et al., 2002).

An important aspect of ERAP1's specificity related to its biological function is its ability to cleave long substrates, in contrast to other aminopeptidases, which generally are restricted to substrates shorter than 4 residues (Chang et al.,

2005). ERAP1's length specificity is characterized by a sharp increase in activity near 8-10 residues, with peptides up to ~16 residues being processed efficiently (Chang et al., 2005; York et al., 2002). This activity matches ERAP1's role in processing antigenic peptide precursors. Cytosolic proteasomal processing generates peptides of 2-25 residues (Kisselev et al., 1999) with C-terminal residues that generally match MHC class I sequence requirements, and TAP1/TAP2, the transporter associated with antigen processing that brings antigen precursors into the ER, prefers of peptides 8-16 residues long (Momburg et al., 1994). Therefore, many transported peptides require additional trimming to fit into MHC class I binding sites, which typically are restricted to binding peptides that are 8-10 residues long (Rammensee, 1995). Post-proteasomal processing by ER aminopeptidases can be considered to be the final contributor to size selection of antigenic peptides. How ERAP1 plays this role is controversial; in addition to ERAP1's intrinsic length dependence (Chang et al., 2005). Peptide hydrolysis in the presence and absence of class I MHC suggest that direct processing of antigenic precursors by ERAP1 while partially bound to MHC class I because MHC can protect peptides from being further cleaved by ERAP1 (Falk et al., 1990; Kanaseki et al., 2006). MHC-protection and peptide products released from ERAP1 can also contribute to the final repertoire of peptides loaded onto class I MHC.

ERAP1 has sequence as well as length specificity. ERAP1 was originally identified as a leucine-specific aminopeptidase based on studies with single

amino acid fluorogenic substrates (Hattori et al., 1999). The activity of ERAP1 towards peptides with various N-terminal residues has been characterized in detail *in vitro*. The N-terminal sequence specificity observed for peptides generally parallels findings from single residue model substrates, although with substantially broader overall reactivity, so that some N-terminal residues are processed efficiently in the context of longer peptides but not as single residue substrates (Chang et al., 2005; Evnouchidou et al., 2008; Hearn et al., 2009; Saveanu et al., 2005; Serwold et al., 2001). An *in vivo* study of ERAP1-mediated generation of T cell epitopes from longer antigenic precursors reported that the N-terminal specificity matches the enzyme's peptide specificity measured *in vitro* (Hearn et al., 2009). A statistical analysis of residues preceding naturally processed epitopes in large epitope databases also concurred (Hearn et al., 2009). Finally, in addition to its N-terminal specificity, ERAP1 is unusual in exhibiting specificity for the substrate's C-terminal residue (Chang et al., 2005) and internal sequence (Evnouchidou et al., 2008).

To clarify the structural basis for ERAP1 specificity and to evaluate possible models for its length dependence, we determined the 3.0 Å crystal structure of ERAP1 bound to the aminopeptidase inhibitor bestatin by X-ray diffraction. The structure reveals an open conformation with a large interior compartment. A shallow groove extending from the active site potentially can accommodate long peptides forming a series of electrostatic and hydrophobic interactions consistent with previously reported substrate preferences.

Experimental characterization of ERAP1's sensitivity to allosteric activation and inhibition by peptides, and comparison with the active site geometries of other M1 family members suggest a mechanism for the length-dependent hydrolytic activity and allows us to propose a molecular model of how ERAP1 trims antigenic peptide precursors.

## MATERIALS AND METHODS

### *Protein expression and purification*

Human recombinant ERAP1 was expressed using a baculovirus insect cell expression system as previously described (Evnouchidou et al., 2008). We used an ERAP1 allelic variant with four substitutions (G346D, G514D, K528R, Q730E) relative to the GenBank reference sequence NP_057526. ERAP1 was purified using Ni-NTA resin (Qiagen), eluted with 100 mM imidazole pH 8.0, concentrated to 1 mg/mL by centrifugal ultrafitration and treated with carboxypeptidase A (Sigma) (molar ratio 5:1) at room temperature in PBS with 0.25 mg/mL protein disulfide isomerase, PDI (Sigma). Carboxypeptidase A was removed with size exclusion chromatography (S200 column, GE Healthcare), and finally ERAP1 was purified by anion exchange chromatography (MonoQ, GE Healthcare). A single band was observed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Enzyme specific activity was monitored using L-AMC hydrolysis, and the specific activity increased as the protein was

purified.

***Crystallization and data collection***

ERAP1-bestatin complexes were crystallized by vapor diffusion in hanging drops after extended storage at 4°C. The crystallization reservoir solution consisted of 14% PEG 8000, 1 mM glutathione (10:1 ratio oxidized to reduced), 0.1 M Bicine, pH 8.6. The drops consisted of equal volumes reservoir and the protein solution (6.5 mg/mL ERAP1, 5:1 molar ratio bestatin, 0.25 mg/mL PDI). Crystals were transferred to reservoir solution containing 30% 2-methyl-2,4-pentanediol and flash frozen in liquid nitrogen. X-ray diffraction data were collected as 1° oscillations at 100 K with 1.08 Å radiation at beamline X29A in the National Synchrotron Light Source. Data were processed and scaled using HKL2000 (Otwinowski, 1997). Data collection and refinement statistics are shown in Table 2.1.

***Structure determination and refinement***

Phases were estimated by molecular replacement using PHASER (McCoy et al., 2007) with LTA4H (PDB 1HS6) and TIFF3 (PBD 1Z5H) as search models. A solution with three molecules, as suggested by Matthews coefficient analysis, was obtained yielding electron density of sufficient quality for model building. To validate the molecular replacement solution, composite omit maps (CNS)

(Brunger, 2007) were generated and inspected using COOT (Emsley and Cowtan, 2004) and O (Jones et al., 1991). Phases were improved in RESOLVE (Terwilliger, 2000) with 3-fold non-crystallographic symmetry real-space averaging. Averaged and unaveraged omit and prime-and-switch maps (Terwilliger, 2000) were utilized for model building. The entire C-terminal domain was different from the search models and was built *de novo*, whereas the N-terminal domains could be modeled based on the search models. Positions of three N-linked glycosylation sites (Asn 70, 154 and 760), and two disulfide bonds (410-449 and 742-749) matched sites of these modifications as observed by mass spectrometry (data not shown). After a complete monomer model was built, a trimer was reconstructed and inspected. Electron density consistent with a five monosaccharide unit chain [Man(a1$_{->}$6),Man(a 1->3)]-Man(b1->4)->GlcNAc(b1->4)->GlcNAc(1-4)-Asn] was observed linked to Asn 70 in two of the three molecules in the asymmetric unit, with a disaccharide unit at the other. At the two other glycosylation sites, no electron density or density consistent with only a monosaccharide was observed. Crystallographic rigid body, positional, B-factor, and TLS refinement was performed in PHENIX (Adams et al., 2010) with NCS restraints between each of the four domains of the trimer. The NCS restraints were optimized to be higher in domain II and lower in domains I, III, and IV. Geometric parameters for bestatin were obtained from the HIC-UP server (Kleywegt and Jones, 1998) During cycles of model building and refinement, the high resolution limit was increased. The inclusion of high resolution data shells

improved electron density maps, cross-validation statistics, and allowed building of some regions, despite having low I/σI and high $R_{sym}$ values.

The completed model consists of ERAP1 residue 46 to residue 934 (of 941 encoded by the expression vector) with five unobservable loops, between residues 416 to 434, 485 to 514, 551 to 556, 863 to 868 and 892 to 904. The second and longest corresponds to an ERAP1-specific insertion not required for enzymatic activity (ES, unpublished data); the others are exposed loops. Ramachandran statistics showed all residues in the allowed regions. Diffraction data and coordinates were deposited into the Protein Data Bank (PDB 3MDJ).

## RESULTS

### *Overall Structure of ERAP1*

ERAP1 is a member of the M1-family of zinc metallopeptidases characterized by GAMEN and $HExxHx_{18}E$ sequence motifs. The family includes a wide variety of soluble and membrane-associated aminopeptidases with various specificities and biological functions (Rawlings et al., 2009). Sequences of the five M1-family aminopeptidases for which crystal structures have been determined are aligned in Figure 2.1: human leukotriene A4 hydrolase (LTA4H), *T. acidophilium* tricorn interacting factor F3 (TIFF3), *E. coli* aminopeptidase N (ePepN), *P. falciparum* aminopeptidase M1 (PfAM1), and *C. psychrerythraea* cold active aminopeptidase (ColAP). These structures reveal increasingly

extensive elaborations of a thermolysin-like fold. Sequence homology is strongest in the catalytic domain containing the conserved GAMEN and HExxH$x_{18}$E motifs, weaker in the N-terminal portion, and essentially absent in the highly variable C-terminal helical region (Figure 2.1).

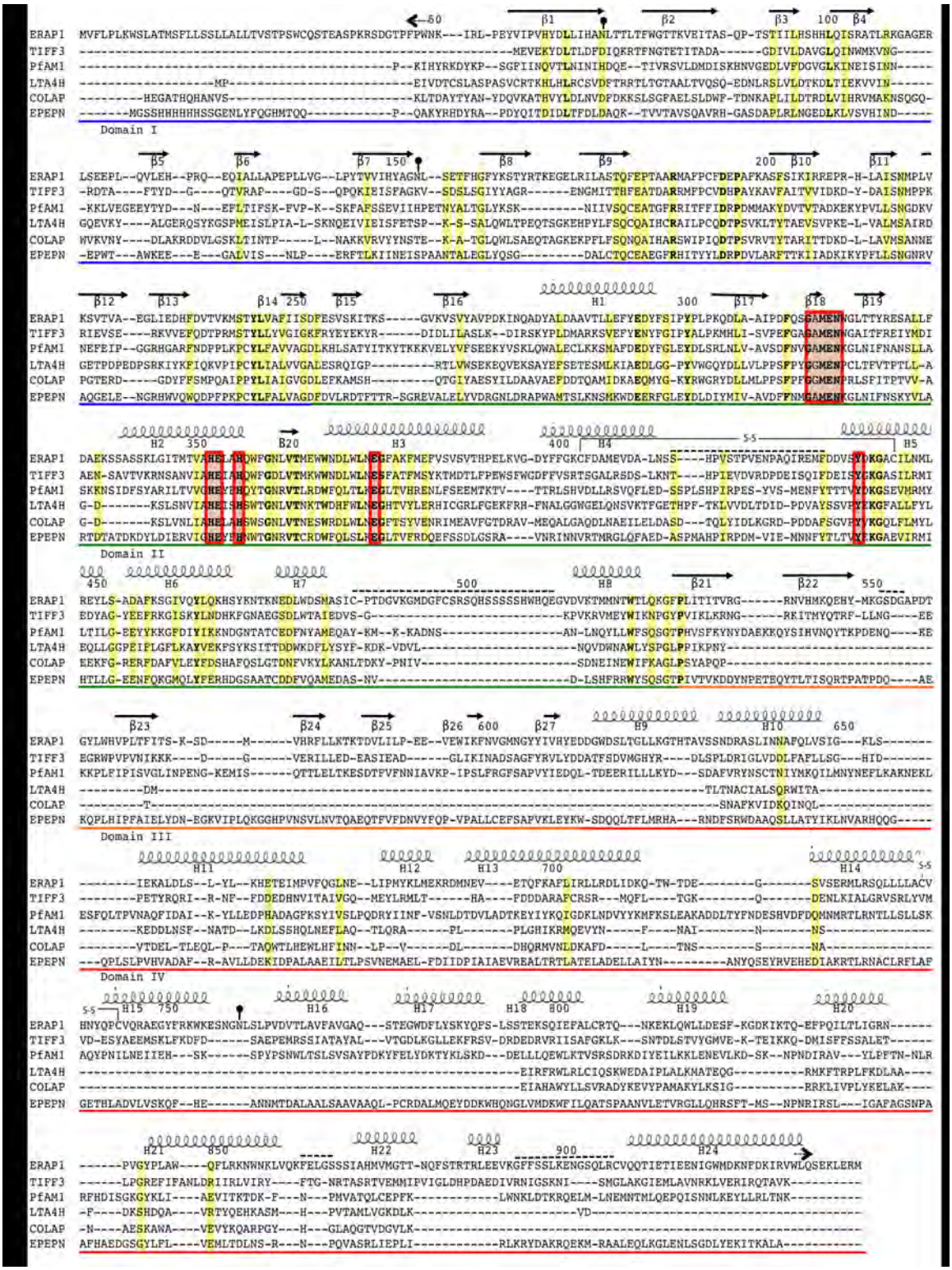

ERAP1
TIFF3
PfAM1
LTA4H
COLAP
EPEPN
Domain I
Domain II
Domain III
Domain IV

**Figure 2.1. Structure Based Alignment M1 Aminopeptidase Family Members.** Numbering corresponds to ERAP1, colored lines indicate domains in ERAP1; arrows and coils represent strands and helices, respectively, and dashes represent loops not observed in crystal structure. Highlighted residues are homologous among the aminopeptidases shown and boxed residues are highly conserved M1 aminopeptidase motifs HExxEx$_{18}$E, GAMEN and Tyr438. Disulfide bonds are shown by connected bars and observed glycosylation sites are indicated by balloon icons. Figures were generated using 3D-COFFEE (Notredame et al., 2000).

The crystal structure of full-length recombinant human ERAP1 bound to the aminopeptidase inhibitor bestatin was determined by X-ray crystallography using cryogenic diffraction data extending to 3.0 Å (Table 2.1). The structure was solved by molecular replacement using LTA4H and TIFF3 as search models, with initial phasing using non-crystallographic symmetry relationships among the three copies of ERAP1 observed in the crystallographic asymmetric unit (Figure 2.2). Electron density was sufficient for model building of C-terminal domain *de novo*. Figure 2.3 shows an example of electron density from data collected. The interactions between the three molecules might provide a model for observed ERAP1 interaction with ERAP2 in pull down experiments (Saveanu et al., 2005).

**Table 2.1. Data Collection and Refinement Statistics for ERAP1**

| | ERAP1-bestatin |
|---|---|
| **Data collection** | |
| Space group | $P2_1$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 70.9, 234.2, 95.6 |
| $\alpha, \beta, \gamma$ (°) | 90, 103.6, 90 |
| Resolution (Å) | 38-2.95 (3.1-2.95)* |
| $R_{sym}$ (%) | 17.8 (76.5) |
| $I / \sigma I$ | 7.5 (1.1) |
| Completeness (%) | 99.8 (98.8) |
| Redundancy | 3.6 (3.4) |
| | |
| **Refinement** | |
| Resolution (Å) | 38-2.95 |
| No. reflections | 63,522 |
| $R_{work}$ / $R_{free}$ (%) | 20 / 26 |
| No. atoms (per trimer) | |
| Protein | 20,556 |
| Ligand/ Zn ion | 66 / 3 |
| *B*-factors (Å$^2$) | |
| main chain/side chain/Zn | 72.6 / 63.3 / 68 |
| Mol A / Mol B / Mol C | 62.8 / 63.5 / 63.4 |
| NCS RMSD (Å) / B values (Å$^2$) | |
| D1 (1-254) | 0.19 / 58.2 |
| D2 (255-527) | 0.14 / 52.3 |
| D3 (528-613) | 0.15 / 53.2 |
| D4-n (614-750) | 0.32 / 68.3 |
| D4-c (751-941) | 0.19 / 105.7 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.005 |
| Bond angles (°) | 0.85 |

*Values in parentheses are for highest-resolution shell.

.

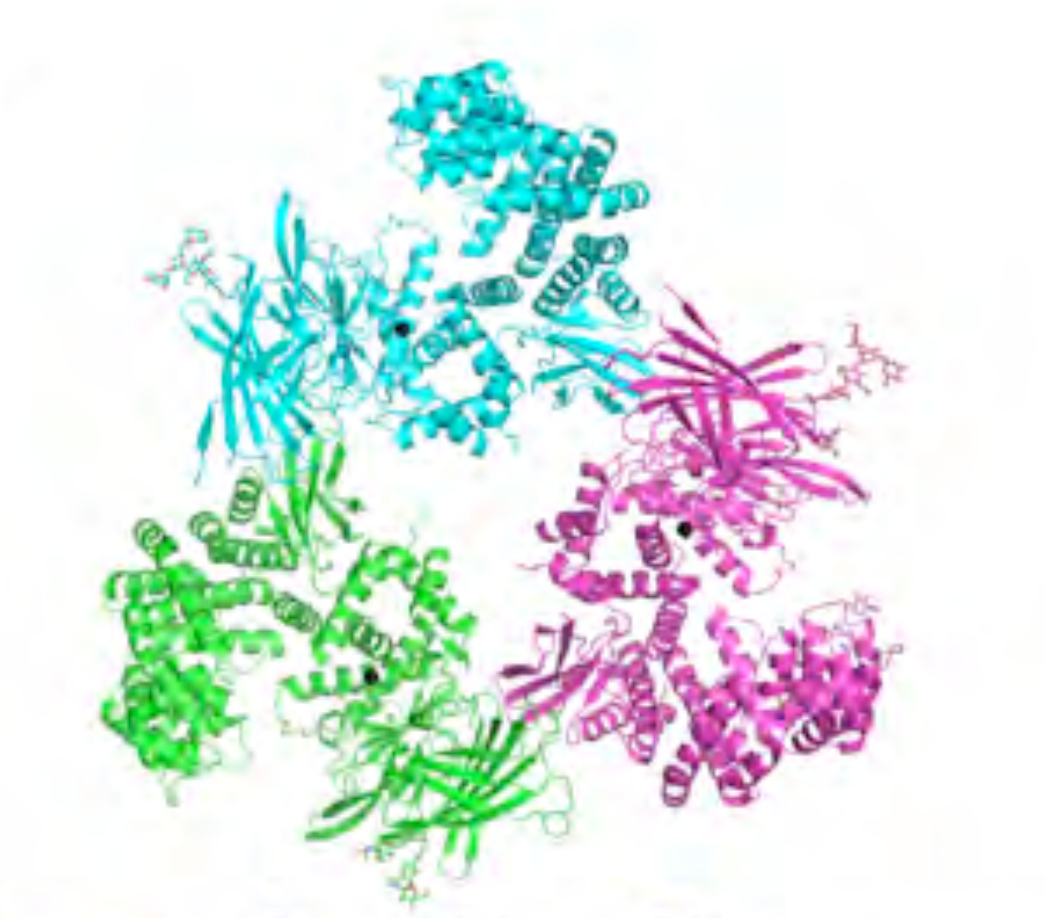
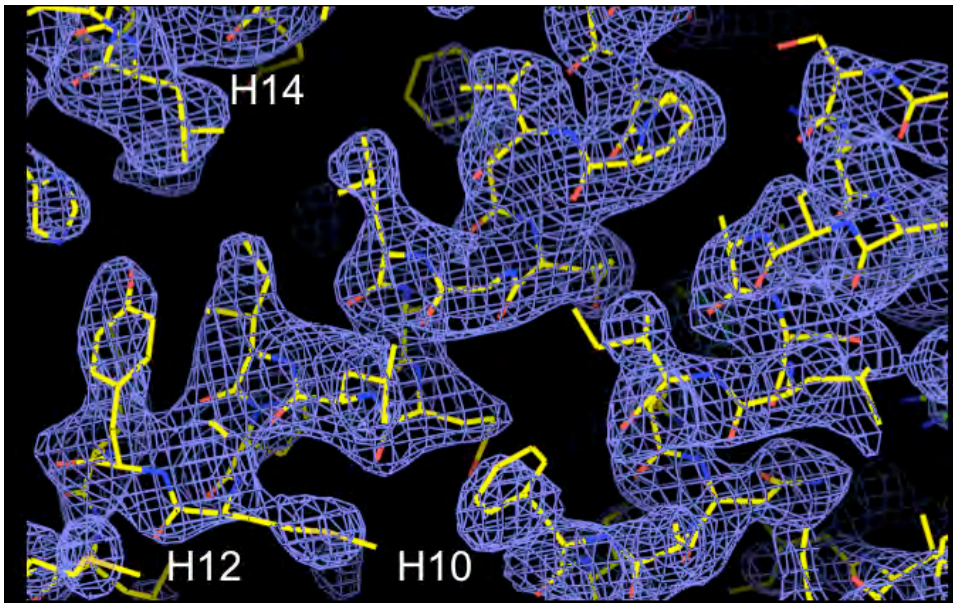


**Figure 2.2. ERAP1 Trimer in the Crystallographic Asymmetric Unit.**

Ribbon representation of an asymmetric unit shows the trimeric arrangement of the crystallographically distinct molecules. ERAP1 main chain is represented in ribbon, the zinc atoms are shown as black spheres, and N-linked sugars and bestatin are shown in stick representation.


H14
H12
H10

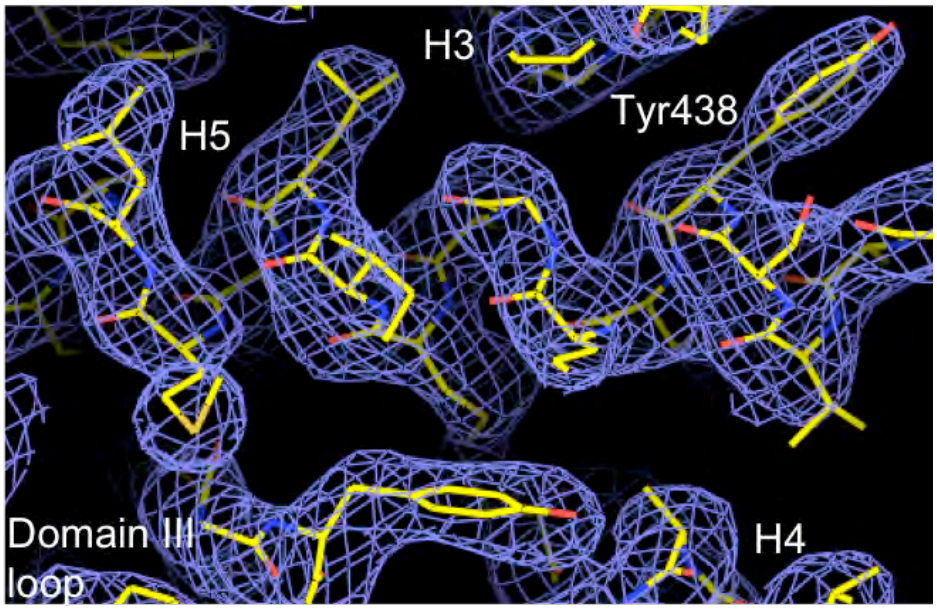

H3
H5
Tyr438
Domain III
loop
H4

**Figure 2.3. Electron Density for ERAP1.** Top. Omit map shows portions of helices 10, 12, and 14 in domain IV. All atoms shown (residues 590:780) were omitted from model calculations. Bottom. Omit map showing Tyr438 on helix 5, helix 3, helix 4 in domain II and a loop from domain III. All atoms shown were omitted from map calculation. Blue mesh represents |2Fo-Fc| electron density, lines represent ERAP1 model.

The crystal structure of ERAP1 reveals four protein domains, with a large cavity between domains II and IV (Figure 2.4). As with other M1-family members, domain II (green in Figure 2.4) is the catalytic domain that carries the zinc atom, GAMEN, and HExxH$x_{18}$E motifs on a thermolysin-like $\alpha\beta$ fold. Domain I (blue) is an all-$\beta$ sandwich domain that docks on top of the thermolysin domain, capping off the active site and providing binding sites for the amino terminus of a substrate peptide. Domain III (orange) is a small $\beta$-sandwich domain between domains II and IV. Among M1-family aminopeptidases, domain IV (red) is the most variable, ranging from a tightly closed spiral of ten helices in LTA4H to seventeen helices in ePepN. In ERAP1, domain IV is composed of sixteen variously sized helices arranged as eight antiparallel Armadillo/HEAT-type helix-turn-helix repeats (Andrade et al., 2001) assembled side-to-side to form a concave surface facing toward the active site (Figure 2.5). Overall, the helices of the C-terminal helical domain are arranged in a large cone or bowl adjacent to the active site, lined by the lateral faces of the even-numbered helices.

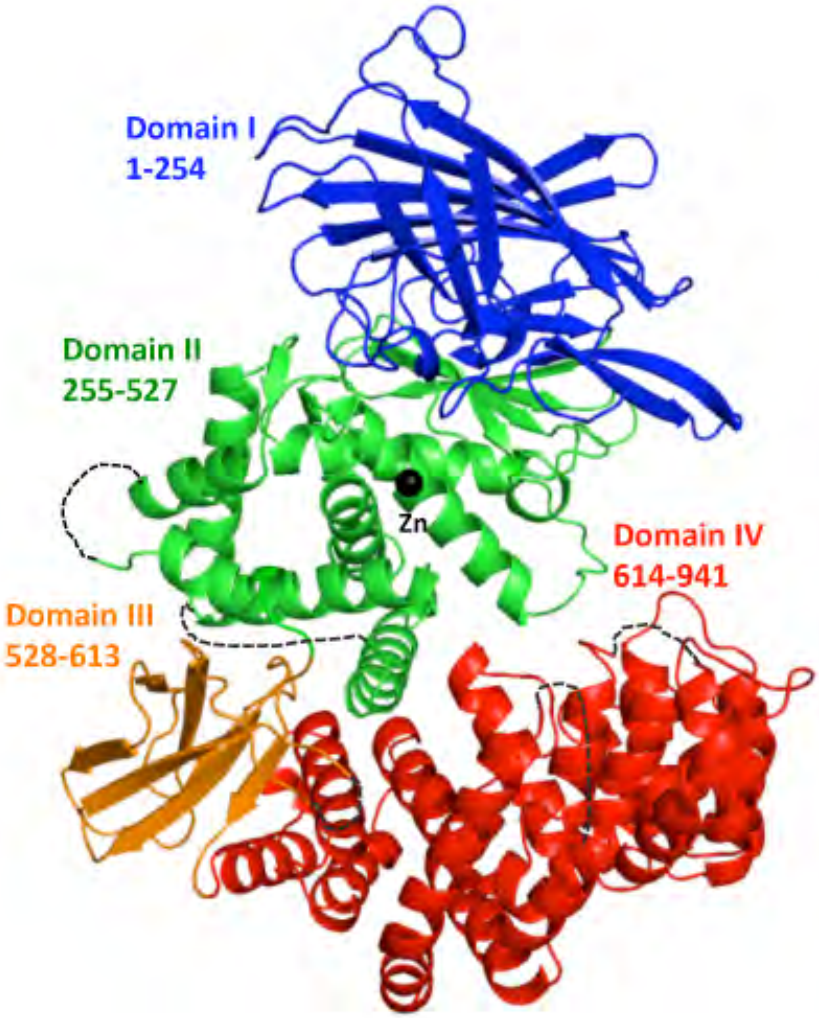

Domain I
1-254
Domain II
255-527
Zn
Domain IV
614-941
Domain III
528-613

**Figure 2.4. Structure of ERAP1**. Overall shape of ERAP1 represented as a ribbon diagram and colored according to domains: blue is domain I, green is domain III, orange is the domain III, and red is domain IV. Dotted lines represent disordered loops. Zinc is represented as a black sphere.

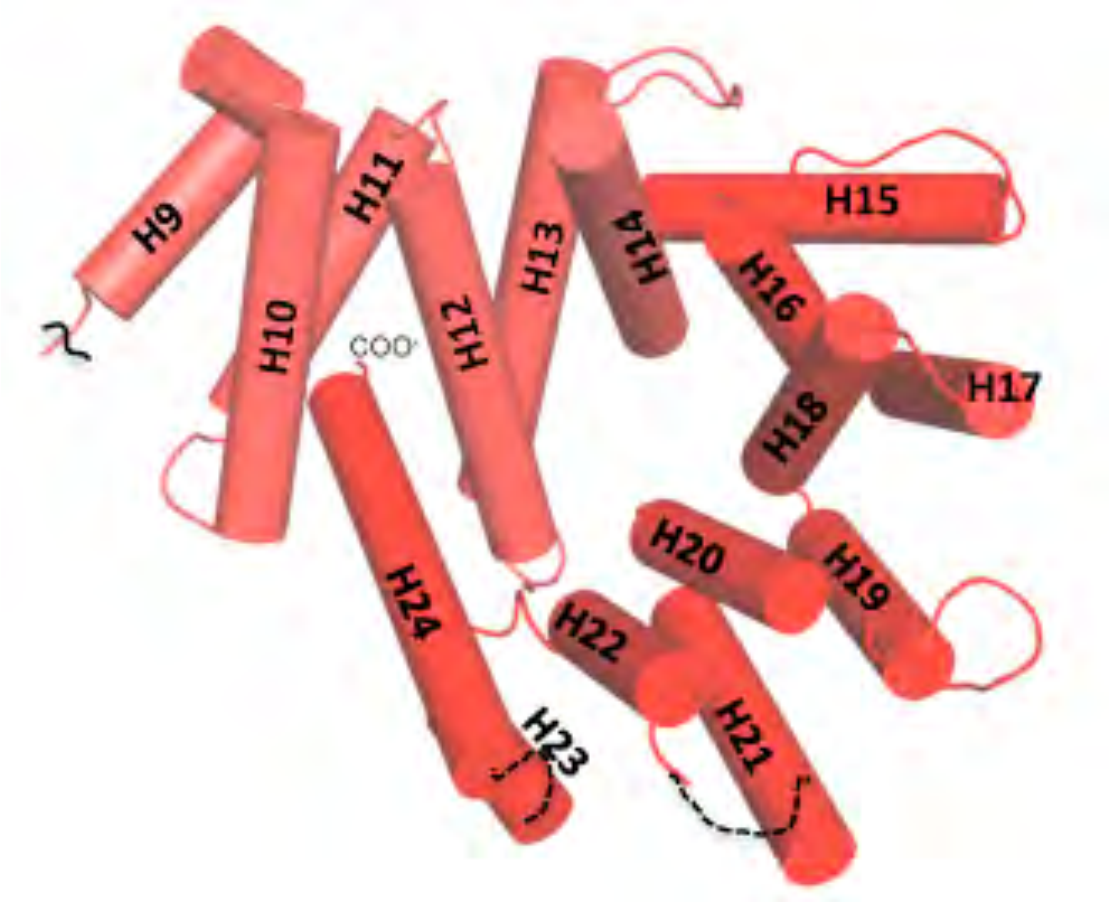

H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
H21
H22
H23
H24
COO⁻

**Figure 2.5. Schematic Diagram of the C-terminal Domain of ERAP1.** Helices are represented as cylinders showing the ARM/HEAT motif spiral forming a large cavity lined by even numbered helices.

A striking feature of the structure is the formation of a deep cavity between domains II and IV. This cavity is the largest seen to date for an aminopeptidase and potentially can provide easy access to the catalytic site for even the longest of the enzyme's substrates. Figure 2.6 shows views of the cavity, in which most of domain I has been removed to allow better visualization of the extent of the cavity. Dimensions of the cavity are shown. This cavity, which extends up to 36 Å from the active site to the interior surface of domain IV, is likely to represent a binding site for long peptide substrates.

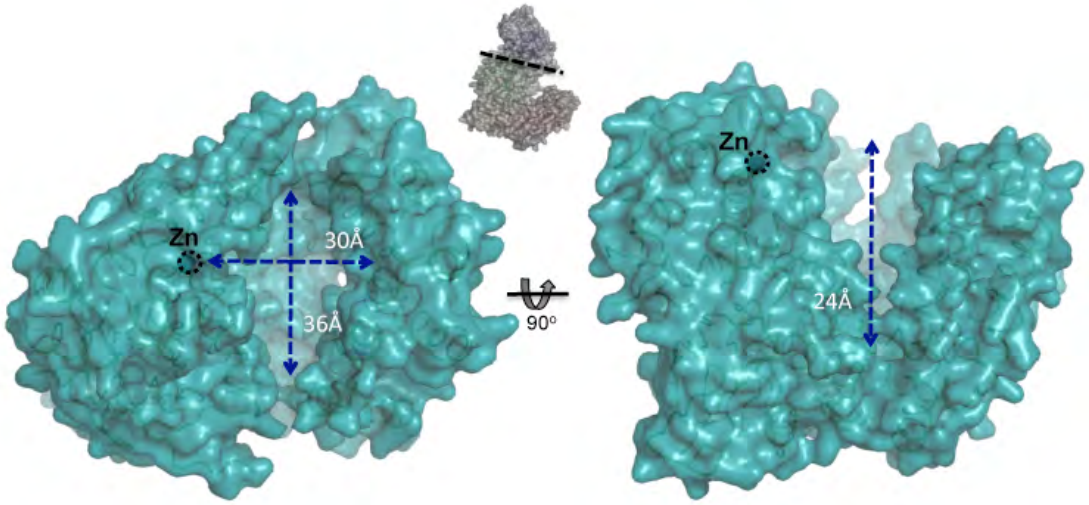

Zn
30Å
36Å
90°
Zn
24Å

**Figure 2.6. Large Cavity Formed by the Catalytic Domain II and C-terminal Domain IV.** The cavity is shown in a surface representation. Inset shows how the N-terminal domain was removed above the dashed black line to provide a clear view of the cavity. Blue dashed lines represent estimated distances across the cavity.

