The Relationship Between Inhibition, Conformation, and Catalysis of the Aminopeptidase ERAP1

Zachary Maben
University of Massachusetts Medical School
THE RELATIONSHIP BETWEEN INHIBITION, CONFORMATION, AND CATALYSIS OF THE AMINOPEPTIDASE ERAP1

A Dissertation Presented

By

ZACHARY MABEN

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

NOVEMBER 15th, 2018

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY
THE RELATIONSHIP BETWEEN INHIBITION, CONFORMATION, AND CATALYSIS OF THE AMINOPEPTIDASE ERAP1

A Dissertation Presented by

ZACHARY MABEN

This work was undertaken in the Graduate School of Biomedical Sciences

(Biochemistry and Molecular Pharmacology)

Under the mentorship of

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

________________________
Osman Bilsel, Ph.D., Member of Committee

________________________
Brian A. Kelch, Ph.D., Member of Committee

________________________
David G. Lambright, Ph.D., Member of Committee

________________________
José A. López de Castro, Ph.D., External Member of Committee

________________________
Anthony Carruthers, Ph.D., Chair of Committee

________________________
Mary Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

November 15th, 2018
Dedication

I dedicate this work to my family, especially to Sarah for support, encouragement, laundry, laughter, tears, and love, and to Anders, for renewing my sense of wonder at the world around us.
Acknowledgments

I am grateful to Larry Stern for providing guidance throughout my training and for supporting the development of new techniques and new strategies to study molecular immunology.

My thanks to James Birtley for assistance, encouragement, and tutelage in all things crystallographic. I also must thank Peter Trenh for interesting scientific discussion on a wide range of topics and insightful suggestions on data analysis and presentation. I am grateful to Richa Arya for brainstorming sessions on ERAP1 and for assistance with experiments, and to Mollie Jurewicz for editing my writing and for contributing her immunological expertise to this project.

Many thanks to the rest of the Stern Lab for all your help and support throughout my time here: Aniuska Becerra, Sydney Blevins, Mauricio Calvo-Calle, John Cruz, Padma Nanaware, Pia Negroni, Loretta Lee, Guoqi Li, Liying Lu, Corrie Painter, Karthik Ramesh, Tessa Santoro, Austin Shen, Lianjun Shen, Inyoung Song, Grant Weaver, Liusong Yin, and Kerstin Zimmermann.

Thanks also to Rachel Bacon and Scott Bacon for guidance regarding in vivo compound efficacy trials, and for game nights.

Lastly, thanks to Chido Kativhu and Ogolouwa Ojelabi for their scientific input and for many years of friendship.
Abstract

ERAP1 is an aminopeptidase that is a component of antigen processing. To distinguish the role of ERAP1 from homologs ERAP2 and IRAP, I identified three specific ERAP1 inhibitors via a high-throughput screen. These compounds inhibit hydrolysis of a decamer peptide, and some inhibit ERAP1 in a cellular assay. These inhibitors enable dissection of ERAP1 mechanism. ERAP1 has been crystallized in two conformations: open and closed. I collected SAXS data on ERAP1 in the presence of various inhibitors. ERAP1 adopts an open conformation in solution, but some inhibitors stabilize the closed form. Compound 3 docks to a distal pocket 28Å from the active site zinc, while DG013 and DG014 bind to the active site. This distal pocket is an allosteric activation site, and allostery is mediated by stabilizing the closed state. I also identified an intermediate step in substrate binding where helix 4a becomes ordered while ERAP1 maintains an open conformation. Helix 4a then rotates and engages substrate when ERAP1 closes. The nonsynonymous SNP rs30187 at position 528 (Lys/Arg) subtly alters ERAP1 activity in vitro and correlates with disease incidence. Position 528 forms a conformation-dependent electrostatic interaction with Glu913 in the closed structure. The energetic contribution of this interaction is stronger for Lys528 than Arg528. Inhibitors that induce closing are more potent for Lys528 than Arg528. I propose a model where either helix 4a stabilization or allosteric site occupancy shift the conformational equilibrium towards a closed state, while substitution at position 528 alters the opening rate.
Table of contents

Dedication ................................................................................................................. iii
Acknowledgments ....................................................................................................... iv
Abstract ....................................................................................................................... v
Table of contents ......................................................................................................... vi
List of Tables ............................................................................................................... ix
List of Figures ............................................................................................................. x
Preface ......................................................................................................................... xii

CHAPTER I .................................................................................................................. 14
Overview of the immune system ................................................................................. 14
Cell-mediated adaptive immune system .................................................................. 15
Class I antigen processing .......................................................................................... 18
  Proteasome .............................................................................................................. 18
  Cytosolic peptidases ............................................................................................... 19
  Transporter associated with Antigen Presentation .............................................. 20
  ER-resident peptidases .......................................................................................... 21
Discovery of ERAP1 .................................................................................................... 22
ERAP1 substrate length preference ........................................................................... 24
M1 family protein structure ....................................................................................... 25
M1 family conformation ............................................................................................. 26
Peptide C-terminal binding site .................................................................................. 27
ERAP1 alters antigen processing and immune function ........................................... 29
ERAP1 polymorphism rs30187 .................................................................................. 32
  SNP rs30187 correlates with disease ................................................................... 33
  SNP rs30187 effect on ERAP1 protein structure and function ............................ 34
  ERAP1 polymorphisms are inherited as discrete alleles ...................................... 35
Small-molecule inhibitors of ERAP1 ........................................................................ 36
Open questions regarding ERAP1 ............................................................................ 39

CHAPTER II ............................................................................................................... 42
ABSTRACT ................................................................................................................... 42
INTRODUCTION ......................................................................................................... 43
RESULTS ...................................................................................................................... 47
  Assay development ............................................................................................... 47
  Inhibitor screen ...................................................................................................... 48
  Inhibitor structure-activity relationship analysis ................................................ 53
  Activator screen ...................................................................................................... 53
Activator structure-activity relationship analysis................................. 55
Molecular docking identifies compound binding sites .......................... 55
Inhibition of intracellular peptide processing .................................... 56
Differential efficacy of compound 3 on ERAP1 alleles .......................... 59
Mechanisms of inhibition or activation .............................................. 61

DISCUSSION .................................................................................. 64
MATERIALS AND METHODS .............................................................. 67
Protein expression and purification ..................................................... 67
High-throughput library screen ......................................................... 68
Confirmatory and selectivity X-AMC hydrolysis assays ........................ 68
Leu-pNA hydrolysis assay ................................................................. 69
Peptide hydrolysis assay ................................................................. 70
In silico Docking .............................................................................. 71
Cellular antigen processing assay .................................................... 72
Data analysis .................................................................................. 72

ACKNOWLEDGEMENTS ................................................................ 73

CHAPTER III .................................................................................. 74
ABSTRACT ..................................................................................... 75
INTRODUCTION ............................................................................. 75

RESULTS ....................................................................................... 79
ERAP1 is open in solution and closes upon active site occupancy ......... 79
Ordering of the H4a helix induced by substrate binding ....................... 89
Domain closure reorients the H5 helix for catalysis ............................. 90
Allosteric activator induces domain closure, promotes catalysis .......... 98
Peptide C-terminus binding sites localize within domain IV ............... 108
Lys/Arg 528 SNP alters domain closure, regulates activity ................. 114

DISCUSSION ............................................................................... 118

MATERIALS AND METHODS .............................................................. 127
Mutagenesis and baculovirus production .............................................. 127
Protein expression and purification ..................................................... 127
Chemical synthesis .......................................................................... 127
Small angle X-ray scattering data collection ...................................... 130
Small angle X-ray scattering data processing and analysis .................. 130
ERAP1 structural modeling ............................................................... 131
Crystallization ................................................................................. 131
X-ray diffraction data collection ....................................................... 132
Structure determination .................................................................... 132
Enzyme activity assays ..................................................................... 133
Negative stain-electron microscopy .................................................... 134
Photocrosslinking and mass spectrometry ........................................ 134
List of Tables

Table 2.1 .........................................................................................................................60
Table 3.1 ..........................................................................................................................80
Table 3.2 ..........................................................................................................................85
Table 3.3 ..........................................................................................................................92
List of Figures

Figure 1.1.................................................................17
Figure 1.2.................................................................37
Figure 2.1.................................................................49
Figure 2.2.................................................................50
Figure 2.3.................................................................51
Figure 2.4.................................................................52
Figure 2.5.................................................................54
Figure 2.6.................................................................57
Figure 2.7.................................................................58
Figure 2.8.................................................................63
Figure 3.1.................................................................82
Figure 3.2.................................................................86
Figure 3.3.................................................................91
Figure 3.4.................................................................93
Figure 3.5.................................................................94
Figure 3.6.................................................................95
Figure 3.7.................................................................99
Figure 3.8.................................................................101
Figure 3.9.................................................................103
Figure 3.10...............................................................104
Figure 3.11...............................................................106

x
Preface

I have contributed to other publications during my time in the laboratory of Lawrence Stern which will not be described within my thesis.

In Yin et al. 2015[1], I performed experiments showing that HLA-DM has the same activity to catalyze peptide exchange from either monomeric or dimeric/aggregated HLA-DR1. This paper concluded that HLA-DM mediates peptide exchange on class II MHC by interacting with unbound MHC, MHC:peptide complex, and inactivated MHC and forming an alternate kinetic pathway to peptide association with MHC.

In Stammogianos et al 2017[2], I helped plan and perform mutation of transient salt bridges within ERAP1, expressed these proteins, and then characterized these proteins via SAXS. This paper showed that disruption of these salt bridges, and by association other transient interdomain electrostatic interactions, alters ERAP1 activity, supporting a model that the closed conformation of ERAP1 is required for activity. The SAXS data in this paper that I collected is the first solution-state determination of ERAP1 conformation, which is open in solution.

In Birtley et al. (under review), I performed SAXS experiments on WT OPCML and a clinically-relevant mutant P95R. This showed that the dimeric OPCML structure determined by X-ray crystallography is also present in solution, and that no structural deformation occurs when the mutation is present. This paper overall
described several structural mechanisms resulting in loss of function of OPCML, an endogenous tumor suppressor.
CHAPTER I

Introduction

Overview of the immune system

As long as there has been life, it is likely there have also been pathogens[3]. Agents of disease (termed ‘pathogens’) subsist in an evolutionary niche infecting other organisms (termed ‘hosts’) and diverting resources to propagate the pathogen, often to the detriment of the host. In response, hosts were selected to have defensive systems that can prevent or reduce damage from pathogen interactions. In multicellular organisms, these host defenses are referred to as immune systems. They consist of multiple mechanisms to sense and respond to pathogens, since pathogens also have selective pressure to circumvent these defenses and have developed numerous strategies to do so[4].

In humans, and indeed in most vertebrates, the immune system utilizes cells and proteins to recognize the presence (and necessarily, the absence) of pathogen infection, and then to direct an immune response against this infection. This immune response often causes damage to infected tissues. Once the infection is cleared, the immune system also plays a role in recovering damaged tissue to homeostasis[5].

Pathogen detection requires sensing molecules that do not normally exist within the host. Some molecules such as double-stranded RNA (dsRNA) are
sensed by interaction with proteins that bind dsRNA without recognition of the oligonucleotide sequence. These broad molecular indicators of pathogen infection are referred to as pathogen-associated molecular patterns (PAMPs)[6]. Because any source of this signal is an unambiguous marker of pathogen activity, the immune system that targets PAMPs is termed the innate immune system, and immune responses against these molecules may be mounted immediately upon detection[7].

Immune systems also recognize more subtle markers of infection. The presence of pathogen-derived proteins (or other non-self proteins) may be distinguished by detecting their primary sequence and their three-dimensional structure. However, the host immune system is not capable of recognizing every undesired protein *ab initio*; the only reference point available initially is the ensemble of host proteins which should be tolerated. Once infection is mounted, immune cells are mobilized to search for unique identifiers of pathogen (termed antigens), and these are used to then develop a more specific response such as antibody production and/or proliferation of immune effector cells. This slower, more focused system is called the adaptive immune system[8].

**Cell-mediated adaptive immune system**

One subsystem is referred to as the cell-mediated adaptive immune system. Specialized cells known as T cells express a highly variable, essentially unique
sensor membrane protein called the T cell receptor (TCR). This protein consists of a dimer of immunoglobulin-fold proteins which are generated when their genes each undergo a DNA splicing event. The spliced region contains some number of randomized nucleotides, which increases the variability of the coding sequence. In the translated protein, this spliced region forms a loop known as CDR3 (CDR1 and CDR2 are within the genome-encoded regions). The CDR loops are the primary contacts when the TCR recognizes an antigen[9].

For T cells, this antigenic molecule is a peptide bound into a groove on a major histocompatibility complex (MHC) protein. There are two classes of MHC, I and II. Canonically, T cells that recognize MHC-I express a coreceptor CD8, and T cells recognizing MHC-II express CD4[4] (although there are exceptions to this trend[10-12]). Each MHC protein has a preferred peptide sequence based on the chemical properties of the binding pockets that anchor the peptide at particular positions. Class I MHC proteins additionally have a peptide length preference of 8-11 amino acids. These two pathways (class I or II) each have distinct cellular processes dedicated to generating and stably binding these peptides to MHC proteins. The steps involved in the generation of these peptides is called antigen processing[4].
Figure 1.1
Overview of canonical class I MHC antigen processing pathway
A protein is translated by the ribosome and carry out some function within the cell. This protein is subsequently degraded by the proteasome into peptide fragments. (colored ovals) Many of these are degraded by proteases within the cytosol. Some fraction survive this process, and a subset of these are transported into the ER by TAP. Some peptides may bind class I MHC (MHC-I) without further processing. (orange oval) Others are too long (blue+red oval), but aminopeptidase activity within the ER by ERAP1 and ERAP2 degrades the N-terminal extension (red) leaving a peptide that is the right length to bind MHC-I. (blue oval) MHC-I then traffics to the cell surface for T cell surveillance.
**Class I antigen processing**

Within adaptive immunity, antigen processing creates a crucial bottleneck. In this process, protein antigens are digested into peptides, some of which have appropriate sequence and length to bind the particular MHC alleles present in a given organism. These bound peptides are then presented to T cells. Changes in processing can create novel peptides that may elicit T cell response. The pathway to convert intact proteins into peptides capable of binding MHC-I molecules involves a group of cellular components. Here I will describe the players within the canonical pathway that produce MHC-I peptides from protein antigens.

**Proteasome**

Many cellular proteins are degraded by the proteasome. This multiunit molecular machine has an internal cavity into which proteins for degradation are threaded. This cavity proteolytically digests the target protein into peptide fragments 3 to 24 amino acids long, with a distribution strongly favoring smaller fragments (median 5-6 amino acids) [13]. While few of these fragments are of ideal length to bind MHC-I directly [13], this proteasomal activity is required for the majority of MHC-I-presented peptides [14]. Proteasome activity is known to change upon signaling from interferon-γ, a pro-inflammatory cytokine. Interferon-γ signaling transcriptionally activates expression of alternate proteasomal
subunits via NF-κB signal transduction[15]. The proteasome complex containing these subunits has been termed the 'immunoproteasome', and it produces peptide fragments with C-termini skewed towards basic or hydrophobic amino acids[16]. This is thought to improve the peptide pool for antigen processing by more closely matching peptide-binding motifs exhibited by MHC-I[15]. Another component activated by interferon-γ signaling is PA28, which binds to the proteasome[17, 18]. This PA28-proteasome complex further alters the efficiency of peptide hydrolysis[17, 19]. It appears that altering the pool of proteasome-derived peptides provides some benefit in immune response[20-22], possibly by retaining peptide epitopes that would normally be degraded or by tailoring peptides to match preferred MHC-I sequence motifs[23].

**Cytosolic peptidases**

In addition to the proteasome, there are cytosolic proteases that subsequently degrade peptides. Products of this degradation feed into metabolic pathways as free amino acids, and are recycled into protein synthesis pathways or are utilized for energy production[24]. Tripeptidyl-peptidase II (TPPII) is a serine protease that produces tripeptide fragments from longer peptides[25]. The role of TPPII in antigen processing is considered minor[26-28]. Thimet oligopeptidase is a zinc endopeptidase that digests peptides into smaller peptide fragments 6-9 amino acids in length[29], and generates some MHC-I peptides in conjunction with another zinc metalloprotease, nardilysin[30].
Leucine aminopeptidase or cytosol aminopeptidase is an M17 family zinc aminopeptidase[31] that is dispensable for normal MHC-I antigen processing in vivo[32]. The cysteine protease bleomycin hydrolase is similarly unnecessary for antigen processing[33]. However, mice lacking both leucine aminopeptidase and bleomycin hydrolase respond more strongly to a specific viral peptide, indicating that these two enzymes have overlapping functions in degrading certain antigenic peptides[33]. Puromycin-sensitive aminopeptidase is an M1 zinc aminopeptidase that may destroy some MHC-binding peptides but is not necessary for normal immune function[34].

Transporter associated with Antigen Presentation

Peptides generated in the cytosol must be delivered to the endoplasmic reticulum (ER), where empty class I MHC proteins are stabilized by specialized chaperone molecules in a complex known as the peptide loading complex (PLC). The transporter associated with antigen processing (TAP) is an ABC transporter protein that associates as part of the PLC[35]. TAP binds peptides 9 to 16 amino acids in length with equivalent affinity, with decreasing affinity for shorter or longer peptides[36]. TAP also is unable to transport peptides with proline at the second position, which is a preferred anchor residue for some MHC-I[37]. It is clear that transport across the ER membrane is a limiting step in the process, as experiments using TAP KO cells result in almost complete loss of surface MHC bound to peptides[38].
The PLC has been structurally characterized bound by two empty MHC molecules[35]. The peptide-binding regions of these MHC proteins are far from the peptide egress pocket of TAP, and transported peptides must pass through a ‘ventral window’ space to bind MHC. There does not appear to be a mechanism to spatially sequester peptides in proximity with MHC in order to promote association. This means that translocated peptides may interact with other molecules present in the ER, among which are peptidases.

**ER-resident peptidases**

Peptidases are present in the ER, although there are many fewer versions than there are cytosolic peptidases. Angiotensin-converting enzyme is a zinc carboxypeptidase that, in addition to its role in blood pressure regulation, processes peptides for both MHC-I and MHC-II pathways[39, 40]. Another source of MHC-binding peptides is signal peptides, which are endogenous N-terminal tags that signal the protein translation system to express a newly-synthesized protein into the ER[41]. Signal peptides are often cleaved from nascent polypeptides during ER translocation by the action of the signal peptidase[41]. Some MHC-I, such as HLA-E, present signal peptides of host proteins for immune surveillance[42].

ER aminopeptidase 1 (ERAP1), the subject of this work, is a zinc aminopeptidase that is a major component of antigen processing[43]. In this
capacity, ERAP1 has a notable substrate length-dependent activity, where peptides longer than 8 amino acids are cleaved preferentially and with higher activity than shorter peptides\cite{44, 45}. ERAP1 also favors substrates with hydrophobic aminoterminal residues, and has complex pattern of substrate sequence preferences at particular positions along the peptide sequence\cite{46}. The structurally and functionally related protein ER aminopeptidase 2 (ERAP2) also preferentially cleaves long peptides, but the substrate sequence specificity of ERAP2 favors basic aminoterminal residues\cite{47}. ERAP1 and ERAP2 appear to act in conjunction to process peptides to a length amenable to MHC-I binding.

**Discovery of ERAP1**

Initial identification of ERAP1 from a human adipocyte cDNA library described it as an M1 leucine aminopeptidase (designated adipocyte-derived leucine aminopeptidase, A-LAP) related by sequence and genomic locus proximity to placental leucine aminopeptidase (PLAP)\cite{48} (otherwise designated insulin-regulated aminopeptidase, IRAP or leucyl-cysteiny1 aminopeptidase, LNPEP). *In vitro* hydrolysis assays using purified components showed that ERAP1(A-LAP) could digest the vasopressure-regulating peptide hormones angiotensin II and kallidin as well as several neuropeptides\cite{49}, though the physiological relevance of this activity was not described.
A second group characterized ERAP1 as a M1 aminopeptidase found in rat cells with the distinguishing feature that it was not inhibited by puromycin, and was termed puromycin-insensitive leucyl-specific aminopeptidase, PILS-AP[50]. This group also noted a strong enzymatic substrate preference for leucine and methionine at substrate aminotermi[50], although their data showing low activity for substrates with aromatic aminotermi is not consistent with later characterization of ERAP1.

