



Published in final edited form as:

Nat Struct Mol Biol. 2011 May ; 18(5): 604–613. doi:10.1038/nsmb.2021.

Structural Basis For Antigenic Peptide Precursor Processing by the Endoplasmic Reticulum Aminopeptidase ERAP1

Tina T. Nguyen¹, Shih-Chung Chang^{2,6}, Irimi Evnouchidou³, Ian A. York^{4,7}, Christos Zikos³, Kenneth L. Rock⁵, Alfred L. Goldberg², Efstratios Stratikos³, and Lawrence J. Stern^{1,5}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA

²Department of Cell Biology, Harvard Medical School, Boston, MA

³National Center for Scientific Research "Demokritos", Aghia Paraskevi, Greece

⁴Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI

⁵Department of Pathology, University of Massachusetts Medical School, Worcester, MA

Abstract

ERAP1 trims antigen precursors to fit into MHC class I proteins. To perform this function, ERAP1 has unique substrate preferences, trimming long peptides while sparing shorter ones. To identify the structural basis for ERAP1's unusual properties, we determined the X-ray crystal structure of human ERAP1 bound to bestatin. The structure reveals an open conformation with a large interior compartment. An extended groove originating from the enzyme's catalytic center can accommodate long peptides and has features that explain ERAP1's broad specificity for antigenic peptide precursors. Structural and biochemical analysis suggest a mechanism for ERAP1's length-dependent trimming activity, whereby binding of long but not short substrates induces a conformational change with reorientation of a key catalytic residue towards the active site. ERAP1's unique structural elements suggest how a generic aminopeptidase structure has been adapted for the specialized function of trimming antigenic precursors.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to: Lawrence J. Stern (lawrence.stern@umassmed.edu).

Contributions: TTN determined the crystal structure, designed and performed enzymatic studies, interpreted results and helped write the paper, SCC expressed ERAP1 in insect cells and designed and performed peptide trimming and allosteric activation experiments, IE designed and performed peptide trimming and allosteric activation and inhibition experiments and interpreted results, IAY interpreted data and helped write the paper, CZ synthesized the L^{Me} peptide, KLR and ALG designed experiments, interpreted data and helped write the paper, ES and LJS designed and performed experiments, interpreted data, and helped write the paper.

⁶Institute of Microbiology and Biochemistry, Department of Biochemical Science and Technology, National Taiwan University, Taipei, Taiwan. (SCC)

⁷Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA. (IAY)

ACCESSION CODE Diffraction data and coordinates were deposited into the Protein Data Bank (PDB 3MDJ).

INTRODUCTION

Most Major Histocompatibility Complex (MHC) class I proteins have strict requirements on the lengths of peptide ligands that can be bound with high affinity, typically 8–10 residues, and therefore peptides destined for presentation by MHC class I molecules must be precisely cleaved before loading¹. Proteasomal degradation and cytosolic aminopeptidase processing are sufficient for the generation of some class I MHC antigens, but many epitopes require an additional processing step in the endoplasmic reticulum (ER) after TAP-mediated transport^{2,3}. A search for peptidases responsible for that activity identified an ER-resident, interferon- γ inducible, metalloaminopeptidase that was termed ERAP1 (ER-associated aminopeptidase 1) or ERAAP (ER aminopeptidase associated with antigen processing)^{4,5}.

Studies of ERAP1-deficient cells have shown that ERAP1 generates many presented peptide epitopes from larger precursors, but it can destroy other epitopes by trimming them below the minimal size needed for MHC class I binding⁶. The peptide repertoire carried by MHC molecules of wild-type and ERAP1-deficient mice differ substantially⁷. Epitope immunodominance hierarchies differ in wild-type and ERAP1-deficient mice infected with LCMV⁸, and T cell responses against many model antigens were reduced in ERAP1-deficient mice⁹. Furthermore, cross-presentation of cell-associated epitopes was lower in ERAP1-deficient mice¹⁰. ERAP1-dependent epitopes were shown to be important in resistance to *Toxoplasma gondii*¹¹.

Allelic variants of ERAP1 have been linked to a number of human diseases, including the autoimmune disease ankylosing spondylitis (AS) (reviewed in ¹²), diabetes¹³, some forms of cervical cancer¹⁴, and hypertension¹⁵.

An important aspect of ERAP1's specificity related to its biological function is its ability to cleave long substrates, in contrast to other aminopeptidases which generally are restricted to substrates shorter than 4 residues¹⁶. ERAP1's length specificity is characterized by a sharp increase in activity near 8–10 residues, with peptides up to ~16 residues being processed efficiently^{6,16}. This activity matches ERAP1's role in processing antigenic peptide precursors. Proteasomal processing generates peptides of 2–25 residues¹⁷, with C-terminal residues that generally match MHC class I sequence requirements, and TAP1/TAP2, the transporter associated with antigen processing that brings antigen precursors into the ER, prefers peptides of at least 8 residues¹⁸⁻²⁰ and up to 16 residues long²⁰. Therefore, many transported peptides require additional trimming to fit into MHC class I binding sites, and processing by ER aminopeptidases may be the final contributor to size selection of antigenic peptides. How ERAP1 plays this role is controversial; in addition to ERAP1's intrinsic length dependence¹⁶, it has been proposed that direct processing of antigenic precursors by ERAP1 while partially bound to MHC class I^{21,22}, and MHC-protection of peptide products released from ERAP1²³, may also contribute to the final repertoire of peptides loaded onto MHC class I.