**Catalytic Site**

The ERAP1 active site is formed at the junction of five secondary structure elements: helix 2 and the adjacent antiparallel helix 3 that carry the HExxH and E residues of the HExxH$x_{18}$E motif, the GAMEN loop, the domain 1 loop, and helix 5 adjacent to the domain II-domain IV interface (Figure 2.7). Bestatin, an $\alpha$–hydroxy $\beta$-amino dipeptide analog that functions as a broad spectrum aminopeptidase inhibitor, binds to the active site of ERAP1 via a bidentate interaction between its 2-hydroxy and amido oxygen atoms and the ERAP1 catalytic zinc atom. The catalytic zinc atom is coordinated by the side chains of His353 and His357 of the HExxH helix and Glu376 of the adjacent helix, with Glu354 next to bestatin's $\alpha$-hydroxyl group (Figure 2.8). The bestatin amino terminus is bound by Glu183 of the domain 1 loop and Glu320 of the GAMEN motif. These interactions are likely to simulate the complex of ERAP1, substrate amide, and water molecule positioned to attack the peptide's scissile bond. In addition, the main chain of GAMEN residues Gly317 and Ala318 interact with bestatin's carboxylate and amide bond, respectively. The Met319 side chain forms part of the S1 pocket, and the Glu320 side chain carboxylate interacts with bestatin's free amino terminus (Figure 2.8). Asn321, the last of the GAMEN residues, forms hydrogen bonds with the neighboring strand that appear to stabilize the local loop conformation (not shown). These interactions are likely to simulate the interaction of ERAP1 with the first few residues of a bound peptide substrate. The last component of the active site is Tyr438 on the interfacial helix.

This tyrosine is conserved in all M1 family aminopeptidases (Rawlings et al., 2009), and has been shown to play an important catalytic role in stabilizing the tetrahedral intermediate formed by water addition to the substrate scissile peptide bond (Thompson et al., 2006). In the ERAP1-bestatin complex, Tyr438 is positioned near the active site but is oriented such that its hydroxyl group is more than 8 Å away from bestatin's amide bond, where it cannot participate in catalysis (Figure 2.8).

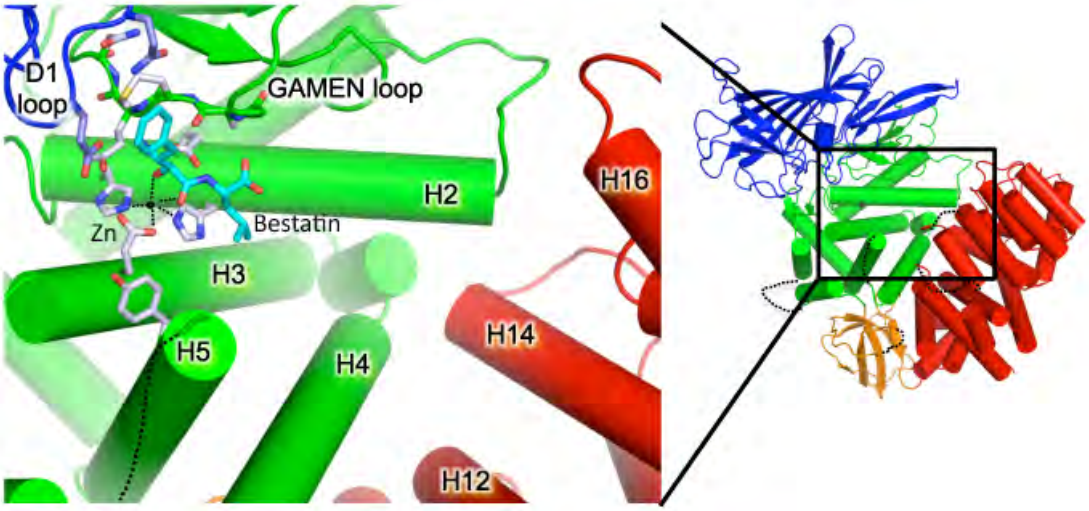

D1
loop
GAMEN loop
H2
H16
Zn
Bestatin
H3
H5
H4
H14
H12

**Figure 2.7. The Active Site of ERAP1.** Structural elements forming the ERAP1 active site. Bestatin is shown with cyan bonds, zinc as a black sphere, and nearby ERAP1 loops, and helices shown as cylinders, are colored as in Figure 2.2. ERAP1 residues in the active site are shown with white bonds. Inset at right shows the entire protein for orientation; in the view shown in panel where helices 19 and 21 that occlude the active site were removed for clarity.

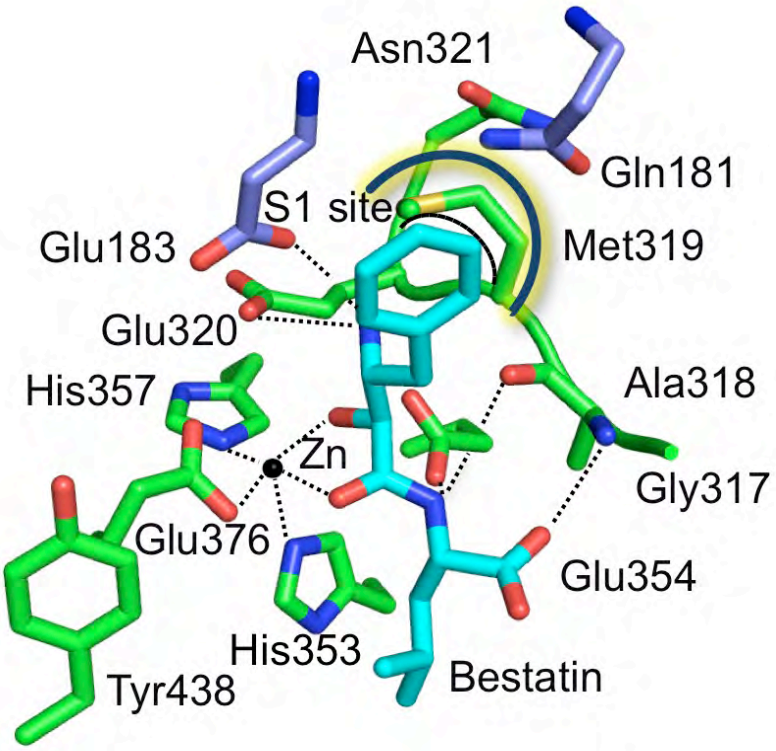

Asn321
Gln181
S1 site
Glu183
Met319
Glu320
His357
Ala318
Zn
Gly317
Glu376
Glu354
His353
Bestatin
Tyr438

**Figure 2.8. The Inhibitor Inside ERAP1 Active Site.** ERAP1 interactions with bound aminopeptidase inhibitor, bestatin. Side chains of conserved catalytic residues of ERAP1 are shown in with bonds colored by domain. Residues from domain I are represented in blue and residues from domain II are in green. Yellow arc indicates S1 site.

Structural alignment of the active sites of M1 aminopeptidases reveals constellations of conserved catalytic residues that are essentially identical to those of ERAP1, with the exception of Tyr438 (Figure 2.9). In all the other M1 structures but one (TIFF3), the analog of Tyr438 is oriented toward the active site, in a position to stabilize the tetrahedral intermediate formed by attack of water at the peptide scissile bond (Addlagatta et al., 2006; Thompson et al., 2006). However, in ERAP1, Tyr438 is found oriented away from the substrate binding site, as a result of reorientation of the interfacial helix and adoption of a different rotameric form than found in the other M1-aminopeptidase structures (Figure 2.9).

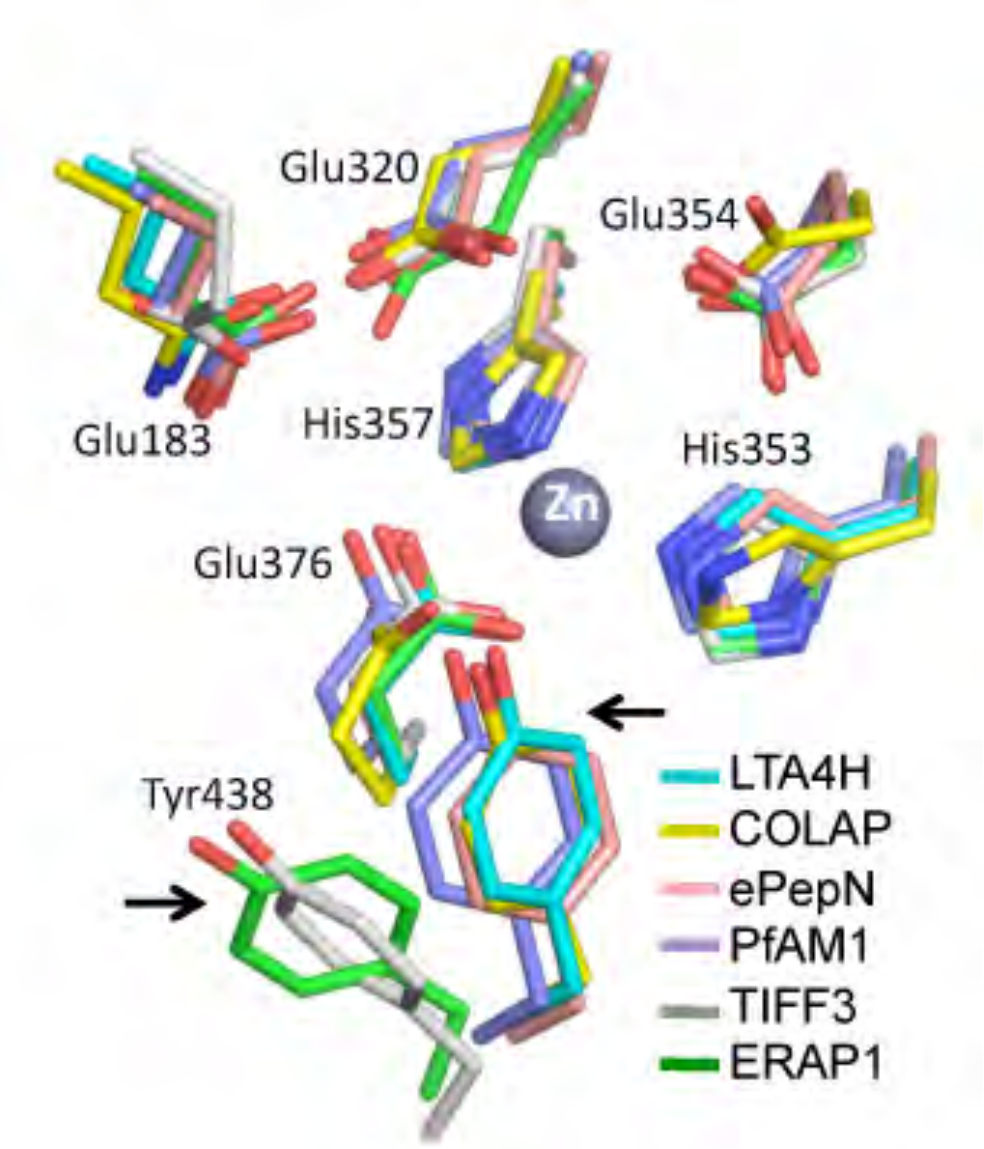

Glu320
Glu354
Glu183
His357
His353
Zn
Glu376
Tyr438
LTA4H
COLAP
ePepN
PfAM1
TIFF3
ERAP1

**Figure 2.9. Conservation of Catalytic Residues Among M1 Family Aminopeptidase Structures.** Conservation shown in overlay of the active sites of human leukotriene hydrolase, LTA4H (cyan, PDB 3FUH), *C. psychrerytheaea* cold active aminopeptidase, COLAP (yellow, PDB 3CIA), *E. coli* aminopeptidase N, ePepN (pink, PDB 2HPT), *P. falciparum* aminopeptidase M1, PfAM1 (purple, PDB 3EBG), *T. acidophilium* tricorn interacting factor F3, TIFF3 (grey, PDB 1Z5H), and human ER aminopeptidase 1, ERAP1 (green). ERAP1 and TIFF3 have an altered orientation of Tyr438 relative to the other aminopeptidases (arrows). Residue numbers correspond to ERAP1.

**Substrate Specificity**

ERAP1 is unusual among aminopeptidases in its ability to accept large peptide substrates up to ~16 residues. Optimal activity is observed for peptides of ~10 residues, while most peptides of 8 or fewer residues are not processed as efficiently (Chang et al., 2005). ERAP1 has a large cavity that provides a potential binding site for such large peptides, formed at the interface between the concave surface of domain II carrying the active site and the concave surface of the domain IV helical bowl (Figures 2.6). A closer view of the cavity in the vicinity of the active site is shown in Figure 2.10. The cavity is relatively narrow near the active site (upper left in Figure 2.10) and widens as it enters domain IV. The wider part of the cavity can accommodate the C-terminal portion of polypeptide substrates bound with their N-terminal regions in position to be processed at the active site. Two possible paths for long peptides are indicated in Figure 2.10. The heads of the arrows are located ~27 Å from the zinc atom, corresponding approximately to the length of an extended 7-8 residue peptide. Due to the sequestered nature of the active site in domain II, and the arrangements of domains I and IV around the mouth of the cavity, it is unlikely that ERAP1 can complete the processing of an epitope peptide precursor while it is bound to MHC class I. Docking experiments show that the closest possible approach of the amino terminus of a peptide bound to a MHC class I protein to the zinc center in ERAP1 is 20 Å, corresponding to at least 6 residues of extended polypeptide.

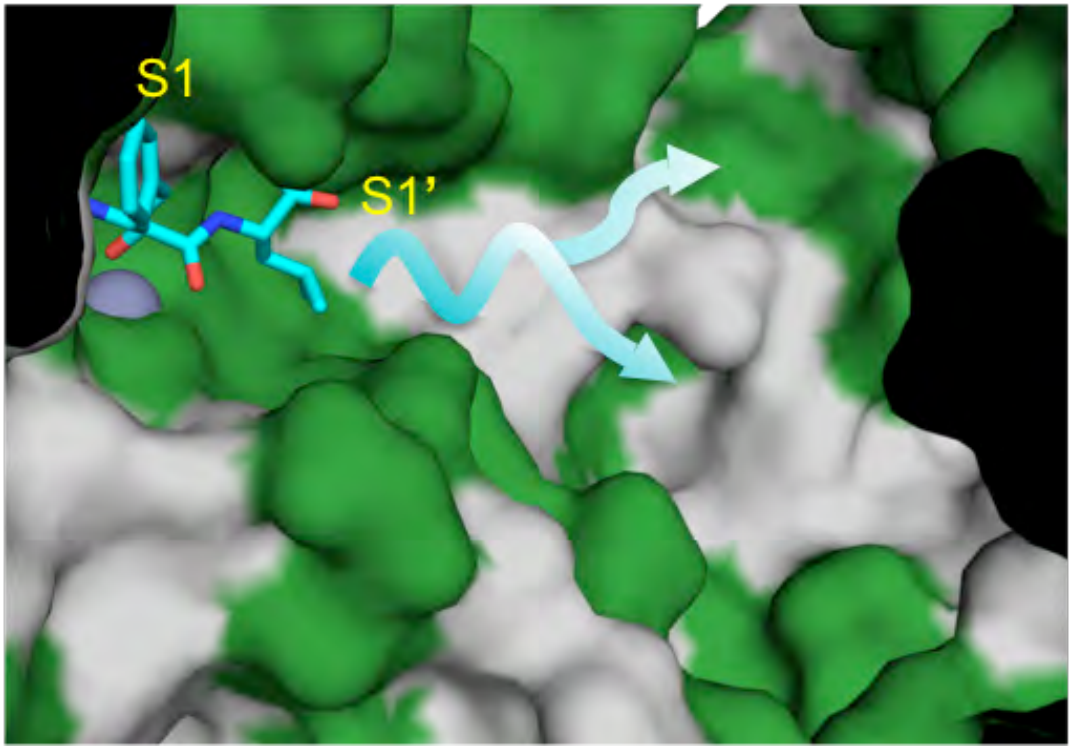

S1
S1'

**Figure 2.10. Surface Polarity of ERAP1 in the Substrate Binding Site.** The substrate binding cavity extends from the active site into domain IV. The surface of ERAP1 near the active site is colored by polarity. The polar residues are green and hydrophobic residues are white. The catalytic zinc is shown as a dark gray sphere, and bestatin has cyan bonds. Expected positions of peptide side chain binding pockets S1 and S1' are indicated. Arrows represent possible paths of peptides extending from the active site into the cavity.

The orientation of bestatin in the aminopeptidase active site can help to identify subsites in the active site region that accommodate peptide side chains (Fournie-Zaluski et al., 2009; Tholander et al., 2008; Thunnissen et al., 2001). Following peptidase convention, we call the site responsible for accommodating the peptide side chain N-terminal to the cleavage site S1, and subsequent positions called S1', S2', etc. Bestatin's phenyl group binds into the aminopeptidase S1 site, with its isobutyl group oriented toward the S1' site (Figure 2.8). In ERAP1, the S1 site is a relatively shallow hydrophobic pocket lined by the side chains of Ser316 and Met319 of the GAMEN loop, and Gln181 and Glu183 of the D1 loop (Figure 2.8). Modeling indicates that a leucine, methionine, or isoleucine would fit closely into the S1 pocket as the first residue of a bound peptide. Larger residues such as tryptophan and arginine can make similar contacts, but only by adopting rotamers that orient the side chain out of the pocket. Smaller residues can be accommodated but would leave voids within the site. This pattern matches the N-terminal specificity observed in peptide hydrolysis and antigen presentation studies where peptides with amino-terminal leucine, isoleucine, and methionine residues were processed most efficiently, while those with amino-terminal tryptophan, arginine, and cysteine were processed less efficiently (Chang et al., 2005; Evnouchidou et al., 2008; Hearn et al., 2009; Saveanu et al., 2005; Serwold et al., 2001). The broader reactivity towards peptides as compared to single residue substrates likely reflects the contribution of binding sites for peptide side chains beyond S1', which can

compensate for non-optimal interactions in the sterically rather permissive S1' pocket.

ERAP2, another ER aminopeptidase also involved in ER antigen processing (Saveanu et al., 2005; Tsujimoto and Hattori, 2005), has a different specificity than ERAP1, with greatly increased preference for fluorogenic amino acid substrates (Chang et al., 2005; Tanioka et al., 2003) and peptides (Saveanu et al., 2005) carrying N-terminal Arg and Lys groups. One difference between ERAP1 and ERAP2 in this region is at position 181, which is Gln in ERAP1 but Asp in ERAP2. Residue 181 forms the top of the S1 pocket (Figure 2.8), and the Gln to Asp substitution would have the effect of lengthening the pocket and making it more acidic. A site specific mutant of ERAP1 carrying a Gln181 to Asp substitution alters its specificity from hydrophobic towards basic amino acids (Goto et al., 2008).

ERAP1 has been shown to be unable to trim peptides with a proline residue as a second amino acid (Hearn et al., 2009; Serwold et al., 2001). Modeling of proline into a mock peptide substrate indicates that the side chain cannot be accommodated into the S1' site without disrupting the GAMEN loop and arrangement of catalytic residues above the nitrogen of the scissile peptide bond.

ERAP1 exhibits sequence specificity for other regions of the substrate beyond its N-terminus, as shown by alanine scanning and library screening studies (Evnouchidou et al., 2008). Without a structure of ERAP1 bound to a

peptide substrate, identification of the sites responsible for this sequence specificity is speculative, however several aspects of the internal cavity are consistent with it being a binding site for long peptide substrates. Figure 2.11 shows a top view of ERAP1, represented as an electrostatic potential surface. Dashed lines show approximate distances of 20 Å and 30 Å from the zinc atom in the active site. The overall negative charge of the cavity in this region is consistent with internal specificity for positively charged residues observed at several sites in a 9-mer peptide substrate (Evnouchidou et al., 2008). Shallow pockets within the overall substrate binding cavity, visible as depressions along the sides of the groove shown in Figure 2.10, potentially correspond to pockets for peptide side chains, and could be responsible for ERAP1's observed preferences for hydrophobic residues at the C-terminal position of peptide 9-mers (Evnouchidou et al., 2008) and 10-mers (Chang et al., 2005). Other binding sites for peptide side chains may be present on the interior faces of helices lining the cavity, as observed for other proteins carrying armadillo/HEAT repeat motifs (Monecke et al., 2009).

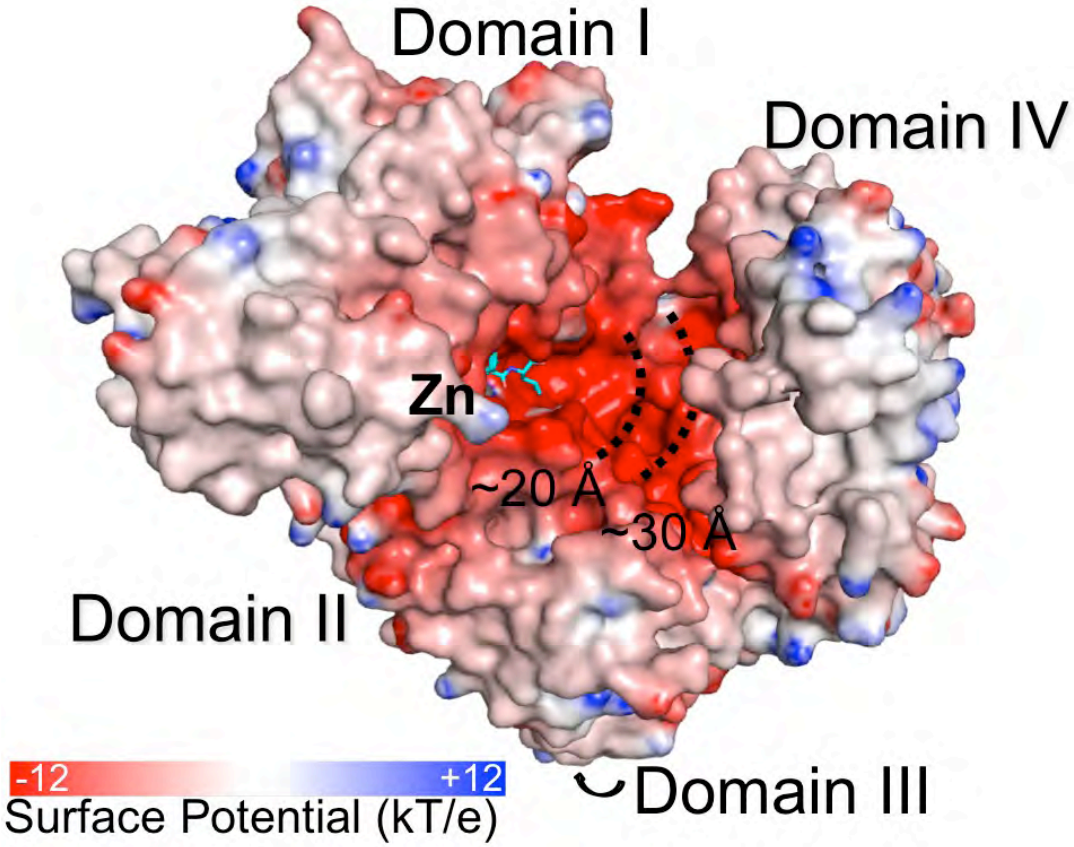

Domain I
Domain IV
Zn
~20 Å
~30 Å
Domain II
Domain III
-12
+12
Surface Potential (kT/e)

**Figure 2.11. Electrostatic Surface Map of ERAP1.** Surface colored from red to blue represents negatively charged to positively charged regions. Dotted lines represent indicated distances from zinc ion (grey sphere). Bestatin colored with cyan bonds. Figure generated using DelPhi (Nicholls and Honig, 1991) and PyMol (DeLano, 2008).

**Domain Motions**

ERAP1 has a significantly more open structure than other M1 family members because of the orientation of its C-terminal helical domain relative to the remainder of the protein (Figure 2.12). In LTA4H and ColAP, the small C-terminal helical domain packs closely against the catalytic domain, with no intervening small β-sandwich domain and no extensive cavity formed. For ePepN, and PfAM1, the helical C-terminal domain is relatively large and similar in overall dimensions to that of ERAP1. However, it packs closely against the active site, forming a closed substrate binding cavity. TIFF3 has a more open domain arrangement somewhat similar to that of ERAP1, but with the C-terminal domain packing closer to the catalytic domains, forming a smaller cavity. In the structure of ERAP1, domain IV extends away from the active site, forming a cavity substantially larger than in the other structures, and more accessible to solvent.

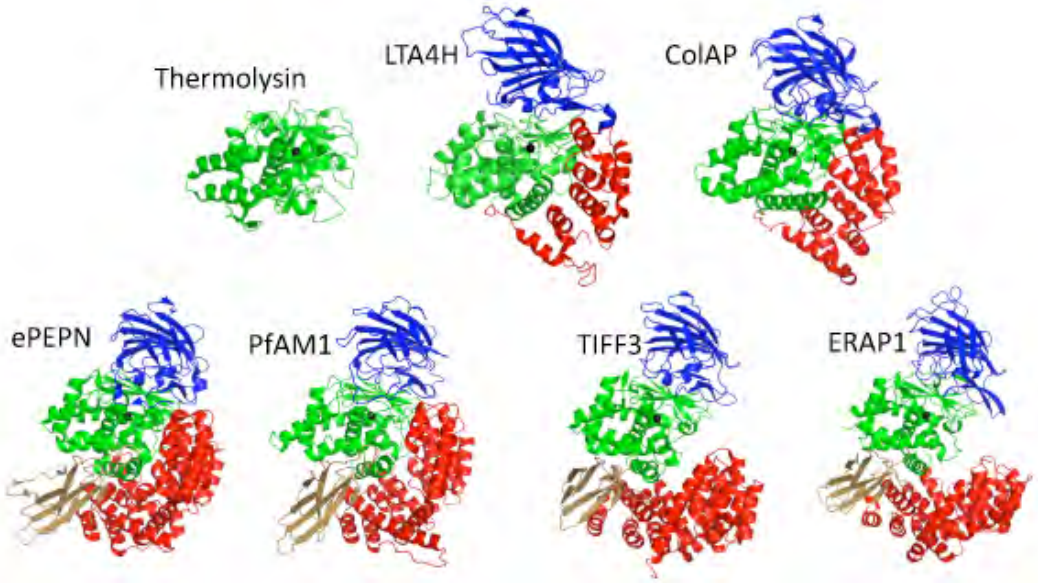

Thermolysin
LTA4H
ColAP
ePEPN
PfAM1
TIFF3
ERAP1

**Figure 2.12. Overall Fold of M1 family of Aminopeptidases.** Structures of M1 aminopeptidases family members and thermolysin are colored by domain. Aminopeptidases are shown in ribbon diagram and colored according to domains: blue is domain I, green is domain II, orange is the domain III, and red is domain IV. Dotted lines represent disordered loops. Zinc is represented as a black sphere.

Among the three copies of ERAP1 in the crystallographic asymmetric unit, the interdomain orientations vary, with the differences mostly affecting the orientation of domain IV relative to the other domains. Figure 2.13 (left) shows these three copies after superposition of domain III, revealing a small motion tending towards a more “closed” conformation. This motion differentially affects the C-terminal domain, with the tops of helices 15-24 moving closer to helices 9-14 and the other protein domains (compare with Figure 2.2). This motion tends towards closing of the cavity. A normal mode analysis (Suhre and Sanejouand, 2004) (Figure 2.13 right), indicates that further closing motions along the same trajectory are accessible, which might allow ERAP1 access to a more closed orientation under certain circumstances.

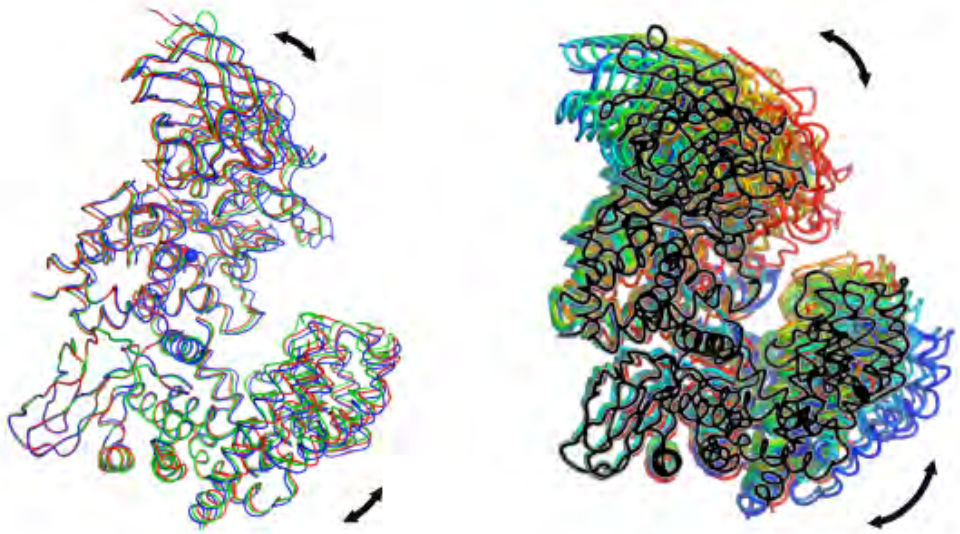


**Figure 2.13. Domain Movements.** Hinge movement suggested by different inter-domain orientations among the three molecules observed in the ERAP1 crystallographic asymmetric unit, shown here as C-alpha traces aligned at domain III (left). The right panel shows samples of ERAP1 motions as indicated by normal mode analysis, shown as C-alpha traces colored blue to red representing extremes of conformational alteration. The ERAP1 structure is overlay in black.

**Length-dependent trimming**

The length dependent trimming experiments were conducted by the laboratories of Alfred Goldberg and Efstratios Stratikos. ERAP1 processes peptides longer than 8-9 residues much more efficiently than shorter peptides (Chang et al., 2005). This activity is unusual among aminopeptidases, which typically are most active in processing shorter peptides, but is consistent with ERAP1's role in antigen processing. One potential mechanism for this effect would be regulation of the enzyme's activity depending on the length of bound peptide substrate. To investigate this possibility, leucine-7-amido-4-methylcoumarin (L-AMC) was used to monitor ERAP1's hydrolytic activity in the presence of peptides of varying lengths (Figure 2.14). The peptide used was a series of peptides based on the antigenic peptide SIINFEKL, which previously has been used to characterize ERAP1 activity *in vitro* (Chang et al., 2005; York et al., 2002) and *in vivo* (Hearn et al., 2009). ERAP1's L-AMC aminopeptidase activity was greatly increased in the presence of increasing SIINFEKL length variants up to eight residues (Figure 2.14 A). These peptides themselves were not long enough to be processed efficiently by ERAP1 (Figure 2.14 A, lower panel). Activation of L-AMC hydrolysis decreased with increasing peptide length for SIINFEKL variants 8 residues or longer, which however were processed efficiently by ERAP1. This phenomenon was not unique to SIINFEKL, since a similar activation occurred with peptides carrying a poly-glycine motif of increasing length (Figure 2.14 B). Thus, they

showed that shorter peptides can act as allosteric activators of ERAP1, converting the enzyme to a more active conformation upon binding.

To investigate this phenomenon without potential complications from hydrolysis of the activating peptide, they used the non-hydrolyzable peptide L(N-Me)VAFKRAF, $L^{Me}$, as an allosteric activator (Figure 2.14 C). The sequence of this peptide was optimized for interaction with ERAP1 based on library screening (Evnouchidou et al., 2008). In this peptide, the amide group of the first peptide bond carries a methyl group that should interfere with the catalytic cleavage of this peptide bond. Accordingly, the $L^{Me}$ peptide is not trimmed by ERAP1 under experimental conditions, although the unmethylated version of the peptide is an excellent ERAP1 substrate and is readily processed (Figure 2.14 C, compare left and right panels). Increasing concentrations of $L^{Me}$ peptide activated ERAP1's hydrolysis of L-AMC substrate (Figure 2.14 D). When the effect of $L^{Me}$ on ERAP1 enzymatic activity was evaluated using a full-length fluorogenic 10-mer substrate (Evnouchidou et al., 2009) rather than L-AMC, a concentration dependent decrease in ERAP1 activity was observed (Figure 2.14 D). The ability of the $L^{Me}$ peptide to activate the hydrolysis of L-AMC but to competitively inhibit hydrolysis of a fluorogenic peptide substrate implies a simple mechanism in which the $L^{Me}$ peptide binds to a regulatory site that partially overlaps with the substrate-binding site, allowing concurrent binding for a small substrate like L-AMC while blocking binding for a larger one.