Up to this point, enzymatic behavior and sequence homology had been established, but the physiological role of this aminopeptidase was not defined. In close succession, two groups independently identified it as a component of MHC-I antigen processing[51, 52]. This protein localized to the ER lumen, and as such was designated ER aminopeptidase 1 or ERAP1, which has become the predominant nomenclature (the term ER aminopeptidase associated with antigen processing, ERAAP, is also used but generally applies only to mice, as Rodentia lack the ERAP2 gene present in other vertebrates).

ERAP1 has also been described as an effector of cytokine receptor shedding (with the acronym ‘aminopeptidase regulator of TNFR1 shedding’ ARTS)[53-56], but the significance of this finding is unclear as other proteases have been reported as sufficient to fulfill this role[57, 58]. ERAP1 is also reported to participate in vasopressure regulation[59] and in macrophage activation[60-62]. These roles will not be described in further detail, but any physiological role involving ERAP1 catalysis may be affected by the results presented in this work.
ERAP1 substrate length preference

An early observation of ERAP1 activity was the relationship with ERAP1 aminoterminal activity and the polypeptide substrate length. York et al. reported that N-terminal extensions of known MHC-I peptide ligands were digested more efficiently by ERAP1 than the unextended MHC-I-binding peptide[44]. This was more thoroughly examined by measuring the hydrolysis of chemically synthesized peptide substrates by purified ERAP1 in vitro, which showed high hydrolysis activity for peptides between 8 and 25 amino acids in length, but greatly decreased hydrolysis for shorter substrates[63]. This minimum substrate length-dependent behavior was unusual compared to other peptidases; however, this behavior was noted to resemble that of MHC-I, which exhibits a minimum peptide length preference of 8 amino acids[63]. This discovery yielded the ‘molecular ruler’ model of ERAP1 activity, where the substrate peptide C-terminus binds an activating site on ERAP1 and increases the hydrolytic activity by some unknown mechanism. Substrates less than 8 amino acids long would be unable to simultaneously occupy the aminopeptidase active site with their N-terminus and bind the activating site simply due to the physical distance between these sites, which would be enforced by the structure of ERAP1[63].
**M1 family protein structure**

By sequence homology and structure, ERAP1 is an M1 family zinc aminopeptidase. These enzymes share an architecture consisting of three domains with a fourth domain present in some family members. The domains are classified numerically from N- to C-terminus.

Domain I is predominantly composed of β-sheets. It is located in close contact with domain II, which holds the active site of the enzyme. Domain II is a thermolysin-like fold, resembling the endopeptidase of the same name from *Bacillus thermoproteolyticus*. The active site coordinates a single zinc ion which interacts directly with the substrate peptide. The thermolysin-like fold has two conserved sequence motifs, HEXXH(X$_{18}$)E and GXMEN, which form much of the active site and the zinc-binding site. While it shares similarity with an endopeptidase, domain I enforces aminopeptidase activity by blocking the substrate binding site beyond the P1 pocket and contributing to an acidic cluster that binds the substrate peptide N-terminus[45, 64].

Domain III is another β-sheet domain that is only present in certain M1 aminopeptidases. It appears to serve a structural role, although this has not been studied extensively. Domain IV is an α-helical domain which in most M1 aminopeptidase crystal structures encloses the active site in a chamber separate from the bulk solvent. For aminopeptidases lacking domain III, domain IV consists of 9-10 helices[65], while for those containing domain III, domain IV
contains 16-17 helices[45, 66]; more helices allow domain IV to encompass a larger volume in the active site-adjacent cavity[64].

**M1 family conformation**

ERAP1 was the first M1 family aminopeptidase to be crystallized in two separate conformations[64]. Previously, most M1 aminopeptidases were crystallized in the closed conformation, with the active site and bound ligands sequestered from solvent. This state implied that an opening step is required to allow substrate binding to the active site and product dissociation from it. One family member, Tricorn-interacting factor F3 (TIFF3) from *Thermoplasma acidophilum*, was crystallized in an open state[67]. In this state, the conserved domain IV is hinged away from the other domains, exposing the active site. One notable change in the active site is the position of tyrosine 351. In this open structure, the tyrosine is not oriented toward the active site[67]. Structures of other M1 family members such as *E. coli* aminopeptidase N (ePepN) found the corresponding tyrosine (amino acid 381 in ePepN) oriented with the sidechain phenol in proximity to the catalytic zinc and contacting the bound substrate analog inhibitor bestatin. For *S. cerevisiae* leukotriene A4 hydrolase, another M1 family member, mutation of the catalytic tyrosine (amino acid 429) to phenylalanine caused a loss of aminopeptidase activity, indicating the importance of this phenol group to catalysis[68]. The sum of these prior
structures suggested that motion of the catalytic tyrosine might coincide with the large-scale conformational change of domain IV.

ERAP1 open and closed crystal structures confirmed that this tyrosine (amino acid 438 in ERAP1) did adopt a conformation-dependent position as observed for M1 family structures[64]. In contrast to this, open and closed structures of the mammalian aminopeptidase N (also designated CD13) showed the catalytic tyrosine (amino acid 472 in porcine aminopeptidase N) remained in the catalytically active position despite the conformation of domain IV[69-71]. This proves that active site perturbation connected to conformational change is not obligatory for M1 family aminopeptidases. Open and closed conformations of IRAP show an intermediate motion of catalytic tyrosine 549[72-74]. It remains to be determined if ERAP1 and IRAP, two of the members of the oxytocinase subfamily of M1 aminopeptidases, are broadly relevant models of M1 family behavior of conformationally activated catalytic site geometry, or if they are the exception.

Peptide C-terminal binding site

The molecular ruler model of ERAP1 activity hypothesized the existence of a dedicated substrate peptide C-terminal binding site that activates ERAP1 hydrolysis. Identification of this site would provide new insight into the mechanism of ERAP1 activation. Understanding the substrate requirements for
ERAP1 activity would also assist in predicting which MHC-binding peptides are present in cells, and might explain the behavior of certain ERAP1 disease-associated polymorphisms (see subsequent sections for additional details).

Crystal structures of ERAP1 fragments (containing domains III and IV only) serendipitously implicated a subsite within domain IV as a peptide substrate C-terminus binding site[75, 76]. In these structures, a C-terminal tag from one unit cell contacts the ERAP1 domain IV of the neighboring unit cell. This site of contact lies in the center of the helical repeats that comprise domain IV, which has been previously suggested as a binding site due to its geometry, its distance from the active site (which is consistent with the molecular ruler model of ERAP1 activity), and its electrostatic properties[64]. However, it remains unclear mechanistically how this site recognizes occupancy of a peptide C-terminus. No experiments are published demonstrating the importance of this site through mutagenesis or other functional experiments. Additionally, the substrate internal sequence specificity of ERAP1, which includes a preference for either hydrophobic or basic amino acids at the substrate C-terminus, is quite broad[46]. One explanation is that there may be multiple substrate C-terminal binding sites to accommodate the broad scope of substrates that ERAP1 hydrolyzes. As such, the crystallized binding site remains plausible but still inconclusive.
ERAP1 alters antigen processing and immune function

The role of ERAP1 on cell-mediated immune responses has been studied in mouse models as well as human samples. ERAP1 knockout (KO) mice have altered immune responses to pathogens[43, 77-79]. Hosts respond to pathogen infections by targeting a specific subset of peptide epitopes, a process termed immunodominance. These preferred epitopes are highly reproducible for model organisms expressing identical MHC proteins. However, ERAP1 KO mice have an altered immunodominance hierarchy. This indicates that some aspect of peptide processing has been altered in these mice. This alteration due to ERAP1 KO was found to impair survival in mice infected with *Toxoplasma gondii*[80]. No defect in survival was observed for the model pathogens lymphocytic choriomeningitis virus (LCMV) and Vaccinia virus[43, 77]. Pathogen-specific survival hypothetically results from differences in the importance of host CD8+ T cell responses in clearing the infection, which is related to the efficiency of MHC-I antigen processing of proteins from a given pathogen.

Another indication of the relevance of ERAP1 to immune response is the observation that ERAP1 is a target of immune evasion by pathogens as well as in oncogenesis. The pathogen human cytomegalovirus (HCMV), which is prevalent in most mammalian species, expresses two microRNAs (miRNAs) to suppress expression of ERAP1[81, 82]. This interferes with MHC-I antigen presentation of viral peptides to surveilling CD8+ T cells specific to HCMV[81]. This pathogen strategy has selected for a host ERAP1 single nucleotide
polymorphism (SNP) within one miRNA binding site that escapes the pathogen-induced decrease in expression[82].

It is also notable that alterations in ERAP1 expression occur in cancer. ERAP1 expression varies depending on the tissue progenitor of the cancer, likely conferring protection by disrupting tissue-specific antigen processing[83]. In two examples, ERAP1 activity destroys tumor-specific peptides for MHC-I presentation in a colon carcinoma cell line (CT26) and confers protection from immune response[84], but in a separate carcinoma cell line (16.113) ERAP1 produces peptide epitopes that induce a CD8+ T cell response[85]. This exemplifies the highly context-specific effect of ERAP1 activity, which is altered depending on the available peptides and the MHC alleles present.

Additional experimentation using syngeneic transplantation identified that the alteration in antigen processing is immunogenic in the absence of pathogen. ERAP1 KO cells transplanted into WT mice stimulate a concerted immune response, as does the opposite transplantation[86, 87]. This same response occurs when the transplanted cells are deficient in ERAP1 activity by application of an in vitro small molecule inhibitor L-leucinethiol rather than through genetic deletion[86]. A large fraction of these responding T cells react to a host peptide originating from the uncharacterized protein Fam49b[86]. This peptide is constitutively expressed but is destroyed by ERAP1 activity. This appears to be an endogenous system present in mice to monitor ERAP1 activity and to generate an immune response if ERAP1 activity is inhibited[86].
These latter examples highlight a critical dichotomy present in the cell-mediated adaptive immune system: the abundance of antigens during and after T cell development.

**T cell development**

As T cells develop, they generate their TCR as described previously. The inherently random nature of this event initially produces many genes that have negligible binding affinity for any MHC:peptide complex. These are not productive as antigen sensors due to their lack of binding sensitivity. Other TCR recombination events generate sensors that are highly reactive to host proteins or peptides (which normally bind MHC even in the absence of pathogen). These receptors are not desirable antigen sensors as they lack specificity.

In order to filter these two categories of TCRs, T cells undergo positive selection (where they must exhibit baseline binding sensitivity to survive) and negative selection (eliminating cells with high binding affinity to host MHC:peptide complexes, i.e. low specificity)[88]. This, in principle, leaves a pool of T cells with TCRs that are not self-reactive but are tuned with affinity just below the threshold for cell activation. A pathogen antigen that activates these T cells may or may not exist, but this information is not available to incorporate into the selection process prior to infection. The only parameters available for filtering during development are the specific host MHC haplotype (MHC molecules are highly polymorphic).
and the specific host peptides present. If either of these components change, T cells selected under one set of criteria may encounter a reactive MHC:peptide.

Additionally, the continuum of T cell reactivity against host antigens permits the possibility for autoimmune disease. Partially self-reactive T cells that survive negative selection during T cell development are generally prevented from activating upon recognition of a host antigen, a condition referred to as tolerance[89]. One model of the breakdown of tolerance states that an injury or inflammatory event such as a pathogen infection activates T cells somewhat nonspecifically and overcomes the suppression of these marginally self-reactive T cells[90, 91]. Once these cells begin a response against a host antigen, a positive feedback loop amplifies the effect.

Among the factors that contribute to loss of tolerance, it is apparent that incidence of autoimmune disease correlates with certain MHC alleles and with certain allelic variants of proteins that perform antigen processing, including ERAP1.

**ERAP1 polymorphism rs30187**

In humans, several prevalent SNPs have been identified within ERAP1. Among these, rs30187 (T/C), a nonsynonymous SNP within codon 528 (lysine/arginine respectively) has been identified as the most significant factor in
disease incidence and is a modulator of enzymatic activity. This SNP is common, with a global minor allele (T, encoding lysine) frequency of 0.4114[92].

**SNP rs30187 correlates with disease**

Polymorphism rs30187 was first identified as a correlate with hypertension[93]. ERAP1 with arginine 528 was shown to have decreased hydrolytic activity compared to lysine 528, leading to the hypothesis that decreased capacity of Arg528 ERAP1 to degrade the hypertensive hormones angiotensin II and kallidin resulted in prolongation of these hypertensive signals and increased prevalence of hypertension in humans carrying the “arginine 528” ERAP1 variant[94].

Subsequent GWAS identified rs30187 as a correlate with ankylosing spondylitis (AS) incidence[95, 96]. AS is an autoimmune disorder with arthritic symptoms that is highly heritable and is associated closely (but not exclusively) with the MHC-I allele HLA-B27[97]. Within the HLA-B27-positive population, ERAP1 SNP rs30187 “T” (encoding lysine 528) is a dose-dependent risk factor for AS incidence[96]. While the disease mechanism may be complex and multifactorial, the correlation of a particular MHC-I along with a particular component of class I MHC antigen processing implicates a CD8$^+$ T cell response in the disease process. SNP rs30187 “T” also correlates with psoriasis[98, 99] and Behçet's disease[100] incidence.
**SNP rs30187 effect on ERAP1 protein structure and function**

Biochemical and enzymatic analysis of recombinant ERAP1 with lysine/arginine substitution at position 528 revealed that the nonsynonymous polymorphism rs30187 significantly alters ERAP1 activity. Both variants express recombinantly in comparable quantities[94, 101]. Structurally, ERAP1 with either lysine or arginine at position 528 (in addition to other polymorphic site variants) have both been crystallized in the open conformation, and no notable alterations were observed[45, 64].

Variation at position 528 is known to alter the enzymatic properties of ERAP1. Arginine 528 ERAP1 has been described in many reports as a hypoactive variant of ERAP1 relative to lysine 528 ERAP1[64, 94, 102]. Peptide hydrolysis experiments under Michaelis-Menten conditions clarified that this altered activity is substrate-dependent; in the most distinct case (decamer peptide substrate LATFPDTLTY), arginine ERAP1 has increased $K_M$ and $k_{cat}$ compared to lysine 528[101]. For some peptides, substrate inhibition is also observed differentially for the different ERAP1 variants[101]. This indicates that the effect of position 528 on ERAP1 catalysis cannot simply be described as increasing or decreasing activity.

Proteomic analysis of MHC-bound peptides is a developing method to assess antigen processing and presentation. Polymorphism rs30187 has been reported to alter the MHC-bound peptidome, where the peptide N-terminal amino acid tends to be bulkier (greater molecular weight) and a weaker ERAP1 substrate
when lysine 528 ERAP1 is expressed relative to arginine 528[103, 104]. ERAP1 polymorphism-dependent differences in the MHC-bound peptidome may explain the disease correlates of these ERAP1 polymorphisms, as a hypothetical peptide might display some propensity for inducing an autoimmune response, and processing of this peptide might be altered by ERAP1 polymorphism. Alternatively, the MHC-I peptidome (all peptides bound to class I MHC in a given cell/tissue/organism) may differ in binding stability or some other parameter in the presence of either polymorphic variant. Alterations in this manner have been observed and may be sufficient to induce autoimmune disease[104].

**ERAP1 polymorphisms are inherited as discrete alleles**

The association of ERAP1 SNPs with disease does not fully describe the relevant ERAP1 alleles present in the human population. Several other ERAP1 SNPs have been identified that associate with disease[95, 96, 100]. These SNPs do not segregate independently from one another, but exhibit genetic heritability as a set of alleles[102, 105]. Classification of ERAP1 alleles and their disease correlation and effect on the MHC peptidome as haplotypes of alleles rather than separate SNPs is ongoing[102, 106-108].
Small-molecule inhibitors of ERAP1

Small-molecule inhibitors of aminopeptidases have been used as molecular probes, purification reagents, and therapeutic agents for several decades. One commonly used inhibitor is bestatin, also known as ubenimex, a dipeptide substrate mimic that was first discovered as a natural product of the bacteria *Streptomyces olivoreticuli*[109]. Another similar inhibitor is amastatin, which is a tetrapeptide mimic isolated from *Streptomyces* ME98-M3 strain[110]. These compounds are broad aminopeptidase inhibitors with ERAP1 inhibition constants between 10-300µM[48, 52, 94]. One interesting observation is the differential inhibition of ERAP1 polymorphic variants at position 528, where these compounds have inhibition constants three- to five-fold greater for arginine 528 ERAP1 than lysine 528 ERAP1[94].

ERAP1 crystal structures were determined with bestatin bound at the active site in both the open and closed conformations[45, 64]. Bestatin is oriented similarly in these two structures, with the hydroxyl and the carbonyl oxygen bracketing the ERAP1 catalytic zinc. (Figure 1.2) In the closed conformation, phenylalanine 433 is visualized in ERAP1, and makes a π-π interaction with the bestatin phenyl group; this phenylalanine is not ordered in the open conformation[45, 64].
Figure 1.2

Bestatin, a dipeptide mimic, binds the active site of ERAP1.
(a) Structural formula of bestatin. (b) Bestatin, green, bound to ERAP1 in the closed conformation, cyan(PDB ID 2YD0)[64]. Hydroxyl and carbonyl oxygens contact catalytic zinc, gray sphere. (c) Bestatin as in (b) bound to ERAP1 in the open conformation, magenta(PDB ID 3MDJ)[45].
Bestatin treatment has been shown to alter immune function. *In vitro* experiments using primary immune cells showed that bestatin boosted lymphocyte proliferation when cells were nonspecifically stimulated with molecules such as phytohemagglutinin (PHA) or the bacterial product lipopolysaccharide (LPS)[111]. Bestatin treatment in mice increases the severity of delayed hypersensitivity reactions[112, 113]. These reactions are a class of immune response mediated by T cells[4]. This immunostimulatory behavior was attributed to the inhibition of aminopeptidases other than ERAP1, as these experiments predated the discovery of ERAP1, but the inhibition of ERAP1 would alter antigen processing during a developing immune response and is a plausible mechanism to produce these results.

Another broad aminopeptidase inhibitor is L-leucinethiol[114]. This mercaptoamine compound is proposed to react with active-site zinc ions present in zinc peptidases[115]. This compound is also likely to react with other redox-active functional groups. Leucinethiol inhibits ERAP1 with submicromolar potency[52, 94].

Leucinethiol has been used in several experiments to test the effect of ERAP1 inhibition on antigen processing and immune function. As noted earlier in this work, leucinethiol-treated cells introduced into mice were shown to be immunogenic[86]. Leucinethiol treatment impaired antigen processing in cultured cells in a MHC allele-specific manner, indicating that MHC proteins have differential preferences for aminoterminally-processed peptides[116].
Zinc-chelating groups such as aminobenzamide have been investigated as ERAP1 inhibitors[117, 118]. Some derivatives have submicromolar potency for ERAP1, but these compounds are also inhibitory to the related oxytocinase subfamily aminopeptidases ERAP2 and IRAP. As these enzymes all function within antigen processing, selectivity is required to parse the functional effects of inhibition.

Phosphinic inhibitors such as DG013 also exhibit nanomolar potency for the oxytocinase subfamily of M1 aminopeptidases[119, 120]. This peptidomimetic scaffold binds the active site with the phosphinic group mimicking the tetragonal transition state of peptide bond hydrolysis. Phosphinic peptides have been crystallized bound to ERAP2 and IRAP to identify the determinants of substrate sequence recognition in these enzymes[72, 74, 119, 121]. These inhibitors have been utilized in some cell culture assays to inhibit aminopeptidases in antigen processing[119]. While some of these phosphinic compounds demonstrate selective inhibition for ERAP2, most have significant potency against all three members of the oxytocinase family.