ERAP1 has sequence as well as length preferences. ERAP1 was originally identified as a leucine-specific aminopeptidase based on studies with single amino acid fluorogenic substrates²⁴, although a wide variety of amino acids can be cleaved. The activity of ERAP1

towards peptides with various N-terminal residues characterized *in vitro* generally parallels findings from single residue model substrates, although with substantially broader overall reactivity, so that some N-terminal residues are processed efficiently in the context of longer peptides but not as single residue substrates^{3,16,25-27}. An *in vivo* study of ERAP1-mediated generation of T cell epitopes from longer antigenic precursors combined with a statistical analysis of residues preceding naturally processed epitopes in large epitope databases, revealed that epitope N-terminal specificity matches ERAP1's peptide specificity measured *in vitro*²⁶. Finally, in addition to its N-terminal specificity, ERAP1 is unusual in exhibiting some preferences as to the substrate's C-terminal residue¹⁶ and internal sequence²⁵.

To clarify the structural basis for ERAP1 specificity, and to evaluate possible models for its length dependence, we determined the 3.0 Å crystal structure of human ERAP1 bound to the aminopeptidase inhibitor bestatin by X-ray diffraction. The structure reveals an open conformation with a large interior compartment. A shallow groove extending from the active site potentially can accommodate long peptides forming a series of electrostatic and hydrophobic interactions consistent with previously reported substrate preferences. Experimental characterization of ERAP1's sensitivity to allosteric activation and inhibition by peptides, and comparison with the active site geometries of other M1 family members, suggest a mechanism for the length-dependent hydrolytic activity and allow us to propose a molecular model of how ERAP1 trims antigenic peptide precursors.

RESULTS

Overall Structure of ERAP1

ERAP1 is a member of the M1-family of zinc metallopeptidases characterized by GAMEN and HExxHx₁₈E sequence motifs. The family includes a wide variety of soluble and membrane-associated aminopeptidases with various specificities and biological functions²⁸. Sequences of the five M1-family aminopeptidases for which crystal structures have been determined are aligned in Fig. 1a. These structures reveal increasingly extensive elaborations of a thermolysin-like fold. Sequence homology is strongest in the catalytic domain containing the GAMEN and HExxHx₁₈E motifs, weaker in the N-terminal portion, and essentially absent in the highly variable C-terminal helical region (Fig. 1a).

The crystal structure of full-length recombinant human ERAP1 bound to the aminopeptidase inhibitor bestatin was determined by X-ray crystallography using cryogenic diffraction data extending to 3.0 Å (Table 1). The structure was solved by molecular replacement using LTA4H and TIFF3 as search models, with initial phasing using non-crystallographic symmetry relationships among the three copies of ERAP1 observed in the crystallographic asymmetric unit (see Experimental Procedures). The interactions between the three molecules (Supplementary Fig. 1) might provide a model for ERAP1's interaction with ERAP2²⁷.

The crystal structure of ERAP1 reveals four protein domains, with a large cavity between domains II and IV (Fig. 1b). As with other M1-family members, domain II (green in Fig. 1b) is the catalytic domain that carries the zinc atom, GAMEN, and HExxHx₁₈E motifs on a

thermolysin-like $\alpha\beta$ fold. Domain I (blue) is an all- β sandwich domain that docks on top of the thermolysin domain, capping off the active site and providing binding sites for the amino terminus of a substrate peptide. Domain III (orange) is a small β -sandwich domain between domains II and IV. Among M1-family aminopeptidases, domain IV (red) is the most variable, ranging from a tightly closed spiral of ten helices in LTA4H to seventeen helices in ePepN (Fig. 1b,c). In ERAP1, domain IV is composed of sixteen variously sized helices arranged as eight antiparallel Armadillo/HEAT-type helix-turn-helix repeats²⁹, assembled side-to-side to form a concave surface facing toward the active site (Fig. 1d). Overall, the helices of the C-terminal helical domain are arranged in a large cone or bowl adjacent to the active site, lined by the lateral faces of the even-numbered helices.

The structure of ERAP1 reported here is more open than for other M1 family members because of the orientation of its C-terminal helical domain relative to the remainder of the protein (Fig. 1c). In LTA4H and ColAP, the small C-terminal helical domain packs closely against the catalytic domain, with no intervening small β -sandwich domain and no extensive cavity formed. For ePepN and PfAM1, the helical domain is relatively large and similar in overall dimensions to that of ERAP1, however it packs closely against the active site, forming a closed substrate binding cavity. TIFF3 has a more open domain arrangement somewhat similar to that of ERAP1, but with the C-terminal domain packing closer to the catalytic domain, forming a smaller cavity. In the structure of ERAP1, domain IV extends away from the active site, forming a cavity substantially larger than in the other structures.

Among the three copies of ERAP1 in the crystallographic asymmetric unit, the interdomain orientations vary with the differences mostly affecting the orientation of domain IV relative to the other domains. Supplementary Fig. 2a shows these three copies after superposition of domain III, revealing a small motion tending towards a more “closed” conformation. A normal mode analysis³⁰ (Supplementary Fig. 2b, Supplementary Movie 1), indicates that further closing motions along the same trajectory are accessible, which would allow ERAP1 access to a more closed orientation under certain circumstances. Further structural evidence of structural changes between open and closed conformations will be discussed below.

A striking feature of the structure is the deep cavity between domains II and IV. This cavity is the largest seen to date for an aminopeptidase and potentially can provide easy access to the catalytic site for even the longest of the enzyme's substrates. The cavity (Fig. 1e), which extends up to 36 Å from the active site to the interior surface of domain IV, is likely to represent a binding site for long peptide substrates.