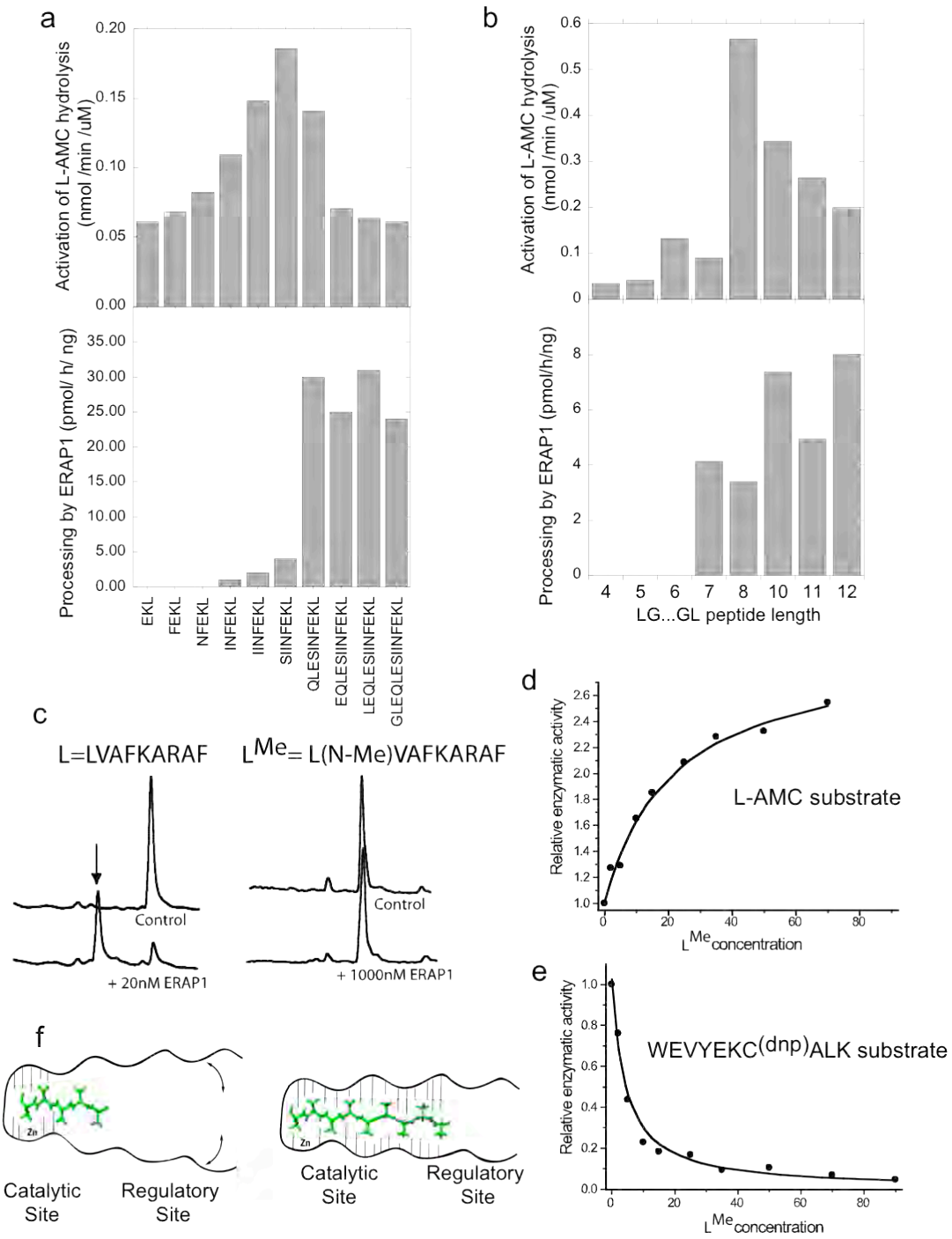

a
Activation of L-AMC hydrolysis (nmol /min /uM)
0.20
0.15
0.10
0.05
0.00
Processing by ERAP1 (pmol/ h/ ng)
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
EKL
FEKL
NFEKL
INFEKL
IINFEKL
SIINFEKL
QLESINFEKL
EQLESIINFEKL
LEQLESIINFEKL
GLEQLESIINFEKL
b
Activation of L-AMC hydrolysis (nmol /min /uM)
0.6
0.5
0.4
0.3
0.2
0.1
0
Processing by ERAP1 (pmol/h/ng)
8
6
4
2
0
4 5 6 7 8 10 11 12
LG...GL peptide length
c
L=LVAFKARAF
L^Me= L(N-Me)VAFKARAF
Control
+ 20nM ERAP1
Control
+ 1000nM ERAP1
d
Relative enzymatic activity
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0 20 40 60 80
L^Me concentration
L-AMC substrate
e
Relative enzymatic activity
1.0
0.8
0.6
0.4
0.2
0.0
0 20 40 60 80
L^Me concentration
WEVYEKC(dnp)ALK substrate
f
Zn
Catalytic Site
Regulatory Site
Zn
Catalytic Site
Regulatory Site

**Figure 2.14. Allosteric Activation of ERAP1.**

A. ERAP1 activation of L-AMC hydrolysis by SIINFEKL variants (top) and degradation of a series of truncated and extended SIINFEKL variants (bottom)

B. ERAP1 activation of L-AMC hydrolysis by LG(n)L peptides (top) and degradation of a series of LG(n)L peptides.

C. N-methylation of the first amide bond in peptide L, an optimized ERAP1 substrate, blocks hydrolysis. Reverse-phase chromatograms of peptides incubated in the presence or absence of ERAP1 shown. Arrow indicates degradation product. $L^{Me}$ is resistant to degradation.

D. ERAP1 hydrolysis of L-AMC is activated by $L^{Me}$, an N-terminally blocked peptide with a sequence optimized for interaction with ERAP1.

E. ERAP1 hydrolysis of a fluorescent peptide substrate is inhibited by $L^{Me}$, the same peptide that activates L-AMC hydrolysis.

F. Model for ERAP1 length dependent cleavage activity. Left. A short peptide (pentamer shown) cannot reach from the catalytic site to the regulatory site, and ERAP1 remains in the lower-activity open conformation. Right. A long peptide (nonamer shown) can reach the regulatory site, inducing conversion to closed conformation with increased peptide hydrolysis activity.

The existence of distinct yet nearby catalytic and regulatory sites within ERAP1, as suggested by the structural and biochemical evidence, allows us to postulate a simple model that can account for ERAP1's unique properties as a length-dependent aminopeptidase (Figure 2.14 E). According to this model, peptides longer than ~8-9 residues can concurrently occupy the active and regulatory sites in ERAP1, leading to the enzyme activation and efficient trimming. Shorter substrates can either occupy the active site or the regulatory site, but not both at the same time. As a result, shorter substrates will be processed at a slower rate. In these model experiments, the very short substrate L-AMC could occupy the active site at the same time that a peptide (<10 residues) occupies the regulatory site, leading to allosteric activation of substrate hydrolysis. Activation of L-AMC by peptides falls off for longer peptides, presumably due to limitations in size within the enzyme's internal cavity. This model explains the efficiency of ERAP1 in trimming long substrates since such substrates can activate their own trimming. Furthermore, this model can explain the slow trimming rate for small substrates that are too short to self-activate the enzyme. Overall, this model is consistent with the biological role of ERAP1 in efficiently removing 1-7 residues from the N-terminus of antigenic peptide precursors with broad specificity but sparing many mature antigenic peptides.

## Ankylosing Spondylitis Associated Polymorphisms

Ankylosing spondylitis (AS) is a chronic inflammatory joint disease with a strong genetic component that is strongly linked to the class I MHC allelic variant HLA-B27. Genome-wide studies in many human populations have shown strong associations outside the MHC region (reviewed in (Brown, 2010)) in particular with a region encompassing the ERAP1 gene (Burton et al., 2007). Figure 2.15 shows sites of eight single nucleotide polymorphisms that are present in allelic variants of ERAP1 associated with altered risk of developing AS: R127P, M349V, D575N, K528R, Q730E, R725Q, V647I, and C736R (Burton et al., 2007; Harvey et al., 2009). Potential effects of some of these variants can be identified by examination of the structure of ERAP1. The ERAP1 polymorphism, K528R, has the highest odds ratio for decreased risk for AS (Reveille et al., 2010) and also has been associated with increased risk for hypertension (Yamamoto et al., 2002). A biochemical study of this variant showed that it reduces enzymatic activity in degrading single residue substrates and peptidic hormones (Goto et al., 2006). This residue is located on an exposed loop linking domains II and III, and potentially could interact with portions of the C-terminal domain after cavity closure as envisioned in Figure 2.15. The polymorphism M349V is associated with altered risk of AS. Position 349 is near the active site, and would be able to interact directly with the peptide substrate. R725Q and Q730E polymorphisms involve domain IV residues exposed on the inner surface of the cavity, and might affect peptide binding or domain closure. R127P creates an unusual PP

sequence at the tip of a loop in domain 1 that extends towards the cavity, and might cause conformational alteration. The V647I polymorphism is located at the end of helix 10 at the interface with helix 24, near the interactions that close domain IV, and would be affected by domain reorientation. The C736R polymorphism could cause a structural change because Cys736 participates in a disulfide bond with Cys743 between helices 14 and 15 of domain IV. D575N is located at the distal end of domain II, oriented away from ERAP1's active site and cavity. Although this position does not appear to be involved directly in substrate binding or catalysis, domain III in TIFF3 has been suggested to regulate domain motions controlling activity (Kyrieleis et al., 2005). These considerations indicate that ERAP1 polymorphisms around the active site, domain junctions, and C-terminal cavity could influence enzymatic activity of ERAP1 and may affect the repertoire of presented self and environmental class I MHC-bound peptides.

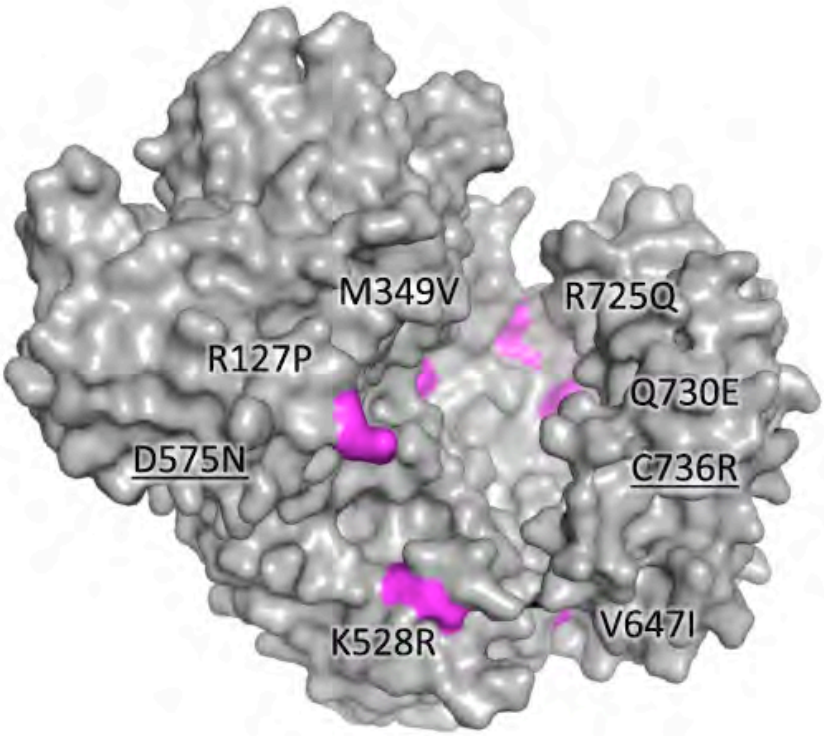

M349V
R725Q
R127P
Q730E
D575N
C736R
K528R
V647I

**Figure 2.15. Ankylosing Spondylitis- Associated Mutations Mapped on Surface of ERAP1.** Surface of ERAP1 is gray, with polymorphisms associated with ankylosing spondylitis shown in pink. Underlined residues are hidden in this view.

## DISCUSSION

The crystal structure of ERAP1 reveals an open, four-domain arrangement with a large solvent-accessible chamber contiguous to a zinc-based aminopeptidase active site. The large cavity appears to be an adaptation specific to ERAP1 and related to its ability to efficiently process peptides that are longer than the substrates of typical aminopeptidases. The ability of short peptides to allosterically activate ERAP1 towards the fluorogenic aminopeptidase substrate L-AMC suggests conversion between (at least) two different conformations dependent on peptide binding, one more and one less active. The length-dependent processing activity characteristic of ERAP1 probably reflects this same activation process, with peptides longer than 7-8 residues able to enhance their own amino-terminal processing. Comparisons with other M1-family aminopeptidases suggest that the structure of ERAP1 bound to bestatin represents an open state not optimized for catalysis, but which could adopt a more closed arrangement by simple rigid-body domain motions. Differences observed between three crystallographically independent copies of the molecule in the crystal and normal mode analysis of the refined structure suggest a trajectory for such a motion. The concave surface of the C-terminal helical domain implies that after closure there would still be a cavity sufficient to accommodate a peptide substrate of up to 13-16 residues

How could conversion between the active and inactive conformations of ERAP1 be regulated? The arrangement of key functional groups in the active site

of ERAP1 is identical to that observed for other M1 family members, except for Tyr438, which is oriented away from the catalytic zinc in ERAP1 but towards it in most other M1 aminopeptidase structures (Figure 2.9). The one other exception is TIFF3, which like ERAP1 is in an open conformation (Kyrieleis et al., 2005), and which also has the Tyr438 analog also oriented away from the active site (Addlagatta et al., 2006). In ERAP1, a small ~30-40º rotation of helix 5 would orient Tyr 438 towards the active site, in a position to participate in peptide bond hydrolysis chemistry. Structural studies of other enzymes related to ERAP1 suggest that domain motions triggered by peptide binding could induce active site changes to enhance peptidase activity as suggested here. In two multidomain zinc metallopeptidases distantly related to ERAP1, human angiotensin-converting enzyme-related carboxypeptidase (ACE2) (Towler et al., 2004) and *E. coli* dipeptidyl carboxypeptidase (Dcp) (Comellas-Bigler et al., 2005), domain closure motions have been associated with active site reorganizations favoring catalysis. In two single-domain zinc metallopeptidases more closely related to ERAP1, thermolysin (Holland et al., 1992) and astacin (Grams et al., 1996), sub-domain motions similarly reorient active site residues for hydrolysis. In the structure of ERAP1, domain closure could reorient helix 5 and Tyr438. Helix 5 is oriented towards the C-terminal domain, sandwiched between helix 3, which carries the second Glu of the $HExxHx_{18}E$ motif, and helix 4, which packs against helix 10 of the C-terminal domain. While no direct contacts between helix 5 and the C-terminal domain are observed in the

structure reported here, a disordered region for which we do not observe electron density extends from the end of helix 5 towards the C-terminal domain. Two other disordered regions are found directly across the cavity, between the tops of helices 21 and 22, and between the tops of helices 23 and 24 (Figure 2.3). Domain closure along the trajectory indicated by normal mode analysis would bring these regions into contact.

A unique aspect of ERAP1 is its preference for long peptide substrates. Low-affinity, low-specificity binding sites on the interior surface of the cavity, similar to those present on nuclear importin-$\alpha$, another armadillo/HEAT repeat protein (Matsuura and Stewart, 2004), might enable efficient peptide capture in the open form. Cavity closure could help to align peptides towards the active site for amino-terminal processing, and differential exposure of interior binding sites between open and closed forms could facilitate peptide binding and release cycles, which would appear to be necessary for sequential removal of peptide N-termini as a long antigenic precursor is processed. This type of mechanism has been suggested for insulin degrading enzyme, a dimeric chambered zinc endopeptidase with “exosites” for binding entrapped peptide at positions distant from the hydrolytic site (Malito et al., 2008; Shen et al., 2006). Recently, ERAP1’s paralog, IRAP, has been shown to process relatively long peptides (Georgiadou et al., 2010). If this model is correct, then IRAP also would be expected to adopt open and closed conformations during catalysis.

During the manuscript preparation, a crystal structure of ERAP1 without bestatin was independently determined by another group (Figure 2.16). That structure reveals ERAP1 in a closed conformation similar to that observed for LTA4H, ColAP, ePepN, and PfAM1, with the C-terminal domain closed into the catalytic domain (Figure 2.16D) and with Tyr438 oriented towards the active site to allow for participation in peptide bond hydrolysis (Figure 2.16E). In the closed conformation, the substrate binding pocket is still large enough to accommodate a 12-14-residue peptide, but it is no longer accessible from the exterior of the protein (Figure 2.16C,F), having been closed off by motion of helices 18-24 in the C-terminal domain (Figure 2.16B,E). The loop connecting helices 21 and 22 contacts the loop between helices H5 and H6, providing a possible route for peptide-induced conformational changes to be communicated to the vicinity of the active site. Together the two structures of ERAP1 provide clear evidence of a large conformational change associated with opening the substrate binding site and coupled to active site reorganization, previously suggested for other M1 family members (Addlagatta et al., 2006). ERAP1 is thus first M1-aminopeptidase for which structures are available in both open and closed conformations, and is likely to serve as a prototype for understanding regulation of enzyme activity in this entire class of proteases.

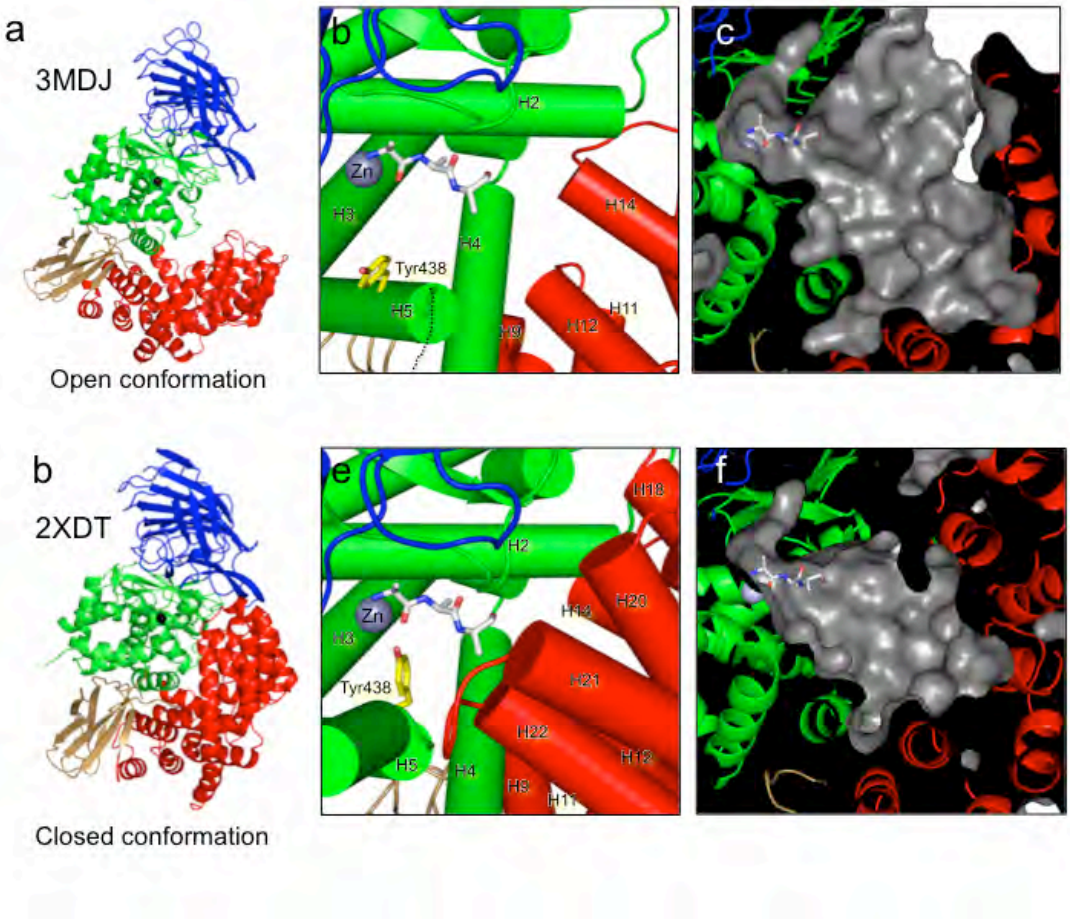

a
3MDJ
Open conformation
b
2XDT
Closed conformation
b
H2
Zn
H3
H4
Tyr438
H5
H14
H11
H12
c
e
H18
H2
Zn
H14
H20
H3
Tyr438
H21
H22
H5
H4
H3
H19
H17
f

**Figure 2.16. Open and Closed Conformations of ERAP1.**

A., B., C.  ERAP1-bestatin in an open conformation as reported here (PDB 2MDJ).

D., E., F., ERAP1 in closed conformation as reported in a recent crystal structure (PDB 2XDT).

A., D. Ribbon diagrams

B. E. The close-up views of active site is shown here with zinc atom shown in gray, model peptide substrate with white bonds, and Tyr438 with yellow bonds. ERAP1 helices are shown as cylinders and colored according to domain.

E,F. Cutaway view of peptide binding cavity, with surface of cavity shown in gray, zinc atom in dark gray, and model peptide substrate with white bonds.

Based on the ERAP1 crystal structure, length and sequence specificity, allosteric regulation by peptides, and comparison with other M1-family proteases, we can describe a likely mechanism for ERAP1 processing of antigenic precursors. (1) Antigenic precursor peptides can access and bind to the groove while the protein is in an open conformation, using interactions at N-terminal S1 and S1' sites but also interactions at other sites in the chamber. Active site residues are not in the optimal arrangement. (2) Peptide binding with occupancy of the regulatory site causes domain closure around the bound peptide and conversion to the optimal active site constellation, possibly with reorientation of Tyr438. Interactions responsible for triggering this change are not yet clear, but likely involve portions of the substrate ~7-8 residues away from the N-terminus, and (directly or indirectly) ERAP1 residues located near domain boundaries. (3) Catalysis. Based on studies of similar metalloproteases, catalytic steps would include: attack of the scissile amide bond by a zinc-bound water, activated by conserved Glu354, to form a tetrahedral intermediate stabilized by interactions with active site residues, followed by collapse of the tetrahedral intermediate with proton transfer to generate a free amino acid and new peptide N-terminus. (4) Release of products after hydrolysis, and substrate rebinding for next cycle. Liberation of the cleaved N-terminal amino acid potentially could occur without conversion to an open form if an exit channel exists in the closed structure, but more likely the protein at least partially converts to an open form after hydrolysis. Interaction of the cleaved peptide product with low-specificity binding sites could

facilitate rebinding in a shifted frame in preparation for the next catalytic cycle. (5) When peptide substrates are trimmed beyond ~8-9 residues, they are too short to engage both regulatory and active sites, and cannot be processed efficiently.

The crystal structure of ERAP1 bound to bestatin provides insight into ERAP1's unusual length specificity and its peptide sequence preferences. These intrinsic properties of ERAP1 probably contribute along with MHC-binding protection to the length and sequence distribution of MHC class I bound antigens. Disease-associated ERAP1 polymorphisms involve residues located near putative substrate binding and regulatory sites. One model for the pathogenesis of AS proposes that particular self peptides presented by HLA-B27 may mimic those of environmental flora, so that T cells originally recognize epitopes in an inflammatory context and subsequently respond to self peptides in an autoimmune fashion. Altering peptide trimming by ERAP1 variants may lead to altered levels of such arthritogenic self peptides or of their mimics, or may lead to altered generation of T cell clones in the thymus. Other studies have linked ERAP1 SNPs to predisposition to diabetes type I, although weakly (log odds 1.07) (Fung et al., 2009), and to the clinical outcome of cervical cancer that is caused by HPV infections (Mehta et al., 2009). These associations could involve constructive or destructive processing of specific peptides, and further studies of these variants are warranted.

Overall, it appears that both the overall conformation of ERAP1 as well as its ability to be activated by long substrates are features that the enzyme has

evolved in order to process large antigenic precursor peptides, while at the same time sparing peptides of length suitable for MHC binding.

# CHAPTER 3

## Structural Analysis of MHCII, HLA-DR1, in Complex with an Environmentally Sensitive Peptide, (4-DAPA)-HA

The work described in this chapter is part of a publication where I am the second author. I did the crystallography and was involved with the Il-2 bioassay to study T cell activation by one of environmentally sensitive probes studied in the publication, Venkatraman P, Nguyen TT, Sainlos M, Bilsel O, Chitta S, Imperiali B, Stern LJ. *Fluorogenic probes for monitoring peptide binding to class II MHC proteins in living cells*. Nat Chem Biol. 2007 Apr;3(4):222-8.

### INTRODUCTION

The MHC-bound antigen repertoire on the cell surface is responsible for eliciting a T cell response that is necessary to defend foreign evasions in organisms. Class II MHC molecules present antigenic peptides at the cell surface. Antigen loading onto class II MHC occurs predominantly in endosomes where the invariant chain gets cleaved and replaced by a peptide by a mechanism that involves HLA-DM. This process is spatially and developmentally regulated (Kleijmeer et al., 2001; Pierre et al., 1997). Real time visualization of this process has been difficult because of the tools available requires many experimental caveats such as not knowing how much is unbound due to the

experimental washes when using radio-labeled or GFP labeled peptides or antibody specific for a particular complex (Day et al., 1995; Murphy et al., 1992).

Class II MHC molecules hold peptides in its peptide binding groove in a polyproline type II conformation through interactions between peptide main chain and MHC main and chain and side chain interactions (Stern et al., 1994). Additionally, the peptide binding groove consist of allele-specific pockets that accommodate side chains of the bound peptide. Class II MHC binds peptides with lengths from 13 to 25 residues long where 9 to 11 residues interact with the MHC molecule and the terminus flanks out on both ends of the peptide binding groove (McFarland and Beeson, 2002). For HLA-DR, the most abundant of the three human class II MHC proteins (HLA-DR, HLA-DP, HLA-DQ), the major determinant of peptide binding is a large hydrophobic pocket (P1 pocket) that can accommodate a hydrophobic or aromatic amino acid side chain near the amino terminus of the peptide (Jardetzky et al., 1990; Sato et al., 2000). Other specificities in the binding groove are at pockets, P4, P6 and P9, with the fine specificity varying between MHC allelic variants (McFarland and Beeson, 2002; Stern et al., 1994).

The side chain specificities of the side chain binding pockets in HLA-DR1 have been well characterized. Specifically, conserved aliphatic and aromatic amino acids line the P1 pocket giving it a non-polar hydrophobic environment (Stern et al., 1994). Taking advantage of the known properties of the P1 pocket, we investigate the possibility developing a fluorogenic probe for that pocket.

There have been reports of solvatochromic probes where an environmental change induces large spectral changes exhibited by the chromaphore. Peptides consisting of the peptide analog probes have been used to phosphotyrosine-binding SH2 (Vazquez et al., 2005), phosphoserine/phospothreonine-binding 14-3-3 domains (Eugenio Vazquez et al., 2004), RNAase assembly (Nitz et al., 2002), delta-opioid antagonist binding (Chen et al., 2006) and electrostatics within protein G (Cohen et al., 2002).

Our studies suggest that replacing the P1 pocket residue of the HA peptide with the solvatochromic probe, 4-DAPA, provides a suitable peptide for antigen processing studies. It demonstrates that the modified HA peptide binds to the class II MHC with the same affinity and activate HA specific T cells similarly as the native HA peptide. It also bound to HLA-DR1 in the same register and did not effect the overall fold of HLA-DR1.

## MATERIALS AND METHODS

All reagents for peptide synthesis were of analytical grade. Resin and amino acids were obtained from AnaSpec (USA). Amino acid and reagents were obtained from Novabiochem and Sigma-Aldrich. 4-DAP, 4-DAPA, were synthesized according to the previously reported methods by the Imperiali laboratory (Eugenio Vazquez et al., 2004; Vazquez et al., 2005).

**Peptide synthesis and purification**

The (4-DAPA)-HA peptide made from HA peptide that was synthesized on a Rink-amide resin using an automatic peptide synthesizer adopting standard Fmoc-based SPPS chemistry by Prasanna Ventraman. The 4-DAPA were then coupled manually to 0.01 to 0.02 mmol resin-bound peptide. Two equivalents of Fmoc-amino acid were mixed with two equivalents each of N, N-Diisopropylethylamine /2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate Methanaminium in 0.30 mL dimethylformamide (DMF) and reacted with the resin for 60 min. The peptides were deprotected (20% piperidine in DMF) and N-acetyl capped using a mixture of pyridine and acetic acid (0.15 M in DMF, 30 min). For peptide 4-DAPA-HA, the remaining Lys and Pro residues were coupled manually after Fmoc-4-DAPA addition and deprotection. Peptides was cleaved from the resin and ether precipitated and washed before purification by reverse phase HPLC (water/0.01% TFA using an acetonitrile/0.01% TFA gradient). The major peak was collected, dried and re-suspended in water at pH 4.0 or in water containing 100 mM urea. MALDI-TOF confirmed the identity of the peptides (Ac-4-DAPA peptide amide: expected 1089.54 Da; observed mass 1090.69 Da). Peptides were stored frozen at -80°C until use.

**MHC-peptide complex formation**

Soluble recombinant DR1 extracellular domain was prepared by expression of truncated HLA-DRA*0101 and HLA-DR1*0101 genes in S2 insect cells cultured in serum-free medium followed by immunoaffinity isolation of the secreted protein as described (Stern and Wiley, 1992). For peptide loading, 7 µM of HLA DR1 was incubated with 30 µM (4-DAPA)-HA in PBS (10 mM $Na_2HPO_4$ containing 137 mM NaCl and 2.7 mM KCl, (pH 7.4), 1 mM PMSF, protease inhibitor cocktail from Sigma Aldrich, 0.02% sodium azide) for 3 days at 37°C. MHC-peptide complexes were isolated by gel filtration using a Superdex S200 column (Pharmacia) with phosphate-buffered PBS containing 0.02% sodium azide buffer. The MHC-peptide complex eluting between 31 and 34 min was collected and concentrated. Non-bound protein tends to aggregate under the incubation conditions and is typically not found in these fractions (Stern et al., 1994). The complex thus purified from free peptide was stable for several months at 4°C. UV-Visible spectra of the purified peptides and the HLA-DR-peptide complex were measured in PBS. Fluorescence excitation and emission spectra were obtained for purified DR-peptide complex and the free peptide, with 5 nm bandpass for excitation /emission. Raman spectra of pure water obtained under the same conditions were generally subtracted from the original spectra. For direct binding measurement in presence of excess peptide, the 4-DAPA-peptide (1 µM) was incubated with 0.125 - 1 µM of HLA-DR1 for 3 days and the fluorescence measured as described above.

**Crystal structure**

HLA-DR1 extracellular domain was produced by expression of the α and β subunits in *E*. *coli* followed by in vitro folding, peptide loading, and final isolation by gel filtration, as described (Frayser et al., 1999). The bacterial superantigen, SEC3-3B2, was expressed in folded form in the *E. coli* periplasm and isolated as described(Andersen et al., 1999). A 1:1 mixture of HLA-DR1-(4-DAPA)-HA and SEC3-3B2 was crystallized by vapor diffusion against 100 mM sodium acetate, 2% PEG 4K and 10% ethylene glycol at 4°C. A 100 µm cubic crystal was transferred to 25% ethylene glycol v/v in mother liquor and cooled to -190°C for data collection at the National Synchrotron Light Source beamline X-29 using 1.1 Å radiation. Diffraction data were processed using HKL2000 (Otwinowski, 1997) and phased using a HLA-DR1-SEC model from the 2.1 Å crystal structure of another peptide complex (PDB 1PYW). Initial unrefined omit maps exhibited clear density for the DAPA side chain that extended only to the middle of the aryl ring, and smeared density for the Pheβ89 side chain. Initial model excluding DAPA and Pheβ89 was improved by several cycles of manual rebuilding, automated refinement using Refmac and TLS (Vagin et al., 2004) or CNS (Brunger, 2007) and addition of ordered water molecules. Omit maps from the improved models exhibited improved 4-DAPA and Pheβ89 side chain density. The 4-DAPA topological parameters and refinement restraints were obtained from the PRODRG server (Schuttelkopf and van Aalten, 2004), and use to include 4-DAPA in the model. Examination of positive and negative electron

density in Fobs-Fcalc difference maps suggested a possible minor alternate orientation of the 4-DAP ring corresponding to a $c_2$ rotation of ~180º, with a corresponding motion of Phe$\beta$89. However, inclusion of this alternate conformation at various occupancy levels did not improve the refinement statistics or appearance of the electron density, therefore it was not included in the final model. Data collection and refinement statistics are shown in Table II. Coordinates for this model have been deposited in the PDB and is available with ID 2IPK.

**Mass Spectrometry**

About 5-6 crystals were looped out of a crystal drop and placed in water to remove excess peptide that might be outside of the crystals. Next the crystals were dissolved in 10% acetic acid in water. The mixture was either purified using C18 zip tip (Millipore) then spotted or spotted directly with *alpha*-cyano-4-hydroxycinnamic acid for Maldi-TOF mass spectrometry analysis.

**T cell activation**

A murine T cell hybridoma carrying the HLA-DR1-restricted HA [316-318] peptide-specific T cell receptor HA1.7 and human CD4, a kind gift of Jerome Bill (Univ. Colorado Health Sciences Center), was used to evaluate the biological activity of the modified HA peptide. This assay was carried out my Mauricio Calvo-Calle and myself. HLA-DR1+ B-lymphoblastoid LG-2 cells (20,000 per well

in 0.1 mL) were pulsed (in triplicate) with the indicated concentration of peptide for 3.5 hours at 37°C, 5% $CO_2$ in RPMI medium supplemented with 10% FBS. After incubation the cells were washed 5 times with fresh medium, and then 10,000 T cells were added in the same medium for an additional incubation of 24 hours under the same conditions as above. Supernatants were removed and assayed for secreted IL-2 using the IL-2 dependent cell line CTLL-2 and a $^{3}H$-thymidine incorporation assay.

## RESULTS

Modeling studies were conducted using the crystal structure of HLA-DR1 in complex with the influenza-derived HA peptide (PKYVKQNTLKLAT) (Stern et al., 1994). The peptide tyrosine that occupies the P1 pocket was replaced with three amino acid analogs that consist of environment-sensitive fluorophores as the side chains. The amino acid analogs are termed 6-dimethylamino-2'-naphthoyl-alanine (DANA/ALADAN) (Cohen et al., 2002; Nitz et al., 2002), 4-N,N,-dimethylaminophthalimido-alanine (4-DAPA) (Eugenio Vazquez et al., 2004), and 6-N,N-dimethylamino-2,3-napthalimido-alanine (6-DMNA) (Vazquez et al., 2005). The amino analogs 4-DAPA and 6-DMNA could be accommodated into the P1 pocket of HLA-DR1 without major distortion of the MHC protein but DANA/ALADAN was too long to fit as suggested by molecular modeling consisting of rigid body docking and torsional angle adjustment.