Open questions regarding ERAP1

The enzyme ERAP1 remains the focus of scrutiny due to its physiological roles (especially within antigen processing) and the extremely complicated relationship between ERAP1 sequence variants and disease incidence. One
crucial aspect of ERAP1 function is the minimal substrate length preference. The molecular ruler model of ERAP1 proposed one model of allosteric activation by sufficiently long substrates, but the location of the site(s) of allosteric activation remains unclear, as is the mechanism in which allosteric binding relates binding to catalytic activation. Binding site identification would help define the amino acid sequence and length parameters that make a peptide a preferred ERAP1 substrate. This understanding would improve the accuracy of antigen processing prediction[122]. It would also test the validity of the molecular ruler model of ERAP1 function and might explain the effect of SNPs on ERAP1 activity.

The conversion of ERAP1 between states is demonstrated by two crystal structures, giving atomic-level detail of two states that ERAP1 may adopt, and computational simulation has predicted additional states that may exist, but the relationship of conformational change with ERAP1 catalysis remains hypothetical. The rates of conformational change are also of interest in order to dissect the enzymatic model of ERAP1 function. SNP-dependent effects on ERAP1 behavior have been proposed but remain untested. More broadly, the role of conformational change for the entire M1 aminopeptidase family is hypothetically relevant but has not been identified experimentally.

One approach currently used to study ERAP1 function is measuring changes in immune function or in antigen processing upon ERAP1 expression or inhibition. Much of this work has used genetic modulation such as gene knockouts[79], or alterations of protein expression by transfection[102] or RNA
interference[123]. While specific, these approaches do not permit perturbation of ERAP1 activity in a time-dependent manner. Quantifying the kinetics of ERAP1 antigen processing would inform the overall rate at which immune responses form. Development of truly selective small molecule inhibitors against ERAP1, ERAP2, or IRAP would allow deeper investigation of their respective roles in homeostasis and disease. Additionally, the use of small-molecule inhibitors to specifically alter ERAP1 activity provides a novel approach to modulate immune responses, both as an experimental tool to perturb in vivo models of disease and potentially as a clinically relevant therapeutic approach.

In this thesis I present the identification and characterization of ERAP1-specific small molecule inhibitors. Additionally, I report the observation of ERAP1 conformational change upon binding substrates and inhibitors, and identify that substitution of lysine or arginine at position 528 alters rates of conformational change.
CHAPTER II

Discovery of specific inhibitors of endoplasmic reticulum aminopeptidase 1

This chapter corresponds to a manuscript in preparation:


Author contributions

I expressed and purified proteins for assays, optimized scalable assays for high-throughput screening, and performed confirmatory activity assays. I also performed peptide hydrolysis assays, in silico docking, and cellular inhibition experiments, analyzed and interpreted data, and wrote this manuscript with input from coauthors.

ABSTRACT

ERAP1 is an endoplasmic reticulum-resident zinc aminopeptidase that plays an important role in the immune system by trimming peptides for loading onto major histocompatibility complex (MHC) proteins. Here we report discovery of the
first inhibitors selective for ERAP1 over its paralogs ERAP2 and IRAP. Compounds 1 (1-(2,5-difluorophenyl)sulfonyl-2-[2-(1H-indol-3-yl)ethyl]guanidine) and 2 (1-[1-(4-acetyl)piperazine-1-carbonyl)cyclohexyl]-3-(4-methylphenyl)urea) are inhibitors of ERAP1’s leucine aminopeptidase activity, and compound 3 (4-methoxy-3-[[2-piperidin-1-yl-5-(trifluoromethyl)phenyl]sulfamoyl]benzoic acid) is an allosteric activator of this activity. All three compounds inhibit ERAP1’s ability to hydrolyze a 10mer peptide substrate, and compounds 2 and 3 inhibit antigen presentation in a cellular assay. Compound 3 displays even higher potency for an ERAP1 variant that correlates with autoimmune disease incidence. These inhibitors provide mechanistic insight into the determinants of specificity for the three related enzymes ERAP1, ERAP2 and IRAP, and offer a new therapeutic approach of specifically inhibiting ERAP1 activity in vivo.

INTRODUCTION

Endoplasmic reticulum aminopeptidase 1 (ERAP1, also known as ARTS1, ALAP, ERAAP, or PILS-AP) is a 107kDa M1 family zinc aminopeptidase (EC 3.4.11.-)[48] ubiquitously expressed in somatic cells[51]. ERAP1 trims peptides in the endoplasmic reticulum prior to their presentation on class I major histocompatibility complex (MHC) proteins. ERAP1 has length specificity unique among aminopeptidases, efficiently trimming peptides longer than ~8 residues but sparing shorter peptides[45], matching the length preferences of MHC[63]. A model for ERAP1’s length dependence has been proposed whereby long but not
short peptide substrates are able to access a regulatory site to activate their own hydrolysis[45, 75]. ERAP1 exhibits broad but defined substrate sequence specificity with preference for bulky hydrophobic P1 and either bulky hydrophobic or basic PΩ'[46, 124].

ERAP1 processing can result in generation or degradation of peptides able to bind MHC, and as a result ERAP1 impacts a large fraction of the overall MHC-bound peptide repertoire[43, 87]. After loading, MHC traffics to the cell surface where are they are assessed by circulating T lymphocytes as part of immune surveillance of bodily tissues for infection or malignancy. Polymorphisms in ERAP1 are associated with increased susceptibility to T-cell mediated autoimmune diseases[125] and cancer[126, 127]. ERAP1 also has been implicated in regulation of blood pressure[59, 93] due to its ability to hydrolyze angiotensin II[49]. Alterations in vasopressure regulation in knockout mice lacking the gene thiol reductase ERp44 implicated this physiological effect stems from unregulated secretion of ERAP1, which is retained in the ER through intermolecular interaction with ERp44[59].

In humans the erap1 gene is found together with paralogs ERAP2 and IRAP in a gene cluster on chromosome 15 and the corresponding proteins form the oxytocinase subfamily of M1 aminopeptidases. Endoplasmic reticulum aminopeptidase 2 (ERAP2, also knows as L-RAP, EC 3.4.11.-) plays a role in antigen processing similar to ERAP1 but with a different substrate peptide sequence preference[124]. Insulin-regulated aminopeptidase (IRAP, EC
3.4.11.3), also known as placental leucine aminopeptidase (PLAP), oxytocinase, and leucyl-cystinyl aminopeptidase, plays an analogous function but in endosomes rather than the endoplasmic reticulum, processing peptides in the class I MHC cross-presentation pathway[128, 129]. IRAP also performs metabolic, neurological and endocrine functions[130-132]. Understanding the specific contributions of the different oxytocinase subfamily aminopeptidases is of interest to both basic and applied research.

ERAP1, ERAP2 and IRAP all have been implicated in processing antigens for presentation by MHC proteins, and several approaches have been used delineate each of their specific effects. Inhibitors with high ERAP1 specificity allow ERAP1 function to be probed in a manner distinct from genetic deletion[133], expression modulation via RNAi knockdown[44], and from less specific inhibitors[116]. While highly specific, genetic mutation or deletion does not give information on the kinetic role ERAP1 plays in antigen processing. RNAi knockdown shares this drawback to some degree, and additionally may be complicated by incomplete knockdown or off-target effects.

One approach that could be used to distinguish the roles of these aminopeptidases is treatment with small molecule inhibitors. However, highly specific inhibitors of ERAP1 have not been identified despite many efforts. Bestatin[109] and amastatin[110] have poor potency for ERAP1[48] and promiscuous inhibition of other aminopeptidases.[134, 135] Leucinethiol is a relatively potent ERAP1 inhibitor[114], but also inhibits many other
peptidases[136]. Other inhibitors act by interaction with the catalytic zinc and surrounding residues, which are well conserved among this enzyme family[117, 137]. Small molecule inhibitors of other M1 aminopeptidases have been identified, notably for PfA-M1, an enzyme expressed by *P. falciparum* during malaria infection[138]. Several potent peptide-based inhibitors of M1 aminopeptidases as well as other zinc proteases have been developed rationally using a phosphinate group as a substrate transition state analog[139-142]. However, as M1 zinc aminopeptidases reaction mechanisms and active site geometry are highly conserved, inhibitors in this class exhibit significant potency for multiple M1 aminopeptidases[120]. These compounds also are limited in their *in vivo* utility due to their similarity to peptides, which are hampered by proteolysis and cell membrane impermeability. As ERAP1, ERAP2 and IRAP each perform different but related immune functions, the effect of broad inhibition may have dire consequences. The clinical implications of either specific or broad inhibition are still unknown but may have a significant impact on treating human disease[143, 144].

In this study we developed a screening strategy to identify small-molecule compounds selective for ERAP1 over other oxytocinase family aminopeptidases ERAP2 and IRAP. Compounds with three different chemical frameworks were identified for optimization and testing using physiologically relevant substrate processing assays. Two of these compounds specifically inhibit ERAP1’s cellular
antigen processing activity. Compounds 1 and 2 target ERAP1’s active site, whereas compound 3 targets an allosteric regulatory site.

RESULTS

Assay development

ERAP1 enzymatic activity is conveniently assayed using the fluorogenic substrate leucine-7-amido-4-methylcoumarin (Leu-AMC), which is analogous to a dipeptide substrate[48]. Since this is shorter than preferred substrates, ERAP1 hydrolysis activity of Leu-AMC is substantially lower than for more physiologically relevant peptides of 8 more residues[45] but can be measured easily by fluorescence in microtiter format HTS assays. A preliminary screen using the LOPAC 1280 library of pharmaceutically relevant compounds identified several redox-active compounds as potent inhibitors of ERAP1, likely due to reaction with ERAP1’s disulfide bonds or free cysteine residues. Additionally, several compounds were identified that activated ERAP1’s Leu-AMC hydrolysis activity, an activity previously observed for short peptide substrates and attributed to binding a hypothesized regulatory site within the overall substrate binding region[45, 75].
Inhibitor screen

In order to identify novel ERAP1-specific inhibitors, we designed and implemented a high-throughput screen of the NIH’s Molecular Libraries Probe Production Centers Network library at the Broad Institute’s Probe Development Center. (Figure 2.1) We screened for compounds that inhibit ERAP1 Leu-AMC hydrolysis activity at a single concentration (10.7µM). Compounds that decreased ERAP1 Leu-AMC hydrolysis activity by 30% were filtered by cheminformatic analysis and by further assays for dose-dependent efficacy. Compounds showing dose-dependent inhibition were then counterscreened for inhibition ERAP2 or IRAP activity. This reduced the pool of candidate compounds to two Leu-AMC inhibitors.

We identified two compounds as specific inhibitors of ERAP1-catalyzed hydrolysis of Leu-AMC. (Figure 2.2b, bottom) These compounds (1 and 2) have distinct chemical scaffolds, where 1 is a phenylsulfonylguanidine and 2 is a piperazinylphenylurea. We tested compounds 1 and 2 in a subsequent confirmatory Leu-AMC hydrolysis assay and they exhibit ERAP1 Leu-AMC IC\textsubscript{50} of 9.2µM and 5.7µM respectively. They are also more than 100-fold more specific for ERAP1 than ERAP2 (Figure 2.2b, bottom) and IRAP (Figure 2.2c, bottom), and in some cases no activity for ERAP2 and IRAP was detected.

We tested the effect of these inhibitors on ERAP1-catalyzed hydrolysis of a decamer peptide WRCYEKMLALK, which is a more physiologically relevant ERAP1 substrate. (Figure 2.3) ERAP1 hydrolyzes the N-terminal tryptophan and
Figure 2.1
Screening pathway identifies three ERAP1-specific compounds. The MLPCN library was screened for alteration of ERAP1 activity. Inhibitors and activators were further screened to remove known problematic scaffolds and moieties as well as peptidomimetic compounds. Remaining compounds were then tested for potency with a cutoff EC$_{50}$ of 10µM. We observed several candidate compounds contained coumarinyl groups and removed these out of concern for false positives. We then counterscreened the compound pools for activity on ERAP2 and IRAP.
Hit compounds are highly specific for ERAP1.

Enzyme activity was measured using dipeptide substrate analog X-7-amido-4-methylcoumarin (X-AMC), where the aminoterminal residue X is leucine (ERAP1, IRAP) or arginine (ERAP2), corresponding to the substrate preference of each enzyme[128]. Hydrolysis was measured over a range of compound concentrations for (a) ERAP1, (b) ERAP2, or (c) IRAP. Activity was normalized to the activity of the enzyme in the presence of DMSO. Shown are representative data (mean ± SD, N=3) from one of two experiments. Data points were fit using a sigmoidal curve with the top or bottom value for inhibitor or activator curves, respectively, constrained to 100%.
Figure 2.3
Hit compounds inhibit ERAP1-catalyzed peptide hydrolysis.
(a) ERAP1 catalyzes hydrolysis of the N-terminal tryptophan from the peptide WRCYEKMALK. MALDI spectrum after 30-minute incubation with ERAP1. Calculated masses are shown at top, and observed peak m/z are displayed on the graph. (b) Hit compounds inhibit ERAP1-catalyzed peptide hydrolysis. Representative data (n=2) normalized to DMSO control condition shows inhibition of peptide hydrolysis. Mean and SD IC₅₀ values derived from eight individual dilution series are reported.
Figure 2.4
Compound 2 structure-activity relationship analysis.
Representative peptide hydrolysis data shown (n=2) demonstrates most alterations at each position decreases potency, except for one R^3 variant, 10. Compound 2 (yellow) data is shown in each graph for comparison. Variants replace the circled group with the moiety depicted in the legends.
releases the nonamer product peptide RCYEKMALK. This assay quantifies product peptide production by LC-MS, and so does not require a non-natural probe for detection. (Figure 2.3a) All compounds inhibited peptide hydrolysis in a dose-dependent manner. (Figure 2.3b) We observed that peptide hydrolysis IC$_{50}$ values closely matched the respective IC$_{50}$ values for Leu-AMC hydrolysis.

**Inhibitor structure-activity relationship analysis**

After determining these compounds inhibit ERAP1-catalyzed hydrolysis of this peptide substrate *in vitro*, we created derivatives to map the interaction of chemical moieties with ERAP1. We selected compound 2 due to its greater ease of synthesis. (Figure 2.4) Variation at most sites resulted in a loss of potency. One variant, wherein a 4-methylphenyl group at the R$_3$ position is replaced with a 4-chlorylphenyl group, showed a two-fold IC$_{50}$ reduction in both Leu-AMC and peptide hydrolysis assays.

**Activator screen**

Based on results from the LOPAC preliminary screen and previous work on allosteric activation of ERAP1[45, 75], we also screened for compounds from the initial screen that increased Leu-AMC activity. We initially selected compounds that increased Leu-AMC hydrolysis by 40% (3 standard deviations) over control for further characterization. (Figure 2.1, on right) After dose-dependent
Figure 2.5
Compound 3 structure-activity relationship analysis.
Representative peptide hydrolysis data shown (n=2) demonstrates most alterations at each position decreases potency, except for one R³ variant, 67. Compound 3 (blue) data is shown in each graph for comparison. Variants replace the circled group with the moiety depicted in the legends.
confirmation, cheminformatic analysis, and selectivity counterscreening as for inhibitors previously, we identified one ERAP1-specific compound, 3.

Compound 3 activates ERAP1 hydrolysis of Leu-AMC (Figure 2.2a, top), with an AC50 of 3.7µM and a maximum activity of 4.1-fold over control. This compound inhibits ERAP2 with an IC50 greater than 200µM, and has no detectable effect on IRAP activity. (Figure 2.2b,c; top) As previously observed for short peptides[45], compound 3 activates Leu-AMC hydrolysis but inhibits peptide hydrolysis. (Figure 2.3)

**Activator structure-activity relationship analysis**

Derivatization of compound 3 produced many compounds with decreased potency. (Figure 2.5) One variant at position R4 (substituting a 1-methyl-piperazinyl group for a piperidinyl group) had equivalent activity to the parental compound. None of the modifications tested resulted in significant improvement in potency. We suggest that this compound might make numerous contacts with ERAP1 across its entire structure, with contacts at many sites required for activity.

**Molecular docking identifies compound binding sites**

The identification of binding sites for these three hit compounds may help in the design of more potent compounds. In order to determine binding sites, we
docked compounds *in silico* to a structure of ERAP1 based on an existing crystal structure[64]. (Figure 2.6) Compounds 1 and 2 are predicted to bind at the active site, making interactions with the catalytic zinc and some catalytic residues in domain II. Both compounds also interact with residues outside the canonical M1 family active site, where 1 interacts closely with domain IV, and 2 makes contacts with portions of domain II distal from the active site. Compound 3 docked in a pocket formed at the interface between domains II and IV in the closed ERAP1 structure, far from the active site.

**Inhibition of intracellular peptide processing**

Among the potential uses of specific ERAP1 inhibitors is the application of these compounds *in vivo* for therapeutic treatment. To evaluate potential utility in this application we tested all hit compounds in a cellular assay modified from a previously described method[44]. We measured the amount of SIINFEKL (a model class I peptide) bound to MHC H-2 Kb at the cell surface using an antibody for this specific complex. (Figure 2.7a) To quantify peptide aminoterminal processing, we compared the amount of MHC:SIINFEKL complex on cells expressing an N-extended SIINFEKL peptide to a control condition where the SIINFEKL peptide is generated without aminopeptidase activity.
**Figure 2.6**
*In silico* docking identifies binding sites for hit compounds.
(a) Hit compounds (shown as spheres) docked to closed ERAP1 docking model (shown as cartoon, colored by domain). (b) Compound binding sites. ERAP1 surface is colored by domain as in (a). Compounds are shown as sticks.
**Figure 2.7**

**Two hit compounds inhibit ERAP1 activity in cells.**

(a) Cartoon depicting experimental design. Cells express SIINFEKL peptide (♦) with N-terminal extension (◆). ERAP1 (◆) must trim the peptide to allow binding to MHC protein (◆). ERAP1 activity is quantified by measuring surface MHC:SIINFEKL complex with a specific antibody. (b) ERAP1 inhibition using the nonspecific zinc peptidase inhibitor leucinethiol (Leu-SH). Samples are gated on GFP+ singlets prior to median fluorescence intensity (MFI) quantification of surface MFI:SIINFEKL complex. (c) Compounds 2 and 3 are active in cell processing assay at 50µM. MFI is normalized to DMSO sample and to a control well with cells expressing ubiquitin-SIINFEKL which is processed independent of ERAP1 activity. Shown are representative data (single well) of two experiments. (d) Compounds 2 and 3 are active in cell processing assay in a dose-dependent manner. Data was processed as in (c). Data points from three separate experiments are plotted together with one global sigmoidal curve calculated for each compound.
Treatment of cells with the nonspecific inhibitor leucinethiol decreases staining intensity only in cells expressing the N-extended peptide. (Figure 2.7b) Compounds 2 and 3 inhibit peptide processing in this cellular assay, while 1 was inactive at 50µM. (Figure 2.7c) The inhibition is dose-dependent, and compound 3 exhibits the highest potency with IC$_{50}$ of 1.04µM. (Figure 2.7d)

**Differential efficacy of compound 3 on ERAP1 alleles**

ERAP1 has several alleles present at high frequency in the human population, some of which correlate with susceptibility or resistance to autoimmune diseases such as ankylosing spondylitis[108]. We tested the effect of compound 3 on different ERAP1 alleles at high substrate concentration, necessitating the use of the dipeptide substrate leucine-p-nitroanilide (Leu-pNA) rather than Leu-AMC, which exhibits significant inner filter effect at concentrations greater than 500µM[45]. Compound 3 activates Leu-pNA hydrolysis at sub-K$_M$ substrate concentrations (Figure 2.8c) as was observed for Leu-AMC, but compound 3 inhibits ERAP1 hydrolytic activity under conditions of high dipeptide substrate concentration, Interestingly, a disease-associated allele (K528/Q725, allele II) is preferentially inhibited with a sub-micromolar IC$_{50}$, 10-fold lower than the disease-protective allele (R528/E730, allele IV). (Table 2.1)
Table 2.1

**Compound 3 exhibits ERAP1 allele-specific IC$_{50}$ difference.**

Several ERAP1 alleles are present in the human population, some at high frequency[108]. ERAP1 hydrolysis activity of leucine-$p$-nitroanilide was measured and datapoints were fit with a sigmoidal curve. Reported are representative derived IC$_{50}$ and confidence intervals from one of two experiments.