Catalytic Site

The ERAP1 active site is formed at the junction of five secondary structure elements: helix H2 and the adjacent antiparallel helix H3 that carry the HExxH and E residues of the HExxHx₁₈E motif, the GAMEN loop, the domain 1 loop, and helix H5 adjacent to the domain II-domain IV interface (Fig. 2a). Bestatin, an α -hydroxy β -amino dipeptide analog that functions as a broad spectrum aminopeptidase inhibitor, binds to the active site of ERAP1 via a bidentate interaction between its 2-hydroxy and amido oxygen atoms and the ERAP1 catalytic zinc atom. The catalytic zinc atom is coordinated by the side chains of His353 and His357 of the HExxH helix and Glu376 of the adjacent helix, with Glu354 next

to bestatin's α -hydroxyl group (Fig. 2b). The bestatin amino terminus is bound by Glu183 of the domain 1 loop and Glu320 of the GAMEN motif. These interactions are likely to simulate the complex of ERAP1, substrate amide, and water molecule positioned to attack the peptide's scissile bond. In addition, the main chain of GAMEN residues Gly317 and Ala318 interact with bestatin's carboxylate and amide bond, respectively. The Met319 side chain forms part of the S1 pocket, and the Glu320 side chain carboxylate interacts with bestatin's free amino terminus (Fig. 2b). Asn321, the last of the GAMEN residues, forms hydrogen bonds with the neighboring strand that appear to stabilize the local loop conformation (not shown). These interactions are likely to simulate the interaction of ERAP1 with the first few residues of a bound peptide substrate.

Structural alignment of the active sites of other M1 aminopeptidases reveals constellations of conserved catalytic residues that are essentially identical to those of ERAP1, with the exception of Tyr438 (Fig. 2c). This tyrosine is conserved in all M1 family aminopeptidases²⁸, and has been shown to play an important catalytic role in stabilizing the tetrahedral intermediate formed by water addition to the substrate scissile peptide bond³¹. In the other M1 structures, the analog of Tyr438 is oriented toward the active site in a position such that its hydroxyl can stabilize the tetrahedral intermediate formed by attack of water at the peptide scissile bond^{31,32}. In the structure of the ERAP1-bestatin complex, Tyr438 is positioned near the active site but is oriented such that its hydroxyl group is more than 8 Å away from bestatin's amide bond, as a result of reorientation of the interfacial helix and adoption of a different rotameric form than found in the other M1-aminopeptidase structures (Fig. 2c). Nonetheless, the side-chain hydroxyl of Tyr438 is essential for catalysis, as Tyr-to-Phe substitution at this position causes a 190-fold reduction in enzymatic activity (Fig. 2d).

ERAP1 is unusual among aminopeptidases in its ability to accept large peptide substrates up to ~16 residues. Optimal activity is observed for peptides of ~10 residues, while most peptides of 8 or fewer residues are not processed as efficiently¹⁶. As described above, ERAP1 has a large cavity that provides a potential binding site for such large peptides, formed at the interface between the concave surface of domain II carrying the active site and the concave surface of the domain IV helical bowl. A closer view of the cavity in the vicinity of the active site is shown in Fig. 2e. The cavity is relatively narrow near the active site (top left in Fig. 2e) and widens as it enters domain IV. The wider part of the cavity can accommodate the C-terminal portion of polypeptide substrates bound with their N-terminal regions in position to be processed at the active site. Due to the sequestered nature of the active site in domain II and the arrangements of domains I and IV around the mouth of the cavity, it is unlikely that ERAP1 can complete the processing of an epitope peptide precursor while it is bound to MHC class I, as previously suggested^{21,22}. Docking experiments show that the closest possible approach of the amino terminus of a peptide bound to a MHC class I protein to the zinc center in ERAP1 is 20 Å, indicating that more than 6 residues of extended polypeptide would need to protrude from the MHC binding site for ERAP1 to trim a preformed MHC-peptide complex.

Substrate Specificity

The orientation of bestatin in the aminopeptidase active site can help to identify subsites in the active site region that accommodate peptide side chains³³⁻³⁶. Two possible configurations for long peptides bound to ERAP1 are indicated in Fig. 2e. The heads of the arrows are located ~ 27 Å from the zinc atom, corresponding approximately to the length of an extended 7–8 residue peptide. Following peptidase convention, we call the site responsible for accommodating the peptide side chain N-terminal to the cleavage site S1, and subsequent positions called S1', S2', etc. Bestatin's phenyl group binds into the aminopeptidase S1 site, with its isobutyl group oriented toward the S1' site (Fig. 2b,e). In ERAP1, the S1 site is a relatively shallow hydrophobic pocket lined by the side chains of Ser316 and Met319 of the GAMEN loop, and Gln181 and Glu183 of the D1 loop (Fig. 2b). Modeling indicates that a leucine or methionine would fit closely into the S1 pocket as the first residue of a bound peptide. Larger residues such as tryptophan and arginine can make similar contacts, but only by adopting rotamers that orient the side chain out of the pocket. Smaller residues can be accommodated but would leave voids within the site. This pattern matches the N-terminal specificity observed in peptide hydrolysis and antigen presentation studies, in that peptides with amino-terminal leucine and methionine residues were processed most efficiently, while those with amino-terminal tryptophan, arginine, and cysteine were processed less efficiently^{3,16,25-27}. The broader reactivity towards peptides as compared to single residue substrates likely reflects the contribution of binding sites for peptide side chains beyond S1' (see below), which can compensate for non-optimal interactions in the sterically rather permissive S1' pocket.