Although modeling suggest that the 4-DAPA or 6-DMNA group may fit into the P1 pocket, it is important that the introduction of these fluorescent reporter group in a modified peptide would not significantly alter the structure of the MHC-peptide complex or its function in eliciting a T cell response. To evaluate these issues, the HA peptide was synthesized with the P1 tyrosine replace with 4-DAPA and termed 4-DAPA-HA. The 4-DAPA-HA peptide bound well to HLA-DR1 with an $IC_{50}$ less than 1 nM in a standard competition binding assay. Fluorescence intensity, Emission $l_{max}$ exhibited a blue shift and fluorescence intensity increased upon binding to HLA-DR1 (Figure 3.1). Lifetime

measurements (measured by Imperiali Lab) were 3.65 nsec for the majority fraction of the bound and 0.2 nsec for the majority fraction of the unbound (4-DAPA)-HA. Additionally, the biological activity of the 4-DAPA-HA was studied. HLA-DR1-positive B cells were treated with HA or 4-DAPA-HA and used to stimulate an HLA-DR1-restricted HA peptide-specific T cell hybridoma. An IL-2 bioassay suggests that HA and 4-DAPA-HA treated B cells caused identical activation (Figure 3.2). To verify that 4-DAPA-HA binds into the peptide binding groove and check for any structural alterations of HLA-DR1 resulting from introduction of the 4-DAPA group, the crystal structure of the HLA-DR1:4-DAPA-HA complex was determined at 2.3 Å resolution, using the bacterial superantigen, SEC3-3b2, to promote crystallization. Data collection and refinement statistics are shown in Table 3.I. The HA and (4-DAPA)-HA peptides adopt essentially identical conformations bound to HLA-DR1 (RMSD 0.24 Å) with differences observed only at the termini of the flexible and solvent-exposed Lys, Val and Lys side chains at the P-1, P2 and P8 positions, respectively (Figure 3.3). As anticipated, the 4-DAP side chain occupies the P1 pocket completely buried by interactions with many aromatic and hydrophobic side chains in this region (Figure 3.4). We had expected that the 4-DAP side chain would displace one or two of the water molecules bond in the P1 pocket that form hydrogen bonds to the Thr89β hydroxyl group, however the structure reveals that 4-DAP predominately adopts the ring-flipped orientation, placing the dimethylamino group on the other side of the ring interacting with the Thrβ89 methyl group, and

not displacing the water molecules (Figure 3.4). Since initial electron density map only showed density up to the middle of the aryl ring in 4-DAP, crystals were tested for peptides. To ensure that the peptide crystallized was 4-DAPA-HA, peptides were extracted from crystals and analyzed by Maldi-TOF mass spectrometry showed no sign of native HA peptide (Figure 3.5). Overall, the introduction of the 4-DAP fluorophore in the context of (4-DAPA)-HA does not significantly alter the binding affinity, biological activity, or binding register compared to that of the HA peptide.

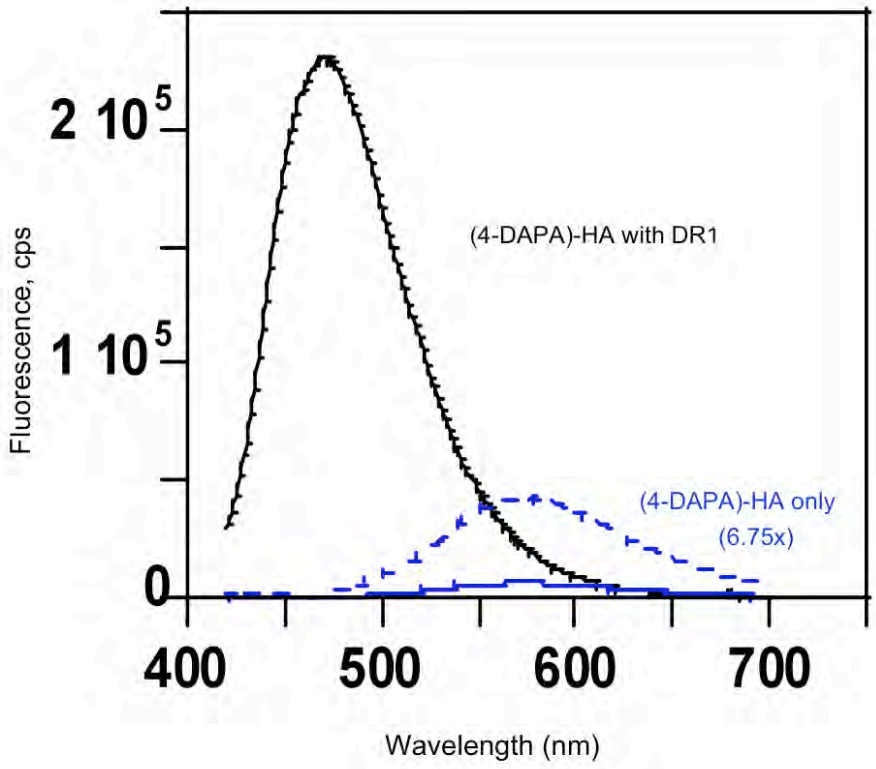

2 10^5
1 10^5
0
Fluorescence, cps
(4-DAPA)-HA with DR1
(4-DAPA)-HA only
(6.75x)
400
500
600
700
Wavelength (nm)

**Figure 3.1. The (4-DAPA)-HA Peptide Exhibits a Spectral Shift when Bound to HLA-DR1.** The (4-DAPA)-HA peptide was incubated with HLA-DR1 in 5 molar excess and the complex formed was purified by gel filtration. Fluorescence of DR-(4-DAPA)-HA was compared with the free peptide.

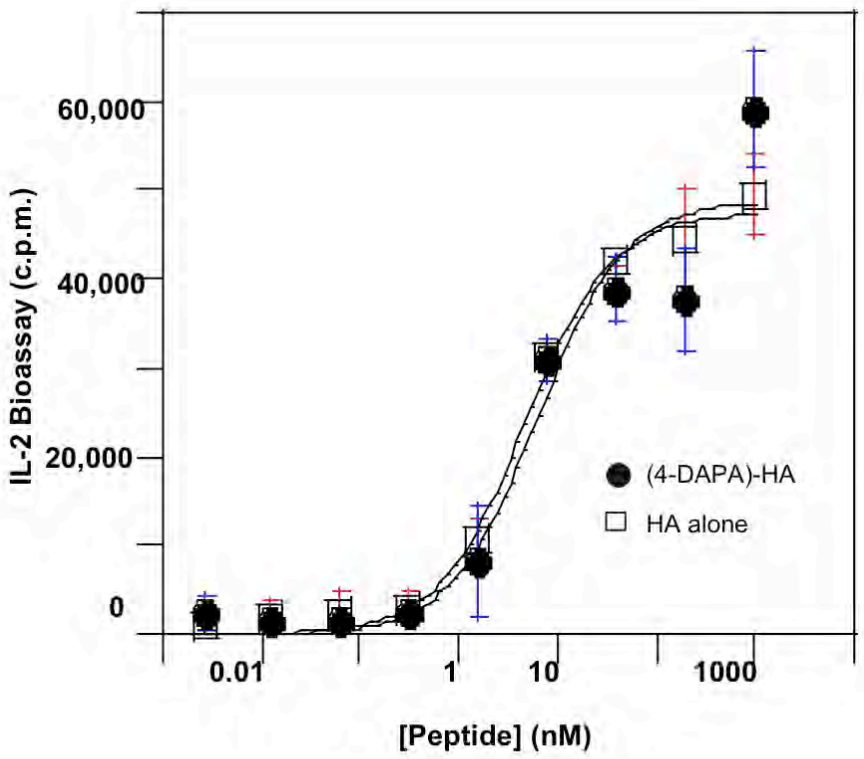

60,000
40,000
20,000
0
IL-2 Bioassay (c.p.m.)
0.01
1
10
1000
[Peptide] (nM)
(4-DAPA)-HA
HA alone

**Figure 3.2. The (4-DAPA)-HA Peptide Stimulates HA Specific T cells Similarly as Compared to Native HA Peptide.** HA (circles) or DAPA-HA (open squares) peptides. T cell activation measured using a bioassay for IL-2 production.

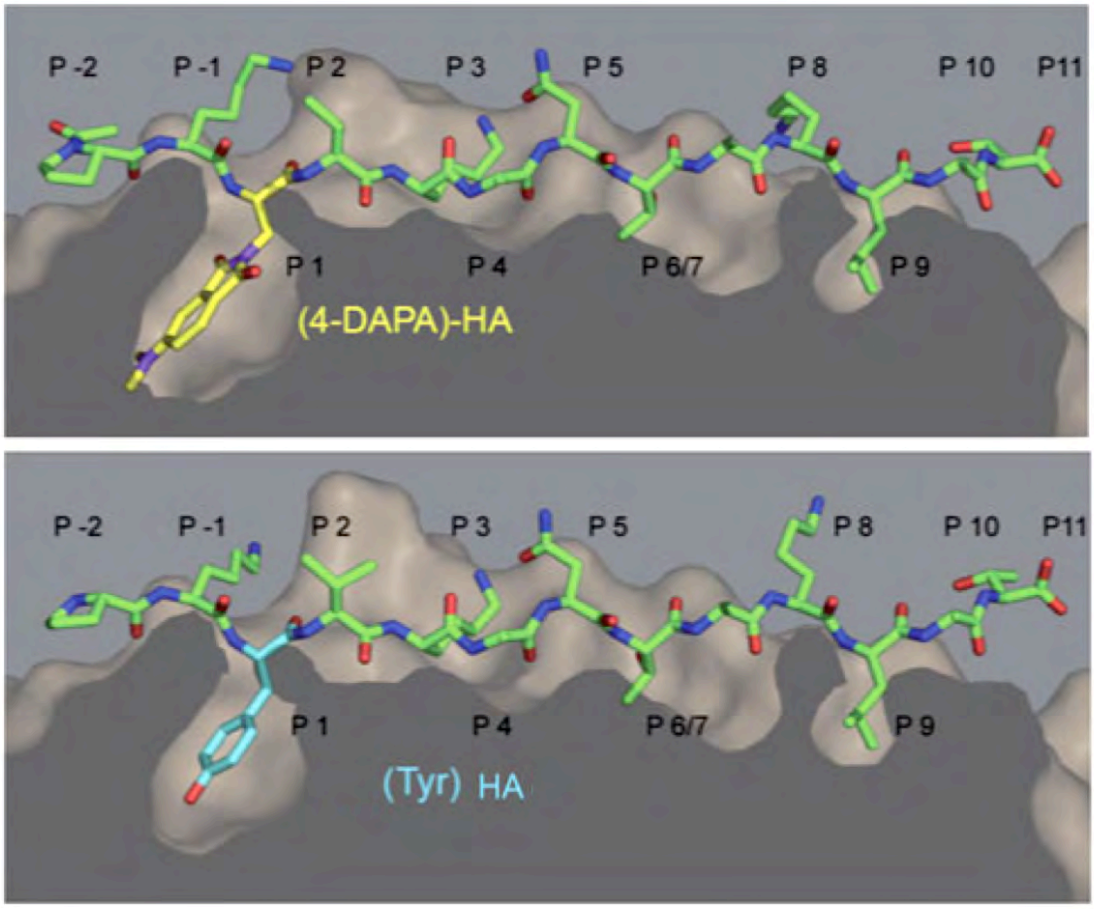

P -2
P -1
P 2
P 3
P 5
P 8
P 10
P11
P 1
P 4
P 6/7
P 9
(4-DAPA)-HA
P -2
P -1
P 2
P 3
P 5
P 8
P 10
P11
P 1
P 4
P 6/7
P 9
(Tyr) HA

**Figure 3.3. The (4-DAPA)-HA Binds into HLA-DR1 in the Same Register as the Native HA Peptide.** The panels show the peptide binding groove of structures of 4-DAPA-HA (top) or HA (bottom) peptide complex with HLA-DR1. HLA-DR1 is shown in surface representation with peptides shown in sticks. (Top) DAPA side chain is shown with yellow bonds extending down into the P1 pocket. (Bottom) Unmodified HA peptide from the crystal structure of the DR1-HA-SEC complex (PDB 1JWU) shown after alignment of MHC peptide binding domain, with tyrosine side chain at the P1 position shown with cyan bonds.

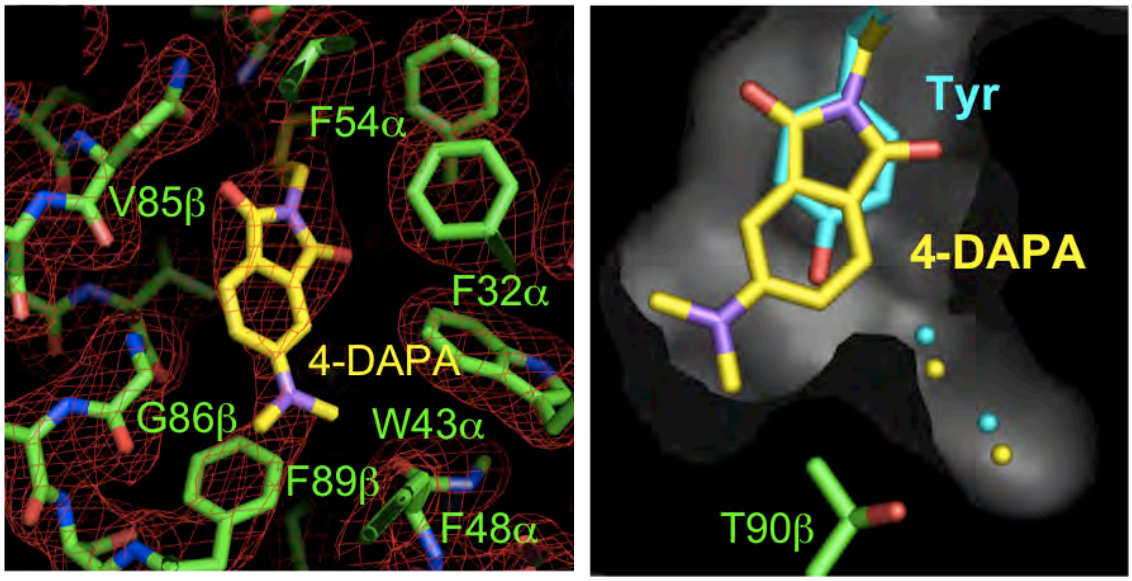

F54α
V85β
F32α
4-DAPA
G86β
W43α
F89β
F48α
Tyr
4-DAPA
T90β

**Figure 3.4. The 4-DAPA Binds into the P1 Pocket.** Left panel shows |2Fo-Fc| omit map of the region around the P1 pocket with all residues shown removed from the model before map calculation. The (4-DAPA) group is colored in yellow. Right panel is a cut off section through the P1 pocket surface in gray, showing the HA peptide tyrosine side chain and the 4-DAPA fluorophore, along with their corresponding ordered water molecules. Unmodified HA peptide tyrosine from the crystal structure of the DR1-HA-SEC complex (PDB 1JWU).

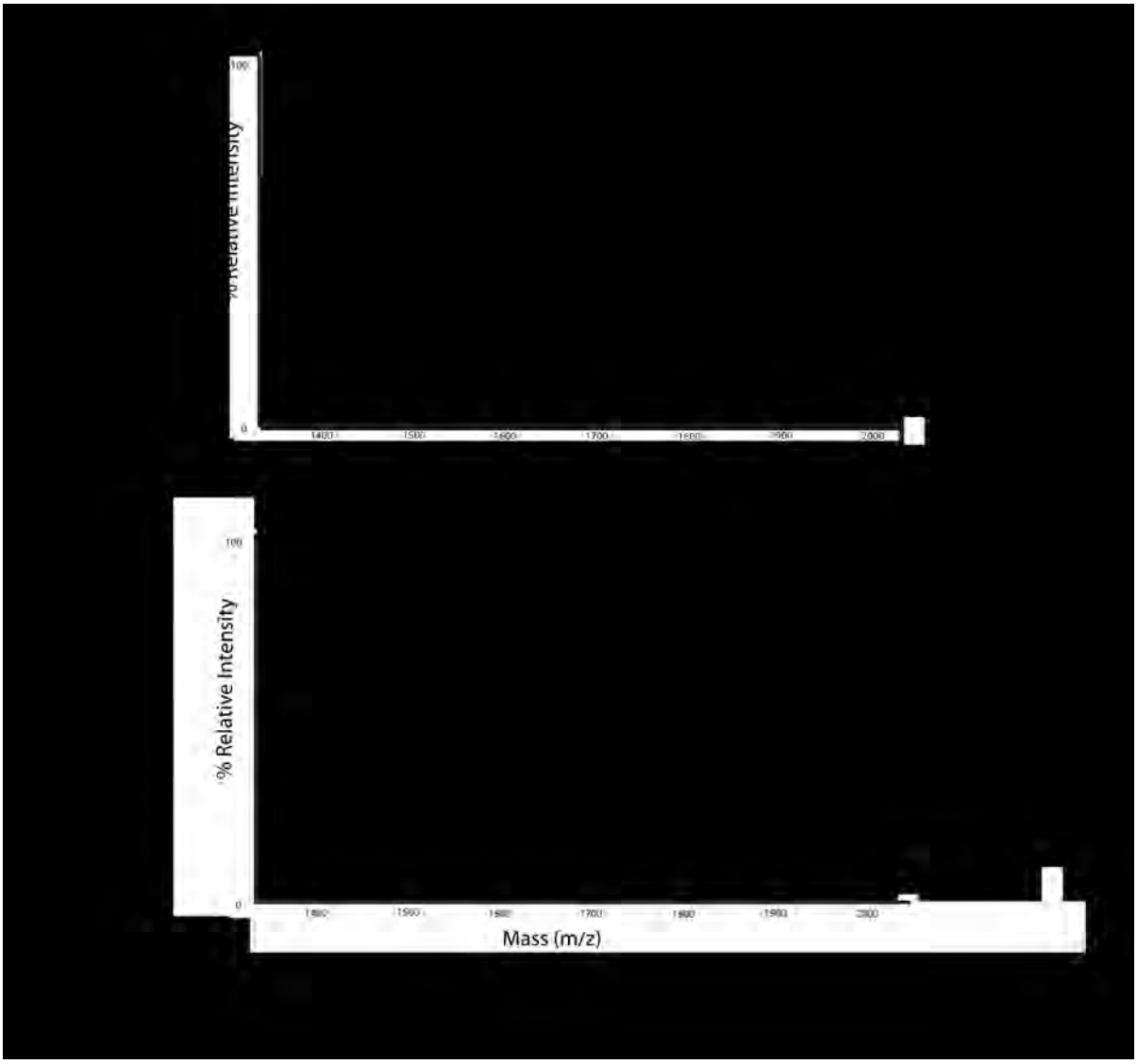

% Relative Intensity
Mass (m/z)

**Figure 3.5. Maldi-TOF Mass Spectrometry Spectra of Peptide Eluted from HLA-DR1/(4-DAPA)-HA Peptide/SEC3-3b2 Crystals.** Top panel shows spectra of sample eluted from crystals and bottom shows a spectra of free purified peptide. Both panels show the expected molecular weight of (4-DAPA)-HA of 1641 Da and sodium conjugated peak.

## DISCUSSION

The work discussed in this chapter is part of the characterization and an application of environmentally sensitive probes (4-DAPA and 6-DMNA) that can be use for antigen processing studies. The 4-DAPA containing peptide binds to HLA-DR1 and activates T cells similarly to that of non-modified peptides. The (4-DAPA) peptide exhibits differential fluorescence properties with increase in lifetimes when bound to HLA-DR1 compared to the unbound form. Even though the structure of (6-DMNA) containing peptide has not been determined, it is expected that (6-DMNA) containing peptide will behave similarly to (4-DAPA) peptide because other characteristics such as lifetime increase, fluorescence intensity increase, peptide binding and T cell activation are essentially identical.

In addition to characterizing these fluorogenic peptides, my coauthors also used the DAPA-HA peptide to study immature and mature DCs that previously used indirect methods of measuring peptide loading by antibodies for specific peptide MHC complexes. Our study using a direct method for measuring bound antigens to an MHC is in agreement with one of two fields of thoughts in regards to DCs as antigen presenting cells. The results suggest that immature DCs can present antigen sufficiently at the cell surface in contrast to other studies that suggest maturation is necessary for DCs to load antigen and present at the cell surface. However, our study does not distinguish the difference between loading of peptide at the cell surface or inside endosomes. It is important to understand

the spatial aspect of antigen loading in DCs to understand their role as antigen presenting cells.

Environmentally sensitive peptides, such as DAPA-HA, can be used to study other protein to peptide interaction such as phosphotyrosine-binding SH2 domain (Vazquez et al., 2005) phosphoserine- and phosphothreonine-binding 14-3-3 proteins (Vazquez et al., 2005). Though these peptides were designed and characterized for HLA-DR1 P1 binding pocket, they can be used in other class II MHC or allelic variants, such as HLA-DR4, HLA-DP and HLA-DQ, that have a pocket similar to the large hydrophobic P1 pocket. Additionally, since these probes are insensitive to pH and can be used in any cell type, they can be used to study HLA-DM peptide binding, regulation or antigen differential compartmentalization in the endosomes. Our characterization shows that the peptide probes consisting of 4-DAPA and possibly 6-DMNA can provide a direct way to measure peptide binding into class II MHC in suitable cellular environments.

Another application for this is to study class I antigen loading. Class I MHC such as H2-$K^d$ (Rammensee et al., 1995) prefers a hydrophobic residue at the $4^{th}$ position away from C-terminus of its peptide substrate. It can be used to study the antigen loading complex associated with class I MHC in the ER.

**Table 3.1. Data collection and refinement statistics for DR1/(4-DAPA)-HA**

| | DR1:(4-DAPA)-HA:SEC |
|---|---|
| **Data collection** | |
| Space group | R3 |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 171.98, 171.98, 121.13 |
| α, β, γ (°) | 90, 90, 120 |
| Resolution (Å) | 34-2.3(2.38-2.3) * |
| $R_{sym}$ or $R_{merge}$ | 0.087(0.363) |
| $I / \sigma I$ | 14(2) |
| Completeness (%) | 100(99.9) |
| Redundancy | 4.6(4.4) |
| **Refinement** | |
| Resolution (Å) | 34-2.3 |
| No. reflections | 282,886 |
| $R_{work}$ / $R_{free}$ | 0.203/0.224 |
| No. atoms | |
| Protein | 5046 |
| Water | 419 |
| *B*-factors | |
| Protein | 48.3 |
| Water | 51.3 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.007 |
| Bond angles (°) | 1.4 |

*Highest-resolution shell is shown in parentheses.

# CHAPTER 4

## Structural Study on the Determinants of the C-terminal Hairpin Loop of the GagP16 Peptide in the Binding Groove of HLA-DR1

## INTRODUCTION

This chapter reports results to further understand the nature of a unique hairpin loop exhibited by an HIV derived peptide, GagP16 using X-ray crystallography and series of peptides designed that differ in sequence at the N-terminus of the peptide.

The class II major histocompatibility complex, human leukocyte antigen (HLA)-DR1, plays an important role in the immune system because HLA-DR1 presents antigenic peptides on cytotoxic cell surfaces for T cell recognition, thus initiating an immune response. It is the most abundant of the three human class II MHC proteins, HLA-DR, HLA-DP, and HLA-DQ. HLA-DR1 contains a 33-kDa $\alpha$ chain and a 28-kDa $\beta$ chain. Each chain includes an IgG-like domain and together the chains make up the binding groove for antigenic peptides (Stern et al., 1994) Furthermore, these antigenic peptides exhibit polyproline type II conformations in the groove, which is enforced by interactions between conserved residues of HLA-DR1 and the backbone of the antigenic peptide (McFarland and Beeson, 2002; Stern et al., 1994) The binding groove stretches at least 9 amino acid residues long with specific residue recognition (Jardetzky et al., 1990). However, HLA-DR1 can present peptides longer than 20 amino acids

where the C- or N-terminus of the peptide hangs out of the binding groove (Chicz et al., 1992; Rudensky et al., 1991). Pockets formed in the peptide binding groove of HLA-DR1 allows for substrate specificities in the P1, P4, P6/7, P9 and P10 pocket (Stern et al., 1994). The P1 pocket allows for large hydrophobic residues such as Tyr, Trp, Phe, Leu and Ile; P4 prefers aliphatic residues, P6 allows small residues such as Gly, Ala, Ser and Pro, P9 prefers aliphatic and P10 allows hydrophobic residues (McFarland and Beeson, 2002; Stern et al., 1994; Zavala-Ruiz et al., 2004a; Zavala-Ruiz et al., 2003). Peptide residue numbering starts at the P1 hydrophobic pocket. Residues towards the N-terminus of the peptide, which are outside the pocket, are numbered P-1, P-2, etc. Residues that are not in the binding pockets of HLA-DR1 shown to be important for T cell binding are the P-1, P3, P5 and P8 residues (Hennecke and Wiley, 2002; Wedderburn et al., 1995). Although structural information on HLA-DR1 in complex with the T cell receptor (TCR) can provide information on T cell recognition, there are limited structures peptide bound MHC complex with TCR.

Bound to HLA-DR1, the HIV gag envelope derived peptide, GagP16 (residues 34-48, PEVIPMFSALEGATP), exhibits a unique C-terminal beta hairpin loop that is required for T cell activation as shown by structural analysis and peptide mutation T cell activation studies (Zavala-Ruiz et al., 2004b). A truncated GagP16 termed GagP13 (PEVIPMFSALEG) does not elicit a T cell response but does bind to T cells more weakly (Norris et al., 2006; Zavala-Ruiz et al., 2004b). Crystallographic studies of HLA-DR1 and gag envelope derived peptides,

GagP13 and GagP16, that were crystallized with the SEC3-3B2 superantigen showed a conventional polyproline conformation up to P10 pocket on HLA-DR1, but GagP16 exhibits an additional unique beta hairpin loop at the C-terminus (Zavala-Ruiz et al., 2004b). A hairpin loop has not been observed in other MHC peptide structures. In structures of class II MHC with longer peptides, the N- and C-terminus of the peptide was observed to flank outside of the peptide binding groove (Madden, 1995; McFarland and Beeson, 2002). Additionally, there were interactions within GagP16 between the Leu at P8 and C-terminal residues, Thr at P13 and Ala at P12 (Zavala-Ruiz et al., 2004b). Understanding the determinants of the C-terminal hairpin loop of GagP16 and the contacts between TCR and HLA-DR1/GagP16 complex will give insight on why the C-terminal beta hairpin loop is necessary for T cell activation.

Here, we study the determinants of the C-terminal hairpin loop of the GagP16 peptide in the binding site of HLA-DR1 by way of X-ray crystallographic analysis of complexes involving a variety of peptides consisting of partial segment of the GagP16 sequence. Our results suggest that the C-terminal hairpin determinants are not within the last seven residues of the peptide and suggest that the interactions between other parts of GagP16 and MHC may contribute to formation of the C-terminal hairpin loop that is essential for T cell response. In addition, we attempted to determine the crystal structure of the HLA-DR1/GagP16 and AC25 TCR complex where AC25 TCR is a T cell receptor cloned from cell lines raised against GagP16.

## MATERIALS AND METHODS

### HLA-DR1 expression, refolding and purification

To get a large quantity of HLA-DR1 (DRA*0101, DRB1*0101) for crystallization, extracellular domains of HLA-DR1 subunits were overexpressed in *E. coli* and purified as inclusion bodies by anion exchange chromatography for refolding as previously described (Frayser et al., 1999). Briefly, both subunits were expressed separately as inclusion bodies from BL21(DE3) grown in LB medium and 100 mg/l ampicilin at 37°C. Bacterial cultures were induced at an $OD_{600}$ of 1.0 by the addition of 1 mM isopropyl-ß-D-thio galactopyranoside (IPTG), and harvested 3 hours later. Cells were washed with 50 mM Tris-HCl pH 8.0 and lysed by resuspending the cell pellets in 50 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.4 mg/mL DNase I, 0.4 mg/mL RNase A, 1 mg/mL lysozyme and brief sonication. Insoluble material was sedimented by centrifugation, and inclusion bodies were washed five times with 20 mM Tris–HCl pH 8.0, 23% (w/v) sucrose, 0.5% (v/v) Triton X-100, 1 mM EDTA and once with 20 mM Tris–HCl pH 8.0, 1 mM EDTA. Inclusion body pellets were resolubilized in 6 M urea, 20 mM Tris-HCl pH 8.0, 2 mM EDTA and 5 mM dithiothreitol (DTT). The $\alpha$ subunit was purified through anion ion exchange, HQ, (Applied Biosystems) in 20 mM Tris-HCl pH 8, 6 M Urea and 1 mM DTT then eluted with an NaCl gradient

from 0 to 1 M. The β subunit was purified in the same way but in a higher pH of 8.5.

The MHC subunits were together refolded by slowly adding the purified subunits in inclusion bodies into cold refolding buffer (100 mM Tris-HCl pH 8, 2 mM EDTA, 0.3 mM oxidized glutathione, 0.3 mM reduced glutathione), at a final protein concentration of 16 mg/l α HLA-DR1 and 16 mg/l ß HLA-DR1, and 5x molar concentration of peptide before a 3 day incubation at 4°C. The refolding mix was then concentrated and buffer exchanged through ultrafiltration before purification through HQ anion exchange column (Applied Biosystem) in 20 mM Tris-HCl pH 8.0 with a linear gradient from 0 to 1 M NaCl. Next, the MHC peptide complex was purified further by size exclusion chromatography (Superdex S200, Pharmacia Biotech) in phosphate buffered saline.

**Superantigen, SEC3-3b2, expression and purification**

The super antigen, SEC3-3b2, was expressed in the periplasm of *E. coli* BL21(DE3) grown in 2xYT medium and 100 mg/l chloramphenicol at 27°C as previously described (Andersen et al., 1999). Bacterial cultures were induced at an $OD_{600}$ of 1.0 by the addition of 1 mM isopropyl-ß-D-thio galactopyranoside (IPTG), and harvested 5 hours later. Protein was extracted via osmotic stress in TES (0.2 M Tris, 0.5 mM EDTA, 20% sucrose) buffer followed by 0.2x TES buffer and purified through a Red A affinity column followed by anion exchange column (MonoQ, GE Healthcare).

**AC25 single chain TCR expression and purification**

AC 25 TCR $\alpha$ and $\beta$ genes were isolated by Phillip Norris and Yuri Sykulev (unpublished data) and from the AC25 T cell (Norris et al., 2004) and were transferred to an *E. coli* periplasm expression system (Maynard et al., 2005) by Zarixia Zavala-Ruiz and Lawrence J. Stern. The TCR construct consist of the V$\alpha$ and V$\beta$ domains of AC25 TCR and linked through a $(Gly_4Ser)4$ connecting the N-terminus of the V$\beta$ domain to the C-terminus of the V$\alpha$ domain similar to that was in PDB 1U3H (Maynard 2005). The covalently bound $\alpha/\beta$ variable domain TCR, AC25scTCR, was expressed in the periplasm of *E. coli* BL21(DE3) grown in 2xYT medium and 100 mg/l chloramphenicol at 27°C. Bacterial cultures were induced at an $OD_{600}$ of 1.0 by the addition of 1 mM IPTG, and harvested 5 hours

later. Cells were pelleted and resuspended in TES buffer. Osmotic stress was induced by 0.2x TES buffer and the supernatant was purified through a Ni-NTA affinity column (Qiagen) in PBS and eluted with 30 mM imidazole. AC25scTCR was further purified by size exclusion column in PBS (Superdex S200, GE Healthcare).

**Circular dichroism and thermal melt**

AC25scTCR unfolding was monitored using a Jasco model J810 CD spectrophotometer in the range of 200-250 nm, 0.2 cm cell length and 2.5 nm bandwidth. AC25scTCR was in 10 mM Tris-HCl, pH 7.5 and 0.6 mg/mL. The temperature was increased from 6 to 70°C and CD measurements were recorded every 6 degrees.

**Peptide synthesis**

Peptides were synthesized using standard solid phase fluorenylmethoxycarbonyl (F-moc) chemistry (Symphony instrument, Protein Technology, Inc.), cleaved and deprotected. Crude peptides were precipitated in diethyl ether two times before lyophilization and then purified by reverse phase high-pressure liquid chromatography (Vydac-C18). Peptides were characterized by matrix- assisted laser desorption ionization time of flight, MALDI-TOF, mass spectrometry. To prevent ionization of the terminus, peptides were aminated at

the C-terminus. Concentrations were determined using the theoretical extinction coefficients.

**Surface plasmon resonance**

The binding experiments were done on a BIAcore3000 using CM-5 carboxymethyldextran-coated gold chips coupled to modified avidin (Neutravidin). HLA-DR1 tetramers complexed with GagP16 or HA or A2-gag (Gift from Yuri Sykulev) was immobilized. AC25scTCR was injected at 2 mL/min at 67 µM. Volumes were adjusted to 1 µL, 0.5 µL, and 0.25 µL to avoid nonspecific binding. As a control, HLA-A2/gag peptide complex was immobilized and specific TCR-D3#1 was injected.