<table>
<thead>
<tr>
<th>Allele</th>
<th>SNP</th>
<th>Frequency (%)</th>
<th>IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>528</td>
<td></td>
<td>mean</td>
</tr>
<tr>
<td>IV</td>
<td>R</td>
<td>E</td>
<td>43</td>
</tr>
<tr>
<td>III</td>
<td>R</td>
<td>Q</td>
<td>0.1</td>
</tr>
<tr>
<td>II</td>
<td>K</td>
<td>Q</td>
<td>27</td>
</tr>
</tbody>
</table>
Mechanisms of inhibition or activation

The ERAP1-specific compounds described in this work display either inhibition or activation of ERAP1-catalyzed hydrolysis of dipeptide substrate analogs. To further assess their behavior, we analyzed the mechanism of inhibition of these three compounds by measuring their perturbation of ERAP1 Leu-pNA-cleaving activity under Michaelis-Menten conditions. Under uninhibited conditions ERAP1 exhibits cooperative behavior for this short substrate, requiring modification to the equations used in data analysis. Compounds 1 (Figure 2.8a) and 2 (Figure 2.8b) competitively inhibit ERAP1 under these conditions (altering the $K_{\text{prime}}$ component of the cooperative Michalis-Menten equation while not affecting $V_{\text{max}}$). This indicates that binding of these inhibitors is mutually exclusive with substrate binding, possibly (but not necessarily) by binding ERAP1 where substrate also binds.

Compound 3 exhibits much more complicated behavior, where activity is increased under low substrate conditions but decreases at high substrate concentration. (Figure 2.8c) This data was fit with a cooperative Michalis-Menten equation with an additional term allowing for substrate inhibition, with $V_{\text{max}}$ constrained to the value observed under uninhibited conditions. At increasing concentrations of compound 3, the Hill coefficient decreases, which may account for the activating behavior at low substrate concentration. Compound 3 may occupy the allosteric binding site with higher affinity than Leu-pNA, causing the apparent mitigation of the cooperative behavior. Additionally, the substrate
[Leu-pNA] (M)

L-pNA hydrolysis rate (s$^{-1}$)

[compound B] (nM)

Residuals

<table>
<thead>
<tr>
<th>[1] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>6.25</td>
</tr>
<tr>
<td>12.5</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
</tbody>
</table>

K$_i$ = 63µM
h = 1.2
K$_{half}$ = 1000µM
V$_{max}$ = 11.4 mol(pNA)*mol$^{-1}$(ERAP1) * s$^{-1}$

[Leu-pNA] (mM)

[compound D] (nM)

Residuals

<table>
<thead>
<tr>
<th>[2] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>2.5</td>
</tr>
<tr>
<td>12.5</td>
</tr>
</tbody>
</table>

K$_i$ = 4.5µM
h = 1.2
K$_{half}$ = 830µM
V$_{max}$ = 15.8 mol(pNA)*mol$^{-1}$(ERAP1) * s$^{-1}$

<table>
<thead>
<tr>
<th>[3] (µM)</th>
<th>h</th>
<th>K$_{half}$ (µM)</th>
<th>K$_i$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.2</td>
<td>740</td>
<td>N/A</td>
</tr>
<tr>
<td>0.1</td>
<td>1.2</td>
<td>750</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0</td>
<td>790</td>
<td>27</td>
</tr>
<tr>
<td>2.5</td>
<td>0.96</td>
<td>730</td>
<td>8.7</td>
</tr>
<tr>
<td>12.5</td>
<td>0.91</td>
<td>710</td>
<td>6.1</td>
</tr>
<tr>
<td>62.5</td>
<td>0.73</td>
<td>1000</td>
<td>6.9</td>
</tr>
</tbody>
</table>

K$_i$ = 63µM
h = 1.2
K$_{half}$ = 1000µM
V$_{max}$ = 11.4 mol(pNA)*mol$^{-1}$(ERAP1) * s$^{-1}$
Figure 2.8
Mechanisms of small compound alteration of ERAP1 activity.
ERAP1 hydrolysis of dipeptide substrate Leu-pNA is altered by hit compounds. ERAP1 exhibits cooperative behavior for Leu-pNA hydrolysis so data were fit with allosteric sigmoidal models. ERAP1 allele II was used for these experiments. 

\[ h = \text{Hill coefficient}, \quad V_{\text{max}} = \text{maximum reaction rate}, \quad K_{\text{half}} = [\text{Leu-pNA}] \text{ at } 0.5 \times V_{\text{max}}, \quad K_I = \text{inhibition constant}. \]

(a) Compound 1 competitively inhibits ERAP1 Leu-pNA hydrolysis. Shown is representative data from one of two experiments. (b) Compound 2 competitively inhibits ERAP1 Leu-pNA hydrolysis. One experiment was performed. (c) Compound 3 induces complex behavior where ERAP1 Leu-pNA hydrolysis is inhibited at high substrate concentration but is activated at low substrate concentration (on right). Data points for conditions containing compound 3 are fit to an allosteric sigmoidal model modified to account for substrate inhibition, with all curves constrained with \( V_{\text{max}} \) equal to 14.78 mol pNA per mol ERAP1 per second (determined from unconstrained fit of activity in the absence of compound 3). One experiment was performed.
inhibition constant decreases more than 10-fold as compound 3 concentration increases. This results in decreased ERAP1 activity at high substrate concentration. Compound 3 may stabilize a form of ERAP1 that retains activity but is more amenable to binding Leu-pNA substrate in a nonproductive manner. Overall, the behavior of compound 3 does not fit in a classical mode of small-molecule modulation.

**DISCUSSION**

We implemented a high-throughput screen for short substrate inhibitors and activators of ERAP1. Counterscreens against ERAP2 and IRAP activity and cheminformatic filtering removed promiscuous compounds and identified three ERAP1-specific hit compounds. These all inhibited hydrolysis of a decamer peptide, similar in length to substrates that ERAP1 digests *in vivo*. Two compounds also inhibited ERAP1 activity in a cellular assay.

Compounds 1 and 2 are novel non-peptidomimetic ERAP1-specific inhibitors that bind ERAP1’s active site. The highly conserved catalytic sites among the oxytocinase subfamily members presents an interesting question of how these compounds maintain specificity while acting as short substrate inhibitors. Docking shows that compound 1 contacts domain IV, which is more variable among the oxytocinase subfamily. Compound 2 contacts domain II primarily, where the active site is located, and interacts directly with the active site zinc atom. However, compound 2 is long enough that it extends 14.8Å from this zinc
atom, interacting with ERAP1 residues outside the catalytic site. These binding modes suggest that specificity may be attained by binding, at least partially, outside the active site.

The discovery of compound 3 identifies the first non-peptide small molecule that allosterically activates ERAP1 hydrolysis activity of Leu-AMC. Previous work showed ERAP1 Leu-AMC hydrolysis activity is increased by peptides two to eight amino acids long[45]. Further research will investigate if these peptides and compound 3 are competing for one binding site. It is possible that multiple allosteric sites are present in ERAP1’s large internal cavity, and activity measurements in the presence of mixtures of activators can determine this. A separate pocket in domain IV has been previously indicated as the binding site for peptides[75, 76].

Hit compound 3 is an aryl sulfonamide. Another aryl sulfonamide has been identified as an IRAP inhibitor, which is predicted by in silico docking experiments to bind the active site zinc[145, 146]. However this mode of inhibition is unlikely for compound 3, as it activates hydrolysis of Leu-AMC, and therefore does not occlude the catalytic site. Docking simulations of compound 3 identified a binding site 28Å from the active site zinc. This indicates that these two aryl sulfonamide compounds do not share a binding site, and have distinct modes of inhibition.

The two opposite effects on Leu-AMC hydrolysis of these three compounds preclude at least two separate binding sites on ERAP1. Interestingly, among the three hit compounds, the activator 3 exhibits the highest potency for ERAP1 in all
 assays, including a cellular assay. Since screening small molecules for modulation of allosteric activation identified a novel hit compound that may be active \textit{in vivo}, we propose that designing future screens to identify activators of enzymes that degrade biological polymers (proteases, glycosidases, and nucleases) may also yield novel hit compounds inaccessible by inhibition screens.

The inhibitory activity of two hit compounds in an intracellular antigen processing assay suggests that use of these compounds can probe the specific role of ERAP1 \textit{in vivo}. Lack of ERAP1 activity is known to elicit a cell-mediated immune response[86, 87], presumably by allowing presentation of antigenic peptides that are normally degraded by ERAP1, but the immunological and possible therapeutic relevance of transient specific inhibition is not well understood. For example, specific ERAP1 inhibition might modulate a mounting immune response by temporarily disrupting the class I MHC peptidome. This disruption would likely impair a T-cell mediated response, which normally expands as T-cells recognize a peptide antigen bound to MHC. Obscuring a peptide antigen in this manner by altering its trimming could be beneficial if this T-cell response has deleterious effects, such as in psoriasis[147, 148] or ankylosing spondylitis[149]. M1 aminopeptidase inhibition has also been correlated with increased antitumor immune response, another potential field of interest[84]. ERAP1 expression is altered in malignancies of different tissue origins[83], and in cases where ERAP1 overexpression protects cancerous cells
from immune detection by degrading antigenic peptides ERAP1 inhibition may produce a productive immune antitumor response.

MATERIALS AND METHODS

Protein expression and purification

Human ERAP1 (UniProtKB Q9NZ08 allele IV with additional SNPs Asp346 and Arg514, and alleles II and III)[108], human ERAP2 (UniProtKB Q6P179 with Asn392), and mouse IRAP (UniProtKB Q8C129 lumenal domain) were expressed recombinantly by baculoviral infection of High Five cells grown in SFX serum free media (Hyclone). All ORFs contain endogenous signal sequence for ER translocation and C-terminal 6His tag. Three days after infection by mixing cells (7 x 10^5 cells per mL) with 1% (v/v) virus preparation, culture supernatant was collected by centrifugation and concentrated and buffer exchanged >100 fold into Ni-NTA binding buffer (50mM phosphate pH8, 300mM NaCl, 10mM imidazole). After filtration, His-tagged protein was captured and eluted from Ni-NTA-agarose resin (Qiagen) in buffer containing 100mM imidazole. Enzyme stocks were characterized by SDS-PAGE, and by measuring Leu-AMC (ERAP1 and IRAP) or Arg-AMC (ERAP2) hydrolysis activity.
High-throughput library screen

High-throughput screen reaction conditions were 7µL per well in 1536 well plate format. Reactions were carried out using 20mM Tris pH 7.5, 100mM NaCl, 0.01% (w/v) BSA, 0.4% (v/v) DMSO, 1ng/µL enzyme, and 10.7µM substrate. After mixing, plates were incubated for 1 hour at room temperature, and then fluorescence emission was measured at 450nm with 380nm excitation. Z'-scores ranged from 0.8-0.94, using 100µM leucinethiol as a positive control for inhibition.

Confirmatory and selectivity X-AMC hydrolysis assays

Reactions were carried out in 100µL per well in black flat-bottom 96 well polypropylene plates (Greiner Bio-One) in 20mM Tris pH 7.5, 100mM NaCl, 0.01% (w/v) BSA. Enzyme (400ng ERAP1, 12.5ng ERAP2 or 40ng IRAP per well) and hit compounds were mixed, and then substrate (final concentration 100µM Leu-AMC for ERAP1 and IRAP, 10µM Arg-AMC for ERAP2) was added to start the reaction. Fluorescence at 380/460nm was measured using a BMG POLARstar OPTIMA once every 10 minutes starting after addition of substrate. Reaction rates were quantified by calculating the rate of fluorescence change. Each plate contained controls with DMSO alone, or with 100µM each leucinethiol and DTT. All reactions were normalized using these two conditions as 100% and 0% activity, respectively. To determine EC$_{50}$, normalized datapoints were fit with a sigmoidal curve constrained to 100% (top) for inhibitors or 100% (bottom) for activators.
**Leu-pNA hydrolysis assay**

To assay the effect of compound 3 on ERAP1 at substrate concentrations greater than substrate $K_M$, we used Leu-pNA rather than Leu-AMC. Reactions were performed in 100µL volumes in clear polystyrene 96-well plates (Corning), 20mM Tris pH7.5, 100mM NaCl, 0.01% (w/v) BSA. Assay conditions used 200ng per well ERAP1 (allele II or III or IV), 7mM Leu-pNA (with allele IV) or 7.5mM Leu-pNA (alleles II and III), and serial dilutions of compound 3 from 200µM to 12nM. Absorbance at 405nm was measured at 5 minute intervals. Reaction rates were calculated as the rate of absorbance change. To determine IC$_{50}$, datapoints were plotted as a function of concentration of compound 3 and fit with a sigmoidal curve.

To determine modes of inhibition for the three hit compounds, ‘allele II’ ERAP1 (200ng per well) was assayed as before but with a range of Leu-pNA concentrations (5mM-39µM). Reactions were carried out as previously described. To determine inhibition constants for compounds 1 and 2, data curves were fit globally to a competitive inhibition Michalis-Menten equation modified to account for cooperative behavior:
Compound 3 displayed activation at low substrate concentration followed by inhibition at high substrate concentration, and was fit to a substrate inhibition model to approximate the observed behavior.

\[
V = \frac{V_{\text{max}} \times S^h}{K_{\text{half}}^h + S^h \times \left(1 + \frac{S}{K_i}\right)}
\]

\(S\) = substrate concentration  \(K_{\text{half}}\) = \(S\) when \(V = 0.5 \times V_{\text{max}}\)

\(h\) = Hill coefficient  \(V_{\text{max}}\) = maximum catalytic rate

\(K_i\) = inhibitory substrate binding constant

**Peptide hydrolysis assay**

Reactions were carried out in 50µL volumes in 96 well V-bottom plate format with 20mM Tris pH7.5, 100mM NaCl, 0.01% (w/v) BSA. Purified ERAP1 (20ng/well) and hit compounds were mixed, and then substrate peptide (WRCYEKMALK, 10µM final concentration) was added to start the reaction. After ten minutes the reaction was stopped by addition of 25µL 1.2% TFA. Product peptide (RCYEKMALK) was quantified using RapidFire LC-MS (Agilent, Pure Honey Technologies) by summing the area under the curve. Each plate contained controls with DMSO alone, or with TFA stop solution added prior to adding ERAP1. All reactions were normalized using these two conditions as 100% and 0% activity, respectively. IC\(_{50}\) was determined by fitting normalized datapoints with a sigmoidal curve constrained to 100% (top) for inhibitors or 100% (bottom) for activators.
**In silico Docking**

Compound structures were generated initially using the PRODRG server[150]. Docking simulations and energy minimization was performed using Maestro Schrödinger version 2017-2[151]. Compounds structures were prepared for docking using LigPrep (pH7.5 ± 0.5) and docked using Glide onto a structure of ERAP1 allele II based on the closed conformation, (PDB ID 2YD0[64]) which was prepared using Prime. A search volume of 76Å×76Å×76Å was used, encompassing the entire enclosed surface of ERAP1 concurrent with the active site (ligand diameter midpoint search volume 40Å x 40Å x 40Å). Compounds 1 and 2 were docked without any excluded volume. Compound 3 docking enforced a 6Å-radius spherical exclusion volume centered on the active site zinc atom, as this compound is known to activate short substrate hydrolysis and so the active site is likely not occluded upon binding compound 3. For compound 3, twelve out of thirteen hits docked at the indicated pocket, with Glide scores of -4.261 to -0.243. The remaining dock solution bound at the catalytic zinc (2.37Å at closest) and was ruled out based on known enzymatic activation effect. Docking solutions were then inspected and one solution per compound was selected for further energy minimization using Embrace, using the OPLS3 force field, PRCG method, 5000 maximum iterations, gradient convergence (convergence threshold 0.05), energy difference mode.
Cellular antigen processing assay

HeLa cells stably expressing H-2 K\(^{b}\) were grown to 90% confluency in 24 well plates and infected with recombinant Vaccinia virus containing a cassette to express the ovalbumin epitope SL8 (SIINFEKL) at the C-terminus of ubiquitin, or a cassette which contains the same epitope preceded by a ER translocation signal sequence and the sequence LEQLE, which are immediately N-terminal to the SL8 epitope in the ovalbumin sequence. Both viral strains also express GFP with an IRES translation start site in line with the SIINFEKL epitope ORF. At the time of infection, cells were treated with inhibitor. After 16-24 hours, the cells were pipetted off the plate and stained with 25D1 antibody specific for SL8 in complex with H-2 K\(^{b}\), followed by Alexa Fluor 647-goat anti mouse secondary antibody (Thermo Fisher). Cells were then fixed in 4% formalin and analyzed by flow cytometry by gating on singlet GFP+ population and quantifying Alexa 647 median fluorescence intensity (MFI).

Data analysis

Data fitting was performed using Graphpad Prism 7[152]. Structural modeling data was analyzed and figures were prepared using PyMOL[153].
ACKNOWLEDGEMENTS

HeLa expressing H-2 K\textsuperscript{b}, 25D1 cells for antibody production, and vaccinia virus strains were generously provided by Kenneth Rock (UMMS). The LOPAC screen was performed with the assistance of the UMMS Chemical Screening facility. The HTS screen data is submitted as PubChem AID 652221, “Broad Institute Small Molecule Probes of ERAP-1 Inhibitor Probe Project”. We are grateful to Patti M. Aha and Catherine Communal for assistance in organization and coordination of this collaborative effort and for assistance in submitting findings to PubChem. The cell-based assay used for this research was derived from a HeLa cell line. Henrietta Lacks, and the HeLa cell line that was established from her tumor cells without her knowledge or consent in 1951, have made significant contributions to scientific progress and advances in human health. We are grateful to Henrietta Lacks, now deceased, and to her surviving family members for their contributions to biomedical research.
CHAPTER III

Large-scale conformational change during ERAP1 catalysis explains allostery and defines mechanism of enzymatic activity of disease-associated polymorphism

This chapter corresponds to a manuscript in preparation:


Author contributions

I engineered, expressed, and purified proteins for SAXS and crystallography. I measured SAXS data and performed data analysis. I performed crystallization trials, optimized hit conditions, looped crystals and collected diffraction datasets, and solved crystal structures. I prepared grids for electron microscopy and analyzed images, picked particles, and analyzed class averages. I performed some enzymatic assays. I performed photocrosslinking reactions and analyzed final LC-MS/MS datasets. I further analyzed in silico docking solutions (initial results presented in Chapter II), and performed in silico analysis of electrostatic interactions with position. Lastly I prepared this manuscript with input from coauthors.
ABSTRACT

The enzyme ERAP1 contributes to the generation of antigenic peptides which are used by the immune system to detect disease. Previous structural studies determined two ERAP1 conformations by crystallography. Prior enzymatic data suggested that conformational alteration occurs during catalysis but the specific nature of this mechanism is only hypothetical. We wanted to determine the role of conformational exchange in the ERAP1 catalytic cycle. We measured the conformational state of ERAP1 and observed changes in protein conformation in the presence of substrates and inhibitors. Structural and enzymatic data suggest that active site structural reconfigurations are physically linked to the closing step, providing a mechanism for allosteria. Polymorphism at position 528 alters the stability of the closed state by affecting the energetics of an interdomain electrostatic interaction. These results clarify the steps required in ERAP1 catalysis and demonstrate the importance of conformational dynamics within the catalytic cycle.

INTRODUCTION

An aminopeptidase designated endoplasmic-reticulum aminopeptidase 1 (ERAP1) is a key determinant of antigen processing and presentation. In this role, ERAP1 activity digests peptides that bind to major histocompatibility complex (MHC) proteins in a process that both creates and destroys peptides
before presentation to T cells\[79]. Polymorphisms within ERAP1 correlate with disease incidence\[127], likely within defined alleles in the human population\[108]. The ERAP1 single nucleotide polymorphism (SNP) with the highest disease association is rs30187, a nonsynonymous SNP with the disease-correlated variant encoding lysine 528 (otherwise arginine 528)\[96]. The exact mechanism resulting in disease correlation is unknown but the effect is highly correlated with the presence of specific MHC alleles\[154, 155], and ERAP1 SNP rs30187 alters the set of peptides bound to MHC\[103, 104]. In biochemical assays, lysine/arginine 528 substitution alters ERAP1-catalyzed hydrolysis of peptides in a substrate-dependent manner\[101].