ERAP2, another ER aminopeptidase also involved in ER antigen processing^{27,37}, has a different specificity than ERAP1, with greatly increased preference for fluorogenic amino acid substrates^{16,38} and peptides²⁷ carrying N-terminal Arg and Lys groups. One difference between ERAP1 and ERAP2 in the S1 pocket region is at position 181, which is Gln in ERAP1 but Asp in ERAP2. Residue 181 forms the top of the S1 pocket (Fig. 2b), and the Gln to Asp substitution would have the effect of lengthening the pocket and making it more acidic. A site specific mutant of ERAP1 carrying a Gln181 to Asp substitution alters its specificity from hydrophobic towards basic amino acids³⁹.

ERAP1 has been shown to be unable to trim peptides with a proline residue as a second amino acid^{3,26}. The location of the S1' site in ERAP1 can be estimated using bestatin as a guide (Fig. 2b,e). Modeling of a mock peptide substrate with proline at the second position indicates that the side chain cannot be accommodated into the expected S1' site of ERAP1 without disrupting the GAMEN loop and arrangement of catalytic residues above the nitrogen of the scissile peptide bond.

ERAP1 exhibits sequence specificity for other regions of the substrate beyond its N-terminus, as shown by alanine scanning and library screening studies²⁵. Without a structure of ERAP1 bound to a peptide substrate, identification of the sites responsible for this sequence specificity is speculative. However, several aspects of the internal cavity are consistent with it being a binding site for long peptide substrates. Fig. 2f shows a top view of ERAP1, represented as an electrostatic potential surface. Dashed lines show approximate distances of 20 Å and 30 Å from the zinc atom in the active site. The overall negative charge

of the cavity in this region is consistent with the observed specificity for positively charged residues at several sites in a 9-mer peptide substrate²⁵. Shallow pockets within the overall substrate binding cavity, visible as depressions along the sides of the groove shown in Fig. 2e, potentially correspond to pockets for peptide side chains, and could be responsible for ERAP1's observed preferences for hydrophobic residues at the C-terminal position of peptide 9-mers²⁵ and 10-mers¹⁶. Other binding sites for peptide side chains may be present on the interior faces of helices lining the cavity, as observed for other proteins carrying armadillo/HEAT repeat motifs⁴⁰.

Comparison with structure of ERAP1 in the closed conformation

During the preparation of this manuscript, a crystal structure of ERAP1 without bestatin was independently determined by another group (Fig. 3). That structure reveals ERAP1 in a closed conformation similar to that observed for LTA4H, ColAP, ePepN, and PfAM1, with the C-terminal domain packed closely to the catalytic domain (Fig. 3e). By aligning the open structure of ERAP1 bound to bestatin (PDB 3MDJ reported here), the closed structure of ERAP1 alone (PDB 2XDT), and the closed structures of LTA4H and ePepN bound to bestatin (PDB 3FUH, 2HPT, respectively) and to tri-peptide substrates or substrate analogs (PDB 3B7T, 2ZZG)^{32,33,35,36}, we were able to model a polyalanine tripeptide into the ERAP1 active site and evaluate the effect of the open-closed conformational change on peptide substrate interactions (Fig. 3b,f). In the closed conformation, the substrate binding pocket is still large enough to accommodate a 12–14 residue peptide, but it is no longer accessible from the exterior of the protein (Fig. 3c,g), having been closed off by motion of helices 18–24 in the C-terminal domain (Fig. 3a,b,e,f). In the closed conformation, Tyr438 orients towards the active site as in the other M1 aminopeptidases, with the phenolic hydroxyl moving $>5 \text{ \AA}$ to allow for participation in peptide bond hydrolysis (Fig. 3b,f). With the exception of Tyr438 on helix H5 and a slight motion of the GAMEN loop, there is no substantial change in any of the active site residues between the open and closed structures. However, residues 417–433, which in the open structure are disordered, fold to extend the H5 helix toward H4 by several turns (Fig. 3b,f), capping off the top of the S1 site and substantially restricting its solvent accessibility in the closed structure (Fig. 3d,h). The extended portion of helix H5, H5', contacts the loop between helices H21 and H22 of the C-terminal domain (Fig. 3f), providing a possible route for peptide-induced conformational changes in the large C-terminal binding cavity to be communicated into the vicinity of the active site.

Together the two structures of ERAP1 (Fig. 3a,e) provide clear evidence of a large conformational change associated with opening the substrate binding site and coupled to active site reorganization, as previously suggested for other M1 family members³². ERAP1 is thus the first M1-aminopeptidase for which structures are available in both open and closed conformations, and is likely to serve as a prototype for understanding regulation of enzyme activity in this entire class of aminopeptidases.