**Crystallization, structure determination and refinement**

Crystals were grown for X-ray diffraction experiments by vapor diffusion method in hanging drops at 4°C by combining 1 µL of protein (10 mg/mL) with 1 µL of well solution (2% polyethylene glycol 4000, 10% ethylene glycol, 100 mM sodium acetate pH 5.4-5.6). The protein solution consisted of equal molar concentration of HLA-DR1/peptide complex and SEC3-3b2. Crystals for data collection were approximated to be 100 µm x 100 µm x 50 µm in size. For data collection, crystals were first transferred into a cryo-protecting crystallization well solution with 25% ethylene glycol and then flash frozen in liquid nitrogen. X-ray diffraction data were collected under cryo conditions from single crystals

containing peptides PKYVKQNTLKLEGATP, PKYVKQNTLLLEGATP or PKYVKQNTLLSEGATP to 2.8, 2.2 and 2.2 Å resolution, respectively, at the National Synchrotron Light Source, beamline X29A, using 1.0898 Å wavelength X-rays. Oscillation images were processed with the program package HKL2000 (Otwinowski, 1997). The structures of the complexes were determined by molecular replacement with the starting model of HLA-DR1/GagP16 in complex with the superantigen, SEC3-3b2, without GagP16 peptide (PDB 1SJH). Model building and rebuilding were done in O and COOT (Jones et al., 1991). Rigid body and positional refinement was performed in PHENIX (Adams et al., 2010). A bulk solvent correction and anisotropic *B*-factor tensor were applied throughout the refinement. Refinement statistics are as described in Table 4.1.

AC25scTCR complexed with HLA-DR1 and GagP16 crystals were grown at 4°C by vapor diffusion method in hanging drops consisting of 1:1 volume ratio of protein solution and crystallization well solution. The protein solution consisted equal molar concentrations of AC25scTCR and HLA-DR1/GagP16 complex at 14.5 mg/mL. The well solution consisted of 21% PEG 3350, 0.1 M Hepes, pH 7.5 and 0.2 M $LiSO_4$. Crystals were estimated to be smaller than 20 µm x 50 µm x 20 µm and transferred into cryo condition consisting of well solution with the addition of 25% glycerol. X-ray diffraction data were collected from a single crystal at the National Synchrotron Light Source, beamline X6A, using 1.0 Å wavelength X-rays and ADSC Q270 CCD detector. Oscillation images were processed with the

program package HKL2000 (Otwinowski, 1997). Molecular replacement method for phase estimation were conducted using Phaser (McCoy et al., 2007).

## RESULTS

To investigate the determinants of the hairpin loop at the C-terminus of GagP16, the following designed peptides, which are hybrids between the HA peptide and GagP16 peptides, were used for crystallographic studies: (1) Hybrid 1, PKYVKQNTLKLEGATP, is a peptide where the last 5 residues is of GagP16 and the first 11 residues of HA, (2) Hybrid 3, PKYVKQNTLLLEGATP, is the same as hybrid 1 with the exception of a Leu at the P8 instead of a lysine because Leu at P8 in GagP16 appears to be involved in some interactions with Ala and Thr. (3) Hybrid 2, PKYVQNTLLSEGATP, is the last 7 residues of GagP16 and the first 9 residues of HA, The crystal structures of each of the complexes were examined for presence of the hairpin loop at the C-terminus. The crystals were grown in the similar conditions as those of HLA-DR1 with GagP16 and SEC3-3b2 to facilitate crystallization. Crystals typically grew less than a week after drops were set up. The protein crystallized in the same space group, R3 and they yielded comparable resolution with the exception of hybrid 1. Initial maps were constructed without peptide in the binding groove and built in with confidence. Arp/Warp was utilized for hints of additional density in the C-terminal region (Cohen et al., 2004). Fo-Fo maps were generated using structure factors of HLA-DR1/HA crystals to see differences in the C-terminal region.

**Determinants of the GagP16 C-terminal hairpin turn**

The crystal structure of HLA-DR1 in complex with hybrid 1, PKYVKQNTLKLEGATP, was determined with the resolution of 2.8 Å. Electron density was observed for first 13 residues, PKYVKQNTLKLEG, of the hybrid 1 peptide. It is not likely to have formed a loop that is not observable because the main chain of the last glycine residue points out of the peptide binding groove instead of pointing towards the binding groove to complete the hairpin turn as exhibited by GagP16 (Figure 4.1). The crystal structure of HLA-DR1 complexed with hybrid 1 peptide suggests that not only the last 5 residues of the GagP16 peptide are responsible for the C-terminal hairpin turn needed for T cell activation.

Hybrid 2 peptide, PKYVKQNTLLLEGATP, in complex with HLA-DR1 diffracted to 2.2 Å. The electron density for this complex yields density for the peptide from the N-terminal proline to the P11 Gly. Similar to the previous structure of hybrid 1 peptide, C$\alpha$ main chain of the Gly points away from the peptide binding groove (Figure 4.2). This data suggest that not only the last 5 residues of the GagP16 peptide but also P8 are important for interactions with the P12 Ala and P13 Thr on the other side of the turn does not dictate the C-terminal hairpin turn needed for T cell activation.

To further study the determinants of the C-terminal hairpin loop we analyzed the crystal structure of HLA-DR1 in complex with peptide hybrid 3,

PKYVQNTLLSEGATP, which diffracted to 2.2 Å. This peptide consists of all four of the residues involved in the hairpin turn in GagP16 in addition to a Leu in the $7^{th}$ position. Electron density suggest that the loop does not form within these residues. Only electron density from the first residue to the Ser at P9, which is the first residue in the observed hairpin turn, was observed. This data shows that the last 7 residues alone does not dictate the hairpin turn. All the residues that appear to be involved in the hairpin turn and C-terminal structure within the peptide are conserved but it is still not sufficient in formation of the hairpin turn (Figure 4.3). These data suggest that the sequence of the residues directly involved in the secondary structure alone does not dictate the formation of the secondary structure observed in GagP16 in the context of HLA-DR1.

PKYVKQNTLKLEGATP

PEVIPMFSALSEGATP

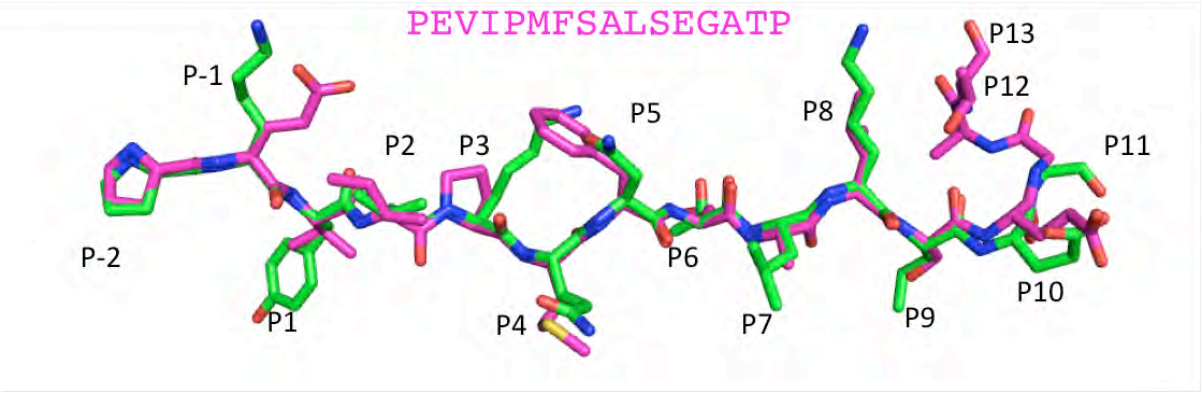
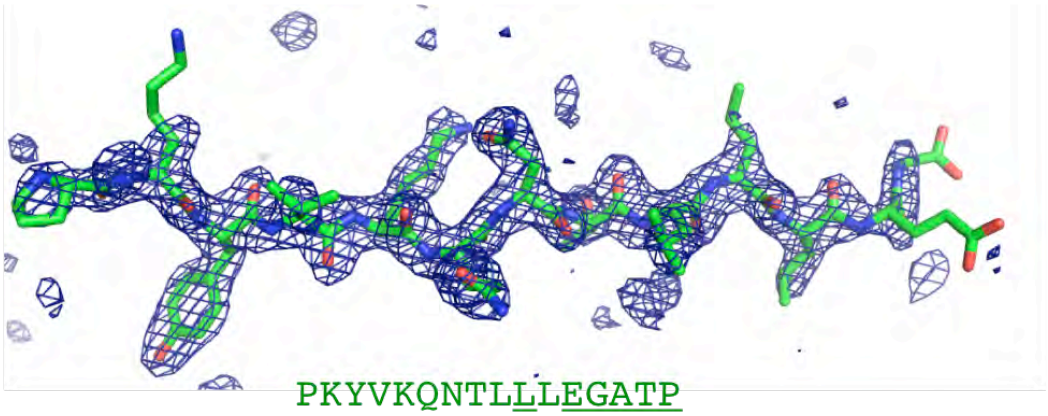


**Figure 4.1. Electron Density of Hybrid 1 Peptide in Binding Groove.**

Top. Omit map density for peptide in peptide binding groove, hybrid 1 (PKYVKQNTLKLEGATP) where underlined residues are same as those in GagP16 peptide. All atoms shown were omitted from model calculations. Bottom. Overlay of hybrid 1 peptide (carbon colored green) with GagP16 peptide (carbon colored magenta) and residues labeled according to their pocket numbers. Blue mesh represents |2Fo-Fc| electron density, sticks represent peptide model. PDB ID used is 1SJE.



PKYVKQNTLLLEGATP

PEVIPMFSALSEGATP

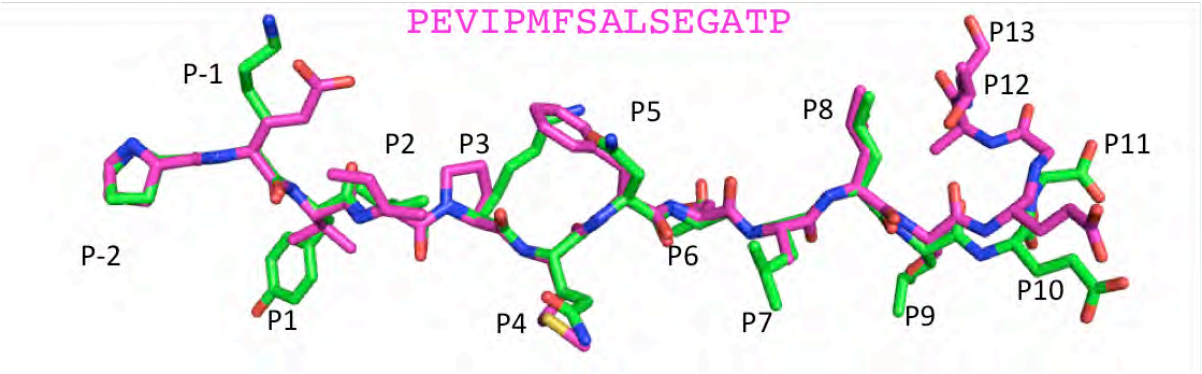
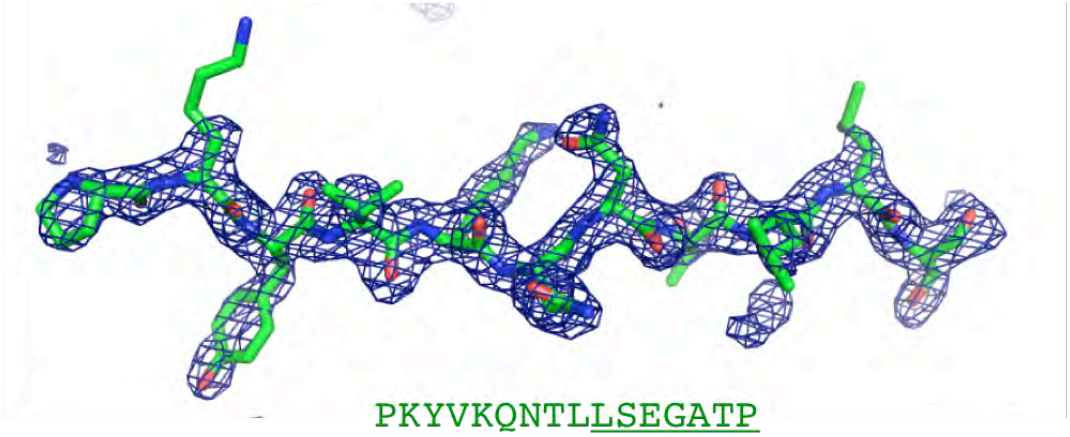


**Figure 4.2. Electron Density of Hybrid 2 Peptide in Binding Groove.**

Top. Omit map density for peptide in peptide binding groove, hybrid 2 (PKYVKQNTLLLEGATP) where underlined residues are same as those in GagP16 peptide. All atoms shown were omitted from model calculations. Bottom. Overlay of hybrid 1 peptide (carbon colored green) with GagP16 peptide (carbon colored magenta) and residues labeled according to their pocket numbers. Blue mesh represents |2Fo-Fc| electron density, sticks represent peptide model. PDB ID used is 1SJE.



PKYVKQNTLLSEGATP

PEVIPMFSALSEGATP

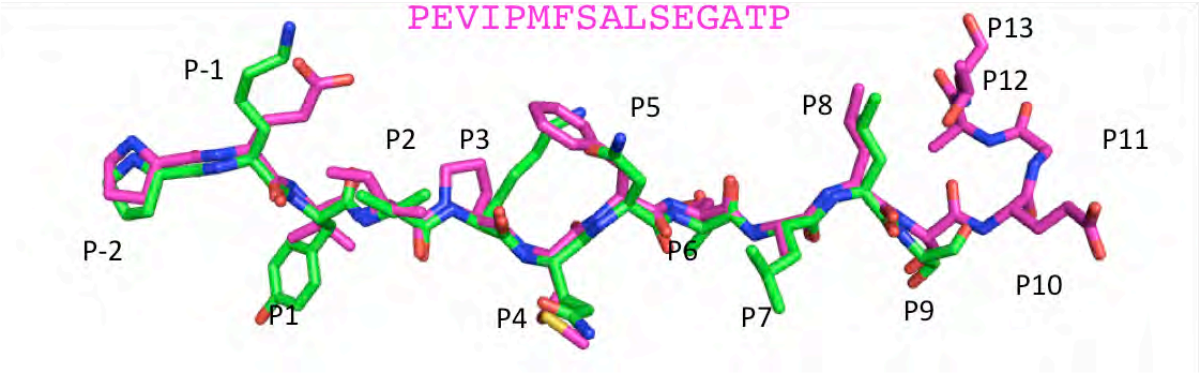


**Figure 4.3. Electron Density of Hybrid 3 Peptide in Binding Groove.**

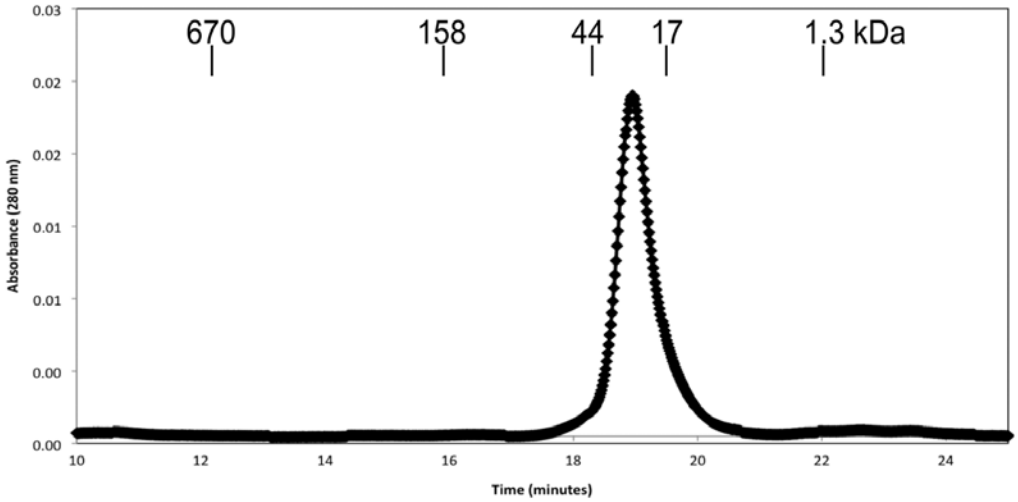
Top. Omit map density for peptide in peptide binding groove, hybrid 2 (PKYVKQNTLLSEGATP) where underlined residues are same as those in GagP16 peptide. All atoms shown were omitted from model calculations. Bottom. Overlay of hybrid 1 peptide (carbon colored green) with GagP16 peptide (carbon colored magenta) and residues labeled according to their pocket numbers. Blue mesh represents |2Fo-Fc| electron density, sticks represent peptide model. PDB ID used is 1SJE.

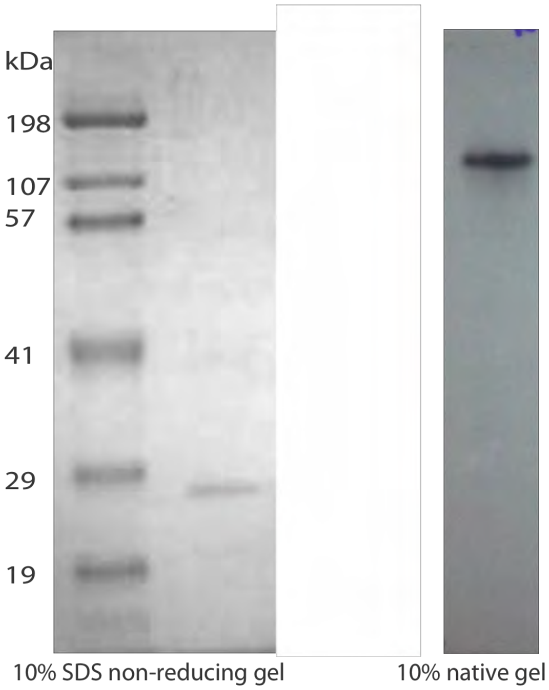
**Studies with AC25 T cell receptor**

To visualize the interaction between the TCR and HLA-DR1 in complex with GagP16, we expressed the variable domains of the AC25 TCR covalently bound by a linker, $(Gly_4Ser)_4$, termed AC25scTCR. Purified AC25scTCR eluted from size exclusion chromatography as a single monomer (Figure 4.5). Circular dichroism and thermal denaturation suggest a melting temperature at 59°C (Figure 4.6). No binding between AC25scTCR and HLA-DR1/GagP16 was detected in surface plasmon resonance (SPR) experiments (Figure 4.7). A possible reason for this result can be due to the expected weak interaction between $\alpha/\beta$ TCR and class II MHC ($K_d$ $10^{-5}$ to $10^{-7}$ M) and hence be difficult to detect (Cole DK, 2007). Binding of HLA-DR1/GagP16 tetramers to AC25 expressing T cells was observed in staining experiments (Norris et al., 2006). Since the T cell activation $ED_{50}$ values of cells expressing AC25 compared to cells expressing HA1.7 TCR are to the same order of magnitude (~1 mM), the concentration dependence for MHC tetramer staining of the two TCRs are similar, and these properties have been linked to the intrinsic MHC/TCR affinity; it is expected that the AC25TCR is likely to bind with the similar affinity as HA.1.7 TCR with $K_d$ ~$30^{-6}$ M (Cochran et al., 2000; Falta et al., 2005; Norris et al., 2004; Stone et al., 2001; Zavala-Ruiz et al., 2004b). Binding of recombinant soluble TCR and immobilized biotinylated MHC-peptide complex have been detected to $K_d$ values of $250^{-6}$ M using SPR (Huseby, E and Stern, L.J. personal communications). Another possible reason for the lack of binding observation

can be due to the clones derived from HIV patients expressing AC25 TCR. The cells can express two Vα strands and this particular Vα does not recognize HLA-DR1/GagP16 (Padovan E, 1995). We crystallized the protein solution consisting of a 1:1 molar ratio of HLA-DR1/GagP16 complex to AC25scTCR. Crystals smaller than 40 µm x 40 µm x 40 µm grew and diffracted to 5 Å. Though low resolution, the overall mode of interaction between the MHC and TCR can be visualized. Using PHASER, molecular replacement using the TCR HA1.7 in complex with HLA-DR1/HA peptide was conducted and no solution were found (Cohen et al., 2004). Matthews coefficient analysis suggest that 2 MHC/TCR complex can be in the crystal with 58% solvent (Matthews, 1968). No solutions were found when searching for both TCR and MHC as their own entity. When searching for a solution with only HLA-DR1, a solution was found with 3 molecules in the asymmetric unit with solution scoring TFZ=9.6, LLG=82, TFZ=24.9, LLG=383, RFZ=3.0 ,TFZ=21.0 ,LLG=726 and LLG=1734. The MHC molecules of the solution leaves little space for TCR packing and electron density maps yielded little or no density for regions other than in those which were modeled for HLA-DR1 (Figure 4.7). The crystals may have been of HLA-DR1/GagP16 alone.

[redacted]




10% SDS non-reducing gel 10% native gel

**Figure 4.4. AC25 Single Chain TCR Purifies as a Monomer.**

Top. Size exclusion chromatograph of AC25scTCR shows single eluted peak. Molecular weight standards are marked above curve. Bottom left. Analysis of AC25scTCR stained by coomassie blue on 10% SDS PAGE in non-reducing conditions shows a single band. Bottom right. The native gel showing AC25scTCR in a single band.

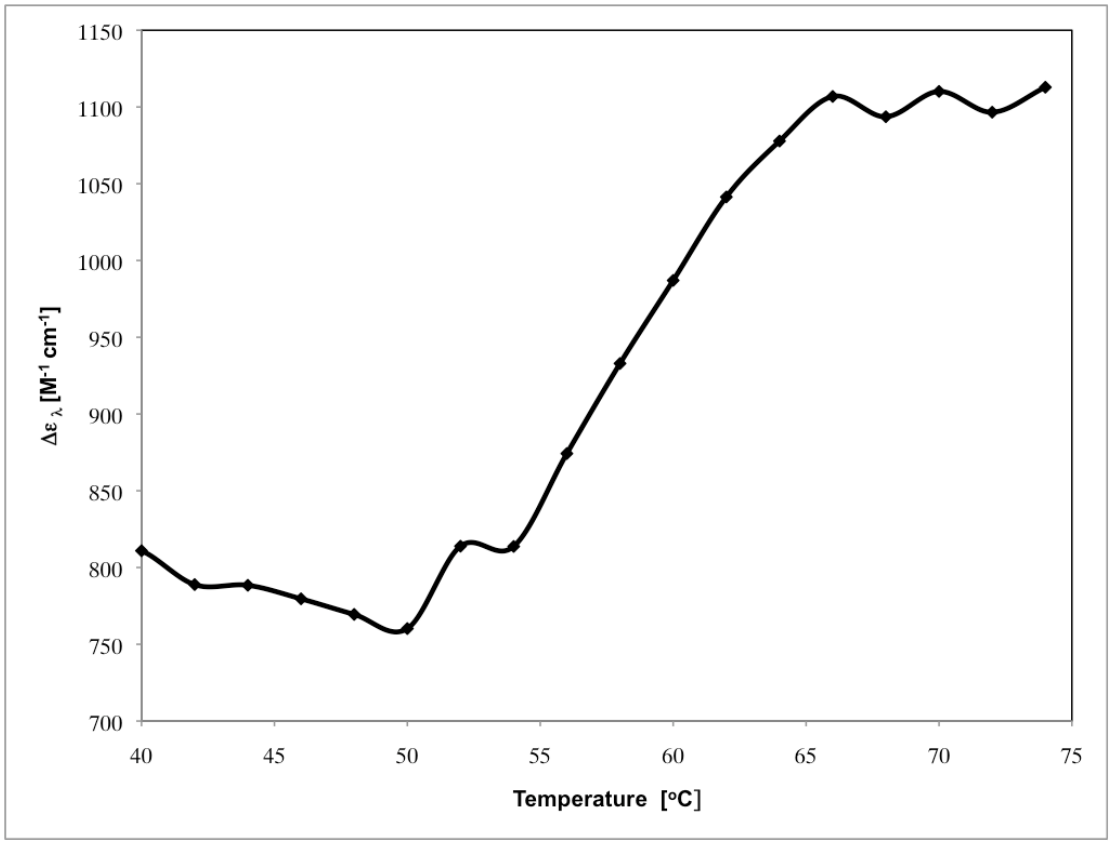



**Figure 4.5. Thermal Melt of AC25 Single Chain TCR Observed by Circular Dichroism.**

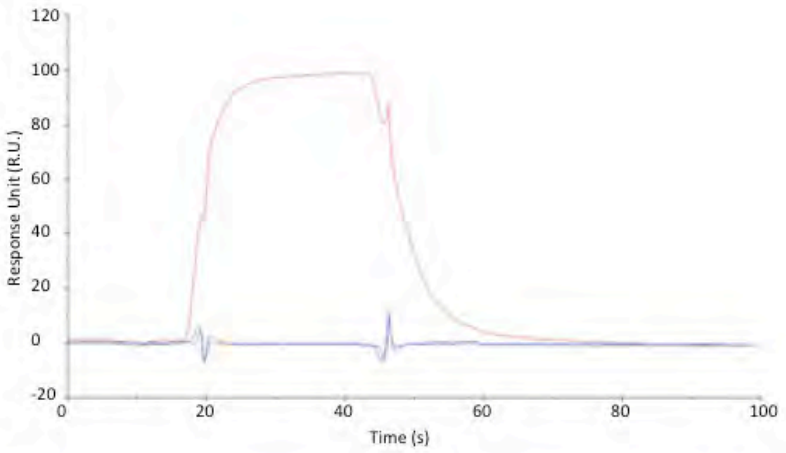
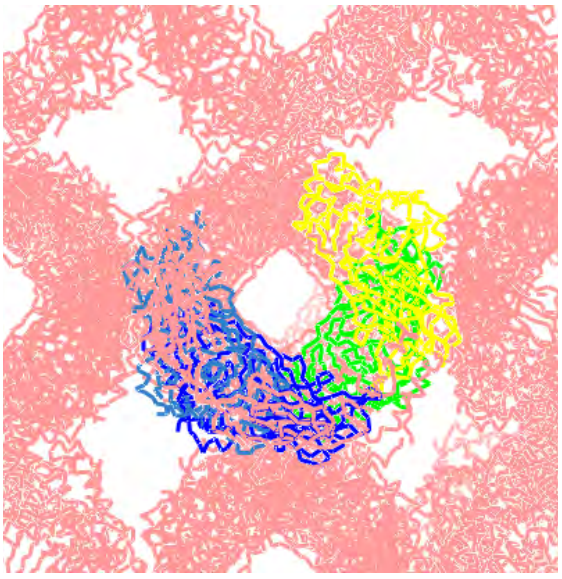

120
100
80
60
40
20
0
-20
Response Unit (R.U.)
0
20
40
60
80
100
Time (s)

**Figure 4.6. Surface Plasmon Resonance Does Not Detect Interaction Between AC25scTCR and HLA-DR1/GagP16.** Red trace indicates positive control between HLA-A2/gag peptide and D3#1 TCR. The blue trace is observation of the HLA-DR1/GagP16 peptide immobilized with AC25scTCR injected. Black line was the same as blue but with HLA-DR1/HA peptide immobilized.



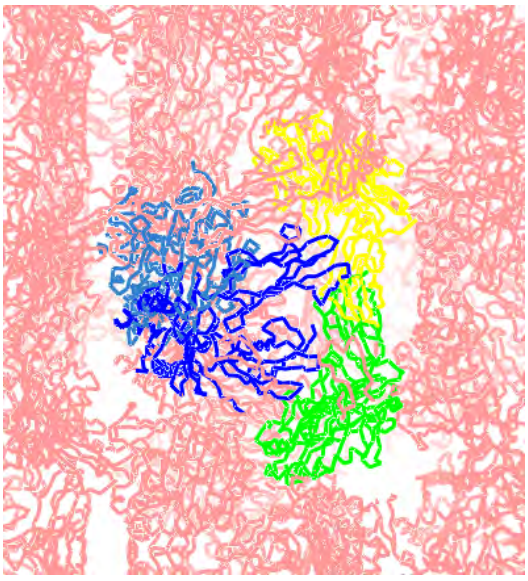


**Figure 4.7. Crystal Packing of HLA-DR1/GagP16.** The three molecules of HLA-DR1 found in the molecular solution shown in blue, green and yellow and the symmetry related molecules are in pink.

## DISCUSSION

The C-terminal loop exhibited by GagP16 in context of HLA-DR1 is also exhibited in the context of the N-terminal domain of the mature HIV-1 capsid protein. The N-terminal region of the peptide exhibits a helix in the protein and in HLA-DR1, this conformation is enforced though main chain interactions to hold a type II polyproline conformation. One possible explanation for the weak electron density for C-terminal hairpin observation is due to limitation of the X-ray data at these resolutions though the resolution of the data sets collected was comparable to the GagP16 crystal structure data, with the exception of hybrid one peptide. The lack of density at C-terminal hairpin turn when the N-terminal region was replaced by the HA peptide suggests that the binding of peptide into the peptide binding groove may promote the formation of the secondary structure outside the binding groove. Peptide alanine scanning experiments suggest that the P13 Thr is critical for T cell response but not for MHC binding. This hairpin turn may place the P13 Thr in the appropriate position for T cell activation. The hairpin formation can be determined not only by the sequence involved in the turn but also by the residues that bind into the MHC. One possible explanation for this is that upon binding into the MHC peptide binding groove anchoring residues holding the peptide down positioning the C-terminus in a conformation for hairpin turn formation. In the peptides with the HA N-terminal region, the peptide binds into the grooves differently because of different anchoring residues and does not set up the appropriate conditions for the hairpin turn to occur. For

example, the structure with the peptide consisting of only the last five residues of GagP16 exhibited a peptide in the binding groove that is less puckered than the GagP16 peptide in the binding groove. Can it be that the residues at P1, P4, P7 and P9 are holding the hybrid peptide in a way that prevents the formation of the beta hairpin turn? The P9 Leu replaces the GagP16 Ser at P9. Possibly, interactions in the pocket hold the peptide down into the groove and not allowing for the correct angle between the first and second residues of the turn to form. These data suggest that the manner in which a peptide binds into the groove and not the sequence alone dictates a T cell response.

The mode of interactions between AC25 and HLA-DR1/GagP16 was not visualized in crystallographic studies. Molecular modeling of HA1.7 TCR onto HLA-DR1/GagP16 suggests that there can be interactions between the last two residues of the peptide with the CDR1 and CDR3 loops of the Vα chain. However, the structure of GagP16 does not have the last proline residue that could prevent this kind of interaction. The crystal structure of the trimeric complex can confirm the mode of interaction and increase our library of class II MHC / TCR interaction motif. The possibility of cloning out the incorrect AC25 TCR is likely because it was not observed in the co-crystallizations and SPR binding was not detected even though expected weaker MHC/TCR affinities of other complexes were observed using SPR. In that case, cloning out another TCR from the TCR clones is necessary to further study the interaction between AC25 TCR and HLA-DR1/GagP16 hairpin loop. Cloning out another TCR from the cell

lines specific for GagP16 can be difficult due to the fragility of the cells and difficulty of revitalizing the cells for our purpose. If we do get another TCR cloned out from the AC25 cell line, one possible way to ensure MHC and TCR interaction is by tethering of the peptide to the TCR through a flexible peptide cross linker. Previous crystal structures of TCR/MHC have linked the peptide through the C-terminus of the peptide and the N-terminus of the TCR (Hennecke and Wiley, 2002). Since the interaction of interest in the C-terminus of the peptide, the linkage will need to be through the N-terminus of the peptide. The studies discussed in this chapter provide insight on the nature of determinants for the C-terminal hairpin observed so far only in the context of the GagP16 peptide in HLA-DR1 but not in other peptides that have all the components of the loop.

**Table 4.1. Data Collection and Refinement Statistics for DR1/hybrid Peptides**

| | HLA-DR1/ PKYVKQNTLKL**EGATP** | HLA-DR1/ PKYVKQNTLL**LEGATP** | HLA-DR1/ PKYVKQNTL**LSEGA** |
|---|---|---|---|
| **Data collection** | | | |
| Space group | R3 | R3 | R3 |
| Cell dimensions | | | |
| *a*, *b*, *c* (Å) | a=172.5 | a=171.3 | a= 173.1 |
| | c=121.1 | c=121.0 | c=121.3 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 50-2.8 | 35-2.2 | 35-2.2 |
| $R_{sym}$ or $R_{merge}$ | 10.5 (26) | 6.8(46) | 7.3(30) |
| $I / \sigma I$ | 6.7(2.7) | 17.2(1.3) | 14.7(2) |
| Completeness (%) | 99.8(99.9) | 100 (97.9) | 93(95.3) |
| Redundancy | 2.7(2.9) | 2.8(2.6) | 3.1(2.9) |
| **Refinement** | | | |
| Resolution (Å) | 35-2.8 | 32-2.2 | 32-2.2 |
| No. reflections | 31,133 | 66,691 | 63,949 |
| $R_{work}$ / $R_{free}$ | 0.17/0.22 | 0.17/0.20 | 0.17/0.20 |
| No. atoms | | | |
| Protein | 5041 | 5034 | 5025 |
| Water | 125 | 474 | 501 |
| *B*-factors | | | |
| Protein | 48.3 | 42.4 | 40.3 |
| Water | 50.6 | 48.8 | 47.3 |
| R.m.s. deviations | | | |
| Bond lengths(Å) | 0.008 | 0.008 | 0.008 |
| Bond angles (°) | 1.1 | 1.1 | 1.1 |

*Highest-resolution shell is shown in parentheses.