ERAP1 is a member of the M1 aminopeptidase family, which is found in all domains of life and plays numerous functional roles\[143]. These proteins contain a thermolysin-like domain\[156] (II) which holds the active site. The M1 aminopeptidase family has a conserved HEAT-motif-repeat domain (IV) forming a bowl-shaped C-terminal structure\[143]. In most crystal structures of this family (54 out of 69 as of October 7, 2018) domain IV is closely associated with domain II and forms an internal pocket that separates the active site from the bulk solvent. This physically blocks substrate binding and product release, and so it was hypothesized that these aminopeptidases change structure in order to accommodate these steps\[143].

The first crystal structures of ERAP1 added some complexity to the mechanistic model of this protein. These structures showed two separate
conformations that ERAP1 adopted during crystallization, which was the first structural determination of one M1 aminopeptidase in two states. One matched the closed conformation that has been observed in other family members[64]. Two other crystal structures captured an open conformation, in which domain IV makes a somewhat-rigid-body translocation away from domain II, exposing the internal cavity to bulk solvent[45, 64]. Simultaneously, the active site reorganizes, with tyrosine 438 rotating away in a position not optimal for catalysis. This suggested a connection exists between conformation and catalytic activity. A similar motion of the catalytic tyrosine 549 was later observed in crystal structures of insulin-regulated aminopeptidase (IRAP)[72-74]. For the mammalian M1 family aminopeptidase CD13, crystallized open structures do not demonstrate conformation-dependent catalytic tyrosine inactivation[69], indicating that this may be a particular feature of the oxytocinase subfamily of M1 aminopeptidases (consisting of ERAP1, ERAP2, and IRAP)[157].

A connection between the active site and a distal binding site within ERAP1 has been theorized previously to explain ERAP1 enzymatic behavior. ERAP1 hydrolysis of dipeptide substrates such as leucine-\(\text{p}\)-nitroanilide (Leu-\(\text{p}\)NA) shows allosteric behavior deviating from Michaelis-Menten parameters, suggesting that multiple short substrate molecules bind simultaneously[45]. Additionally, catalysis of short substrates is increased by the presence of short peptides 3-14 residues in length[158, 159]. These findings, along with data showing preferential hydrolysis of substrates longer than 8 amino acids[44],
support a model that an optimal substrate peptide simultaneously binds the active site and some regulatory site that activates catalysis[63].

We examined the conformational state of ERAP1 in solution and observed ERAP1 conformational changes upon substrate and inhibitor binding. Crystal structures of two substrate-mimic inhibitors bound to ERAP1 provide new snapshots of ERAP1 during the substrate-binding process, conceptually connecting a sensor of active site occupancy, the catalytic tyrosine, and a disulfide bond that acts as a fulcrum to connect 6-10Å changes near the active site to a ~20Å motion of domain closure. We also observed that a small molecule binding far from the active site at an interdomain junction induces closure. We propose that this site is a substrate length sensor that additionally stabilizes the closed state, which mechanistically indicates that the phenomenon of ERAP1 substrate minimal length preference relates to the capacity of a substrate to stabilize the closed conformation. Finally, we propose that the disease-related polymorphic position 528 (lysine/arginine) alters ERAP1 activity by energetically contributing to the closed conformation as it forms a transient interdomain electrostatic interaction with glutamate 913.
RESULTS

*ERAP1 is open in solution and closes upon active site occupancy*

ERAP1 has been observed to adopt open and closed conformations in X-ray crystal structures[45, 64], but whether these states interconvert in solution, and the relevance of the conformational change to catalysis had not been determined. In order to begin to investigate conformational transitions of ERAP1 in solution we used small-angle X-ray scattering (SAXS). In initial experiments we used ERAP1 isoform 2 allele IV (Appendix I) as previously used for structural characterization[45] in the presence or absence of the broad-range aminopeptidase inhibitors bestatin[109] or leucinethiol (LeuSH)[114], (Appendix II) which target the active site zinc and enzyme S1 and S1’ subsites[45, 64, 109, 114, 134]. In the presence of either of these inhibitors, ERAP1 becomes more compact, with the radius of gyration $R_g$, decreasing by ~2Å. (Figure 3.1a) To investigate the effect of a more physiological substrate on ERAP1 conformation, we used the octamer peptide, SIINFEKL. This peptide is generated efficiently from longer substrates by ERAP1 *in vitro*[63] and *in vivo*[43], and itself is a suboptimal ERAP1 substrate[44]. In the presence of a saturating amount of SIINFEKL, ERAP1 adopts a closed conformation. (Figure 3.1a)

To examine the effect of active-site occupancy on ERAP1 conformation in more detail, we used a tight-binding tripeptide analog DG013 ($IC_{50}$ 33nM[119]), which has a phosphinic group expected to mimic the tetrahedral intermediate that would form at the scissile peptide bond during catalysis. (Appendix II)
### Table 3.1

Crystallographic data for structures determined in this work. Statistics for the highest resolution shell shown in parentheses. Both datasets were processed using STARANISO to account for anisotropy, which generated an ellipsoidal envelope to identify usable reflections. Subsequent statistical parameters such as completeness, resolution, and mean I/σ(I) were calculated using these parameters. The resolution range and completeness were determined by analyzing the distribution of the I/σ(I) values, identifying the highest resolution shell with a completeness greater than 90%. The Wilson B-factor was calculated using the I/σ(I) values from the highest resolution shell. The R-free and R-meas values were calculated using only the reflections in the highest resolution shell. The Ramachandran favored, allowed, and outliers were determined using the PROCHECK program. The rotamer outliers and clashscore were calculated using the WHAT CHECK program. The average B-factor was calculated using the reflections in the highest resolution shell.
as R-factors are treating ellipsoidal datasets as spherical, resulting in unusually large deviations from expected values. Note that the set of reflections used for R-free calculations is capped at ~2000 based on the prior determination that this value is statistically sufficient[160].
Figure 3.1
ERAP1 adopts a closed conformation when bound to substrate and substrate-mimic inhibitors. (a) ERAP1 \(R_g\) decreases in the presence of saturating inhibitor and substrate, indicating ERAP1 adopts a more compact structure in these conditions. (b) DG013 binds ERAP1 active site. Polder map (gray mesh) contoured at 3.5\(\sigma\) shows DG013 electron density at ERAP1 active site. (c) DG013-bound ERAP1 crystal structure is in the closed conformation. Shown are cartoon or surface representations of structure with domains colored. Surface cutaway reveals DG013 bound at active site. (d) Representative ERAP1 SAXS curve collected using SAXS/WAXS data collection method. Data is from ERAP1 reference sequence (isoform I) in the absence of inhibitor or substrate. Inset, Guinier plot demonstrating \(R_g\) determination using linear fit of low resolution scattering data \((q < 0.04 \text{ Å}^{-1})\) (e) DG013 induces closed ERAP1 conformation as measured by \(R_g\). Each point is from a separate dataset collected and processed individually, with error of linear fit shown. Models based on previously determined ERAP1 crystal structures in two conformations (2YD0 and
3MDJ chain B) are displayed for reference above, and calculated $R_g$ values of these two structures are indicated with dotted lines. (f) ERAP1 experimental SAXS curve shown in Figure 1B (black) overlaid with calculated SAXS curves generated from three structural models. A model derived from molecular dynamics simulation fits best. Fit residuals for these three models are shown below. (g) Selected ERAP1 models from MD simulation demonstrate the relationship between angle theta and structural conformation. Respective theta angles shown below. (h) DG013 induces closed ERAP1 conformation as measured by SAXS curve fitting to structural models from molecular dynamics. Models that fit experimental SAXS data with $\chi^2 < 1$ are shown as points on the graph, forming a distribution of models that each fit scattering data within error.
We determined the X-ray crystal structure of ERAP1 bound to DG013, and confirmed that DG013 binds as expected to the ERAP1 active site (Figure 3.1b, Table 3.1, Table 3.2), with the enzyme adopting the closed conformation. (Figure 3.1c, Figure 3.2) Structural determination required special consideration as the X-ray diffraction dataset collected was highly anisotropic. Using the STARANISO webserver we processed the diffraction data to account for this anisotropic behavior and increased the effective resolution of the dataset. Model building was then aided by the high degree of non-crystallographic symmetry present in the asymmetric unit. (Figure 3.2)

To facilitate detailed SAXS analysis, we used a different ERAP1 construct (isoform 1, allele II, Appendix I) which lacks a nine-residue unstructured C-terminal tail and recombinant myc tag not visualized in the X-ray structure that complicates the matching of experimental scattering data to molecular models. We collected SAXS/WAXS data at beamline 16-ID at NSLS-II, Brookhaven National Laboratory; inclusion of wide-angle WAXS data allows direct measurement of the water scattering profile (maximum intensity at \( q = 2\text{Å}^{-1} \)) producing high quality background-subtracted scattering curves. We initially compared data collected by standard SAXS or SAXS/WAXS using the low-resolution \( R_g \) analysis.

In the absence of inhibitors, ERAP1 in solution adopts an open conformation, with radius of gyration (\( R_g \)) similar to that calculated for the open conformer observed by X-ray crystallography, as previously reported[2]. (Figure 3.1d)
<table>
<thead>
<tr>
<th>ERAP1 atom chain A</th>
<th>DG013 atom chain A</th>
<th>Interatomic distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu 183 OE1</td>
<td>N09</td>
<td>2.5</td>
</tr>
<tr>
<td>Gly 317 N</td>
<td>O12</td>
<td>2.8</td>
</tr>
<tr>
<td>Glu 320 OE1</td>
<td>N09</td>
<td>3.0</td>
</tr>
<tr>
<td>Glu 354 OE2</td>
<td>O01</td>
<td>3.1</td>
</tr>
<tr>
<td>Glu 376 OE1</td>
<td>O02</td>
<td>2.9</td>
</tr>
<tr>
<td>Glu 376 OE2</td>
<td>N09</td>
<td>4.0</td>
</tr>
<tr>
<td>Lys 380 NZ</td>
<td>N36</td>
<td>4.0</td>
</tr>
<tr>
<td>Phe 433 CD1</td>
<td>C06</td>
<td>3.5</td>
</tr>
<tr>
<td>Phe 433 CE2</td>
<td>C05</td>
<td>3.5</td>
</tr>
<tr>
<td>Tyr 438 OH</td>
<td>O02</td>
<td>2.4</td>
</tr>
<tr>
<td>ZN</td>
<td>O02</td>
<td>2.5</td>
</tr>
<tr>
<td>ZN</td>
<td>O01</td>
<td>2.7</td>
</tr>
</tbody>
</table>

**Table 3.2**
Intermolecular contacts between DG013 and ERAP1, chain A (PDB ID 6M8P)
Figure 3.2
ERAP1 (Δ exon 10 loop):DG013 closed crystal structure overview.
(a) Four ASU per unit cell. (b) Ribbon view of 22 ERAP1 monomer in ASU colored by chain. (c) view rotated 90 degrees relative to (b), ASU consists of two
stacked 11mer rings, here each 11mer ring colored green and magenta respectively. (d) As in (c) colored as chainbow. (e) As in (d) showing two monomers. (f) ERAP1 monomers are in closed conformation, one chain shown as chainbow.
observed identical $R_g$ from samples of this isoform collected by the two methods. (Figure 3.1e, open circles) In the presence of DG013, ERAP1 $R_g$ decreases by ~2Å, an effect observed in multiple measurements and protein preparations. (Figure 3.1e, yellow circles) Comparison with $R_g$ calculated for structural models derived from open and closed crystal structures (PDB ID 3MDJ and 2YD0 respectively, dotted lines in Figure 3.1e) shows that the ~2Å $R_g$ shift upon ERAP1 binding DG013 is consistent with a shift of the conformational equilibria from open to closed.

The improved SAXS/WAXS buffer subtraction method improves data quality at higher q-range, allowing higher resolution analysis in which the full range of scattering data are fit to experimental or computational atomic models. The full-profile data fit better to a model based on an open crystal structure than to one based on the closed crystal structure. (Figure 3.1f, cyan and gray symbols) To explore other conformers, we used a set of 45 structural models selected from a molecular dynamics simulation[2] that sampled a range of opening angles theta 54-74°. (Figure 3.1g) Of these the best fitting model had theta=67°,(Figure 3.1f, red symbols) similar to that of the open crystal structure, but with a lower $X^2$ value.

To account for uncertainties in the experimental data, we considered all models with $X^2$<1. This provided a cluster of models fitting the experimental ERAP1 data, with theta angles bracketing the value calculated for the open crystal structure. (Figure 3.1h, open symbols) When we performed the same
analysis for ERAP1 bound to DG013, we observed a separate cluster of models with smaller theta angles, similar to but slightly larger than that of the closed structure. (Figure 3.1h, yellow symbols) Overall, the SAXS analysis shows that ERAP1 adopts an open conformation in solution, and closes upon binding substrate or substrate-like inhibitors.

**Ordering of the H4a helix induced by substrate binding**

To dissect structural features related to substrate binding from those related to the open-closed transition, we crystallized ERAP1 in the presence of a decamer phosphinic inhibitor, DG014 (Appendix II) under crystallization conditions used previously to crystallize the open structure. (Appendix I, Table 3.1, Figure 3.3, Table 3.3) DG014 shares DG013’s phosphinic group and N-terminal residues but is seven amino acid residues longer, and similarly inhibits ERAP1 with high potency. (Figure 3.4) In solution, DG014 stabilizes ERAP1 in a closed conformation as does DG013. (Figure 3.5) We submit that this crystal structure of the substrate-analog inhibitor DG014 bound to open-conformation ERAP1 is an intermediate step in substrate binding, which is prevented from proceeding to the closed conformation by stabilization through crystal packing.

The X-ray crystal structure shows that DG014 binds as expected with the phosphinic group at the active site, with only the first five residues sufficiently structured to allow confident modeling, and ERAP1 in the open conformation.
(Figure 3.6a,b) Relative to previously determined structures of the open conformation for ERAP1 bound to bestatin[45], conformational changes induced by peptide binding are apparent. A region between helices 4 and 5 (residues 426-433) that was not visualized in previous open conformation structures[45, 64] is now ordered and forms a short helix 4a. (Figure 3.6c) Interactions between the hydroxyl group of the threonine at DG014 position 3 and the sidechain of ERAP1 aspartate 435 constrain the C-terminal end of helix 4a. (Figure 3.6d) Bestatin does not make this interaction and so does not stabilize helix 4a. (Figure 3.6e) In the absence of inhibitor, aspartate 435 is unbound and helix 4a is disordered. (Figure 3.6f) Thus, we propose that helix 4a becomes ordered in conjunction with initial binding of substrate to ERAP1.

**Domain closure reorients the H5 helix for catalysis**

Helix 4a, once ordered, must further shift to adopt the configuration present in the closed structure. Phenylalanine 433 makes substantial contacts with the substrate P1 side chain, as observed previously in the closed-state structure of ERAP1 bound to bestatin[64]. In the open state, phenylalanine 433 is oriented away from the active site, and reorientation is required for substrate interaction. Conversion between the DG014-bound (open) and DG013-bound (closed) structures requires translation and rotation of helix 4a which brings phenylalanine 433 in contact with substrate P1 side chain. (Figure 3.6g)
**Figure 3.3**  
**ERAP1 (allele IV, isoform 2):DG014 open crystal structure overview.**  
(a) Asymmetric unit has three ERAP1 monomers, colored by chain. (b) View as in (a) colored by domain. (c) DG014 N-terminal homophenylalanine is bonded to active site residues and catalytic zinc. (d) Interatomic contacts between DG014 and ERAP1.
<table>
<thead>
<tr>
<th>ERAP1 atom chain A</th>
<th>DG014 atom chain D</th>
<th>Interatomic distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu 183 OE1</td>
<td>2X0 1 N10</td>
<td>3.1</td>
</tr>
<tr>
<td>Ala 318 N</td>
<td>7GA 2 O</td>
<td>2.7</td>
</tr>
<tr>
<td>Glu 320 OE2</td>
<td>2X0 1 N10</td>
<td>3.1</td>
</tr>
<tr>
<td>Glu 354 OE1</td>
<td>2X0 1 O12</td>
<td>3.4</td>
</tr>
<tr>
<td>Glu 354 OE2</td>
<td>2X0 1 O12</td>
<td>3.1</td>
</tr>
<tr>
<td>Glu 376 OE1</td>
<td>2X0 1 O13</td>
<td>2.7</td>
</tr>
<tr>
<td>Glu 376 OE2</td>
<td>2X0 1 N10</td>
<td>2.3</td>
</tr>
<tr>
<td>Lys 380 NZ</td>
<td>Thr 3 O</td>
<td>3.6</td>
</tr>
<tr>
<td>Glu 383 OE1</td>
<td>Phe 4 O</td>
<td>3.9</td>
</tr>
<tr>
<td>Glu 383 OE2</td>
<td>Phe 4 O</td>
<td>3.8</td>
</tr>
<tr>
<td>Tyr 399 OH</td>
<td>Phe 4 O</td>
<td>4.0</td>
</tr>
<tr>
<td>Asp 435 OD2</td>
<td>Thr 3 OG1</td>
<td>2.4</td>
</tr>
<tr>
<td>ZN</td>
<td>2X0 1 O12</td>
<td>2.3</td>
</tr>
<tr>
<td>ZN</td>
<td>2X0 1 O13</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**Table 3.3**
Intermolecular contacts between DG014, chain D, and ERAP1, chain A (PDB ID 6MGQ)
Phosphinic pseudopeptides inhibit ERAP1 hydrolysis of leucine-AMC.

Shown are normalized Leu-AMC hydrolysis inhibition curves for DG013, DG014, and DG023. (Appendix II) Points were fit to sigmoidal functions with best fit IC$_{50}$ as reported. Representative data from one of two experiments. Note that apparent cooperativity of DG023 inhibition did not reproduce in subsequent experiments.
Figure 3.5
Phosphinic pseudodecapeptide inhibitor DG014 stabilizes more closed ERAP1 conformation in solution.
(a) SAXS experiments on ERAP1 (allele II) shows the protein becomes more compact and more similar to the closed ERAP1 conformation as determined by Rg, as in Figure 3.1e. (b) The SAXS curve collected in the presence of DG014 best fits to structural models which have decreased theta angles (are more closed) compared to data in the absence of inhibitor, as in Figure 3.1h.
Figure 3.6
Helices 4a, 5, and 4 rotate and translate between open and closed conformations.