Length-dependent trimming and subsite activation

ERAP1 processes peptides longer than 8–9 residues much more efficiently than shorter peptides (Fig. 4a,b and reference¹⁶). This activity is unusual among aminopeptidases, which

typically are most active in processing shorter peptides¹⁶, but is consistent with ERAP1's role in antigen processing. One potential mechanism for this effect would be regulation of the enzyme's activity depending on the length of bound peptide substrate. To investigate this possibility, we used a small fluorogenic substrate, leucine-7-amido-4-methylcoumarin (L-AMC), to monitor ERAP1's hydrolytic activity in the presence of peptides of varying lengths (Fig. 4c,d). We used a series of peptides based on the antigenic peptide SIINFEKL, which previously has been used to characterize ERAP1 activity *in vitro*^{6,16} and *in vivo*²⁶. ERAP1's aminopeptidase activity on L-AMC was increased several-fold in the presence of increasing length variants from four (FEKL) up to eight (SIINFEKL) residues (Fig. 4c, closed bars) over the basal LAMC hydrolysis rate (Fig. 4c, open bar). These activating peptides themselves were not long enough to be processed efficiently by ERAP1, as assayed in a separate experiment (Fig. 4a). Activation of L-AMC hydrolysis decreased with increasing peptide length for variants of 11 residues (QLESIIINFEKL) or longer (Fig 4c), which however were processed efficiently by ERAP1 (Fig. 4a). A similar pattern of activation occurred with a different series of peptides carrying a poly-glycine motif, LGnL, of increasing length (Fig. 4d). The optimum length for activation of L-AMC hydrolysis in this series was eight residues. As with the SIINFEKL series, this was shorter than the length needed for efficient hydrolysis of the peptides themselves (Fig. 4b). (A more detailed analysis of this activity will be presented elsewhere, SCC and ALG, unpublished results). Thus shorter peptides not themselves hydrolyzed by ERAP1 can act as activators for the hydrolysis of L-AMC, presumably by converting the enzyme to a more active form upon binding.

We investigated the hydrolysis reaction further using steady state kinetics. To follow hydrolysis of a full-length peptide, we used the internally quenched decapeptide substrate, WRVYEKC^{dnp}ALK41. This peptide contains a tryptophan that can be liberated from the dinitrophenol (DNP) quencher by aminopeptidase activity so that the reaction can be followed by measuring tryptophan fluorescence over time. To follow hydrolysis of a short substrate, we used dipeptide analog substrates, L-AMC or L-pNA (leucine pnitroanilide). The initial rate versus substrate concentration profile for the long peptide substrate followed simple Michaelis-Menten kinetics (Fig. 5a). By contrast, the dipeptide substrates exhibited sigmoidal velocity curves (Fig. 5b,c), which did not follow simple Michaelis-Menten kinetics (dashed curves) but were described well using an equation describing multi-site binding (solid curves), with apparent Hill coefficients of 2.1 and 2.3 for L-AMC and L-pNA, respectively. Analysis of the L-AMC data required an inner-filter effect correction term to account for reduction in product fluorescence intensity at high substrate concentrations (see Methods).

We evaluated whether the ERAP1 activity was due to an oligomeric form of the enzyme because allosteric interactions among monomeric units in an oligomeric enzyme commonly lead to sigmoidal velocity curves such as those observed for L-AMC and L-pNA⁴². However, dynamic light scattering (Fig. 5d) and gel filtration (Fig. 5e) experiments indicate that ERAP1 is predominantly monomeric even at concentrations significantly higher than typical enzymatic assay conditions. The peak of enzymatic activity co-elutes with monomeric ERAP1 by gel filtration (Fig. 5e, bottom panel). Moreover, a plot of L-AMC

hydrolysis activity versus ERAP1 concentration (Fig. 5f) is linear throughout the range 8.6 to 46 nM, suggesting that activity is not dependent on oligomerization in this concentration range. Thus, ERAP1 appears to be active as a monomer, so the substrate activation pattern observed for L-AMC and L-pNA is not likely due to allosteric interactions among subunits of an oligomeric enzyme. Instead, we explored the possibility that the interacting sites might be present as subsites within the same monomeric enzyme.

To investigate the mechanism of peptide activation of L-AMC hydrolysis without potential complications from hydrolysis of the activating peptide, we used the nonhydrolyzable decapeptide L(N-Me)VAFKARAF (L^{Me}) as an activator. The sequence of this peptide was optimized for interaction with ERAP1 based on library screening²⁵ and the unmethylated version of this peptide (L-peptide) is an excellent ERAP1 substrate and is readily processed (Fig. 6a). In the L^{Me} peptide, the amide group of the first peptide bond carries a methyl group that blocks catalytic cleavage, and the L^{Me} peptide is not trimmed by ERAP1 under experimental conditions (Fig. 6b). Increasing concentrations of both L- and L^{Me} -peptides activated ERAP1's hydrolysis of L-AMC (Fig. 6 c,d). A non-hydrolyzable peptide with a sequence designed for minimal ERAP1 interaction, R(N-Me)VAEAAFAE, did not activate L-AMC hydrolysis (not shown). We measured initial rates of reaction in the presence of various concentrations of L-AMC and L^{Me} in order to investigate the kinetic basis of the activation mechanism. Increasing L^{Me} concentrations resulted in a marked decrease in the sigmoidal character of the substrate velocity curve (Fig. 6e, Supplementary Fig. 3). Kinetic analysis revealed a mixed-mode activation mechanism with increasing apparent V_{max} and decreasing Hill coefficient with increasing L^{Me} concentration (Supplemental Fig. 3). When the effect of L^{Me} on ERAP1 enzymatic activity was evaluated using a full-length fluorogenic 10-mer substrate, WEVYEKC^{dnp}ALK 41, rather than L-AMC, a concentration-dependent decrease in ERAP1 activity was observed (Fig. 6f). Thus L^{Me} acts as an activator for hydrolysis of L-AMC but as an inhibitor for the full-length WEVYEKC^{dnp}ALK peptide. These results suggest that a small substrate like L-AMC can bind into the catalytic site even when the regulatory subsite is occupied, but a full-length peptide substrate cannot, implying spatial proximity between the regulatory and catalytic sites.