**Table 4.2. Data Collection Statistics for AC25scTCR with DR1/GagP16**

| | HLA-DR1/ GagP16 AC25scTCR |
|---|---|
| **Data collection** | |
| Space group | I4122 |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | a=b=125.4 |
| | c=440.3 |
| $\alpha, \beta, \gamma$ (°) | 90, 90, 90 |
| Resolution (Å) | 50-5 |
| $R_{sym}$ or $R_{merge}$ | 28(56) |
| $I / \sigma I$ | 5(1.6) |
| Completeness (%) | 91(70) |
| Redundancy | 6.2(5.7) |

*Highest-resolution shell is shown in parentheses.

# CHAPTER 5

## Conclusions and Future Directions

The purpose of my studies was to enhance our structural and functional understanding of the molecules involved in antigen presentation. **I**n Chapter 2, I described the structural characterization the ER aminopeptidase, ERAP1 for the first time with insights on how peptides can bind into the peptide binding cavity formed by the catalytic domain II and the helical C-terminal domain IV. In Chapter 3, the (4-DAPA)-HA peptide was characterized in the context of HLA-DR1 to show that it can be used in class II MHC antigen processing studies. Chapter 4 provided insights on the determinants of the unique C-terminal hairpin loop exhibited by an HIV gag derived peptide in complex with class II MHC.

### Mechanism for Activation of ERAP1

The structure of ERAP1 provides a first glance of this aminopeptidase and the 8$^{th}$ protein represented in the M1 family of metallopeptidases. This structure supports the idea that even though M1 aminopeptidase family members exhibit low sequence identity, they exhibit structural homology. Though the M1 aminopeptidases were categorized because of their thermolysin like fold and conserved HExxEx$_{18}$H and GXMEN motif and they appear to share structural homology at the N-terminal domain and in the C-terminal domain (Addlagatta et al., 2006; Bauvois et al., 2008; Fournie-Zaluski et al., 2009; Goettig et al., 2005;

Haeggstrom, 2000; Holland et al., 1992; Ito et al., 2006; Tholander et al., 2008; Thunnissen et al., 2001). We have identified Tyr438 as an important residue for activity suggesting that ERAP1 can be in an inactive open conformation. The differences between molecules of the asymmetric unit and normal mode analysis suggest an opening and closing motion between the catalytic domain II and the helical domain IV. In addition, another structure of ERAP1, which was deposited while our manuscript, was in preparation shown an even more closed form than our analysis. ERAP1 is the first M1 aminopeptidase to exhibit two conformations in crystal structures. Our studies support a mechanistic model where peptide binding occurs in the open conformation with interactions in the large cavity between the peptide substrate or activators and the C-terminal domain residues of ERAP1. A conformational change can then occur where the C-terminal domain comes into closer proximity with the catalytic domain and as a result of that, H5 rotates placing Tyr438 into a position for substrate hydrolysis. More analysis will be needed to determine the exact interactions between peptide substrate and the activation sites of ERAP1.

**ERAP1 and ERAP2 molecular interactions**

The work in Chapter 2 explains the N-terminal substrate preference differences in ERAP1 and ERAP2. ERAP1 and ERAP2 activity differs in substrate peptide N-terminal residues preferences. ERAP1 prefers hydrophobic residues and ERAP2 prefers basic residues in the N-terminal position (Hattori et

al., 1999; Tanioka et al., 2003). Since the two aminopeptidases are more than 50% homologous, the structure of ERAP1 provides a better model for ERAP2 than other M1 aminopeptidases. Molecular modeling of residues in the S1 site explains the difference in substrate preference in ERAP1 and ERAP2. ERAP1 has residue Gln181 in place of residue Asp198 in ERAP2. Asp198 in ERAP2 allows for charged residues such as Arg and Lys to bind into the pocket where as Gln181 in ERAP1 allows for smaller uncharged residues. Additionally, Gln 181 may play a structural role in the active site because mutation studies show that a Gln to Ala at position 181 result in loss of aminopeptidase activity (Goto et al., 2008). Both structures of ERAP1 show Gln181 is in a position to interact with the conserved GAMEN motif near the active site and can be important in holding the GAMEN loop in place for activity.

Our studies provide a possible model for ERAP1 and ERAP2 interactions. ERAP1 has been shown to interact with ERAP2 in co-immunoprecipitation and sucrose density gradient centrifugation studies (Saveanu et al., 2005). The concerted activity of the two enzymes allow for efficient processing of a peptide found in HLA-B27 HIV infected patients (Saveanu et al., 2005). The epitope precursor peptide consists of hydrophobic and positively charged N-terminal residues. About 10% of ERAP1 and 20% of ERAP2 were suggested to be in a complex; this raises the question of the importance of this interaction (Saveanu et al., 2005). The interaction between ERAP1 and ERAP2 can result in ERAP1 inhibition, activation and can be a mechanism for ER retention. Although, the

asymmetric unit of ERAP1 in the open conformation consists of three molecules, a homotrimer has not been observed in gel filtration studies or dynamic light scattering studies. The interactions between the molecules in the asymmetric unit involve residues in domains I and III. Domain III forms a hinge between the catalytic domain II and the C-terminal domain IV that makes the peptide binding cavity. Conformational changes in domain III can translate to domain II and IV opening and closing. Since ERAP1 and ERAP2 are homologous proteins with ~50% identity and more than 75% homology, the interactions between the molecules within the asymmetric unit can be used to predict the mode of interactions between ERAP1 and ERAP2. Further exploration of the interactions between ERAP1 and ERAP2 and enzymatic studies of the two enzymes can provide an understanding of the significance of the partnership.

**Conformational change in ERAP1**

In the studies described in Chapter 2, ERAP1 shows an open conformation that is likely to be inactive because Tyr438 faces away from the other catalytic residues. Does this open conformation allow for longer products to escape before further N-terminal residue cleavage? Since the open conformation was crystallized with the presence of the aminopeptidase inhibitor, bestatin, it is possible but unlikely that ERAP1 exhibits this conformation without an inhibitor. The introduction of bestatin is unlikely to force Tyr438 to point away from the active site because structures of other M1 aminopeptidases, LTA4H and pepN,

with and without bestatin have shown no changes in this residue (Addlagatta et al., 2006; Parnas et al., 1996). Further exploration of the ERAP1 conformations can explain their role in substrate binding and unique product formation. Additionally, there may be a conformation dependency for ERAP1 interaction. Understanding the role of conformational changes may lead to identification binding partners and catalytic mechanism.

**Other oxytocinase subfamily members of M1 aminopeptidase**

ERAP1 together with ERAP2 and IRAP forms the oxytocinase subfamily of aminopeptidase. ERAP1 and IRAP were shown to trim N-extended SIINFEKL to different lengths. ERAP1 trims LSIINFEKL to SIINFEKL and IRAP trims LSIINFEKL to IINFEKL and INFEKL (Georgiadou et al., 2010). The structure of IRAP in comparison with the structures of ERAP1 can provide an explanation of how ERAP1 can stop aminopeptidase activity at 8 or 9 residues long whereas IRAP can cleave down to 6 or 7 residues long. Does IRAP or ERAP2 exhibit an open and closed conformation to the same extent as ERAP1? Further exploration of closely homologous family members can provide a better understanding of the importance of conformational changes for enzymatic activity.

**Class II MHC antigen presentation studies**

The studies in Chapter 3 show that the fluorogenic peptide, (4-DAPA)-HA, binds to HLA-DR1 in the same register and with the same affinity as to native HA peptide. It also activates HA specific T cells, HA1.7, similarly as compared native HA. This environmentally sensitive peptide is suitable for use in cellular studies that can shed light on unclear aspects of class II MHC antigen processing. For instance, it remains unclear on when the interactions of HLA-DM and HLA-DO with HLA-DR1 occur. Does HLA-DO inhibit the loading of peptides assisted by HLA-DM by binding to HLA-DR1 forming a tertiary complex? How are the antigens spatially controlled in the endosomes? Further analysis of peptide loading *in vivo* can enhance our understanding of the antigen loading processes that occurs in the endosomes. Since these probes provide a direct measurement of peptide binding into MHCs and are insensitive to pH and can be used in any cell type, they can be used to study HLA-DM and HLA-DO peptide binding regulation or antigen differential compartmentalization in the endosomes. Studies of weakly to strongly binding peptide labeled with these probes can show real time compartmental regulation of peptides and the roles of HLA-DM and HLA-DO in peptides with different affinities to class II MHC.

**Class II MHC, peptide and TCR modes of interaction**

Our studies in Chapter 4 suggest that N-terminal residues of the gag derived peptide effect the formation of the C-terminal hairpin loop in the context of HLA-DR1. These studies suggest that primary sequence alone is not enough to predict T cell activation in residues flanking the peptide binding groove. Currently, there are 12 structures of class II MHC, peptide and $\alpha/\beta$ TCR complexes in the protein database. About half of the complexes display the conventional diagonal mode of binding where the TCR CDR loops predominantly interact with the center of the peptide binding groove and the residue corresponding to the P7 pocket displayed by TCR HA.1.7 to HLA-DR1/HA peptide (Hennecke and Wiley, 2002). The other complexes display a mode of binding where the TCR CDR loops display a peptide N-terminus bias binding motif as in the TCR Ob.1A12-HLA-DR2-myelin basic protein derived peptide (Maynard et al., 2005). Modeling HLA-DR1/ GagP16 into the conformations suggest that it is likely to bind either in the conventional diagonal binding conformation where additional interactions between the C-terminal residues and the V$\beta$1 and V$\beta$3 CDR loop can be achieved with the GagP16 but not with GagP13. Since binding of the N-terminal residues to HLA-DR1 affects the formation of the loop, the question arises of whether it is the hairpin loop or the residue sequence that is important for T cell recognition. Though alanine scans of the C-terminal residues were done suggesting that the residues in that region are important for T cell recognition, it does not differentiate between whether the

loop or the sequence is important for T cell activation. The trimeric crystal structure of HLA-DR1/GagP16 with AC25 TCR could reveal how this mode of interaction with the last five residues of GagP16 are important for a T cell response. It will be interesting to observe different modes of interaction between the GagP13 with TCR and GagP16 with TCR that can differentiate between T cell activation and T cell binding.

**Future Studies in Antigen Processing**

The work described in this thesis provides insights for further studies to better understand antigen processing. The structural characterization of ERAP1 provides the basis for further exploration to understand why polymorphisms observed in patients with ankylosing spondylitis can affect the antigen presentation that can affect T cell responses. The ERAP1 structure also the basis for substrate design to understand the substrate length specificity of ERAP1 that can be applied to the understanding of the mechanism of the other M1 aminopeptidase family members. Further studies on the ERAP1 peptide binding site and binding partners can help understand ERAP1's effect in conditions such as cancer, diabetes, hypertension and ankylosing spondylitis and explore the possibility of ERAP1 as a therapeutic drug target. The structural characterization of (4-DAPA)-HA in complex with HLA-DR1 provides groundwork for further studies using (4-DAPA)-HA as a reagent for studying in real time and allowing visualization of peptide binding in a cell. The studies

involving the GagP16 peptides suggested the C-terminal loop does not depend solely on the sequence and factors other than sequence can contribute to the formation of the loop and therefore eliciting a T cell response. Further studies are necessary to understand other determinants of the GagP16 C-terminal loop and how it interacts with TCR. Future studies involving ERAP1, (4-DAPA)-HA peptide, and GagP16 will provide an increase in understanding of antigen presentation, processing and autoimmunity that can lead to potential drug development and targets.

**APPENDICES**

## A.1. IRAP Crystallization and Data Collection

Crystallization trails were executed using recombinant soluble IRAP (without the transmembrane domain) using protein that was provided by the laboratory of Peter van Endert. IRAP was treated with carboxypeptidase A to cleave C-terminal ends and gel purified using Superdex S200 is PBS. Protein was concentrated to 10 mg/mL concentration in 10 mM Tris-HCl buffer, pH 7.5. Small crystal plates (100 µm x 40 µm x ~5 µm) formed after ten days of incubation at 4°C in hanging drops. Well solution consisted of 0.1 M Tris-HCl, pH 8.0, 18% PEG 8K, and 1 mM glutathione with the ratio of 10:1 oxidized to reduced form. Crystals were frozen in 30% 2-methyl-2,4-pentanediol and 1 of 10 crystals shot diffracted beyond 3 Å. X-ray diffraction data were collected as 1° oscillations at 100 K with 1 Å radiation at beamline, 14-ID-B, at the Advance Photon Source. Data were processed and scaled using HKL2000 (Otwinowski, 1997). Data collection statics is shown in table below. Matthews coefficient analysis estimates 6 molecules in the asymmetric unit (Matthews, 1968). Molecular replacement was tried with side chain truncated ERAP1 (PDB 3MDJ) as the search model using MolRep (Vagin et al., 2004). Since ERAP1 crystallized with 3 molecules in the asymmetric unit, we tried the trimer and the monomer as search models. Lock rotation function analysis suggests that it has a three fold in the asymmetric unit. Only searches with monomers yielded a solution (contrast= 3.6, Rfactor= 0.578), which did not appear to be correct because the composite

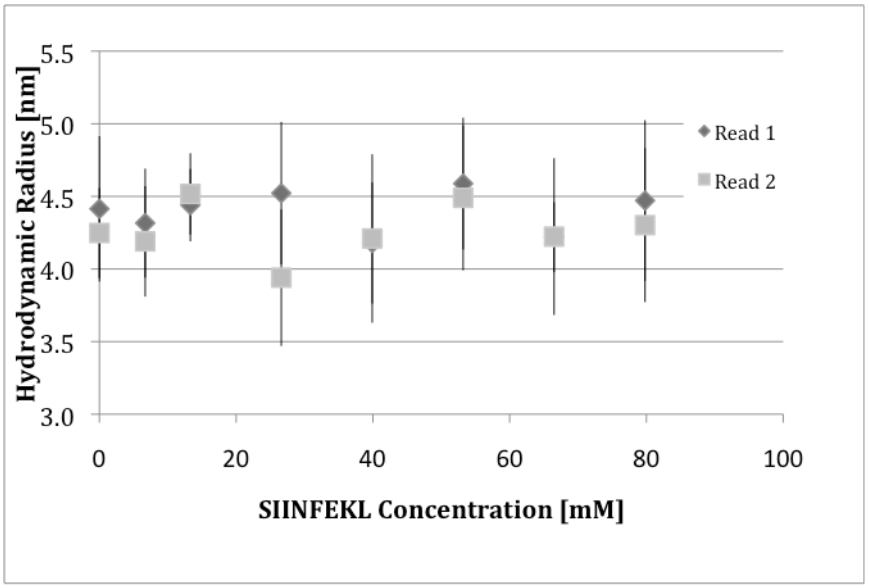
omit map calculated from this solution was not interpretable. Experimental phasing will be needed for structure determination.

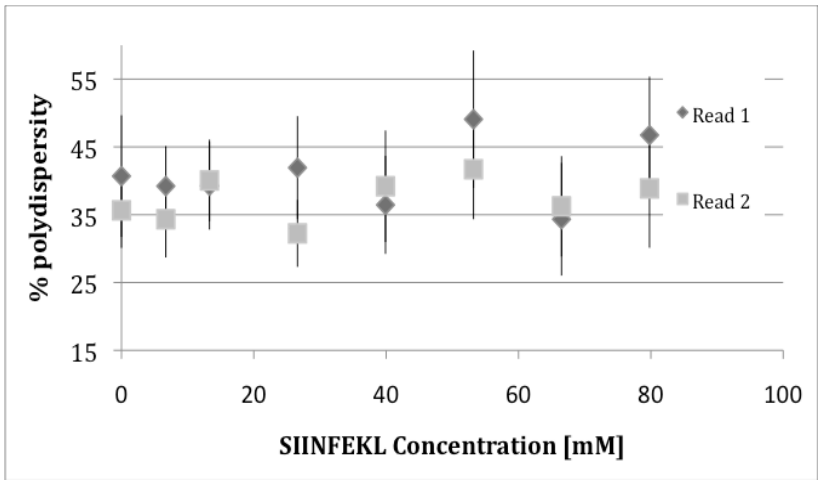
| | IRAP |
|---|---|
| **Data collection** | |
| Space group | *C2* |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 204.4, 147.2, 207.2 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 93.3, 90 |
| Resolution (Å) | 43-3 (3.2-3)* |
| $R_{sym}$ (%) | 33 (58) |
| $I / \sigma I$ | 2.7 (1.1) |
| Completeness (%) | 97 (85.8) |
| Redundancy | 1.9 (1.5) |

* Highest resolution shell

## A.2. Hydrodynamic Studies of ERAP1

To study conformational changes in ERAP1, dynamic light scattering experiments were conducted in the presence of L-AMC activator, SIINFEKL. The results suggest that the noise within the experimental set up is as much as the variation in hydrodynamic radius. There appears not to be a concentration dependence of hydrodynamic radius or sample homogeneity change according to DLS. The difference in radius of gyration between open and closed forms of ERAP1 is about 2 Å, which is less than the variation between the readings.

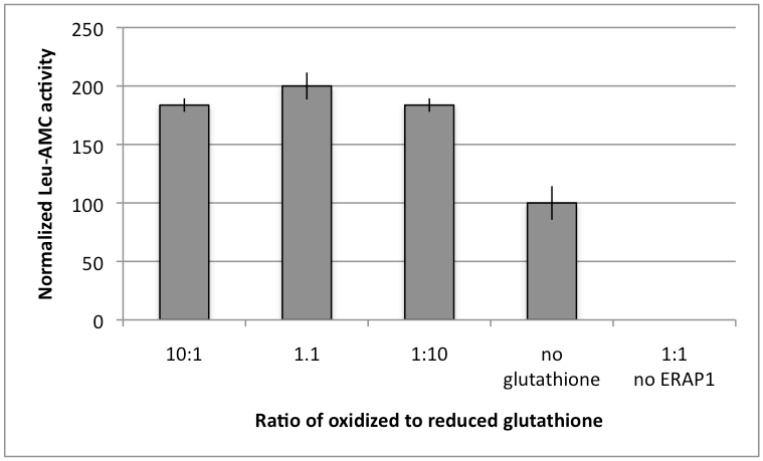


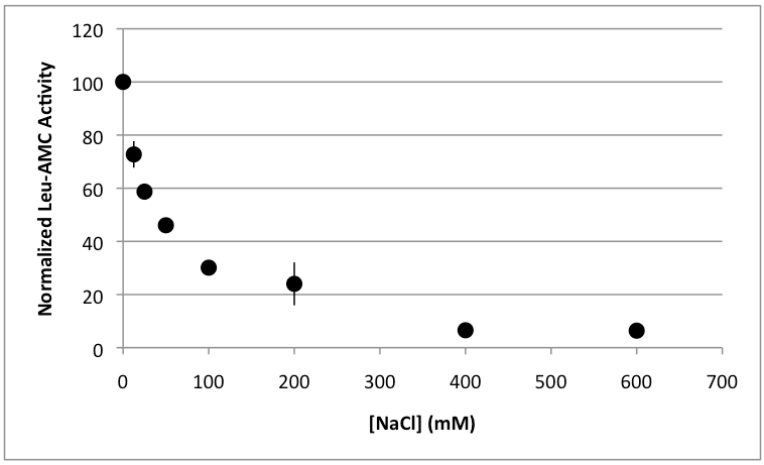

% polydispersity
55
45
35
25
15
0
20
40
60
80
100
SIINFEKL Concentration [mM]
Read 1
Read 2

### A.3. Studies of ERAP1 L-AMC Hydrolysis

Proteins from preps that have higher specific activity tend to produce ERAP1 crystals. Not all preps yielded crystals. To increase signal to noise ratio in our enzymatic assay for ERAP1, we studied the effect of glutathione and NaCl on the assay. To study the effect of reducing environment on ERAP1 L-AMC activity, ERAP1 activity with reaction buffer consisting of 10:1, 1:1, 1:10 ratio of oxidized to reduced glutathione. Total concentration of glutathione was 1 mM. Though it is not clear whether oxidizing glutathione or reducing glutathione conditions were more favorable for L-AMC hydrolysis, it was clear that 1 mM glutathione increased the activity of ERAP1.

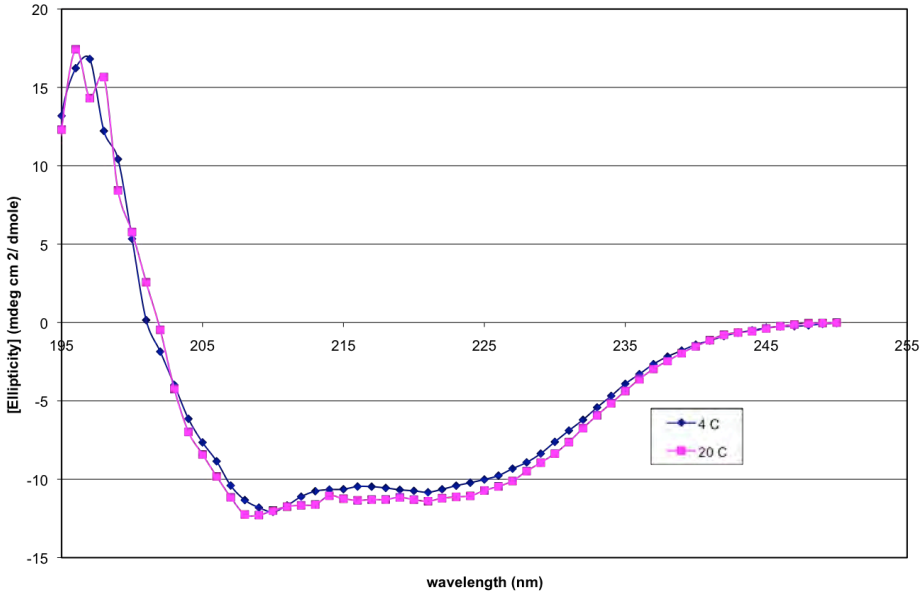
ERAP1 appears to have a concentration dependence decrease of activity by NaCl. The original reaction buffer consisted of 100 mM NaCl which is only ~30% of what the activity could have been without any NaCl. These experiment suggest that the addition of 1 mM glutathione and elimination of NaCl can increase the signal to noise of the L-AMC hydrolysis assay.


Normalized Leu-AMC activity
250
200
150
100
50
0
10:1
1.1
1:10
no glutathione
1:1 no ERAP1
Ratio of oxidized to reduced glutathione


Normalized Leu-AMC Activity
120
100
80
60
40
20
0
0
100
200
300
400
500
600
700
[NaCl] (mM)

## A.4. Circular Dichroism Analysis of ERAP1

The secondary structure of ERAP1 was examined using circular dichroism. Samples of ERAP1 at 1.5 µM in 5 mM NaPi pH 7.5 spectra was obtained with a Jasco model J810 CD spectrophotometer between 195-250 nm using a 0.2 cm CD cell and 2.5 nm bandwidth. Data suggest that there is little if any secondary structural change due to temperature differences. The 222 nm point indicative of helical content seems to have increased slightly at 20°C.



## A.5. ERAP1 Studies with Mass Spectrometry to Identify Glycosylation Sites and Disulfide Bonds

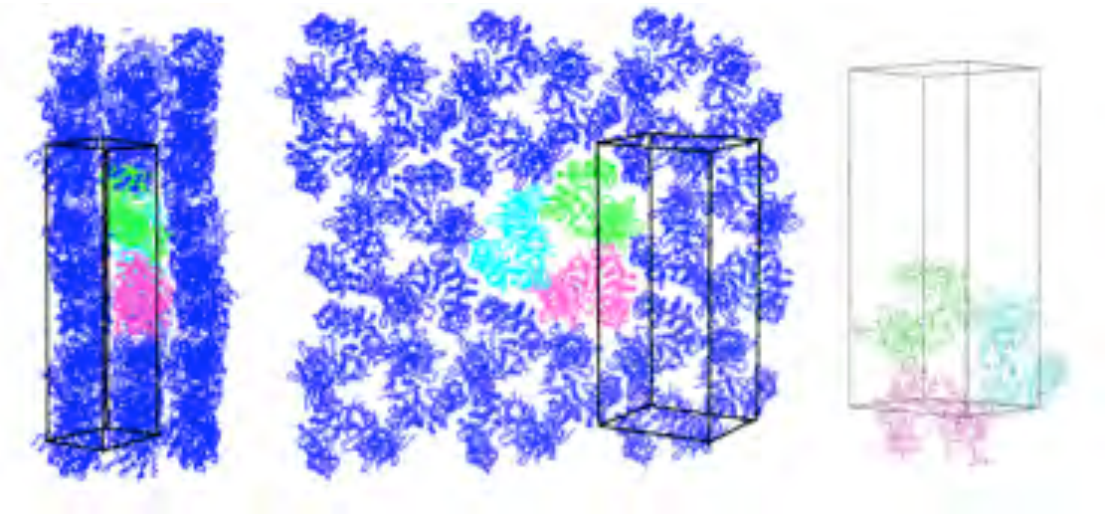
To study the glycosylation sites and disulfide bonding pattern in ERAP1, ERAP1 was submitted for mass spectrometry analysis (Proteomic Mass Spectrometry Lab, UMASS Medical School). Samples of Endo F treated and not treated ERAP1 were examined to determine glycosylation sites. To determine disulfide bonds, ERAP1 was alkylated with iodoacetamide, reduced with dithiothreitol and modified with methyl methanethiosulfonate. Samples were digested with trypsin and pepsin before mass spectrometry analysis.

```
1   MVFLPLKWSL ATMSFLLSSL LALLTVSTPS WCQSTEASPK RSDGTPFPWN KIRLPEYVIP
61  VHYDLLIHAN LTTLTFWGTT KVEITASQPT STIILHSHHL QISRATLRKG AGERLSEEPL
121 QVLEHPRQEQ IALLAPEPLL VGLPYTVVIH YAGNLSETFH GFYKSTYRTK EGELRILAST
181 QFEPTAARMA FPCFDEPAFK ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV
241 KMSTYLVAFI ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL EFYEDYFSIP
301 YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS ASSKLDITMT VAHELAHQWF
361 GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT HPELKVGDYF FGKCFDAMEV DALNSSHPVS
421 TPVENPAQIR EMFDDVSYDK GACILNMLRE YLSADAFKSG IVQYLQKHSY KNTKNEDLWD
481 SMASICPTDG VKGMDGFCSR SQHSSSSSHW HQERVDVKTM MNTWTLQRGF PLITITVRGR
541 NVHMKQEHYM KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI LPEEVEWIKF
601 NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI NNAFQLVSIG KLSIEKALDL
661 SLYLKHETEI MPVFQGLNEL IPMYKLMEKR DMNEVETQFK AFLIRLLRDL IDKQTWTDEG
721 SVSERMLRSE LLLLACVHNY QPCVQRAEGY FRKWKESNGN LSLPVDVTLA VFAVGAQSTE
781 GWDFLYSKYQ FSLSSTEKSQ IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG
841 RNPVGYPLAW QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE VKGFFSSLKE
901 NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLEH DPEADATGLE RMLESRGPFE
961 QKLISEEDLN MHTEHHHHHH
```

Sequenced highlighted in red were observed in mass spectrometry analysis. Bold residues are modified residues.

Analysis observed N-linked sugars at residues Asn 69, 154, 414, and Cys 193, 404, 486 and 806 are involved in disulfide bonds.

**A.6. ERAP1 Trimer Packing in the Unit Cell.**



## A.7. ERAP1 Expression in Hi5 cells and Purification

Modified from protocol by Efstratios Stratikos

Outline:

Express ERAP1 in Hi5 cells, concentrate and buffer exchange in concentrator, Purify with NiNTA, Concentrate for size exclusion purification, size exclusion purification, buffer exchange for ion exchange and purify in anion exchange Avoid freeze thaw process, aminopeptidase inhibitors and chelators.

ERAP1 expression:

1) Hi5 cells growth in serum free Excell 405 medium with antibiotic/antimycotic at 27°C.
2) Cells are split one day before infection.
3) Infect Hi5 cells at 0.5 million cells for 72 hrs at 27°C in shaker flask. Other flasks tend to cause cell death during infection. Test virus to cell ratio to determine optimal amount of virus to infect with.
4) Harvest by centrifugation at 4000 x g for 20 minutes. ERAP1 will be in the supernatant.
5) Concentrate supernatant and change buffer of supernatant by diluting concentrated protein by 10 fold and reconcentrate (1 L to 100 mL) in 4°C
   - Used spiral concentration system with 30 kDa MWCO
   - Exchange buffer to binding buffer
     - 50 mM Sodium Phosphate pH 8.0, 300 mM NaCl, 5 mM imidazole
6) Filter concentrated ERAP1 through 0.45 µm filter.
7) Bind supernatant to pre-equilibrated (same buffer) NiNTA resin (Qiagen) overnight in 4°C

NiNTA purification:

1) Wash column with 5 CV of starting buffer (50 mM NaP pH 8.0, 300 mM NaCl, 5 mM imidazole)
2) First elution step with 5 CV (50 mM NaP pH 8.0, 300 mM NaCl, 15 mM imidazole)
3) Second elution step with 5 CV (50 mM NaP pH 8.0, 300 mM NaCl, 30 mM imidazole)
4) Third elution step with 5 CV (50 mM NaP pH 8.0, 300 mM NaCl, 100 mM imidazole)
5) Check protein all elution and flow through for ERAP either by L-AMC or western or SDS PAGE.

For crystallography, ERAP1 is concentrated to 1 mg/mL and buffer exchanged to PBS for cleavage by carboxypeptidase A (Sigma) overnight (1:1 molar ratio) at room temperature.

Size Exclusion Chromatography:

Use Superdex S200 (GE Healthcare) column and PBS as the running buffer. Major peak is ERAP1 which elutes around 100 kDa according to standards. Buffer exchange pooled fractions by dialysis against 20 mM Tris-HCl pH 7.5 for anion exchange chromatography.

Anion Exchange Chromatography:

Using MonoQ column with starting buffer as 20 mM Tris pH 7.5 and elution buffer as 20 mM Tris-HCl pH 7.5, 1 M NaCl. The gradient is over 10 CV. There are usually multiple peaks (2 major peaks). All peaks are ERAP1 with comparable specific activity.

Average yield for this prep and purification scheme is about 0.5 mg/mL with the carboxypeptidase A treatment and 1 mg/mL without carboxypeptidase A treatment.

## A.8. Fluorometric L-AMC Assay for Aminopeptidase Activity

Modified from protocol by Efstratios Stratikos

Substrate: L-AMC (Bachem)
Stock 50mM in DMSO, store at -20°C
ERAP1 stock is 0.09 mg/mL in 10 mM Tris pH 7.5, 100 mM NaCl (stored at 4°C)
Better activity observed at pH 8 (Optimum pH for ERAP1)

Dilute substrate 500x to a final concentration of 100 µM in reaction buffer.
(20 mM Tris pH 8, 100 mM NaCl buffer)

Before reading with microplate reader, add 5 µl of ERAP1 to each well.
Add 95 µL of 100 µM Leu-AMC right before reading. (This can be adjusted in reactions in which we do not want to dilute the samples).
Measure using by exciting at 380 nm measure emissions at 460 nm.
Set gain using well where you expect highest read out.

Initial activity is the slope of the curve (of early linear part).
You can calibrate the activity using aminomethylcoumarin (AMC) as a standard (so that the activity is product released per time unit) but this is generally not needed, just compare activity between different samples for purification purposes. Use at least 10 ng of ERAP1.

**A.9. Expression of HLA-DR1 in *E. Coli* (Fermentation)**

Developed Mia Rushe

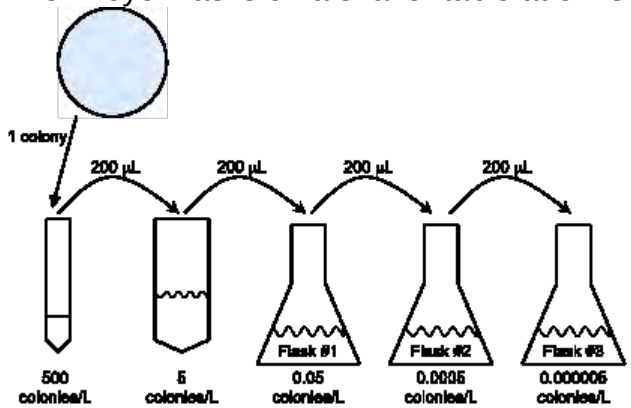
***Procedure*:**

Day 1:

1) Streak an LB-amp plate in the evening with BL21 cells of interest from frozen stocks. Incubate at 37°C overnight.

Day 2:

1) Put streaked plate at 4°C in morning, sealed with parafilm.
2) Autoclave assembled fermentor filled with 10 L of 1 x LB broth, 1 L of distilled $H_2O$, 1 L of 1 x LB broth, at least 150 mL 20% glucose, and 3-125 mL Erlenmeyer flasks.
3) In the early evening, add 83 mL sterilized LB broth to one of the sterile flasks. Add 83 mL Ampicillin (50 mg/mL) and 1.2 mL 20% glucose. **If the cells are kanamycin resistant, substitute 83 mL of a 50 mg/mL stock of kanamycin sulfate for ampicillin. If cells are $pLys^S$, add 83 mL chloramphenicol (35 mg/mL).**
4) Mix well and take a 1 mL blank sample for $OD_{600}$ measurements. You may want to add 1 mL 20% $NaN_3$ to prevent growth in the blank sample. Keep this blank in a plastic cuvette.
5) The remaining LB solution will be split into different containers. Pipet 2 mL into a sterile 15 mL conical tube, 20 mL into a 50 mL conical tube, and 20 mL into each of the remaining sterile Erlenmeyer flasks. Label the flasks 1, 2 and 3.
6) Pick one colony from your plate. Add it to your 15 mL tube and vortex vigorously. Take 200 mL of that tube and add it to the 50 mL conical tube and vortex vigorously. Take 200 mL of that tube and add it to Erlenmeyer flask #1 and swirl a lot. Take 200 mL from that flask and add it to Erlenmeyer flask #2 and swirl a lot. Take 200 mL of that flask and add it to Erlenmeyer flask #3.
7) Grow the 3 Erlenmeyer flasks on a shaker table at 37°C overnight.