(a) DG014, cyan, is shown bound to ERAP1 active site. Polder map (gray mesh) contoured at 3.5σ shows DG014 density in chain A. (b) ERAP1 was crystallized in the open conformation with DG014. Domains I(blue), II(green), III(orange), and IV(magenta) are shown as cartoon. DG014 is shown as cyan spheres. (c) Helix 4a is ordered along the bound inhibitor. Polder map (gray mesh) contoured at 2.8σ. (d) Asp435, immediately C-terminal to helix 4a, contacts DG014 Thr3. (e) Bestatin does not contact Asp435 when bound to ERAP1 active site, and helix 4a is unstructured (PDB ID 3MDJ). (f) ERAP1 Asp435 lacks interaction partners when no ligand is bound at ERAP1 active site. Helix 4a is not ordered in this structure (PDB ID 3QNF). (g) Helix H4a tilts and rotates between open and closed conformations. Helix 4a is colored as N-to C-terminal blue-red rainbow. Helix 4a
tilts 35° between open and closed conformations (angle between helix central axes, black lines). Rotation along the helix central axis is depicted using Phe433 sidechain for reference. DG013 homophenylalanine-phosphinyl group is shown bound at active site zinc. Phe433 in closed conformation contacts P1 homophenylalanine (hPhe) of DG013 in a T-shaped π-π interaction. (h) Phe433 motion occurs in conjunction with Tyr438 repositioning. Helix 4a moves as helix 5 twists, positioning Tyr438 for catalysis. (interaction with DG013 shown as dashed line) (i) Helix 5 rotation aligning Tyr438 for catalysis[64] is coupled to motion of helix 4 through a disulfide bond. Arrows show relative positions of central helical axes. (j) Helices 10 and 12 are adjacent to helix 4, and move farther between open and closed conformations. (k) Quantification of incident angles between open and closed crystal structures for the three M1 family aminopeptidases indicated. For PDB entries containing multiple chains, chain A was used.
Another key structural transition between open and closed conformations is the repositioning of helix 5. It was noted previously that tyrosine 438 near the end of helix 5 is oriented away from substrate in the open conformation, but moves towards the active site in the closed conformation[45, 64]. Based on mechanistic studies of M1 family aminopeptidases[161] and mutagenesis studies of ERAP1[45], tyrosine 438 is expected to play an important role in the catalytic mechanism by stabilizing the oxyanion that forms in the transition state with water attack on the scissile bonds of a peptide substrate. Indeed, in the ERAP1-DG014 (open) complex, (Figure 3.6h, cyan) as in previous open ERAP1 crystal structures, tyrosine 438 is oriented away from the active site, and moves to interact with the peptide bond transition state mimic phosphinic group in the ERAP1-DG013 (closed) complex (Figure 3.6h, magenta), in a position analogous to the the previously determined bestatin-bound ERAP1 closed structure[64]. Rotation and tilt of helix 5, which repositions tyrosine 438, is coupled to reorientation of phenylalanine 433 and motion of the helix 4a. (Figure 3.6h) This mechanism connects phenylalanine 433 motion, which senses substrate occupancy at the S1 pocket, with tyrosine 438 motion, which optimizes the active site for catalysis.

This conformational transition that couples phenylalanine 433 and substrate binding to tyrosine 438 and optimization of active site configuration is regulated by domain closure. Helix 5 motion is transmitted to domain IV through a disulfide bond between residues cysteine 443 (helix 5) and cysteine 404 (helix 4). (Figure
3.6i) Helix 4 forms the edge of domain II, where it lies along helices 10 and 12 at the edge of domain IV. (Figure 3.6j) Helix 5 acts as a lever to mechanically connect active site occupancy with domain closure, with rotation of helix 4a at the N-terminal side of helix 5 is propagated to domain IV via helix 4. Helix 4 motion is further propagated to the domain interface with the C-terminal side of helix 5 acting as a fulcrum. (Figure 3.6j, compare open (cyan) to closed (magenta)) The disulfide bond between helix 4 and helix 5 is unique to ERAP1 and ERAP2 among the M1 family of zinc aminopeptidases and is highly evolutionarily conserved among ERAP1 orthologs. (Figure 3.7) For the two other M1 family members that have been crystallized in both open and closed conformations, insulin-regulated aminopeptidase (IRAP) and CD13 (also known as aminopeptidase N) helix 4 motion is not as coupled to motion of helix 5. (Figure 3.6k)

**Allosteric activator induces domain closure, promotes catalysis**

An interesting aspect of ERAP1’s catalytic activity is its substrate-length dependent allosteric modulation. The N-methylated non-hydrolyzable peptide L(N-Me)VAFKARAF activates ERAP1 hydrolysis of single amino acid substrates such as leucine-amidomethylcoumarin (Leu-AMC), while inhibiting processing of a full-length 10-mer peptide[45]. Leu-AMC, a synthetic fluorogenic dipeptide substrate commonly used to assess peptidolytic activity, is digested by ERAP1 much less efficiently than peptides greater than 8 residues in length. Additionally,
Figure 3.7
Evolutionary evidence of a disulfide bond between helices 4 and 5 inducing substrate length dependent aminopeptidase activity in ERAP1 and ERAP2. (a) ERAP1 and ERAP2 alone among M1 family zinc aminopeptidases have a disulfide bond linking H4 and H5. Sequence alignment of domain II of M1 zinc aminopeptidases in Homo sapiens. PSA, puromycin-sensitive aminopeptidase. (b) Phylogram of M1 zinc aminopeptidases in Homo sapiens. (c) ERAP1 H4-H5 disulfide is evolutionarily conserved. Sequence alignment of domain II of ERAP1 from selected model organisms. Alignments were performed using Clustal Omega[162] and UniprotKB protein sequence designations. (IRAP gene name is LNPEP, CD13 gene name is aminopeptidase N)
peptides that are too short to be processed efficiently themselves activate processing of other suboptimal short substrates. For example, the hexapeptide INFEKL inhibits ERAP1 processing of full-length peptide, but activates processing of shorter peptides, thus shifting ERAP1 length specificity so that it preferentially trims 4-6 residue peptides over 10-12mers[45].

These data were interpreted in terms of a mechanism whereby the allosteric modulators bind within the overall substrate-binding cavity, inducing conversion to an active conformation[45]. To investigate the relationship between conformational states in solution and allosteric modulation of ERAP1 enzymatic activity, we characterized ERAP1 bound to a non-peptidomimetic small molecule allosteric modulator of ERAP1, compound 3) (see Chapter II). (Figure 3.8a) Compound 3 inhibits ERAP1 hydrolysis of a decamer peptide, (Figure 3.8b) yet acts as an allosteric activator of ERAP1 and increases hydrolysis activity of the dipeptide substrate Leu-AMC. (Figure 3.8c) By SAXS/WAXS we observed that compound 3 induced a closed conformation in solution. $R_g$ and theta values for ERAP1 bound to compound 3 were similar to those for ERAP1 bound to DG013 and the other small molecule inhibitors (Figure 3.8d-f). This implicates the closed conformation as the induced ‘active conformer’ hypothesized to explain minimal substrate length preference and allosteric activation of short substrates.

We also probed the effect of compound 3 on ERAP1 conformation by negative-stain electron microscopy (NS-EM), which allows for observation of protein conformation by a method complementary to scattering and diffraction
Figure 3.8
An allosteric small molecule compound 3 induces ERAP1 domain closure and promotes catalysis.
(a) Structural formula of compound 3. (b) Compound 3 inhibits long peptide hydrolysis. Substrate 10mer peptide has sequence WRCYEKMALK. Quantification of product peptide (RCYEKMALK) by LC-MS and summing area-under-curve (AUC) shows inhibition in product formation in the presence of compound 3. Data was normalized to uninhibited condition, datapoints shown are duplicate measurements from two separate experiments. (c) Compound 3 activates hydrolysis of dipeptide substrates leucine-p-nitroaninide (Leu-pNA) and leucine-amidomethylcoumarin (Leu-AMC) Shown are representative data from one of several experiments. (d-f) ERAP1 adopts closed conformation in the presence of compound 3, as measured by (d) Rg analysis, and (e) fitting to MD models. (f) Residual plot of best structural model fit to SAXS data in the presence of compound 3, with crystal structure model fits shown for reference. (g) Selected 2D class averages (4 out of 32, top panels) from negative stain electron microscopy (NS-EM) on ERAP1 aligned with 2D projections of MD models previously used for SAXS analysis (bottom panels). (h) NS-EM datasets of ERAP1 alone (black) or in the presence of saturating 3 (blue). Contingency table analysis comparing number of particles ≥60° vs. <60° shows a significant difference in the distribution of ERAP1 conformation between presence and absence of compound 3. (p < 1×10^{-15}, X^2 = 895, N = 52019, df = 1)
Figure 3.9
**NS-EM 2D class average model fitting**
ERAP1 NS-EM 2D class averages fit best to MD models. (on left) Models generated from crystal structures (on right) fit with worse scores. Plotted are best fit scores for 2D class averages generated from particles of ERAP1 with no inhibitor (black) or in the presence of saturating 3 (blue). Ideal fit score is -1.
ERAP1 closes at pH lower than 7.
SAXS data was collected for ERAP1 isoform 2 construct (as in Figure 3.1a) with a range of pH conditions using citrate buffer (pH 5, 6) or Tris buffer (pH 7, 8, 9). Aggregation observed at pH 5 likely occurred as ERAP1 transitioned through its isoelectric point (calculated pI = 5.7).
methods[164]. Observed particles were grouped into 2-D class averages and aligned with 2-D projections of models derived from crystal structures and molecular dynamics simulation[164], as in the SAXS fitting, (Figure 3.8g, Figure 3.9) A histogram of particle count as a function of model opening angle theta shows a bimodal conformation distribution, with closed and intermediate-open conformations populated similarly in the absence of compound 3. (Figure 3.8h) The closed conformation of ERAP1 is favored at acidic pH, (Figure 3.10) and the low pH of the negative stain reagent uranyl formate may alter the distribution of open and closed states for ERAP1 particles observed by NS-EM, increasing the population of closed states at a single molecule level. Despite this, in the presence of compound 3, ERAP1 even further populates the closed conformation relative to the ensemble in the absence of compound 3, as observed by NS-EM. (Figure 3.8h)

We were not able to localize the site(s) of compound 3 binding to ERAP1 by X-ray crystallography, but an unguided in silico docking experiment placed compound 3 in a pocket at the junctions of domains II, III, and IV within a closed conformation ERAP1 model (see Chapter II). (Figure 3.11a,b) The pocket changes geometry between the open and closed conformations with more residues contacted (compare Figure 3.11a and 3.10c) and greater interface surface area (Figure 3.11d) for the closed conformation[165]. We propose that occupancy of this site preferentially stabilizes the closed conformation, which allosterically stabilizes helix 4a and positions phenylalanine 433 and tyrosine 438
Figure 3.11
Analysis of \textit{in silico} compound 3-binding pocket.
(a) Table of ERAP1 residues that form compound 3-binding pocket in closed ERAP1 structure (based on PDB 2YD0). (within 4Å) (b) Structural model showing compound 3 (stick representation with compound 3 carbons colored blue) with surrounding ERAP1 residues listed in (a). (c) Table of ERAP1 residues that form compound 3-binding pocket in open ERAP1 structure (PDB 3MDJ chain A aligned with domain IV of closed structure). (within 4Å) (d) Interface surface area between compound 3 and ERAP1 is greater for closed structure, as determined by PISA[165].
to bind and hydrolyze substrate. This mechanism of stabilizing the active site of ERAP1 allosterically though conformational change would explain the activating effect compound 3 has on suboptimal short substrate activity.

**Peptide C-terminus binding sites localize within domain IV**

ERAP1 exhibits side chain specificity at the C-terminal as well as N-terminal end of peptide substrates, suggesting contact (s) occur between enzyme and substrate beyond the five residues of DG014 for which ordered peptide density was observed. To identify peptide C-terminal binding sites, we used a photocrosslinking approach. We developed a crosslinker probe peptide inhibitor, designated DG023, consisting of a peptide sequence based on an ERAP1 substrate[101] but with the nonhydrolyzable phosphinic group replacing the first peptide bond, and the unnatural amino acid p-benzoyl-L-phenylalanine (BPA)-amide at the C-terminus[166]. (Figure 3.12a) UV irradiation (350nm) of BPA generates a carbene that can form a covalent bond with nearby molecules[166]. This compound inhibits ERAP1 hydrolysis activity. (Figure 3.4) We performed crosslinking reactions in this manner using DG023 and purified ERAP1, and identified three crosslinking sites by LC-MS/MS. (Figure 3.12b, Figure 3.13, Figure 3.14, Figure 3.15, Figure 3.16) Two of the sites, leucines 677 and 838, lie in domain IV next to a subsite previously identified as a possible site of substrate peptide C-terminal binding by X-ray crystallography[64, 75, 76]. (Figure 3.12c) The third site, leucine 686, borders domains II, III, and IV at the *in silico* docking
**Figure 3.12**
**Substrate C-terminus binding sites identified by crosslinking.**
(a) Crosslinker peptide sequence is based on sequence of characterized ERAP1 substrate[101]. First two amino acids are phenylalanine and leucine analogs connected by a phosphinic group ‘p’[119]. Peptide C-terminus bears a photocrosslinker p-benzoylphenylalanine (Bpa)[166]. After crosslinking, trypsin cleaves probe peptide leaving Bpa-amide as a 268 Da adduct at the crosslink site. (b) Crosslinked ERAP1 residues reside in domain IV, magenta surface. Orthogonal views shown of closed ERAP1 structure(PDB ID 2YD0). Crosslinked amino acids shown as sticks with colored boxes. On right, a cutaway view (cutaway plane indicated by dotted line on left) reveals crosslink sites as red surface. (c) Close views of modeled Bpa-amide in proximity to crosslinked ERAP1 leucines 838 and 677, labeled on left and shown in red. (d) Close view of modeled Bpa-amide in proximity to crosslinked ERAP1 leucine 686, as in (c). This pocket is formed between domains II, III, and IV. (green, orange, magenta surface respectively) (e) Compound 3 docked *in silico* to same interdomain pocket shown in (d)
Figure 3.13
Crosslinked ERAP1 tryptic digest fragments identified by LC-MS/MS.
Identified peptides are highlighted gray, yellow, or cyan. Crosslinked leucine residues are boxed.
Figure 3.14
ERAP1 peptide 666-685 crosslink identification.
(a) Peptide identification parameters. (b) MS/MS spectra. (c) Fragmentation table. All data generated from Scaffold4 output.
Figure 3.15
ERAP1 peptide 666-689 crosslink identification.
(a) Peptide identification parameters. (b) MS/MS spectra. (c) Fragmentation table. All data generated from Scaffold4 output.
Figure 3.16
ERAP1 peptide 829-854 crosslink identification.
(a) Peptide identification parameters. (b) MS/MS spectra. (c) Fragmentation table. All data generated from Scaffold4 output.
site identified for compound 3. (Figure 3.12d) The homologous site is accessed by long peptides bound to mammalian aminopeptidase N (CD13), a structural homolog of ERAP1 involved in neuropeptide processing[71]. The leucine 686 pocket changes geometry between the open and closed conformations as noted previously (Figure 3.11) with greater buried surface area observed for the closed conformation[165]. Increased buried surface area correlates with greater energetic stability by decreasing the energetic requirement of solvation[165], and in this case the energetic contribution of binding differs between two conformations. We propose that this interdomain site bordered by leucine 686 is a novel peptide C-terminus binding site of ERAP1. We propose that occupancy of this site by a ligand such as a portion of a peptide substrate or allosteric modulator such as compound 3 promotes ERAP1 catalytic activity by increasing the energetic stability of the closed conformation.

**Lys/Arg 528 SNP alters domain closure, regulates activity**

Numerous genome-wide association studies (GWAS) have linked a common ERAP1 polymorphism at position 528 with incidence of diseases such as psoriasis, Behçet’s disease, and most strongly, with ankylosing spondylitis[96, 155, 167]. The disease-correlated variant codes for lysine at this position, compared to arginine in the protective ERAP1 allele. This seemingly conservative substitution nevertheless has an effect on immune function, presumably by altering the presentation of a peptide that stimulates an
autoimmune response[103]. Proposed mechanisms for this effect include alterations in protein expression[168] (although this may be due to linkage with other SNPs and not directly resulting from rs30187[169]) and altered protein conformational flexibility[159].

As reported previously[101, 159], we find that arginine 528 distinctly alters enzymatic activity compared to lysine 528, increasing the enzyme catalytic rate ($k_{cat}$) but decreasing apparent substrate affinity ($K_M$). (Figure 3.17) This trend is similar, albeit with lesser magnitude, to that observed when mutations that disrupt interdomain salt bridges were introducted into ERAP1[2]. Examination of the ERAP1 structure and electrostatic calculations show that a positively charged residue at position 528 forms a long-range electrostatic interaction with glutamate 913 in domain IV. This interaction would contribute preferentially to the closed conformation, as the pair of amino acids moves far apart in the open conformation. (Figure 3.17b) Calculation of the electrostatic potential between positions 528 and 913 shows that arginine 528 has a weaker energetic contribution to the closed conformation than lysine 528. (Figure 3.17c) These data supports a mechanism where the closed conformation is required for catalysis and ERAP1 variants with arginine 528 are less able to access this conformation due to the loss of a stabilizing interdomain interaction.

If this model is correct, one prediction is that inhibitors which stabilize the closed conformation of ERAP1 would have weaker affinity for arginine 528 than lysine 528, while inhibitors that do not stabilize the closed ERAP1 conformation
Lys528
Arg528

PDB ID 2YD0 - closed
Lys528
Asn414
Glu913

PDB ID 3MDJ chain A - open
Arg528
Asn414
Glu913

C

<table>
<thead>
<tr>
<th></th>
<th>Lys 528</th>
<th>Arg 528</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding energy</td>
<td>-0.35</td>
<td>-2.74</td>
</tr>
<tr>
<td>G (kJ/mol)</td>
<td>-0.87</td>
<td>-6.80</td>
</tr>
<tr>
<td>( \Delta G )</td>
<td>-5.93</td>
<td>-1.41</td>
</tr>
<tr>
<td>( \Delta G_{Lys-Arg} ) (kJ/mol)</td>
<td>-4.52</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.17
Substitution at position 528 (lysine or arginine) alters ERAP1 activity by making a conformationally-dependent interdomain electrostatic interaction. (a) Substitution at position 528 alters ERAP1 activity. Shown is combined data (Lys528, 6 experiments. Arg528, 4 experiments) globally fit to an allosteric sigmoidal curve. Fit parameters are shown in table below. Asterisks indicate statistical significance by extra-sum-of-squares F test. ( * = p < 0.05, ** = p < 0.0001) (b) Position 528 makes a long-distance electrostatic interaction with glutamate 913. The interaction brackets asparagine 414. N-acetyl-glucosamine in crystal structure shown in gray. Views shown are in closed and open conformations. Closest interatomic distances between Lys/Arg528 and Glu913 are 7.6Å(closed) and 13.6Å(open), respectively. (c) Lys528 stabilizes closed ERAP1 conformation more than Arg528. Pairwise electrostatic potential between residues 528 and 913 shown in gray highlight. Potential energy was calculated using DelPhiForce webserver. Models for input were generated using PDB ID 2YD0 (Lys528/closed) or 3MDJ chain A (Arg528/open) and computationally mutating position 528 to complete the matrix.
should have equivalent affinity for the two variants. This trend was observed when aforementioned artificial electrostatic mutations were introduced into ERAP1 and the resulting proteins demonstrated differential inhibition by DG013[2]. We found that a small-molecule inhibitor, compound 2 (see Chapter II), (Figure 3.18a) does not close ERAP1 when present at saturating concentrations. (Figure 3.18b) We observed equivalent potency of this inhibitor for the two ERAP1 variants (lysine/arginine 528). (Figure 3.18c) In contrast, DG013 (Figure 3.18d) and compound 3 (Figure 3.18e) both affect lysine 528 ERAP1 with greater potency (lower IC$_{50}$/AC$_{50}$ respectively) than arginine 528 ERAP1. While compound 2 and DG013 both act as inhibitors of the short substrate Leu-AMC used in these experiments, their inhibition mechanisms are distinct due to their differential conformational effect. The differential affinity of ERAP1 arginine 528 and lysine 528 variants for compounds that preferentially stabilize the closed state (DG013 and compound 3) are consistent with a model that ERAP1 domain closure is altered by polymorphic variation at position 528.

**DISCUSSION**

Together these data let us build a more complete mechanistic model of ERAP1 catalysis. (Figure 3.19) In this conceptualization, ERAP1 begins in the open state, with helix H4a unstructured (E$_{O}$). Substrate binds this state (E$_{O}$S), H4a ordering occurs (E$_{O}'$S), and then simultaneously, H4a rotates and domain IV
Small molecules that induce closing have altered potency for ERAP1 variants, while those that maintain the open conformation have equal potency for ERAP1 variants.