We evaluated the proximity between regulatory and catalytic sites in an additional experiment. Processing by ERAP1 of the LGnL series (Fig. 4b) was re-evaluated in the presence of (50 μ M) INFEKL, a 6-mer SIINFEKL variant that is not processed appreciably by ERAP1 (Fig. 4a) but is able to activate L-AMC hydrolysis (Fig. 4c). In the presence of INFEKL, the LGnL peptide processing profile was shifted towards smaller peptides, with stimulation of processing of peptides 4–6 residues long and inhibition of processing of peptides longer than 10 residues (Fig. 4e). INFEKL thus acts as an activator or an inhibitor of LGnL hydrolysis, depending on the length of the peptide substrate. Apparently, as the length of activating peptides is increased, progressively more of the substrate binding groove becomes occupied, and consequently only shorter substrates can access and be hydrolyzed by the catalytic site. These results further support the idea of spatial proximity of regulatory and catalytic sites within a single binding cleft, with interaction between them dependent on the length of the substrate.

DISCUSSION

The crystal structure of ERAP1 bound to the aminopeptidase inhibitor bestatin reveals an open, four-domain arrangement with a large solvent-accessible chamber contiguous to a zinc-based aminopeptidase active site. The large cavity appears to be an adaptation specific to ERAP1 and related to its ability to efficiently process peptides that are longer than the substrates of typical aminopeptidases. The ability of short peptides to allosterically activate ERAP1 towards the fluorogenic aminopeptidase substrate, L-AMC, suggests conversion between (at least) two different forms of the enzyme dependent on peptide binding, with one more and one less active. Kinetic analysis suggests that conversion between the forms is regulated by occupancy of a regulatory subsite that is distinct from but near the enzyme active site. Comparison with a recent crystal structure of ERAP1 in a closed form suggests that the structure of ERAP1 bound to bestatin represents an open state not optimized for catalysis, but which could adopt a more closed arrangement by simple rigid-body domain motions. The concave surface of the C-terminal helical domain ensures that after closure there would still be a cavity sufficient to accommodate a peptide substrate of up to 13–16 residues.

A model for how subsite activation within the ERAP1 binding cleft can lead to length-dependent cleavage of peptide substrates is shown in Fig. 7. In this model, binding to a regulatory site promotes a conformational change to a closed conformation with increased enzymatic activity. Short peptide substrates cannot reach from the catalytic site to the regulatory site (Fig. 7a), and so are not able to stabilize the closed conformation. As a result, they will be processed at a slower rate. Peptides longer than ~8–9 residues can concurrently occupy the catalytic and regulatory sites (Fig. 7b), leading to the stabilization of the closed, active conformation and efficient trimming. Short fluorogenic substrates cannot reach from the catalytic site to the regulatory site (Fig. 7c), and are processed inefficiently, unless catalysis is accelerated by occupancy of the regulatory site (Fig. 7d). The regulatory site can be occupied by the fluorogenic substrate itself if present at high concentration, leading to positive cooperativity and sigmoidal activation kinetics, or by binding of an activator peptide, provided that the peptide does not interfere sterically with substrate binding. For a full-length substrate that occupies both catalytic and regulatory sites, no cooperativity and hyperbolic (Michaelis-Menten) kinetics are observed. Similarly a non-hydrolyzable peptide activator of L-AMC kinetics acts as an inhibitor for hydrolysis of a full-length peptide, since in this case occupancy of the regulatory site prevents substrate binding. This model explains the efficiency of ERAP1 in trimming long substrates since such substrates can activate their own trimming by binding to both the regulatory and catalytic sites. Furthermore, this model can explain the slow trimming rate for small substrates that are too short to self-activate the enzyme. Occupancy of both catalytic and regulatory sites within the overall extended peptide binding site is thus an important determinant for efficient trimming. Similar subsite activation models have been proposed to explain the length preference of bovine RNase A43, and the sigmoidal velocity plots observed for the monomeric *B. subtilis* PI-PLC when assayed with short-chain lipid analog substrates⁴⁴. The location of the regulatory site(s) within the ERAP1 structure is not yet clear in the absence of structural data for a bound peptide complex, but one possible site might be the inner surface of top of helices H20 and

H22, which in the closed conformation would contact an extended version of the tripeptide model shown in Fig. 3, influencing interactions between helix H22 and helix H5 near the active site (see below). Overall, this model is consistent with the biological role of ERAP1 in efficiently removing 1–7 residues from the N-terminus of antigenic peptide precursors, with broad specificity but sparing many mature antigenic peptides.

How could conversion between open and closed conformations regulate ERAP1's enzymatic activity? The arrangement of key functional groups in the active site of ERAP1 is identical to that observed for other M1 family members, except for Tyr438, which is oriented away from the catalytic zinc in ERAP1 but towards it in most other M1 aminopeptidase structures (Fig. 2c). The one other exception is TIFF3, which like ERAP1 is in an open conformation⁴⁵, and which also has the Tyr438 analog also oriented away from the active site³². In the closed form of ERAP1, a small ~30–40° rotation of helix H5 orients Tyr438 towards the active site, in a position to participate in peptide bond hydrolysis chemistry. Helix H5 is oriented towards the C-terminal domain, sandwiched between helix H3, which carries the second Glu of the HExxHx₁₈E motif, and helix H4, which packs against helix H10 of the C-terminal domain. In the open conformation, no direct contacts between helix H5 and the C-terminal domain are observed, although a disordered region for which we do not observe electron density extends from the end of helix 5 towards the C-terminal domain. Two other disordered regions are found directly across the cavity, between the tops of helices H21 and H22, and between the tops of helices H23 and H24 (Fig. 1d). In the closed conformation, these regions make contact with the disordered region upstream of helix H5 folding and packing against the loops between helices H21 and H22 and between H23 and H24. This interaction appears to result in reorientation of helix H5, motion of Tyr438, and capping of the S1 site, all changes which would favor increased catalytic activity. Structural studies of other enzymes related to ERAP1 suggest that domain motions triggered by peptide binding could induce active site changes to enhance peptidase activity as suggested here. In two multidomain zinc metallopeptidases distantly related to ERAP1, human angiotensin-converting enzyme-related carboxypeptidase (ACE2)⁴⁶ and *E. coli* dipeptidyl carboxypeptidase (Dcp)⁴⁷, domain closure motions have been associated with active site reorganizations favoring catalysis. In two single-domain zinc metallopeptidases more closely related to ERAP1, thermolysin⁴⁸ and astacin⁴⁹, subdomain motions similarly reorient active site residues for hydrolysis.