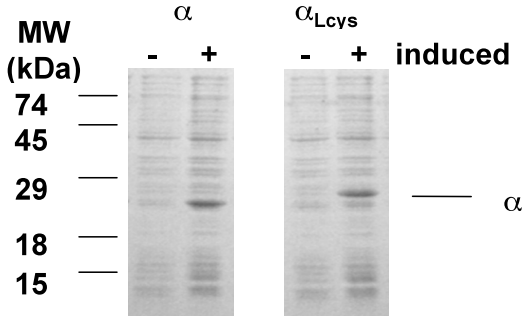


Day 3:

1) Measure the absorbance at 600nm of each flask compared to the blank you saved. Pick the one that is less than 1.0 $OD_{600}$, but still has measurable growth in it. (0.7 is about ideal.)
2) Set up the fermentor: Add 10 mL of Ampicillin, 50 mg/mL, 100 mL 20% glucose (sterile), and 200 mL antifoam A. **If the cells are kanamycin resistant, substitute 10 mL of a 50 mg/mL stock of kanamycin sulfate for ampicillin. If the cells are $pLys^S$, also add 10 mL 35 mg/mL chloramphenicol.** Mix well. Remove 1 mL as a blank and store in a plastic cuvette, possibly with 1 mL 20% $NaN_3$.
3) Run the mixer on the fermentor at ~700 rpm, and run the air at 15 psi and 5 LPM. Watch for excessive foaming and add antifoam if it is necessary. **If you have never set the fermentor up before, ask for help!!
4) Seed the fermentor with 10 mL of the overnight flask you selected.
5) Take samples and check the $OD_{600}$ every hour until you get close to 1.0, and then check more frequently. Do not overgrow!
6) When the $OD_{600}$ reaches 1.0, take a 1 mL sample "not induced." Spin down and re-suspend in 1x urea-SDS loading buffer. Freeze at -20°C.
7) Add IPTG to 0.75 mM final concentration. I use 10 mL of 0.75M IPTG stock (1000 x). Allow to grow for 2-5 hours with IPTG.
8) Take a 1 mL sample "induced." Spin down and re-suspend in 1x urea-SDS loading buffer. Freeze at -20°C until you are ready to check the "not induced" and "induced" samples on a gel by SDS-PAGE.
9) Spin down all the cells at 8000 x g and discard the supernatant. For inclusion bodies, you may freeze the pellet at -20°C, or proceed with the inclusion body prep through the DNAse step and freeze at -20°C.

**To run the 12.5% SDS-gel:**

Take the 1 mL uninduced/induced bacterial pellets and resuspend in 40 µL of water. Add 10 µL 5X Reduced SDS loading buffer with 6 M urea. Boil for 2 minutes and spin down lysed cells. Load 10 µl of both samples onto the gel. Look for an induced band around 25 kDa.



**A.10. Isolation of Class II MHC Subunits from Bacterial Pellets**

Modified protocol that was developed by Mia Rushe

**Solutions:**

**DNAse solution** (can be made and stored in freezer)

75 mM NaCl
50% glycerol
2 mg/mL DNAse (Sigma D-5025)

**Sucrose Solution** (if no DTT can be stored on bench; same for following three solutions)

50 mM Tris-HCl, pH 8.0
25% sucrose
1 mM EDTA
0.1% $NaN_3$
10 mM DTT (add just before use)

**Deoxycholate/Triton**

1% Na Deoxycholate
1% Triton x100
20 mM Tris-HCl, pH 7.5
100 mM NaCl
0.1% $NaN_3$
10 mM DTT (add just before use)

**Triton Solution**

0.5% Triton x100
50 mM Tris-HCl, pH 8.0
100 mM NaCl
1 mM EDTA
0.1% $NaN_3$
1 mM DTT (add just before use)

**Tris Solution**

50 mM Tris-HCl pH 8.0
1 mM EDTA
0.1% $NaN_3$
1 mM DTT (add just before use)

**Urea Solution** (make up fresh each time)

8 M urea
20 mM Tris-HCl, pH 8.0
0.5 mM EDTA
30 mM DTT

**4 M $MgCl_2$**

**0.5 M EDTA pH 8.0**

*Protocol for pellet of 10 liter fermentor culture

***Procedure (for pellet from 10 L culture)*:**

1. Spin down cells at 5000 x g and collect supernatant into a container, sterilize with 1% Wescodyne for 10-20 minutes, and then dump down the sink.

2. With a rubber spatula, re-suspend fresh bacteria into a single plastic container with ~200 mL Sucrose Solution.

3. Chop the solution briefly in a homogenizer or polytron. **Do not sonicate!

4. If cells are not pLys$^S$, add 1 mg dry lysozyme per mL suspension (0.2 g) and stir for 10 minutes. If cells are pLys$^S$, just stir for 10 minutes.

5. While stirring, add 500 mL Deoxycholate-Triton Solution. Solution will become very viscous due to cell lysis and DNA release.

6. Add 1 mL of 4 M $MgCl_2$ Solution to make 5 mM final concentration.

7. Add 2 mL DNAse Solution. Stir until the solution is the viscosity of water.

8. Freeze overnight at -20°C, or until ready to complete the prep.

9. Thaw solution in warm water bath.

10. Stir an additional 10 minutes after thawing to allow the DNAse to work again.

11. Spin down in 2 centrifuge bottles at 8000 x g for 20 minutes. Discard supernatant.

12. Re-suspend pellets in 300 mL or more each Triton Solution. Chop briefly, keeping the pellets on ice as much as possible. Spin down at 8000 x g for 20 minutes and discard supernatant.

13. Repeat step 12 three or more times.

14. Re-suspend pellets into 300 mL or more each Tris-HCl Solution. Chop briefly, keeping the pellets on ice as much as possible. Spin down at 8000 x g for 20 minutes and discard supernatant.

15. Repeat step 14 two or more times.

16. Re-suspend/dissolve the pellets and chop in ~200 mL Urea Solution.

17. Spin down at 20°C, 15,000 x g for 30 minutes. Filter through a 0.2 µm filter.

18. Freeze at -70°C until ready to purify by Urea HQ.

**Subunit Purification:**

Column: HiTrap Q XL column, GE Healthcare
Buffers: A: 8 M urea, 20 mM Tris, 1 mM DTT
B: 8 M urea, 20 mM Tris, 1 M NaCl, 1 mM DTT
For $\alpha$-subunits: pH 8.0
$\beta$-subunits: pH 9.0

Preparing inclusion bodies:

After urea solubilized inclusion bodies are thawed, adjust the pH of the b-subunit to 9.0. The $\beta$-subunits bound best at pH 9.0 on Applied Biosystem HQ column.

1) Equilibrate the column with 1-5 column volumes of buffer A.

2) Inject or load protein.

3) Wash with at least 2-5 column volumes of buffer A. You'll know it's enough when the absorbance stabilizes again.

4) Run a gradient from 0% B to 50% B (0-500 mM NaCl) over 10 column volumes.

5) If there is more than one major peak eluting (as often is the case for $\beta$-subunits), analyze fractions by SDS-PAGE or possibly by setting up test refoldings if you're really nervous about it. Usually all fractions of the majority peak of protein is $\beta$-subunit.

6) Pool peak fractions. For $\beta$-subunits adjust the pH back to 8.0. Add EDTA to a final concentration of 5 mM. Measure the concentration by UV absorbance scan against a urea (buffer A) blank. For $\alpha$-subunits: 1.3 $OD_{280}$ nm = 1 mg/mL, and for $\beta$-subunits: 1.7 $OD_{280}$ nm = 1 mg/mL. Store in convenient aliquots at -80°C.

Notes:
Components of the urea buffer (EDTA) seem to interfere with protein binding. Cleaning the column with 1 M NaCl/ 1 M NaOH after every couple of runs helps to regenerate the column capacity. However, column capacity becomes significantly lower with use, despite complex regeneration.

The inclusion bodies give inconsistent, messy, broad looking chromatograms, but this purification is necessary to get rid of an unknown component that affects protein folding.

**A.11. Preparation of Ab-Protein A Immunoaffinity Column**
Developed by Mia Rishe

**Materials:**
200 mM Borate: 1.24 g /100 mL, titrate to pH 9.0 with NaOH
200 mM Ethanolamine: 1.2 mL/100 mL, titrate to pH 8.0
Antibody solution, enough to have 10 mg/5 mL matrix
Protein A-coupled matrix; IPA-400 Fast-Flow.
PBS + 0.02% $NaN_3$
Dimethylpimelimidate (20 mM final, (MW=259), 5.18 mg/mL)
*Advice:* Use a fresh bottle (e.g. unopened) every time

**Procedure:**

1) Mix Antibody and Protein A matrix in PBS, preferably with as small a volume as possible (ideally ~15 mL if using 5 mL matrix; can be scaled up accordingly). This can be done in a polypropylene centrifuge tube, or directly in the gravity column body (just make sure the ends are sealed) Rotate at room temperature 1 hour or longer. Remove 15 mL of suspension into tube labeled “TOTAL”.

2) Spin 2000 rpm 5 min, remove and save supernatant and pellet. Or, if using the column, let the solution drain out. Remove 15 mL of supernatant into tube labeled “NON-BOUND”.

3) Resuspend pellet in 50 mL borate solution, spin/drain, remove supernatant.

4) Resuspend pellet in 15 mL fresh borate solution. Remove 10 mL into tube labeled “BOUND”. Add 20 mM final dimethylpimelimidate as solid (77.7 mg for 15 mL) and dissolve. Rotate 30 minutes at room temperature (going longer will over-couple the matrix)

5) Spin/drain to remove supernatant and check that the pH of the supernatant is > 8.0.

6) Resuspend pellet in 50 mL ethanolamine, rotate 30 minutes room temperature or longer, spin/drain off supernatant.

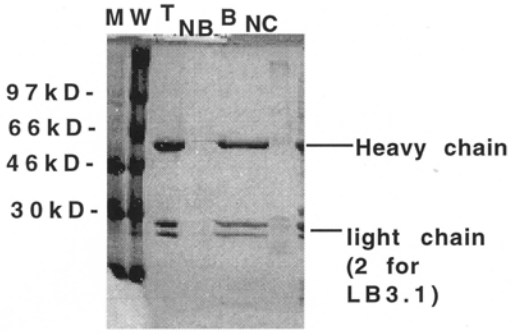
7) Wash pellet with 50 mL ethanolamine, then 50 mL borate, then 50 mL PBS

8) Resuspend pellet in 15 mL PBS. Remove 15 mL of suspension into tube labeled “NON-COUPLED”. Store matrix at 4°C in PBS + $NaN_3$.

**Analysis:**

Add 10 mL 2x Laemlli buffer to each fraction, boil, and spin in microfuge. Run 10% SDS-PAGE of MW markers, 1 mg mouse IgG, and 10 mL of each fraction taken. Gel should show heavy and light chains in TOTAL and BOUND, and nothing in NON-BOUND and NON-COUPLED. If there is more protein in NON-BOUND than BOUND, try a different binding buffer (PBS works fine for LB3.1/L243). If there is any protein in NON-COUPLED, cycle the matrix through all binding and elution conditions that will be used in the affinity purification procedure, and next time use fresh pimelimidate!

Sample of analysis gel:



**A.12. Refolding and Purification of HLA-DR1**

Developed by M. Rushe

Refolding of Subunits into Empty DR1:

Refolding Mix:

25% glycerol
20 mM Tris-HCl, pH 8.5
1 mM EDTA
2 mM glutathione, reduced form
0.2 mM glutathione, oxidized form

(glutathione should be added just before refolding)

Anion exchange purified alpha and beta subunits solubilized in 8 M urea

1) Chill refolding mix to 4°C before adding glutathione. Refolding should be performed in a container close to the volume of refolding to minimize exposure to air. If desired, sparge the folding mix with argon for about 20 minutes.

2) Add glutathione and stir until just dissolved with a magnetic stirrer (about 30-45 minutes).

3) While rapidly stirring the folding mix, very slowly add purified subunits with a pipet. Add 2 mg of each subunit per 1 liter of refolding. If refolding peptide-loaded DR1, add 4 mg of each subunit and 5 M excess of peptide to the refolding mix assuming that 1 mg/L is the concentration of protein.

4) After all components are well mixed, cover tightly and store at 4°C for at least 36 hours. Week long refolding incubation tends to yield more complexes.

**Immunoaffinity Purification of Empty Class II MHC from *in vitro* Folding Reactions:**

1) Concentrate folding mixture in the Amicon spiral filter concentrator 10 fold. Save the flow-through in case the filter membrane breaks. If you routinely check the recovery with an ELISA, you will see if this happened by the presence of DR1 in the sample of flow-through you test.

2) Exchange the folding buffer by filling the concentrator vessel (5 fold dilution) with PBS and concentrating. Repeat this step two more times. It is important to get rid of the glutathione in the folding mix or it will trash the immunoaffinity column!

3) Flow the concentrated folding mix over the immunoaffinity column by gravity. The column is composed of the anti-MHC class II antibodies LB3.1 covalently coupled to Protein A.

4) Wash with at least 10 column volumes of PBS.

5) Pre-elute the column with at least two column volume 10 mM NaPi pH 7.

6) There are two ways to elute. Traditionally, the high pH method is used to elute class II from an immunoaffinity column.

High pH: Elute the protein with 50 mM CAPS pH 11.5. Collect 1 mL fractions into tubes containing 300 µl of 300 mM NaPi pH 6 (to neutralize). About 5 column volumes should be sufficient. (Note: The lab experience is that the high pH method seems to yield more protein)

Low pH: Elute the protein with 50 mM glycine pH 3. Collect 1 mL fractions into tubes containing 300 µl of 300 mM Tris-HCl pH 8 (to neutralize). About 5 column volumes should be sufficient.

Prepare a “blank” (300 µl of 300 mM NaPi pH 6 plus 1 mL 50 mM CAPS pH 11.5, for example) to zero the spectrophotometer and also to check that the pH of the combined buffers is neutral.

7) Neutralize the column with 300 mM NaPi pH 6 (if high pH elution was used), or 300 mM Tris pH 8 (if low pH elution was used).

8) Wash column with at least 10 column volumes of PBS and 0.02% $NaN_3$ and store in refrigerator.

9) Since the capacity of an immunoaffinity column varies greatly, you should repeat steps 1-7 with the “flow-through” until no more protein elutes from the column.

10) To clean the column after you are finished:
5 CV 50 mM CAPS pH 11.5
5 CV 50 mM glycine pH 3
5 CV 50 mM CAPS pH 11.5
5 CV 50 mM glycine pH 3
10-20 CV PBS/0.02% $NaN_3$

11) Pool the fractions containing DR1 based on their absorbance at 280 nm, concentrate and switch into a desired buffer (usually PBS). To avoid aggregation, store empty DR1 below 1 mg/mL.

12) Check the prep for peptide binding activity:

Incubate 5-10 µg DR1 with and without 5-molar excess HA peptide at 37°C for at least 48 hours.

Example:
DR1 stock 1 mg/mL

| +Ha sample: | no peptide sample: |
|---|---|
| 5 µl DR1 | 5 µl DR1 |
| 2 µl 0.25 mM HA peptide | |
| 13 µl PBS/0.02% $NaN_3$ | 15 µl PBS/0.02% $NaN_3$ |

final DR1 concentration: 0.25 µg/mL = 5 µM
final peptide concentration: 25 µM

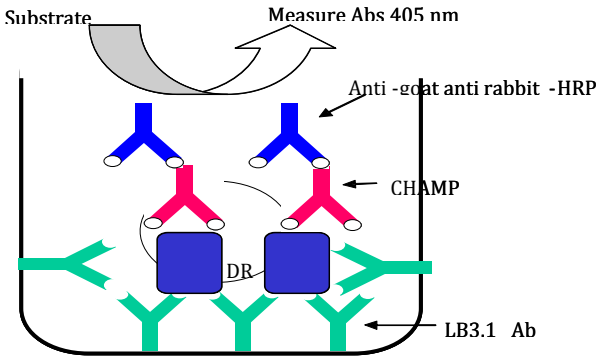
Analyze by 12.5% SDS-PAGE: Run a boiled / not-boiled version of each sample
Example:
Add 5 µl of 5X reducing SDS sample buffer to each tube. Load 12 µl into “non-boiled” lane.
Boil remaining 13 µl for 5 minutes and pulse in microfuge. Load 12 µl into “boiled” lane.

Ha-loaded class II should give a ~45 kDa band on an SDS gel that dissociates into $\alpha$ and $\beta$ subunits when boiled. “Empty” class II is not SDS stable and will dissociate into $\alpha$ and $\beta$ subunits, for the boiled and not-boiled samples. You can further purify the immunoaffinity eluted empty class II MHC by size exclusion chromatography to obtain monomeric DR1. There will be a lot of aggregated protein that elutes at a large molecular weight, but you should also see a peak at

~45 kDa corresponding to a class II $\alpha\beta$ heterodimer. The oligomeric distribution varies from prep-to-prep.

## A.13. HLA-DR1 ELISA

Protocol from Mia Rushe




***Materials:***

1. Block Solution: 3% BSA in PBSZ
2. Dilution Solution: 0.3% BSA
   0.1% Triton X-100
   PBS (filter and store at 4°C)
3. TBST: 25 mM Tris-HCl
   137 mM NaCl
   2.7 mM KCl
   0.05% Triton X-100
4. Alternatively you can PBST instead of TBST: PBS with 0.05% Triton X-100.
5. ABTS solution: 5 mL 10X ABTS buffer (from Roche-made and frozen at -20°C in 5 mL aliquots)
   45 mL dd$H_2O$
   1 ABTS tablet (from Roche also)
   *Unused portion can be stored in dark at 4°C for 2 weeks.
   Or you can use the ready-to-use-ABTS solution that Roche sells!!!

***Protocol:***

1) Coat 96-well plate with anti-DR monoclonal antibody LB3.1 (2 mg/mL stock in PBS + 0.02% $NaN_3$). Dilute 1:1000 in PBSZ and put 100 µL in each well. Incubate at 37°C for 2 hours or 4°C overnight.

2) Block plates with block solution for at least 4 hours at room temperature. (It is okay to store the plates in the cold room at this stage and use them later, they are good for about 3 weeks.)

3) Put 50 µL of dilution solution in every well (It takes awhile to pipet samples and standards into wells to prevent them from drying out.)

4) Standards: Do 2-fold dilutions in *duplicate* starting with 100 ng/ well: 100 ng, 50, 25 , 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, 0.19, 0.1. Dilute samples into dilution solution. (0.3% BSA, 0.1% Triton X-100, in PBS)

5) Samples: Put appropriate amount of sample on. For tissue culture sup, put 50 and 5 µL (add and additional 45 µL tot eh wells that only 5µL of sample, so that you have 100 µL total in each well).

6) Incubate standards and samples at 37°C for 1 hour or at room temperature for 2 hours or overnight at 4°C.

7) Wash 3 times with PBST or TBST. (Be careful when washing the first 2 times, you have different samples in each well and you don't want the wells to spill into one another.)

8) Add 100 µL/well of rabbit CHAMP diluted 1:25,000 in dilution solution. Incubate for 1.5 hours at room temperature or 1 hour at 37°C or overnight at 4°C.

9) Wash wells 3 times with PBST or TBST.

10) Add 100 µL/ well of goat anti-rabbit IgG-peroxidase diluted 1:4,000 in dilution solution. It is very important that no azide is present in this step, it inhibits the HRP. Incubate 0.5 to 2 hours at room temperature. (Background increases with if this step is left at 4°C overnight.)

11) Wash wells 3 times with PBST or TBST.

12) Add 200 µL of ABTS/ well and read absorbance at 405 nm.

13) Work up data.

**ELISA fit-four**

equation used to fit data:
line 1: ((m1-m2)/(1+ (m0/m3)^m4)) + m2;
inital guesses: m1=0.7; m2=0; m3=2; m4=-0.5
partial derivatives:
line 2: 1/(1+(m0/m3)^m4)
line 3: 1-(1/(1+(m0/m3)^m4))
line 4: (m1-m2)*(1+(m0/m3)^m4)^(-2)*(m4*(m0/m3)^(m4-1))*m0/m3^2
line 5: (m2-m1)*(1+(m0/m3)^m4)^(-2)*(m0/m3)^m4*ln(m0/m3)

**inverse fit-four**

c1=m3*((m1-m2)/(c0-m2)-1)^(1/m4)
an alternate equation for fitting ELISA data

## A.14. Expression and Purification of the Superantigen, SEC3-3B2

Protocol by Zarixia Zavala-Ruiz

**Expression:**

1) Several dilutions of 15 mL culture were grown overnight (2xYT media, 100 µg/mL ampicillin, 2% glucose) at 37°C.
2) Fifteen mL were used to inoculate two 2 L flask (each containing 500 mL of 2xYT media and 100 µg/mL ampicillin). Cells were grown for about 2-4 hours at 37°C until $OD_{600}$ was about 1 (you can induce when $OD_{600}$ is anywhere between 0.8-1.2)
3) Cells were induced for about 5hrs (induction can be between 4-6 hrs).
4) Cells were spin down at 5000 rpm for 15 minutes.
5) The cells were resuspended in 10 mLs of ice-cold TES buffer (100 mLs of TES buffer: 3.15 g of Tris-HCl, 0.186 g of EDTA, 17.1 g of sucrose, pH 8.0). Suspension was left at 4°C overnight. Alternatively, it can be incubated on ice for 1 hour.
6) Next day, 15 mL of ice-cold 0.2x TES buffer were added to the suspension and left on ice for 45 minutes (should not be left on ice for more than 1 hr).
7) Cells were spin down for 10 minutes at 7000 rpm.
8) The supernatant was transferred to a reusable tube and it was spin for 20 minutes at 12000 rpm.
9) The yellowish liquid was dialyzed against 2 L of 20 mM Tris-HCl pH 7.5 for ~8 hrs at 4°C. Fresh buffer was used to dialyze overnight.

**Red-A purification:**

1) Approximately 6 mL of Red-A agarose were transferred to a low-pressure column.
2) It was washed with 2 CV of 0.5M NaOH, 6 M Urea to remove excess dye (this step is only done when the resin is new).
3) Column was equilibrated using 10 CV of 20 mM Tris-HCl pH 7.5.
4) Protein was added slowly, and the flow through was collected.

5) Column was washed with 50 mL of 20 mM Tris-HCl pH 7.5.
6) The protein was eluted with 25 mL of 20 mM Tris-HCl, 0.5 M NaCl pH 7.5. (Fractions of 5 mL were collected)
7) The column was regenerated using 25 mL of 0.5 M NaOH, 6 M Urea.
8) Column was stored in water + 0.02% azide.
9) Fractions were checked on an SDS gel.
10) Protein buffer was exchanged to 20 mM Tris-HCl pH 8.5 for ion-exchange purification.

**Ion-exchange-purification:**

1) Mono Q purification was performed using 20 mM Tris-HCl pH 8.5 as buffer and eluting with 1 M NaCl.
2) The protein comes out at about 30% NaCl. Fractions are checked, and the buffer is exchange to 10 mM Tris-HCl pH 7.5 for crystallization.

Note: Using HQ, the protein will come out 85-90% pure, with Mono-Q is 95% pure. Purity is important for crystallographic purposes.

## A.15. Expression and Purification of the T cell Receptor, AC25scTCR

**Expression:**

10) Several dilutions of 15 mL culture were grown overnight (2xYT media, 100 µg/mL Chloramphenical, 2% glucose) at 27°C.
11) Fifteen mL were used to inoculate two 2 L flask (each containing 500 mL of 2xYT media and 100 µg/mL Chloramphenical). Cells were grown for about 2-4 hours at 27°C until $OD_{600}$ was about 1 (you can induced when $OD_{600}$ is anywhere between 0.8-1.2)
12) Cells were induced for about 5 hrs (induction can be between 4-6 hrs).
13) Cells were spin down at 5000 rpm for 15 minutes.
14) The cells were resuspended in 10 mL of ice-cold TES buffer (100 mL of TES buffer: 3.15 g of Tris-HCl, 0.186 g of EDTA, 17.1 g of sucrose, pH 8.0). Suspension was left at 4°C overnight. Alternatively, it can be incubated on ice for 1 hour.
15) Next day, 15 mL of ice-cold 0.2x TES buffer were added to the suspension and left on ice for 45 minutes (should not be left on ice for more than 1 hr).
16) Cells were spin down for 10 minutes at 7000 rpm.
17) The supernatant was transferred to a reusable tube and it was spin for 20 minutes at 12000 rpm.
18) The yellowish liquid was dialyzed against 2 L of 20 mM Tris-HCl pH7.5 for ~8 hrs at 4°C. Fresh buffer was used to dialyze overnight.

**Purification of AC25scTCR for Crystallization with HIS tag**

NiNTA purification:

1. Equilibrate NiNTA (Qiagen) with 3-5 CV 20mM Tris-HCl pH 7.5
2. Bind equilibrated AC25scTCR sample onto column. Save flow through.
3. Wash with 1-2 CV 20 mM Tris-HCl pH 7.5. Collect wash.
4. Elute with 5 CV 20 mM Tris-HCl pH 7.5 with 10 mM imidazole.

5. Elute with 5 CV 20 mM Tris-HCl pH 7.5 with 30 mM imidazole.
6. Elute with 5 CV 20mM Tris-HCl pH 7.5 with 100 mM imidazole.
7. Elute with 2 CV 20mM Tris-HCl pH7.5 with 300 mM imidazole.
8. Run SDS PAGE gel of wash, flow through and fractions.
9. Pool fractions and concentrate for gel filtration.

Gel Filtration can be done with Superdex S200 in PBS

1. Filter samples before injection and be sure that all solutions are degassed and filtered.
2. Run SDS PAGE to determine which fractions to pool
3. Pool fractions, buffer exchange to 10 mM Tris-HCl pH 7.5 and concentrate to at least 10 mg/mL for crystallization.

## A.16. ERAP1 Protein Construct

Sequence

```
1   MVFLPLKWSL ATMSFLLSSL LALLTVSTPS WCQSTEASPK RSDGTPFPWN KIRLPEYVIP
61  VHYDLLIHAN LTTLTFWGTT KVEITASQPT STIILHSHHL QISRATLRKG AGERLSEEPL
121 QVLEHPRQEQ IALLAPEPLL VGLPYTVVIH YAGNLSETFH GFYKSTYRTK EGELRILAST
181 QFEPTAARMA FPCFDEPAFK ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV
241 KMSTYLVAFI ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL EFYEDYFSIP
301 YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS ASSKLDITMT VAHELAHQWF
361 GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT HPELKVGDYF FGKCFDAMEV DALNSSHPVS
421 TPVENPAQIR EMFDDVSYDK GACILNMLRE YLSADAFKSG IVQYLQKHSY KNTKNEDLWD
481 SMASICPTDG VKGMDGFCSR SQHSSSSSHW HQERVDVKTM MNTWTLQRGF PLITITVRGR
541 NVHMKQEHYM KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI LPEEVEWIKF
601 NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI NNAFQLVSIG KLSIEKALDL
661 SLYLKHETEI MPVFQGLNEL IPMYKLMEKR DMNEVETQFK AFLIRLLRDL IDKQTWTDEG
721 SVSERMLRSE LLLLACVHNY QPCVQRAEGY FRKWKESNGN LSLPVDVTLA VFAVGAQSTE
781 GWDFLYSKYQ FSLSSTEKSQ IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG
841 RNPVGYPLAW QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE VKGFFSSLKE
901 NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLEH DPEADATGLE RMLESRGPFE
961 QKLISEEDLN MHTEHHHHHH
```

Expected Molecular Weight: 111964.8 Daltons

Expected Extinction Coefficient at 280 nm: 171,200 $M^{-1}$ $cm^{-1}$

Protein pI: 5.84

## A.17. HLA-DR1 Protein Construct

Sequence

Alpha-Chain

```
1   MIQEEFKMIK EEHVIIQAEF YLNPDQSGEF MFDFDGDEIF HVDMAKKETV WRLEEFGRFA
61  SFEAQGALAN IAVDKANLEI MTKRSNYTPI TNVPPEVTVL TNSPVELREP NVLICFIDKF
121 TPPVVNVTWL RNGKPVTTGV SETVFLPRED HLFRKFHYLP FLPSTEDVYD CRVEHWGLDE
181 PLLKHWEFDA END
```

Beta-Chain

```
1   MIQEEFKMGD TRPRFLWQLK FECHFFNGTE RVRLLERCIY NQEESVRFDS DVGEYRAVTE
61  LGRPDAEYWN SQKDLLEQRR AAVDTYCRHN YGVGESFTVQ RRVEPKVTVY PSKTQPLQHH
121 NLLVCSVSGF YPGSIEVRWF RNGQEEKAGV VSTGLIQNGD WTFQTLVMLE TVPRSGEVYT
181 CQVEHPSVTS PLTVEWRAEN D
```

Expected Molecular Weight:
Alpha-Chain: 22,528.5 Daltons
Beta-Chain : 23,449.2 Daltons

Expected Extinction Coefficient at 280 nm:
Alpha-Chain: 28,085 $M^{-1}$ $cm^{-1}$
Beta-Chain: 39,420 $M^{-1}$ $cm^{-1}$

Protein pI:
Alpha-Chain: 4.67
Beta-Chain: 5.47

## A.18. Superantigen, SEC3-3b2, Protein Construct

Sequence

```
1   MESQPDPMPD DLHKSSEFTG TMGNMKYLYD DHYVSATKVK SVDSFFKWDL IYNISDKKLK
61  NYDKVKTELL NEDLAKKYKD EVVDVYGSNY YVNCYFSSKD NVGKVTGGKT CMYGGITKHE
121 GNHFDNGNLQ NVLVRVYENK RNTISFEVQT DKKSVTAQEL DIKARNFLIN KKNLYEFNSS
181 PYETGYIKFI ENNGNTFWYD MMPAPGDKFD QSKYLMMYND NKTVDSKSVK IEVHLTTKNG
```

Expected Molecular Weight: 27,814.2 Daltons

Expected Extinction Coefficient at 280 nm: 37945 $M^{-1}$ $cm^{-1}$

Protein pI: 6.53

## A.19. AC25 single chain T cell receptor Protein Construct

Sequence

```
1   MKYRGQILLP TAAAGLLLLA AQPAMADSVT QLDSQVPVFE EAPVELRCNY SSSVSVYLFW
61  YVQYPNQGLQ LLLKYLSGST LVKGINGFEA EFNKSQTSFH LRKPSVHISD TAEYFCAVSP
121 RSNTGKLIFG QGTTLQVKPG GGGSGGGGSG GGGSGGGGSM GVTQTPKHLI TATGQRVTLR
181 CSPRSGDLSV YWYQQSLDQG LQFLIQYYNG EERAKGNILE RFSAQQFPDL HSELNLSSLE
241 LGDSALYFCA SSPRTGAGSD TQYFGPGTRL TVLAASGADH HHHHH
```

Expected Molecular Weight: 30485.1 Daltons

Expected Extinction Coefficient at 280 nm: 30,370 $M^{-1}$ $cm^{-1}$

Protein pI: 7.28

## REFERENCES


Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Addlagatta, A., Gay, L., and Matthews, B.W. (2006). Structure of aminopeptidase N from Escherichia coli suggests a compartmentalized, gated active site. Proc Natl Acad Sci U S A *103*, 13339-13344.

Aki, M., Shimbara, N., Takashina, M., Akiyama, K., Kagawa, S., Tamura, T., Tanahashi, N., Yoshimura, T., Tanaka, K., and Ichihara, A. (1994). Interferon-gamma induces different subunit organizations and functional diversity of proteasomes. J Biochem *115*, 257-269.

Alfonso, C., Liljedahl, M., Winqvist, O., Surh, C.D., Peterson, P.A., Fung-Leung, W.P., and Karlsson, L. (1999). The role of H2-O and HLA-DO in major histocompatibility complex class II-restricted antigen processing and presentation. Immunol Rev *172*, 255-266.

Andersen, P.S., Lavoie, P.M., Sekaly, R.P., Churchill, H., Kranz, D.M., Schlievert, P.M., Karjalainen, K., and Mariuzza, R.A. (1999). Role of the T cell receptor alpha chain in stabilizing TCR-superantigen-MHC class II complexes. Immunity *10*, 473-483.

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., and Bork, P. (2001). Comparison of ARM and HEAT protein repeats. J Mol Biol *309*, 1-18.

Azimzadeh, O., Sow, C., Geze, M., Nyalwidhe, J., and Florent, I. (2010). Plasmodium falciparum PfA-M1 aminopeptidase is trafficked via the parasitophorous vacuole and marginally delivered to the food vacuole. Malar J *9*, 189.

Bacik, I., Cox, J.H., Anderson, R., Yewdell, J.W., and Bennink, J.R. (1994). TAP (transporter associated with antigen processing)-independent presentation of endogenously synthesized peptides is enhanced by endoplasmic reticulum insertion sequences located at the amino- but not carboxyl-terminus of the peptide. J Immunol *152*, 381-387.

Bauvois, C., Jacquamet, L., Huston, A.L., Borel, F., Feller, G., and Ferrer, J.L. (2008). Crystal structure of the cold-active aminopeptidase from Colwellia psychrerythraea, a close structural homologue of the human bifunctional leukotriene A4 hydrolase. J Biol Chem *283*, 23315-23325.