(a) Structural formula of compound 2. (b) ERAP1 conformation remains open in the presence of compound 2, determined by $R_g$ analysis and model fitting of SAXS/WAXS data. (c) Compound 2 does not show ERAP1 variant-specific potency. Leucine-AMC hydrolysis rate was measured and normalized to control condition without compound 2. Data was fit to a sigmoidal curve with constrained top and bottom = 100 and 0% activity respectively. Significance calculated using...
two-tailed ANOVA. (d) DG013 exhibits greater potency for Lys528 ERAP1 than Arg528 ERAP1. Leucine-AMC hydrolysis rate was measured and normalized to control condition without DG013. Data was fit as in (c). Significance calculated using two-tailed ANOVA. (*, p<0.0001) (e) Compound 3 exhibits greater potency for Lys528 ERAP1 than Arg528 ERAP1. Leucine-AMC hydrolysis rate was measured and normalized to control condition without compound 3. Data was fit to a sigmoidal curve with constrained bottom = 100% activity. Significance calculated using two-tailed ANOVA. (*, p<0.0001) (f) Structural formula of compound 1. (g) ERAP1 conformation remains open in the presence of compound 1, determined by R_g analysis and model fitting of SAXS/WAXS data.
Figure 3.19
Conceptual reaction coordinate diagram of ERAP1 catalytic cycle.
closes \((E_C)\), with phenylalanine 433 and tyrosine 438 engaging substrate for hydrolysis. The reaction products (one free amino acid and the remaining peptide) are then released by opening \((E_CP \rightarrow E_C^P)\) and dissociation \((E_O + P)\). In the absence of contrary evidence, this model depicts product dissociation as energetically equivalent to substrate dissociation. This assumption may not be valid, as the ‘register’ of the peptide bound to ERAP1 would not match between these two states. Alternatively, release of the hydrolyzed N-terminal amino acid may occur separately from release of the C-terminal peptide product. Future experiments are indicated to determine the nature of product release. We also do not know the order of binding events for long substrate peptides, which may engage either the allosteric C-terminal site or the active site initially. This order of events may even differ for substrates of differing sequences and lengths, and so more mechanistic studies are warranted to dissect this open question.

Previous reports of conformational change upon substrate or inhibitor binding in M1 family aminopeptidases relied on indirect methods such as measurements of enzyme aminopeptidase activity[69] or competitive binding assays[71]. While useful, these results may be complicated to design and interpret. Our direct structural observations delineate between effects on activity and effects on conformation. Wider study of the relationship between conformation state and catalysis of M1 aminopeptidases would inform the mechanisms of the enzyme family and any specific features unique to the oxytocinase subfamily.
One caveat of using SAXS to determine conformational state in this system is the difficulty in distinguishing mixtures of states from single conformations. In theory this is possible, however the changes in scattering between open and closed states are relatively small within this system. We fit SAXS data to single molecular models, as one-model fits described the data within error ($\chi^2<1$). Fitting SAXS data as a weighted ensemble of two models nominally improved the fit, yielding even lower $\chi^2$ values, but this may be due to fitting of noise in the SAXS data and was not, in our opinion, justifiable. More sophisticated statistical analysis may allow further investigation of this ensemble using SAXS[170].

We expect that ERAP1 molecules transition throughout an energy landscape between open and closed states, as was defined by molecular dynamics simulation previously[2, 159], and that molecules in solution populate an ensemble of these major states, and intermediates therein, in a dynamic equilibrium. Electron microscopy may be used to observe protein conformation on a per-particle basis, and our negative stain-EM data of ERAP1 identified that molecules generally populate two distributions (one ‘closed’ and one ‘semi-open’) with conformations observed bracketing each peak, as opposed to a sharp distribution where each conformation is defined by a single model. This supports the model of a more continuous energy landscape. Cryo-EM data collection, which would avoid the acidic pH from negative staining with uranyl formate, would more closely match the buffer conditions used for SAXS and would provide a complementary perspective on ERAP1 conformation.
The prior identification of compounds 2 and 3 as nonpeptidomimetic ERAP1-specific inhibitors opened questions regarding their mechanism of action and determinants of specificity. While characterization is ongoing, the conformational effects induced by inhibitors permits classification of three inhibition mechanisms: (1) inhibition by substrate mimicry (bestatin, leucinethiol, DG013) causing closed conformation stabilization, (2) inhibition by open conformation stabilization (compound 2), and (3) allosteric inhibition by closed conformation stabilization (compound 3). Cocrystallization of IRAP with a phosphinic peptidomimetic inhibitor stabilized a closed structure[72], consistent with mechanism (1). Development of small molecule M1 aminopeptidase inhibitors that utilize other inhibition mechanisms may be possible, but would require concerted exclusion of peptidomimetic moieties which are a principle focus of rational inhibitor design in the field currently[143].

Our data describes a mechanism where polymorphism at position 528 (lysine/arginine) alters ERAP1 enzymatic activity by changing the opening rate of the closed conformation. In this case, ERAP1 containing lysine 528 has a slower opening rate than arginine 528. Conformational effects of polymorphisms have been proposed previously based on MD simulations of ERAP1 variant conformation[159]. These previous in silico studies support a model where arginine 528 ERAP1 does not adopt as wide an open conformation as lysine 528, and instead converts between ‘closed’ and ‘semi-closed’ states. Our SAXS data of these two variants finds no appreciable difference in conformation state due to
substitution of position 528, with both adopting an open conformation. In the presence of saturating concentrations of small molecules, the two variants changed conformation in the same manner and to the same degree. Using inhibitor potency as a probe, we found evidence that the key difference between variants at position 528 is not the equilibrium state at saturation but in the rates of exchange within that equilibrium.

Direct observation of rates of conformational change would further test this model. This is possible using techniques such as continuous flow time-resolved SAXS (trSAXS) where ERAP1 and a ligand (substrate, inhibitor, or activator) flow continuously through a mixing chamber and X-ray scattering data is measured at particular locations along the flow path, which are correlated with post-mixing time as determined by the solution flow rate[171]. This technique requires large quantities (tens of milligrams) of protein, which currently limits application of trSAXS to this project. Alternatively, spectroscopic techniques such as single-molecule Forster resonance energy transfer (FRET) might permit the observation of conformational exchange rates. This method requires coupling fluorescent probes to ERAP1 which may alter ERAP1 activity, and so is not without its own caveats.

The consequence of an enzyme enclosing a substrate, such as occurs with M1 aminopeptidases, is an effective increase in substrate binding affinity by kinetically trapping the substrate into the active site. For substrates that innately exhibit affinity to bind the open form, alteration of this trapping step (such as an
increased opening rate, as we predict for ERAP1 arginine 528 relative to lysine 528) would have little effect on apparent affinity. For substrates with lower affinity, however, a faster opening rate may allow substrate dissociation. We predict that the reported decreased activity of arginine 528 ERAP1 is exacerbated for substrates with weaker substrate affinity for the open conformation.

Our findings still do not explain the role that other polymorphic sites play. Some sites such as position 127 (arginine/proline) lie near domain junctions, and may contribute to conformation. Other polymorphic sites such as the important position 730 (glutamine/glutamate) are far from interdomain regions and may alter ERAP1 activity by a separate mechanism, possibly related to substrate binding. The possibility and potential utility of inhibiting particular ERAP1 alleles in vivo remains to be seen, but in heterozygous patients with ankylosing spondylitis[96] or psoriasis[99, 172](lysine 528 increases risk of incidence of both diseases), this might modulate the underlying immune reaction by disrupting antigen processing. Great care should be taken when adjusting immune reactions in this manner, but the benefit may outweigh the risk[173].
MATERIALS AND METHODS

Mutagenesis and baculovirus production

Sequence variants of ERAP1 were generated as gene cassettes in pFastBac plasmid using Agilent QuikChange II XL site-directed mutagenesis. Positive clones were used to make bacmids and baculoviral stocks were prepared following Bac-to-Bac baculoviral system protocol.

Protein expression and purification

High Five cells were infected with baculoviral stocks and cultured for three days at 27°C. Cells were then pelleted and supernatant was concentrated to ~100mL and buffer exchanged >100-fold into 50mM Tris pH8, 300mM sodium chloride, 10mM imidazole. Samples were bound to Ni-NTA-agarose resin, washed with 10mM imidazole buffer, then washed with 20mM imidazole buffer, then eluted with 100mM imidazole buffer. Samples were then purified by anion exchange chromatography and size exclusion chromatography.

Chemical synthesis

Synthesis of phosphinic peptide DG013: Phosphinic pseudopeptide DG013 was synthesized as described previously[119].
Synthesis of phosphinic peptide DG014: Phosphinic pseudopeptide DG014 was synthesized by applying standard solid phase peptide synthesis, on trityl alcohol lanterns (15 µmol/pin) using a Fmoc chemical protocol. A solution of acetyl chloride in dry dichloromethane (1:10 v/v) at room temperature was used to afford the trityl chloride lanterns. Attachment of the first aminoacid Fmoc-Tyr (tBu)-OH (30 µmol/pin) was performed by using N,N-diisopropylethylamine (18 µL/pin) in dry dichloromethane (0.4 mL/pin) at room temperature for 12 h.[174] The loading amount of Fmoc-Tyr (tBu)OH was evaluated to be 12 µmol/pin, after cleavage from the polymer-support with 0.5% trifluoroacetic acid (TFA)/dichloromethane (room temperature, 1h). Fmoc deprotection was performed with a solution of 20% piperidine in N,N-dimethylformamide over 1h for each cycle of the synthesis. Fmoc protected aminoacids (45 µmol/pin), 1-hydroxybenzotriazole (45 µmol/pin) and diisopropylcarbodiimide (45 µmol/pin) in dichloromethane/N,N-dimethylformamide (6/1) (0.4 ml/pin), were used for the coupling steps and each coupling reaction was allowed to proceed for 5h. Coupling of the building block Boc- (R)hPhep[PO (OAd)CH₂] (R,S)LeuOH (23 µmol/pin)[119], was performed using the coupling conditions described above (36 µmol/pin of each reagent). Deprotection and removal of the final pseudodecapeptide from the solid support was accomplished by using a solution of TFA/dichloromethane/triisopropylsilane/H₂O 39/58/2/1 for 2h at room temperature. After concentration in vacuo, the crude product was precipitated in cold diethyl ether. DG014 was obtained after purification by analytical RP-HPLC.
and characterized by mass spectroscopy [ESMS m/z (z = 1): calcd for [C_{62}H_{90}N_{9}O_{18}P+H]^+ 1281.4; found: 1281.5].

**Synthesis of phosphinic peptide DG023:** The phosphinic pseudoundecapeptide DG023 was prepared by conventional solid-phase peptide synthesis, on Rink amide lanterns (8 µmol/pin), using the Fmoc strategy. Fmoc deprotection and aminoacid coupling steps were performed as described above for DG014. After the introduction of Fmoc-Bpa-OH all synthetic steps were performed in light-protected conditions. For the introduction of phosphinic pseudodipeptidic sequence, the building block Boc- (R)Phe[PO (OAd)CH\textsubscript{2}] (S)LeuOH was synthesized in three steps starting from the R-stereoisomer of the Boc-protected aminophosphinic analog of phenylalanine[158, 175]. The phosphinic pseudodipeptide Boc (R)Phe)[PO (OH)CH\textsubscript{2}] (S)LeuOEt was prepared as previously described[176], and obtained in a stereochemically pure form after 2 recrystallizations with AcOEt. Subsequent adamantylation of the phosphinic group and saponification of the C-terminal ethyl ester group afforded the final building block Boc (R)Phe[PO (OAd)-CH\textsubscript{2}] (S)LeuOH[176], which was incorporated in the last step of the solid phase synthesis. For the coupling of the aforementioned building block, 16 µmol/pin were used by using standard coupling conditions. The final pseudoundecapeptide was cleaved and deprotected from the solid support in presence of a solution of trifluoroacetic acid (TFA)/H\textsubscript{2}O/triisopropylsilane 95/2.5/2.5 over 2 h at room temperature. The solution of deprotected peptide was concentrated in vacuo and the residue was
treated with cold dry diethyl ether. DG023 was obtained after purification by analytical RP-HPLC and characterized by mass spectroscopy [ESMS m/z (z = 1):
calculated for \([C_{79}H_{104}N_{17}O_{14}P +H]^+\) 1546.8; found: 1546.9].

Small angle X-ray scattering data collection

Purified protein samples were mixed with inhibitor, if indicated, and concentrated in rinsed Centricon 10kDa MWCO 0.5mL centrifugal concentrators equilibrated in 50mM HEPES pH7.5, 200mM NaCl, 0.02% (w/v) NaN₃. Concentrator retentate and flowthrough were stored and used as sample and buffer respectively during SAXS data collection. SAXS data was collected for each sample at three concentrations (generally 4mg/mL, 2mg/mL, and 1mg/mL). Small molecule additives were included at the following concentrations: 2mM bestatin, 30µM leucinethiol with 30µM dithiothreitol, 120µM SIINFEKL peptide, 100µM DG013, 200µM compound 3, 200µM compound 2.

Small angle X-ray scattering data processing and analysis

Scattering curves were buffer subtracted using matched buffer scattering curves. The concentration series was then compared, scaled, and merged manually using PRIMUS[177] in ATSAS v2.5.2 and SCÅTTER (version 3.0g). Merged SAXS curve \(R_g\) was calculated using AutoRG. Minimal ensemble search
was performed using the FOXS webserver[178] with data fit to single structural models of ERAP1.

**ERAP1 structural modeling**

Structural models were generated by two methods. For models derived from crystal structures (PDB ID 2YD0, and 3MDJ chains A/B/C), missing portions of polypeptide and complete high mannose N-glycans were added by rounds of simulation using the AllosMod webserver[179]. Alternately, structural models were sampled from a prior molecular dynamics simulation of ERAP1[2].

**Crystallization**

Purified protein samples were prepared in 10mM Tris or 10mM HEPES buffer. Crystal trials screens were set up and hits were optimized. For DG013-bound ERAP1 cocrystal, final crystallization conditions were: sitting drop vapor diffusion, 1.3M ammonium sulfate, 100mM MES pH5.5, 15mg/mL ERAP1 preincubated with saturating DG013, grown at 25C, dehydrated by changing well solution to 2M ammonium sulfate one day prior to looping, cryoprotected with lithium sulfate. For DG014-bound ERAP1 cocrystal, final crystallization conditions were: hanging drop vapor diffusion, 15% PEG8000, 100mM Tris pH8.60, 3% D-sorbitol, 7.5mg/mL ERAP1 preincubated with saturating DG014, grown at 4C,
cryoprotected with ethylene glycol. Crystals were mounted on loops and frozen by plunging in liquid nitrogen in preparation for data collection.

**X-ray diffraction data collection**

Crystal diffraction data was collected in a 180° arc about a static crystal (PDB ID 6MGQ) or a 360 arc about a crystal transiting along a defined vector (PDB ID 6M8P). X-ray wavelengths used were 1.110Å (PDB ID 6MGQ) and 0.979Å (PDB ID 6M8P).

**Structure determination**

X-ray diffraction data was integrated using XDS version June 1, 2017[180]. Data is highly anisotropic and so was processed using STARANISO[181]. Initial molecular replacement was found using 2YD0 (PDB 6M8P) or 3MDJ (PDB ID 6MGQ) as search models. For PDB ID 6M8P, initial diffraction data was integrated as P2_12_1 but did not yield MR solution in Phaser-MR[182]. Allowing other solutions within point group yielded solutions in spacegroup P2_12_1 with 22 chains placed in two 11-mer rings. For PDB ID 6MGQ, a suitable solution was found by searching separately for (domains I+I+III) and (domain IV). Inhibitor chemical structures were initially built using the PRODRG webserver[150]. Restraints were then generated using eLBOW[183], and models were built
through cycles of refinement[184] and manual rebuilding[185]. Omit maps were generated using Polder[186].

**Enzyme activity assays**

ERAP1 activity was measured consistently in 96-well format, 100µL reactions containing 20mM Tris pH7.5, 100mM NaCl, 0.02% sodium azide, with 0.1% (w/v) bovine specific antigen (BSA) present to block ERAP1 binding to plastic (observed even for ‘non-binding’ polypropylene plates). For peptide hydrolysis, ERAP1 (0.2ng/µL) was mixed with inhibitor (25µM) and then incubated with substrate peptide (10µM) for 10 minutes. Reactions were stopped by addition of trifluoroacetic acid, 0.4% (v/v) final concentration. Plates were frozen and shipped to PureHoney Technologies for quantification of substrate and product by LC-MS and summing area-under-curve of respective peaks.

For leucine-amidomethylcoumarin hydrolysis, ERAP1 (2ng/µL) was mixed with inhibitor if indicated and reactions were begun by addition of Leu-AMC substrate (100µM). Product formation (7-amino-4-methylcoumarin) was quantified by measuring change in fluorescence (380nm excitation/460nm emission) over 20 minutes at 25C using BMG POLARstar OPTIMA.
Negative stain-electron microscopy

ERAP1 was prepared at 6µg/mL alone (in 10mM HEPES pH7.5, 200mM NaCl), or at 7.5µg/mL with 150µM compound 3. Protein was then blotted on 400-mesh copper grids with Formvar/carbon film and stained with 0.75% (w/v) uranyl formate (pH4.5). Serial images were collected using a Tecnai F20 microscope at 200kV (154 and 113 micrographs with and without compound 3 present, respectively). Images were processed using EMAN 2.21a [187]. Particles were picked using a neural network, yielding 19798 and 32221 particles with and without compound 3 present, respectively. Particles were then classified automatically into 32 2-D class averages and, using the “e2classvsproj” module within EMAN2, aligned with 2-D projections of MD models or crystal-structure-based models used in SAXS analysis[164]. The theta value of each best model and the alignment score for each class average are reported in Figure 3.9.

Photocrosslinking and mass spectrometry

Purified ERAP1 (allele IV, isoform 2) was incubated in a V-bottom 96-well plate in the presence or absence of DG023 crosslinking peptide inhibitor (100µM DG023, 20µM ERAP1) and irradiated for 30 minutes on ice using a long-wavelength UV lamp (Blak Ray B100AP/R, UVP). Samples were then denatured, deglycosylated enzymatically, and separated by SDS-PAGE. Excised gel bands were trypsin-digested and analyzed by LC-MS/MS at the UMass Proteomic Core Facility. Crosslinks were detected by searching the SwissProt human proteome
database using Sequest (IseNode from Proteome Discoverer 2.1.1.21, fragment ion mass tolerance of 0.050 Da, parent ion tolerance of 10.0PPM) for a +268 Da modification on leucine, serine, tryptophan, or methionine. One identified crosslinked ERAP1 peptide was excluded from analysis as it appeared in the dataset generated in the absence of DG023. Crosslinked peptides were identified with 75% peptide confidence or greater, and 2.74% prophet FDR (spectra). Data analysis and figure preparation was performed using Scaffold 4[188].

Electrostatic potential calculation

Structural models of ERAP1 domains II-IV (lacking domain I due to limits on number of atoms allowed) were prepared for analysis using PDB2PQR, which calculated atomic charges using AMBER force field[189]. The output of PDB2PQR was then input to the DelPhiForce webserver[190] to calculate pairwise electrostatic potential. Settings used were: pairwise interactions mode, Gaussian dielectric distribution method, 3.0 grids/Å, 0.2M salt concentration. To test the effect of substitution at position 528, computational mutagenesis was performed on existing structural models, using the respective residue rotamers found in the alternate ERAP1 crystal structure.
Statistics

For SAXS analysis, radii-of-gyration are presented as means of each curve fit with the respective fit error plotted. Fitting of SAXS data with structural models used $X^2$ to quantify goodness-of-fit as well as model/data ratios to qualitatively demonstrate residual deviation. For statistical analysis of $EC_{50}$, sigmoidal curves were fit to eight concentrations of inhibitor, in triplicate. $\log EC_{50}$, standard error of measurement, and number of points (24) from two experiments were analyzed for statistical significance using two-tailed ANOVA.

Figure design

Structural figures were prepared using Pymol[153]. Graphical data were prepared with Graphpad Prism 7.0c[152].