The open conformation of ERAP1 may be important for catalysis because it can facilitate long substrate approach and capture. Low-affinity, low-specificity binding sites on the interior surface of the cavity, similar to those present on nuclear importin- α , another armadillo/HEAT repeat protein⁵⁰, might enable efficient peptide capture in the open form. Cavity closure could help to align peptides towards the active site for amino-terminal processing, and differential exposure of interior binding sites between open and closed forms could facilitate peptide binding and release cycles, which would appear to be necessary for sequential removal of peptide N-termini as a long antigenic precursor is processed. This type of mechanism has been suggested for insulin degrading enzyme, a dimeric chambered zinc endopeptidase with “exosites” for binding entrapped peptide at positions distant from the hydrolytic site^{51,52}. Recently, ERAP1's paralog, IRAP, has been

shown to process relatively long peptides⁵³. If this model is correct, then IRAP also would be expected to adopt open and closed conformations during catalysis.

Based on the ERAP1 crystal structures, length and sequence specificity, allosteric regulation by peptides, and comparison with other M1-family proteases, we can describe a likely mechanism for ERAP1 processing of antigenic precursors. (1) Antigenic precursor peptides can access and bind to the groove while the protein is in an open conformation, using interactions at N-terminal S1 and S1' sites but also interactions at other sites in the cavity. Active site residues are not in the optimal arrangement. (2) Peptide binding with occupancy of the regulatory site causes domain closure around the bound peptide and conversion to the optimal active site constellation. Interactions responsible for triggering this change are not yet clear, but are likely to involve portions of the substrate ~7–8 residues away from the N-terminus, and (directly or indirectly) ERAP1 residues located near domain boundaries. (3) Catalysis. Based on studies of similar metalloproteases, catalytic steps would include: attack of the scissile amide bond by a zinc-bound water to form a tetrahedral intermediate stabilized by interactions with active site residues, followed by collapse of the tetrahedral intermediate with proton transfer to generate a free amino acid and new peptide N-terminus. (4) Release of products after hydrolysis and substrate rebinding for next cycle. Most likely liberation of the cleaved N-terminal amino acid requires at least partial conversion to the open form after hydrolysis. Interaction of the cleaved peptide product with low-specificity binding sites could facilitate rebinding in a shifted frame in preparation for the next catalytic cycle. (5) When peptide substrates are trimmed beyond ~8–9 residues, they are too short to engage both regulatory and active sites, and cannot be processed efficiently.

Ankylosing spondylitis (AS) is a chronic inflammatory joint disease with a strong genetic component that is strongly linked to the MHC allelic variant HLA-B27. Genome-wide studies in many human populations have shown strong associations outside the MHC region (reviewed in 12), that encompasses the ERAP1 gene⁵⁴. Identified disease-associated ERAP1 polymorphisms include residues located near putative substrate binding and regulatory sites. Fig. 8 shows sites of eight single nucleotide polymorphisms that are present in allelic variants of ERAP1 associated with altered risk of developing AS^{54,55}. Potential effects of some of these variants can be identified by examination of the structure of ERAP1. M349V is located in the active site, and R725Q and Q730E are exposed on the inner surface of the C-terminal cavity that could affect substrate sequence or length specificity. Other polymorphisms, R127P, K528R, D575N and V647I, at domain junctions could indirectly affect specificity or enzymatic activity by altering the conformational change between open and closed forms. Altered peptide trimming by ERAP1 variants may lead to altered levels of arthritogenic self peptides or of their mimics, or may lead to altered generation of T cell clones in the thymus. Structure-guided studies of the specificity and activity of these ERAP1 variants are clearly warranted.

The crystal structure of ERAP1 bound to bestatin provides insight into ERAP1's unusual length specificity and its peptide sequence preferences. These intrinsic properties of ERAP1 with MHC-binding protection contribute to the length and sequence distribution of MHC class I bound antigens. Overall, it appears that both the overall conformation of ERAP1 as well as its ability to be activated by long substrates are features that the enzyme has evolved

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Drs. Evangelia Livaniou, and Persefoni Klimentzou for assistance with chemical synthesis of the L^{Me} peptide, Loretta Lee for cell culture, Liying Lu for assistance with baculovirus, protein expression and purification, Daniel Panne, Zarixia Zavala-Ruiz and Eric Schreiter for help with cryocrystallography, Howard Robinson, Annie Heroux, and Vivian Stojanoff for assistance with data collection. Data for this study were measured at beamlines X6, X25, and X29 of the National Synchrotron Light Source, supported by the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. Resources of the UMass Medical School Mass Spectrometry Facility and Diabetes and Endocrine Research Center were used. Funding for this work was provided by the Sixth Framework Programme of the European Union (Marie Curie International Reintegration grant 017157 to E.S.) and the National Institutes of Health grants AI-38996 (LJS). LJS and KLR are members of the UMass DERC (DK32520).