Beninga, J., Rock, K.L., and Goldberg, A.L. (1998). Interferon-gamma can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase. J Biol Chem *273*, 18734-18742.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. Nature *329*, 506-512.

Blanchard, N., Gonzalez, F., Schaeffer, M., Joncker, N.T., Cheng, T., Shastri, A.J., Robey, E.A., and Shastri, N. (2008). Immunodominant, protective response to the parasite Toxoplasma gondii requires antigen processing in the endoplasmic reticulum. Nat Immunol *9*, 937-944.

Blanchard, N., Kanaseki, T., Escobar, H., Delebecque, F., Nagarajan, N.A., Reyes-Vargas, E., Crockett, D.K., Raulet, D.H., Delgado, J.C., and Shastri, N. (2010). Endoplasmic reticulum aminopeptidase associated with antigen processing defines the composition and structure of MHC class I Peptide repertoire in normal and virus-infected cells. J Immunol *184*, 3033-3042.

Boulanger, D.S., Oliveira, R., Ayers, L., Prior, S.H., James, E., Williams, A.P., and Elliott, T. (2010). Absence of tapasin alters immunodominance against a lymphocytic choriomeningitis virus polytope. J Immunol *184*, 73-83.

Braciale, T.J., Braciale, V.L., Winkler, M., Stroynowski, I., Hood, L., Sambrook, J., and Gething, M.J. (1987a). On the role of the transmembrane anchor sequence of influenza hemagglutinin in target cell recognition by class I MHC-restricted, hemagglutinin-specific cytolytic T lymphocytes. J Exp Med *166*, 678-692.

Braciale, T.J., Morrison, L.A., Sweetser, M.T., Sambrook, J., Gething, M.J., and Braciale, V.L. (1987b). Antigen presentation pathways to class I and class II MHC-restricted T lymphocytes. Immunol Rev *98*, 95-114.

Brown, M.A. (2010). Genetics of ankylosing spondylitis. Curr Opin Rheumatol *22*, 126-132.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat Protoc *2*, 2728-2733.

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J.*, et al.* (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet *39*, 1329-1337.

Cascio, P., Hilton, C., Kisselev, A.F., Rock, K.L., and Goldberg, A.L. (2001). 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. EMBO J *20*, 2357-2366.

Chang, S.C., Momburg, F., Bhutani, N., and Goldberg, A.L. (2005). The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. Proc Natl Acad Sci U S A *102*, 17107-17112.

Chen, Y.L., Law, P.Y., and Loh, H.H. (2006). Sustained activation of phosphatidylinositol 3-kinase/Akt/nuclear factor kappaB signaling mediates G protein-coupled delta-opioid receptor gene expression. J Biol Chem *281*, 3067-3074.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., and Strominger, J.L. (1992). Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature *358*, 764-768.

Cochran, J.R., Cameron, T.O., and Stern, L.J. (2000). The relationship of MHC-peptide binding and T cell activation probed using chemically defined MHC class II oligomers. Immunity *12*, 241-250.

Cohen, B.E., McAnaney, T.B., Park, E.S., Jan, Y.N., Boxer, S.G., and Jan, L.Y. (2002). Probing protein electrostatics with a synthetic fluorescent amino acid. Science *296*, 1700-1703.

Cohen, S.X., Morris, R.J., Fernandez, F.J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V.S., Kleywegt, G.J., and Perrakis, A. (2004). Towards complete validated models in the next generation of ARP/wARP. Acta Crystallogr D Biol Crystallogr *60*, 2222-2229.

Cole DK, P.N., Boulter JM, Sami M, Bell JI, Gostick E, Price DA, Gao GF, Sewell AK, Jakobsen BK. (2007). Human TCR-binding affinity is governed by MHC class restriction. J Immunol. *178*, 5727-5723.

Collins, E.J., Garboczi, D.N., Karpusas, M.N., and Wiley, D.C. (1995). The three-dimensional structure of a class I major histocompatibility complex molecule missing the alpha 3 domain of the heavy chain. Proc Natl Acad Sci U S A *92*, 1218-1221.

Comellas-Bigler, M., Lang, R., Bode, W., and Maskos, K. (2005). Crystal structure of the E. coli dipeptidyl carboxypeptidase Dcp: further indication of a ligand-dependent hinge movement mechanism. J Mol Biol *349*, 99-112.

Craiu, A., Akopian, T., Goldberg, A., and Rock, K.L. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. Proc Natl Acad Sci U S A *94*, 10850-10855.

Cui, X., Hawari, F., Alsaaty, S., Lawrence, M., Combs, C.A., Geng, W., Rouhani, F.N., Miskinis, D., and Levine, S.J. (2002). Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding. J Clin Invest *110*, 515-526.

Day, P.M., Esquivel, F., Lukszo, J., Bennink, J.R., and Yewdell, J.W. (1995). Effect of TAP on the generation and intracellular trafficking of peptide-receptive major histocompatibility complex class I molecules. Immunity *2*, 137-147.

Degen, E., and Williams, D.B. (1991). Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility molecules. J Cell Biol *112*, 1099-1115.

DeLano, W.L. (2008). The PyMOL Molecular Graphics System. DeLano Scientific LLC, Palo Alto, CA, USA. http://www.pymol.org.

Deng, L., Langley, R.J., Brown, P.H., Xu, G., Teng, L., Wang, Q., Gonzales, M.I., Callender, G.G., Nishimura, M.I., Topalian, S.L., and Mariuzza, R.A. (2007). Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. Nat Immunol *8*, 398-408.

Denzin, L.K., and Cresswell, P. (1995). HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. Cell *82*, 155-165.

DiBrino, M., Parker, K.C., Shiloach, J., Turner, R.V., Tsuchida, T., Garfield, M., Biddison, W.E., and Coligan, J.E. (1994). Endogenous peptides with distinct amino acid anchor residue motifs bind to HLA-A1 and HLA-B8. J Immunol *152*, 620-631.

Dick, T.P., Bangia, N., Peaper, D.R., and Cresswell, P. (2002). Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. Immunity *16*, 87-98.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Eugenio Vazquez, M., Rothman, D.M., and Imperiali, B. (2004). A new environment-sensitive fluorescent amino acid for Fmoc-based solid phase peptide synthesis. Org Biomol Chem *2*, 1965-1966.

Evnouchidou, I., Berardi, M.J., and Stratikos, E. (2009). A continuous fluorigenic assay for the measurement of the activity of endoplasmic reticulum aminopeptidase 1: Competition kinetics as a tool for enzyme specificity investigation. Anal Biochem *395*, 33-40.

Evnouchidou, I., Momburg, F., Papakyriakou, A., Chroni, A., Leondiadis, L., Chang, S.C., Goldberg, A.L., and Stratikos, E. (2008). The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. PLoS One *3*, e3658.

Falk, K., Rotzschke, O., and Rammensee, H.G. (1990). Cellular peptide composition governed by major histocompatibility complex class I molecules. Nature *348*, 248-251.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., and Rammensee, H.G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature *351*, 290-296.

Falta, M.T., Fontenot, A.P., Rosloniec, E.F., Crawford, F., Roark, C.L., Bill, J., Marrack, P., Kappler, J., and Kotzin, B.L. (2005). Class II major histocompatibility complex-peptide tetramer staining in relation to functional avidity and T cell receptor diversity in the mouse CD4(+) T cell response to a rheumatoid arthritis-associated antigen. Arthritis Rheum *52*, 1885-1896.

Firat, E., Saveanu, L., Aichele, P., Staeheli, P., Huai, J., Gaedicke, S., Nil, A., Besin, G., Kanzler, B., van Endert, P., and Niedermann, G. (2007). The role of endoplasmic reticulum-associated aminopeptidase 1 in immunity to infection and in cross-presentation. J Immunol *178*, 2241-2248.

Fournie-Zaluski, M.C., Poras, H., Roques, B.P., Nakajima, Y., Ito, K., and Yoshimoto, T. (2009). Structure of aminopeptidase N from Escherichia coli complexed with the transition-state analogue aminophosphinic inhibitor PL250. Acta Crystallogr D Biol Crystallogr *65*, 814-822.

Frayser, M., Sato, A.K., Xu, L., and Stern, L.J. (1999). Empty and peptide-loaded class II major histocompatibility complex proteins produced by expression in Escherichia coli and folding in vitro. Protein Expr Purif *15*, 105-114.

Frentzel, S., Pesold-Hurt, B., Seelig, A., and Kloetzel, P.M. (1994). 20 S proteasomes are assembled via distinct precursor complexes. Processing of LMP2 and LMP7 proproteins takes place in 13-16 S preproteasome complexes. J Mol Biol *236*, 975-981.

Fruci, D., Giacomini, P., Nicotra, M.R., Forloni, M., Fraioli, R., Saveanu, L., van Endert, P., and Natali, P.G. (2008). Altered expression of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in transformed non-lymphoid human tissues. J Cell Physiol *216*, 742-749.

Fruci, D., Niedermann, G., Butler, R.H., and van Endert, P.M. (2001). Efficient MHC class I-independent amino-terminal trimming of epitope precursor peptides in the endoplasmic reticulum. Immunity *15*, 467-476.

Fung, E.Y., Smyth, D.J., Howson, J.M., Cooper, J.D., Walker, N.M., Stevens, H., Wicker, L.S., and Todd, J.A. (2009). Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. Genes Immun *10*, 188-191.

Galocha, B., Hill, A., Barnett, B.C., Dolan, A., Raimondi, A., Cook, R.F., Brunner, J., McGeoch, D.J., and Ploegh, H.L. (1997). The active site of ICP47, a herpes simplex virus-encoded inhibitor of the major histocompatibility complex (MHC)-encoded peptide transporter associated with antigen processing (TAP), maps to the NH2-terminal 35 residues. J Exp Med *185*, 1565-1572.

Garbi, N., Tan, P., Diehl, A.D., Chambers, B.J., Ljunggren, H.G., Momburg, F., and Hammerling, G.J. (2000). Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. Nat Immunol *1*, 234-238.

Georgiadou, D., Hearn, A., Evnouchidou, I., Chroni, A., Leondiadis, L., York, I.A., Rock, K.L., and Stratikos, E. (2010). Placental leucine aminopeptidase efficiently generates mature antigenic peptides in vitro but in patterns distinct from endoplasmic reticulum aminopeptidase 1. J Immunol *185*, 1584-1592.

Goettig, P., Brandstetter, H., Groll, M., Gohring, W., Konarev, P.V., Svergun, D.I., Huber, R., and Kim, J.S. (2005). X-ray snapshots of peptide processing in mutants of tricorn-interacting factor F1 from Thermoplasma acidophilum. J Biol Chem *280*, 33387-33396.

Goto, Y., Hattori, A., Ishii, Y., and Tsujimoto, M. (2006). Reduced activity of the hypertension-associated Lys528Arg mutant of human adipocyte-derived leucine aminopeptidase (A-LAP)/ER-aminopeptidase-1. FEBS Lett *580*, 1833-1838.

Goto, Y., Tanji, H., Hattori, A., and Tsujimoto, M. (2008). Glutamine-181 is crucial in the enzymatic activity and substrate specificity of human endoplasmic-reticulum aminopeptidase-1. Biochem J *416*, 109-116.

Grams, F., Dive, V., Yiotakis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stocker, W. (1996). Structure of astacin with a transition-state analogue inhibitor. Nat Struct Biol *3*, 671-675.

Haeggstrom, J.Z. (2000). Structure, function, and regulation of leukotriene A4 hydrolase. Am J Respir Crit Care Med *161*, S25-31.

Hahn, M., Nicholson, M.J., Pyrdol, J., and Wucherpfennig, K.W. (2005). Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. Nat Immunol *6*, 490-496.

Hammer, G.E., Gonzalez, F., James, E., Nolla, H., and Shastri, N. (2007). In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. Nat Immunol *8*, 101-108.

Harvey, D., Pointon, J.J., Evans, D.M., Karaderi, T., Farrar, C., Appleton, L.H., Sturrock, R.D., Stone, M.A., Oppermann, U., Brown, M.A., and Wordsworth, B.P. (2009). Investigating the genetic association between ERAP1 and ankylosing spondylitis. Hum Mol Genet *18*, 4204-4212.

Hattori, A., Matsumoto, H., Mizutani, S., and Tsujimoto, M. (1999). Molecular cloning of adipocyte-derived leucine aminopeptidase highly related to placental leucine aminopeptidase/oxytocinase. J Biochem *125*, 931-938.

Hearn, A., York, I.A., and Rock, K.L. (2009). The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. J Immunol *183*, 5526-5536.

Hennecke, J., and Wiley, D.C. (2002). Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. J Exp Med *195*, 571-581.

Holland, D.R., Tronrud, D.E., Pley, H.W., Flaherty, K.M., Stark, W., Jansonius, J.N., McKay, D.B., and Matthews, B.W. (1992). Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. Biochemistry *31*, 11310-11316.

Ito, K., Nakajima, Y., Onohara, Y., Takeo, M., Nakashima, K., Matsubara, F., Ito, T., and Yoshimoto, T. (2006). Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from Escherichia coli and conformational change of methionine 260 involved in substrate recognition. J Biol Chem *281*, 33664-33676.

Janeway, C.A., Jr., Travers, P., Walport, M., Scholmchik, M. (2001). Janeway's Immunobiology - Fifth Edition, 7 edn (New York: Garland Science Publishing).

Jardetzky, T.S., Gorga, J.C., Busch, R., Rothbard, J., Strominger, J.L., and Wiley, D.C. (1990). Peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. EMBO J *9*, 1797-1803.

Jiang, X., Zhou, L., Wu, Y., Wei, D., Sun, C., Jia, J., Liu, Y., and Lai, L. (2010). Modulating the substrate specificity of LTA4H aminopeptidase by using chemical compounds and small-molecule-guided mutagenesis. Chembiochem *11*, 1120-1128.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A *47 ( Pt 2)*, 110-119.

Kanaseki, T., Blanchard, N., Hammer, G.E., Gonzalez, F., and Shastri, N. (2006). ERAAP synergizes with MHC class I molecules to make the final cut in the antigenic peptide precursors in the endoplasmic reticulum. Immunity *25*, 795-806.

Kang, K., Park, B., Oh, C., Cho, K., and Ahn, K. (2009). A role for protein disulfide isomerase in the early folding and assembly of MHC class I molecules. Antioxid Redox Signal *11*, 2553-2561.

Kazeto, H., Nomura, S., Ito, N., Ito, T., Watanabe, Y., Kajiyama, H., Shibata, K., Ino, K., Tamakoshi, K., Hattori, A.*, et al.* (2003). Expression of adipocyte-derived leucine aminopeptidase in endometrial cancer. Association with tumor grade and CA-125. Tumour Biol *24*, 203-208.

Kisselev, A.F., Akopian, T.N., Woo, K.M., and Goldberg, A.L. (1999). The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes.

Implications for understanding the degradative mechanism and antigen presentation. J Biol Chem *274*, 3363-3371.

Klar, D., and Hammerling, G.J. (1989). Induction of assembly of MHC class I heavy chains with beta 2microglobulin by interferon-gamma. EMBO J *8*, 475-481.

Kleijmeer, M., Ramm, G., Schuurhuis, D., Griffith, J., Rescigno, M., Ricciardi-Castagnoli, P., Rudensky, A.Y., Ossendorp, F., Melief, C.J., Stoorvogel, W., and Geuze, H.J. (2001). Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. J Cell Biol *155*, 53-63.

Kleywegt, G.J., and Jones, T.A. (1998). Databases in protein crystallography. Acta Crystallogr D Biol Crystallogr *54*, 1119-1131.

Korman, A.J., Auffray, C., Schamboeck, A., and Strominger, J.L. (1982). The amino acid sequence and gene organization of the heavy chain of the HLA-DR antigen: homology to immunoglobulins. Proc Natl Acad Sci U S A *79*, 6013-6017.

Kropshofer, H., Vogt, A.B., Thery, C., Armandola, E.A., Li, B.C., Moldenhauer, G., Amigorena, S., and Hammerling, G.J. (1998). A role for HLA-DO as a co-chaperone of HLA-DM in peptide loading of MHC class II molecules. EMBO J *17*, 2971-2981.

Kyrieleis, O.J., Goettig, P., Kiefersauer, R., Huber, R., and Brandstetter, H. (2005). Crystal structures of the tricorn interacting factor F3 from Thermoplasma acidophilum, a zinc aminopeptidase in three different conformations. J Mol Biol *349*, 787-800.

Lamb, C.A., and Cresswell, P. (1992). Assembly and transport properties of invariant chain trimers and HLA-DR-invariant chain complexes. J Immunol *148*, 3478-3482.

Leach, M.R., Cohen-Doyle, M.F., Thomas, D.Y., and Williams, D.B. (2002). Localization of the lectin, ERp57 binding, and polypeptide binding sites of calnexin and calreticulin. J Biol Chem *277*, 29686-29697.

Liljedahl, M., Winqvist, O., Surh, C.D., Wong, P., Ngo, K., Teyton, L., Peterson, P.A., Brunmark, A., Rudensky, A.Y., Fung-Leung, W.P., and Karlsson, L. (1998). Altered antigen presentation in mice lacking H2-O. Immunity *8*, 233-243.

Madden, D.R. (1995). The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol *13*, 587-622.

Malito, E., Ralat, L.A., Manolopoulou, M., Tsay, J.L., Wadlington, N.L., and Tang, W.J. (2008). Molecular bases for the recognition of short peptide substrates and cysteine-directed modifications of human insulin-degrading enzyme. Biochemistry *47*, 12822-12834.

Matsuura, Y., and Stewart, M. (2004). Structural basis for the assembly of a nuclear export complex. Nature *432*, 872-877.

Matthews, B.W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

Mattson, D.H., Shimojo, N., Cowan, E.P., Baskin, J.J., Turner, R.V., Shvetsky, B.D., Coligan, J.E., Maloy, W.L., and Biddison, W.E. (1989). Differential effects of amino acid substitutions in the beta-sheet floor and alpha-2 helix of HLA-A2 on recognition by alloreactive viral peptide-specific cytotoxic T lymphocytes. J Immunol *143*, 1101-1107.

Maynard, J., Petersson, K., Wilson, D.H., Adams, E.J., Blondelle, S.E., Boulanger, M.J., Wilson, D.B., and Garcia, K.C. (2005). Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. Immunity *22*, 81-92.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McFarland, B.J., and Beeson, C. (2002). Binding interactions between peptides and proteins of the class II major histocompatibility complex. Med Res Rev *22*, 168-203.

Mehta, A.M., Jordanova, E.S., Corver, W.E., van Wezel, T., Uh, H.W., Kenter, G.G., and Jan Fleuren, G. (2009). Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. Genes Chromosomes Cancer *48*, 410-418.

Minami, M., Bito, H., Ohishi, N., Tsuge, H., Miyano, M., Mori, M., Wada, H., Mutoh, H., Shimada, S., Izumi, T., and et al. (1992). Leukotriene A4 hydrolase, a bifunctional enzyme. Distinction of leukotriene A4 hydrolase and aminopeptidase activities by site-directed mutagenesis at Glu-297. FEBS Lett *309*, 353-357.

Momburg, F., Roelse, J., Hammerling, G.J., and Neefjes, J.J. (1994). Peptide size selection by the major histocompatibility complex-encoded peptide transporter. J Exp Med *179*, 1613-1623.

Momburg, F., and Tan, P. (2002). Tapasin-the keystone of the loading complex optimizing peptide binding by MHC class I molecules in the endoplasmic reticulum. Mol Immunol *39*, 217-233.

Monecke, T., Guttler, T., Neumann, P., Dickmanns, A., Gorlich, D., and Ficner, R. (2009). Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. Science *324*, 1087-1091.

Morrice, N.A., and Powis, S.J. (1998). A role for the thiol-dependent reductase ERp57 in the assembly of MHC class I molecules. Curr Biol *8*, 713-716.

Murphy, D.B., Rath, S., Pizzo, E., Rudensky, A.Y., George, A., Larson, J.K., and Janeway, C.A., Jr. (1992). Monoclonal antibody detection of a major self peptide. MHC class II complex. J Immunol *148*, 3483-3491.

Nicholls, A., and Honig, B.H. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve Poisson-Boltzmann equations. J. Comput. Chem. *12*, 657-662.

Nitz, M., Ling, C.C., Otter, A., Cutler, J.E., and Bundle, D.R. (2002). The unique solution structure and immunochemistry of the Candida albicans beta -1,2-mannopyranan cell wall antigens. J Biol Chem *277*, 3440-3446.

Nocek, B., Mulligan, R., Bargassa, M., Collart, F., and Joachimiak, A. (2008). Crystal structure of aminopeptidase N from human pathogen Neisseria meningitidis. Proteins *70*, 273-279.

Norris, P.J., Moffett, H.F., Brander, C., Allen, T.M., O'Sullivan, K.M., Cosimi, L.A., Kaufmann, D.E., Walker, B.D., and Rosenberg, E.S. (2004). Fine specificity and cross-clade reactivity of HIV type 1 Gag-specific CD4+ T cells. AIDS Res Hum Retroviruses *20*, 315-325.

Norris, P.J., Stone, J.D., Anikeeva, N., Heitman, J.W., Wilson, I.C., Hirschkorn, D.F., Clark, M.J., Moffett, H.F., Cameron, T.O., Sykulev, Y.*, et al.* (2006). Antagonism of HIV-specific CD4+ T cells by C-terminal truncation of a minimum epitope. Mol Immunol *43*, 1349-1357.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Orr, H.T., Lopez de Castro, J.A., Lancet, D., and Strominger, J.L. (1979). Complete amino acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. 2. Sequence determination and search for homologies. Biochemistry *18*, 5711-5720.

Otwinowski, Z.a.M., W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology *276*, 307-326.

Padovan E, C.G., Dellabona P, Giachino C, Lanzavecchia A. (1995). Dual receptor T-cells. Implications for alloreactivity and autoimmunity. Ann N Y Acad Sci. *7*, 66-70.

Parham, P., Lawlor, D.A., Lomen, C.E., and Ennis, P.D. (1989). Diversity and diversification of HLA-A,B,C alleles. J Immunol *142*, 3937-3950.

Parnas, B.L., Durley, R.C., Rhoden, E.E., Kilpatrick, B.F., Makkar, N., Thomas, K.E., Smith, W.G., and Corley, D.G. (1996). Isolation and structure of leukotriene-A4 hydrolase inhibitor: 8(S)-amino-2(R)-methyl-7-oxononanoic acid produced by Streptomyces diastaticus. J Nat Prod *59*, 962-964.

Peer, W.A., Hosein, F.N., Bandyopadhyay, A., Makam, S.N., Otegui, M.S., Lee, G.J., Blakeslee, J.J., Cheng, Y., Titapiwatanakun, B., Yakubov, B.*, et al.* (2009). Mutation of the membrane-associated M1 protease APM1 results in distinct embryonic and seedling developmental defects in Arabidopsis. Plant Cell *21*, 1693-1721.

Pierre, P., Turley, S.J., Gatti, E., Hull, M., Meltzer, J., Mirza, A., Inaba, K., Steinman, R.M., and Mellman, I. (1997). Developmental regulation of MHC class II transport in mouse dendritic cells. Nature *388*, 787-792.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics *50*, 213-219.

Rammensee, H.G. (1995). Chemistry of peptides associated with MHC class I and class II molecules. Curr Opin Immunol *7*, 85-96.

Rammensee, H.G., Friede, T., and Stevanoviic, S. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics *41*, 178-228.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2009). MEROPS: the peptidase database. Nucleic Acids Res *38*, D227-233.

Reche, P.A., Glutting, J.P., Zhang, H., and Reinherz, E.L. (2004). Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics *56*, 405-419.

Reinherz, E.L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y., Hussey, R.E., Smolyar, A., Hare, B., Zhang, R.*, et al.* (1999). The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science *286*, 1913-1921.

Reits, E., Neijssen, J., Herberts, C., Benckhuijsen, W., Janssen, L., Drijfhout, J.W., and Neefjes, J. (2004). A major role for TPPII in trimming proteasomal degradation products for MHC class I antigen presentation. Immunity *20*, 495-506.

Reveille, J.D., Sims, A.M., Danoy, P., Evans, D.M., Leo, P., Pointon, J.J., Jin, R., Zhou, X., Bradbury, L.A., Appleton, L.H.*, et al.* (2010). Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet *42*, 123-127.

Riberdy, J.M., Newcomb, J.R., Surman, M.J., Barbosa, J.A., and Cresswell, P. (1992). HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature *360*, 474-477.

Roche, P.A., and Cresswell, P. (1990). Invariant chain association with HLA-DR molecules inhibits immunogenic peptide binding. Nature *345*, 615-618.

Rock, K.L., York, I.A., and Goldberg, A.L. (2004). Post-proteasomal antigen processing for major histocompatibility complex class I presentation. Nat Immunol *5*, 670-677.

Ross, S.A., Scott, H.M., Morris, N.J., Leung, W.Y., Mao, F., Lienhard, G.E., and Keller, S.R. (1996). Characterization of the insulin-regulated membrane aminopeptidase in 3T3-L1 adipocytes. J Biol Chem *271*, 3328-3332.

Rudensky, A., Preston-Hurlburt, P., Hong, S.C., Barlow, A., and Janeway, C.A., Jr. (1991). Sequence analysis of peptides bound to MHC class II molecules. Nature *353*, 622-627.

Rudolph, M.G., Stanfield, R.L., and Wilson, I.A. (2006). How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol *24*, 419-466.

Saric, T., Chang, S.C., Hattori, A., York, I.A., Markant, S., Rock, K.L., Tsujimoto, M., and Goldberg, A.L. (2002). An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. Nat Immunol *3*, 1169-1176.

Sato, N., Nabeta, Y., Kondo, H., Sahara, H., Hirohashi, Y., Kashiwagi, K., Kanaseki, T., Sato, Y., Rong, S., Hirai, I.*, et al.* (2000). Human CD8 and CD4 T cell epitopes of epithelial cancer antigens. Cancer Chemother Pharmacol *46 Suppl*, S86-90.

Saveanu, L., Carroll, O., Lindo, V., Del Val, M., Lopez, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G., and van Endert, P.M. (2005). Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat Immunol *6*, 689-697.

Saveanu, L., Carroll, O., Weimershaus, M., Guermonprez, P., Firat, E., Lindo, V., Greer, F., Davoust, J., Kratzer, R., Keller, S.R.*, et al.* (2009). IRAP identifies an endosomal compartment required for MHC class I cross-presentation. Science *325*, 213-217.

Schlosstein, L., Terasaki, P.I., Bluestone, R., and Pearson, C.M. (1973). High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med *288*, 704-706.

Schomburg, L., Kollmus, H., Friedrichsen, S., and Bauer, K. (2000). Molecular characterization of a puromycin-insensitive leucyl-specific aminopeptidase, PILS-AP. Eur J Biochem *267*, 3198-3207.

Schumacher, T.N., Kantesaria, D.V., Heemels, M.T., Ashton-Rickardt, P.G., Shepherd, J.C., Fruh, K., Yang, Y., Peterson, P.A., Tonegawa, S., and Ploegh, H.L. (1994). Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. J Exp Med *179*, 533-540.

Schuttelkopf, A.W., and van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr *60*, 1355-1363.

Serwold, T., Gaw, S., and Shastri, N. (2001). ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. Nat Immunol *2*, 644-651.

Serwold, T., Gonzalez, F., Kim, J., Jacob, R., and Shastri, N. (2002). ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature *419*, 480-483.

Shen, Y., Joachimiak, A., Rosner, M.R., and Tang, W.J. (2006). Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. Nature *443*, 870-874.

Sherman, M.A., Weber, D.A., and Jensen, P.E. (1995). DM enhances peptide binding to class II MHC by release of invariant chain-derived peptide. Immunity *3*, 197-205.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature *368*, 215-221.

Stern, L.J., and Wiley, D.C. (1992). The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. Cell *68*, 465-477.

Stone, J.D., Cochran, J.R., and Stern, L.J. (2001). T-cell activation by soluble MHC oligomers can be described by a two-parameter binding model. Biophys J *81*, 2547-2557.

Suh, W.K., Derby, M.A., Cohen-Doyle, M.F., Schoenhals, G.J., Fruh, K., Berzofsky, J.A., and Williams, D.B. (1999). Interaction of murine MHC class I molecules with tapasin and TAP enhances peptide loading and involves the heavy chain alpha3 domain. J Immunol *162*, 1530-1540.

Suhre, K., and Sanejouand, Y.H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res *32*, W610-614.

Tanioka, T., Hattori, A., Masuda, S., Nomura, Y., Nakayama, H., Mizutani, S., and Tsujimoto, M. (2003). Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. J Biol Chem *278*, 32275-32283.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr *56*, 965-972.

Tholander, F., Muroya, A., Roques, B.P., Fournie-Zaluski, M.C., Thunnissen, M.M., and Haeggstrom, J.Z. (2008). Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. Chem Biol *15*, 920-929.

Thompson, M.W., Archer, E.D., Romer, C.E., and Seipelt, R.L. (2006). A conserved tyrosine residue of Saccharomyces cerevisiae leukotriene A4 hydrolase stabilizes the transition state of the peptidase activity. Peptides *27*, 1701-1709.

Thunnissen, M.M., Nordlund, P., and Haeggstrom, J.Z. (2001). Crystal structure of human leukotriene A(4) hydrolase, a bifunctional enzyme in inflammation. Nat Struct Biol *8*, 131-135.

Towler, P., Staker, B., Prasad, S.G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., Dales, N.A.*, et al.* (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. J Biol Chem *279*, 17996-18007.

Townsend, A., Elliott, T., Cerundolo, V., Foster, L., Barber, B., and Tse, A. (1990). Assembly of MHC class I molecules analyzed in vitro. Cell *62*, 285-295.

Tsujimoto, M., and Hattori, A. (2005). The oxytocinase subfamily of M1 aminopeptidases. Biochim Biophys Acta *1751*, 9-18.

Uebel, S., and Tampe, R. (1999). Specificity of the proteasome and the TAP transporter. Curr Opin Immunol *11*, 203-208.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr D Biol Crystallogr *60*, 2184-2195.

van Endert, P.M., Tampe, R., Meyer, T.H., Tisch, R., Bach, J.F., and McDevitt, H.O. (1994). A sequential model for peptide binding and transport by the transporters associated with antigen processing. Immunity *1*, 491-500.

van Lith, M., van Ham, M., Griekspoor, A., Tjin, E., Verwoerd, D., Calafat, J., Janssen, H., Reits, E., Pastoors, L., and Neefjes, J. (2001). Regulation of MHC class II antigen presentation by sorting of recycling HLA-DM/DO and class II within the multivesicular body. J Immunol *167*, 884-892.

Vazquez, M.E., Blanco, J.B., and Imperiali, B. (2005). Photophysics and biological applications of the environment-sensitive fluorophore 6-N,N-dimethylamino-2,3-naphthalimide. J Am Chem Soc *127*, 1300-1306.

Watanabe, Y., Shibata, K., Kikkawa, F., Kajiyama, H., Ino, K., Hattori, A., Tsujimoto, M., and Mizutani, S. (2003). Adipocyte-derived leucine aminopeptidase suppresses angiogenesis in human endometrial carcinoma via renin-angiotensin system. Clin Cancer Res *9*, 6497-6503.

Wedderburn, L.R., Searle, S.J., Rees, A.R., Lamb, J.R., and Owen, M.J. (1995). Mapping T cell recognition: the identification of a T cell receptor residue critical to the specific interaction with an influenza hemagglutinin peptide. Eur J Immunol *25*, 1654-1662.

Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C., and Davis, M.M. (2002). Two-step binding mechanism for T-cell receptor recognition of peptide MHC. Nature *418*, 552-556.

Yamamoto, N., Nakayama, J., Yamakawa-Kobayashi, K., Hamaguchi, H., Miyazaki, R., and Arinami, T. (2002). Identification of 33 polymorphisms in the adipocyte-derived leucine aminopeptidase (ALAP) gene and possible association with hypertension. Hum Mutat *19*, 251-257.

Yan, J., Parekh, V.V., Mendez-Fernandez, Y., Olivares-Villagomez, D., Dragovic, S., Hill, T., Roopenian, D.C., Joyce, S., and Van Kaer, L. (2006). In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules. J Exp Med *203*, 647-659.

York, I.A., Brehm, M.A., Zendzian, S., Towne, C.F., and Rock, K.L. (2006). Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance. Proc Natl Acad Sci U S A *103*, 9202-9207.

York, I.A., Chang, S.C., Saric, T., Keys, J.A., Favreau, J.M., Goldberg, A.L., and Rock, K.L. (2002). The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. Nat Immunol *3*, 1177-1184.

Zavala-Ruiz, Z., Strug, I., Anderson, M.W., Gorski, J., and Stern, L.J. (2004a). A polymorphic pocket at the P10 position contributes to peptide binding specificity in class II MHC proteins. Chem Biol *11*, 1395-1402.

Zavala-Ruiz, Z., Strug, I., Walker, B.D., Norris, P.J., and Stern, L.J. (2004b). A hairpin turn in a class II MHC-bound peptide orients residues outside the binding groove for T cell recognition. Proc Natl Acad Sci U S A *101*, 13279-13284.

Zavala-Ruiz, Z., Sundberg, E.J., Stone, J.D., DeOliveira, D.B., Chan, I.C., Svendsen, J., Mariuzza, R.A., and Stern, L.J. (2003). Exploration of the P6/P7 region of the peptide-binding site of the human class II major histocompatability complex protein HLA-DR1. J Biol Chem *278*, 44904-44912.