Data availability

Crystal structures of ERAP1 bound to DG013 or DG014 are deposited in the wwPDB, PDB ID 6M8P or 6MGQ respectively, with corresponding X-ray diffraction datasets deposited as datasets 605 and 606, respectively, in the SBGrid Data Bank.
ACKNOWLEDGEMENTS

Some SAXS data was collected at the SIBYLS beamline at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. Some SAXS data was collected at the LiX beamline, and one X-ray diffraction dataset was collected at beamline 17-ID-2 (FMX). These beamlines are both part of the Life Science Biomedical Technology Research resource, primarily supported by the National Institute of Health, National Institute of General Medical Sciences (NIGMS) under grant P41 GM111244, and by the DOE Office of Biological and Environmental Research under grant KP1605010, with additional support from NIH under grant S10 OD012331. As a National Synchrotron Light Source II facility resource at Brookhaven National Laboratory, work performed at the LSBR is supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences Program under contract number DE-SC0012704. One X-ray diffraction dataset was collected at beamline X25 at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the United States Department of Energy, Division of Materials Sciences and Division of Chemical Sciences under contract DE-AC02-
98CH10886. Electron micrographs were collected at Brandeis University Electron Microscopy Core Facility by Mike Rigney.
CHAPTER IV
Conclusions and future directions

Summary

ERAP1 exhibits complex catalytic behavior which is necessary for its role within MHC-I antigen processing. This behavior includes diverse substrate sequence preferences, a preferred minimum substrate length, and allosteric activation within a monomeric enzyme. This work identified two specific small-molecule inhibitors of ERAP1-catalyzed short substrate hydrolysis by library screening and counterscreening against the related enzymes ERAP2 and IRAP. An additional screen/counterscreen for activators of short substrate hydrolysis identified one ERAP1-specific compound which inhibits peptide hydrolysis by ERAP1. Two of these three hit compounds inhibited ERAP1 activity within cells, suggesting their possible utility within \textit{in vivo} or \textit{ex vivo} experiments to specifically target ERAP1.

These compounds, along with previously developed inhibitors, were used as molecular probes to identify three modes of inhibition. A crucial parameter to delineate inhibitor classification is the effect of inhibitor binding on ERAP1 conformational state. Using small-angle X-ray scattering, we observed that peptides and peptidomimetic inhibitors caused ERAP1 to adopt a closed
conformation, converting from the open conformation which is favored in solution in the absence of inhibitor. For non-peptidomimetic inhibitors, the short-substrate hydrolysis activator also stabilized the closed form. In contrast, ERAP1 maintained an open conformation in the presence of the two short-substrate inhibitors. Categorizing small-molecule binding sites broadly into “active site” or “non-active site” and comparing with the conformation state induced by binding separates the compounds tested into the three classes: active site closer (DG013, bestatin); allosteric site closer (compound 3), and (putatively) active site opener (compounds 1 and 2).

The crystal structures of open and closed ERAP1 bound to peptidomimetic inhibitors DG014 and DG013 respectively identified a novel structural state of ERAP1. In this state the protein adopts a open conformation but with helix ‘4a’ (previously found to be unstructured in the open conformation[45, 64]) organized along one side of the active site. Comparison of these two new structures along with existing ERAP1 structures demonstrated that helix 4a must move after becoming ordered to fully engage the substrate. This new intermediate along the substrate-binding pathway introduces an additional step in substrate binding which should be included in future kinetic modeling.
Refinements to the ERAP1 catalytic model

The observation of ERAP1 conformational equilibria in the presence of inhibitors, the intermediate step in substrate binding prior to enzyme closing, and measurements of ERAP allosteric activity all inform a more detailed mechanistic model of catalysis. I propose that helix 4a is initially stabilized when a compound binds the active site that contacts Asp435. This intermediate state may then convert to a closed conformation while helix 4a simultaneously rotates to engage the substrate P1 sidechain with Phe433.

In cases where the substrate P1 sidechain does not favor interaction with the aromatic sidechain of Phe433, I predict that the reverse rate (ERAP1 opening and H4a rotating Phe433 away from the substrate) remains fast, and substrate may then dissociate. If substrate interaction with Phe433 is favored, however, the reverse rate slows and ERAP1 favors the closed state, permitting catalysis.

Product release may be ordered, where the hydrolyzed amino acid exits the active site through a partially closed state due to loop motion at the interdomain interface, which would disengage Phe433 and the opening rate would then accelerate.

Regarding ERAP1 peptidomimetic inhibitors, I propose that the approximately 1000-fold weaker potency of bestatin relative to DG013 is due to the contribution of Asp435. As shown in Chapter III, bestatin does not contact Asp435 in PDB ID 3MDJ, and so the rate of conversion to the closed state is much slower, which results in more dissociation events occurring. However, bestatin binding still
shifts the ERAP1 conformational equilibrium to the closed state as observed by SAXS (Figure 3.1a), and as is consistent with the previously observed polymorphism-specific potency of bestatin.[94]

This model predicts that mutation of Asp435 would impair ERAP1-catalyzed hydrolysis of long peptide substrates (I predict dipeptide substrates are also unable to interact with Asp435). This position and surrounding ones are well conserved evolutionarily, with either glutamate or aspartate present across vertebrates (...D(D/E)VSY...). (Figure 3.6c) Mutagenesis and enzymology studies can test this prediction.

I also observed that compound 3 binding favors the closed conformation. As the docking site for compound 3 is far from both Asp435 and the active site, I conclude that occupancy at this site independently contributes sufficient energetic stabilization of the closed conformation to shift the ERAP1 conformational equilibrium to favor the closed state. Conformational rate change caused by allosteric site occupancy may not be equivalent to the effect from Asp435 engagement, as compound 3 is less potent than DG013. (Figure 3.18, compare d and e) Measurement of on- and off-rates for compound 3 by kinetic analysis would clarify the relative impact of Asp435 and the allosteric site as effectors of ERAP1 conformational change.

Additional ERAP1 enzymatic activity experiments involving mixtures of compound 3 and DG013 could indicate if one site is dominant over the other. For
example, measurements of DG013 IC$_{50}$ in the presence of saturating compound 3 would test if both compounds may bind simultaneously. More potent IC$_{50}$ in this condition would be consistent with increased energetic stability of the closed conformation, while less potent IC$_{50}$ would suggest that ERAP1 simultaneously bound to DG013 and compound 3 has a faster opening rate than “DG013 alone”-bound ERAP1.

**Inhibition via blocking enzyme conformation change**

It has not escaped my attention that, of the three ERAP1-specific compounds identified, none fall within the ‘active site closer’ category, which is populated by only peptidomimetic compounds thus far. Closed conformation stabilization was also observed when saturating concentrations of the peptide SIINFEKL was present, which presumably occupies the active site. This indicates that conformational dynamism between open and closed states occurs during catalysis. It is possible that a comparable conformational effect occurs during the catalytic cycle of other members of the oxytocinase subfamily and for the M1 family altogether.

The development of peptidomimetic inhibitors and other compounds designed as substrate mimics may be unable to yield highly specific compounds for any individual member of this family due to commonalities in substrate sequence preferences, especially between ERAP1 and IRAP[46]. Based on the data
presented in this work, I predict inhibitors developed in this manner would induce conformational closing at saturation. However, the identification of compounds 1 and 2 as inhibitors that do not stabilize the closed conformation identifies a new class of inhibitory compounds that modulate ERAP1. It is possible that other non-closing inhibitory compounds might demonstrate similar selectivity for other members of the M1 family of aminopeptidases. Investigation of the binding sites of compounds 1 and 2 would clarify the basis of their selectivity.

The related M1 aminopeptidase IRAP has been the target of several efforts to develop small-molecule inhibitors. IRAP inhibition has been shown to improve memory in rodent models\cite{191}. The IRAP inhibitors HA08\cite{192}, HFI-419\cite{193}, and an aryl sulfonamide compound 8\cite{194} have been developed, optimized, and characterized, and work is ongoing to localize the binding sites of these compounds. Cocrystallization has not been successful in some cases such as HA08, wherein IRAP crystals do not form when the inhibitor is present in the drop (personal communication, A. Hallberg). I observed a comparable inhibition of ERAP1 crystallization in drops with compound 2 present (data not shown), though failure to crystallize is not conclusive evidence of structural alteration. It is possible that these IRAP inhibitors stabilize an open conformation, and similar methods used to examine ERAP1 conformation would likely be useful in studying IRAP conformational states given the similarities of the two proteins. One possible complication is the difference in oligomeric state: ERAP1 is a monomer,
while IRAP forms stable homodimers. Domain motion might be more difficult to determine when the scattering particle has increased in size.

**Further scrutiny of ERAP1 helix 4a dynamics**

Helix 4a was determined in this work to transiently order in response to substrate binding. This is notable behavior among M1 aminopeptidases, where other crystallized family members have an ordered helix 4a despite substrate presence or conformational state. Further dissection of this flexibility in ERAP1 would help to understand the mechanism of this first catalytic step. Techniques such as hydrogen-deuterium exchange (HDX) could be used to identify the lability of mainchain amide hydrogens, which may be used as a readout of secondary structure, testing whether helix 4a unravels, or if it maintains secondary structure but is untethered along the ERAP1 active site. This distinction would further clarify the energetics involved in the reaction mechanism. It would also identify the uncrystallized structures present within the open ERAP1 structural ensemble, which would assist rational inhibitor design to target these states.
ERAP1-specific inhibitors as immune response modulators

ERAP1-specific compounds 2 and 3 inhibit ERAP1 activity within HeLa cells. Many more steps lie ahead to demonstrate an effect in vivo, but these compounds have potential as tools to temporally modulate antigen processing. The ramifications of this may be positive or negative. Previous studies using bestatin as an in vivo immune modulator yielded immunostimulatory effects[111-113]. The effects of dose concentration, dose duration, dose timing (at homeostasis, during primary immune response, during contraction of immune response, during memory immune response), and mode of delivery (localized or systemic) may each contribute uniquely in the outcome of specific ERAP1 inhibition. Additionally, the assessment of inhibition (changes in MHC-I peptidome, changes in specific or global T cell activation, general physiological criteria) and the timescale of this assessment (short- or long-term) may greatly impact conclusions regarding the efficacy and safety of ERAP1-specific inhibition. With these compounds available, I look forward to further investigation of their potential as immune modulators in the laboratory and in the clinic.
APPENDIX I

Primary sequences of ERAP1 constructs used in this work

Inserted sequences highlighted yellow
Polymorphic site position 528 highlighted cyan

ERAP1 allele II $\Delta_{\text{exon 10}}$ loop

MVFLPLKWSLAIMSFLSSLLALLTVSTPSWCSQSTEASPKRHHHHHHHHHSD
ENLYFQGTPFPWNKIRLPEYVIPVHYDILLIHANLIHFTTWFQTTKEITASQPTSTII
LHSHHLQISRALRKGAGERSLEEPLQVIERQIEVALLAPEPLLVLGPLYTVVIPHYAGNLSETFHGFYKSYRTKGERLRLASTQFEPTAARMAFPCDFEPAFKASFSIKIRREPRLAISNMPLVKVTVAEGLIEDFDVTVMSTYLVAFIISDFESVSKITKSGKVSYYAVPDHIPQDAVDALAAVTLLEFYEDYFSIPYPLPKQDLAAIPDFQSGAMENWGLTTTYRESALLFDAEKSSASSKLGITMTVAHELAVQWFGNLVTMEWOWNLWNNEGFACMFDVVVSUHPEGVDYFGKCDAVMALNSHHPVSTPVENPAQIREMFDDVSYDKACINLMLREYLSADAKSFQIVYQLKHSYKNTKNEDLWDSMASIGGGGVDSLKTMMINTWTLQKGFLITITVRGRNVHKMQEHYMKGSDGAPDTGYLWHPLFITSKSDMVHRFLKKTQDVLILEEVEWIKFNVGMNGYYIVHYEDEDGWDSLTLGKGTHTAVSSNDRASLNNFQVSIGKLSIEKALDLSLYLKHEIMPVQGGLNELIPYMLKMEKDMEVETQFKAFLIRLRLDIDKQTWTDGVSERMLRSLQLLLACVNYPCVQRAEGYFRGWKESNGLLSPDVTLAVFAVQSTEGWDFLYSKYQFSLTEKSEQIEFALCRTNKQKEKLQWILLESFKGDJKITQFEPQILTLIGRNPGYPLAWQFLRKWNKLVQKFLGGSSIAHMMVMGTTNQFSTRTRLEEVKGFSSLKENGSQLRCVQQTIETIEENIGWMDKNFDKIRVWLQSEKLER
ERAP1 allele IV, isoform 2

MVFLPLKWSLAIMSFLLALLTVSTPSWCQSTEASPKRSDGTPFPWNKIRLPEYVIPVHYDLLIHANLTTLTWFWTGTKVEITASQPTSTIIILHSHHLQISRATLRKGAGERLSEEPLQLVEHPRQEQLALLAPEPLLVLGPLYTIIIHYAGNILSETFHGFYKSTYRTKEGELRILASTQEFTARMAFPCFDEPAFKASFSIKIRREPRHLAISNMPLVKSVTVAEGLEDHFIVEDTVKMSTYLVAFIISDFESVSKITSDKGKSVKSVYAVPDINQADYALDAAVTLLEFYEDYFSIPYPLKQDLAIPDFQSGAMENWGGTTYRESALLFDAEKSASSKLDITMTVAHELAFQWGLLVTMEWWNLWLNEGFAKFMIFVSVSVTHPELVGYFFGKCFDAMEVDALNSSHVPSTVHENPAIQREMFDVDSYDKGACILNMLREYLSADAFKSGIVQYLQKHSYKNTKEDLWDSMASICPTDGVKGMDFCQRSQSQHSSSSSHWHQERDVKTMMNTWTLQRGFLITITVGRNVHMKQEHYMKGSDGAPDTUGYLVHLWPLTFITSKSDMVHRPLLTTKTDVILPEEEWIKFNVGMNGYYIVHYEDDGWDSTLGKLKGTHTAVSSNDASLINNAFQLVSIGKLSIEKALDSLYLKHETEIMPVFGQLNELIPMYKLMERDMNEVTQFKAFLIRLLRDLDIKQTWTDEGSVSRMLRSELLLACVHNQPCVQAEGYFRKWESNGNLSLPVDVTLAVFAVGAQSTEGWDFLYSKYQFSQLSTKEKSIQIEFACRTQNKEKLQWLLDESFKGDKKTQEFQILTLGRNPVGYPLAWQFLRKNWNLKVQFELGSSIAHMVMGGTNQFQSTRRLEEVKGFSSLKENSGLRCVQQITIEENIGWMDKNDKIRVVLQSEKEHDPEADATGELMLERGPFKEQLISEEEDLNMTEHMH
ERAP1 allele II, isoform 1

MVFLPLKWSLAIMSFLSLLALLTVSTPSWCQSTEASPKRSGDTPFPWNKIRL
PEYVIPHYDLLIHANLTTLTWFGETKVEITASQPSSTIIILHSQHLQISRATLTKGA
GERLSEEPLQVEHPQEQIALAAPPLEPLLGLPYTVIHYAGHLSETFHPFYKST
YRTKEGELRILASTQFEPTAARMAFPCFDEPAFKASFSIKIRREPRHLAISNMPL
VKSVTVAEGLIEDHDVFVTKMSTYLVAFIISDFVESVKTSVGKVSVYAVPDGIN
QADYALDAAVTLELYEDYSIPYPLKQDLAIAIPDFQSGAMENGLTYYRESA
LLFDAEKLSSASSKLGITMTVAHELAQWFGNLVTMEWNDLWNLNEGFACKME
FVSVSVTHPELVGDFYQFQCFDAMEVDALNSSHVPSTVOPENPAQIEMFDDLY
SYDKGACILDMLREYLSADAFKSIGIVYLFQKHSYKNTKNEDLWMSACPTDGVK
KMDGFCSRSQHSSSSSHWHQEGVDVKTMMNWTQKXFGPLITITVRGRN
HMQEHYMKGSDGAPDTQGNYVPLTFITSKSDMVHRLLLKTTKTDVILPEEEV
WIKFNVGMNGYYYIVHYEDEDGWDSLGGKGTKHTAVSSNDLASLINNAFQLVSIG
KLIEKALDSLLYLKHETEIMPVFQGLNELIPMKLEKRDNMEVETQFKAFLRIL
LRDLIDKQTWTDEGSVSRMLRSQLLLACVHNQPCVQRAGEYFRKWKESNGNL
SLPVDVTLAVFAVGAQSTEGWDFLYSKYQFSLSEKSQIEFALCRTQNKE
KLWLLLDESFKGDKIKTQEPQPQITLIGRNPGYPLAQFLRKNWNKLQKLFEL
GSSSIAHMVMGTNNQFSTRRLEEVKGFFSSLKENGSQLRCVQQITETIEENIG
WMDKNFDKIRVWLQSEKLEMHHHHHH
ERAP1 allele III, isoform 1

MVFLPLKWSLALLTVSTPSWCQSTEASPKRSDGTPFPWNKIRL
PEYVIPVHYDDLIIHANLTLTFGTGKTVEITASQPTSTIIILHSHHLQISRATLRKGA
GERLSEEPLQVLEHRHQIALLPOLLVLGPLYTVIHYAGNLSETFHGKYKST
YRTKEGELRILASTQFEPTAARMAFPCFDEPAFKAIFSIIKIRREPRHLAISNMPL
VKSKEYAGLEDHFDTVKMSTYVAIIIDSFVSKIKGKVKSQYAVPDKIN
QADYALDAAVTLLEFYEDYSIPYPLQDLAAIPDFQSGAMENWGLTTYRESA
LLFDAEKKSSASSKLGLITMTVAHELAWGFNLVTMEWNDWNLNEGFAKFMF
FVSVSVTSELKVLGDYPDFGKCDAMEVDALNSSHVPSTVENPAIQREMDDV
SYDKGACILNMLREYLSADAIFSIVYIQKHSYKNTKNEDLWDSMASICPTDG
VKGMDGESRSQHSSSSSSHWHEGVDVKTMNWTQRFQPLTITVRGRNV
HMKQEHYMKGSDGAPDYGLYWPLHTSNSKSDMVHRLLLKTVDILPEEEVE
WIKFNVGMNGYIVHYEDDGWDSTGLLKGTHTAVSSNDRSLINNAFQLVSIG
KLSIEKALDSLYLKHETEIMPQFGQNLIPMYKLMKMDNVEFTQFKAFLIRL
LRDLIDKQWTDEGSVSEMLRSSYQLLCAVHNYQPCVQAEGYFRKWKESN
GNLSLPVDTLAVFAVQAPSSTEGWDLYSKYQFSLSSTEKSQIEFALCRTQNEKE
KLQWLLDESFKGDKIKTQEFPIQLTGIRNPVYGPLAWQFLRNWKNLQKFL
GSSSIAHMVMGTNQFSTRRLEEVKGFFSSLKENGSQLRCVQQTIEEENIG
WMDKNFDKIRVWQLQSEKLERMHHHHHH
APPENDIX II

Structural formulas of small molecules used in this work

Figure S2. ERAP1 inhibitors used in these studies.

1. Bestatin
Bestatin is a dipeptide analog commonly used as a nonspecific aminopeptidase inhibitor.

2. LeuSH
LeuSH is a metalloprotease inhibitor that reacts with active site metal ion.

3. SIINFEKL peptide
The octamer peptide SIINFEKL is a model T cell epitope and has been characterized as an ERAP1 substrate.

DG013 and DG014 and DG023 are peptidomimetic inhibitors (trimer, decamer, and unamer respectively) where the first peptide bond is replaced by a nonhydrolyzable tetragonal transition-state mimic phosphinic group. DG023 also has an unnatural photocrosslinker amino acid (L-4-benzoyl-phenylalanine) as its C-terminus.

Compounds 2 and 3 were previously characterized as highly specific inhibitors of ERAP1-catalyzed peptide hydrolysis (in preparation).
REFERENCES


92. rs30187.
101. Evinouchidou, I., et al., Cutting Edge: Coding single nucleotide polymorphisms of endoplasmic reticulum aminopeptidase 1 can affect antigenic peptide generation in vitro.


122. Gfeller, D. and M. Bassani-Sternberg, Predicting Antigen Presentation,ÄIWhat Could We Learn From a Million Peptides? Frontiers in Immunology, 2018. 9(1716).


152. GraphPad Software, I., *Prism 7 for Mac OS X*, 2017: La Jolla California USA.


188. Proteome Software, I., Scaffold Suite Application Software Version 4.8.6, 2018: Portland, OR.