ONLINE METHODS

Protein expression and purification

Human ERAP1 was expressed using a baculovirus system as previously described²⁵. We used an allelic variant with four substitutions (G346D, G514D, K528R, Q730E) relative to the GenBank reference sequence NP_057526. ERAP1 was isolated using Ni-NTA resin (Qiagen), concentrated to 1 mg ml⁻¹ by centrifugal ultrafiltration, and treated with carboxypeptidase A (Sigma) (molar ratio 5:1) at room temperature in PBS with 0.25 µg ml⁻¹ protein disulfide isomerase, PDI (Sigma). After digestion, ERAP1 was further purified by size exclusion and anion exchange (S200 and MonoQ, GE Healthcare). A single band was observed by SDS-PAGE. Enzyme L-AMC hydrolysis specific activity increased as ERAP1 was purified. For studies of the role of Tyr438, wild-type and mutant HA-tagged ERAP1 were expressed in 293F cells from a pTRACERCMV2-based plasmid and isolated by anti-HA and size exclusion chromatography.

Peptides

Synthetic peptides (Genscript) were purified by reversed-phase HPLC, with mass confirmed by mass spectrometry. The N-methylated analog L(N-Me)VAFKARAF was synthesized in-house on a trityl polystyrene resin with the Fmoc chemistry using an N-methylated valine analog (Fmoc-N-Me-Val-OH, from Bachem GmbH).

Crystallization and data collection

ERAP1 was crystallized by vapor diffusion in hanging drops at 4°C from 14% (v/v) PEG 8000, 1 mM glutathione 10:1 oxidized:reduced, 0.1 M Bicine, pH 8.6 using 6.5 mg ml⁻¹ ERAP1, 5:1 molar ratio bestatin, 0.25 µg ml⁻¹ PDI. Crystals were transferred to reservoir solution containing 30% (v/v) 2-methyl-2,4-pentanediol and flash frozen in liquid nitrogen.

X-ray diffraction data were collected at 100 K with 1.08 Å radiation at NSLS beamline X29A, processed and scaled using HKL200056.

Structure determination and refinement

Phases were estimated by molecular replacement using PHASER57 and LTA4H (PDB 1HS6) and TIFF3 (PDB 1Z5H) as search models. A solution with three molecules was obtained yielding electron density of sufficient quality for model building. Phases were improved in RESOLVE58 with 3-fold non-crystallographic symmetry real-space averaging. Averaged and unaveraged composite omit and prime-and-switch maps58 were utilized for model building (Supplementary Fig. 4). The entire C-terminal domain was different from the search models and was built *de novo*, whereas the N-terminal domains could be modeled based on the search models. Positions of three N-linked glycosylation sites (Asn70, 154 and 760), and two disulfide bonds (410–449 and 742–749) matched sites of these modifications observed by mass spectrometry (data not shown). Electron density consistent with a five monosaccharide unit chain was observed linked to Asn70 in two of the three molecules in the asymmetric unit, with a disaccharide unit at the other. At the two other glycosylation sites, no electron density or density consistent with only a monosaccharide was observed. Crystallographic rigid body, positional, B-factor, and TLS refinement was performed in PHENIX59 with NCS restraints optimized for each domain. Final RMSD between the 3 molecules is 0.006 Å (Supplementary Table 1). The inclusion of high resolution data shells improved electron density maps, cross-validation statistics, and allowed building of some regions, despite having low $I/\sigma I$ and high R_{sym} values. The completed model consists of ERAP1 residues 46–934 (of 941 encoded by the expression vector) with five missing exposed loops: 416–434, 485–514, 551–556, 863–868 and 892–904. Structure figures were made using Pymol60.

Normal mode analysis

Normal mode analysis was performed by the Elastic Network Model server (<http://www.igs.cnrs-mrs.fr/elnemo/>) using the following run parameters: NMODES=7, DQMIN=400, DQMAX=400, DQSTEP=20.

Peptide trimming assay

SIINFEKL variants (150 μM) were incubated with ERAP1 (3.5 μg ml⁻¹) at 37°C in 50 mM Tris-HCl, pH 7.8 with 0.25 g L⁻¹ protease-free bovine serum albumin (Sigma). Reactions were terminated by adding 0.6% (v/v) trifluoroacetic acid (J.T.Baker). The peptide-containing supernatant was analyzed by C18 RP-HPLC using an acetonitrile gradient in 0.03% (v/v) TFA. For LGnL peptides, 50 μM of peptide were incubated with 1.0 μg ml⁻¹ ERAP1 for 1h at 37°C. The percentage of the substrate cleaved was calculated by integration of the area under each peptide peak.

Fluorescence assay of aminopeptidase activity

Aminopeptidase activity was followed using the fluorogenic single-residue substrate L-AMC or the internally quenched fluorescent substrate WEVYEKC^{dnp}ALK41. L-AMC (100

μM) was incubated either at 37°C with $0.75 \mu\text{g ml}^{-1}$ ERAP1 in the presence or absence of $100 \mu\text{M}$ SIINFEKL variants, or at 25°C with $3.3 \mu\text{g ml}^{-1}$ ERAP1 in the presence of $100 \mu\text{M}$ LGnL, or at 37°C with $0.87 \mu\text{g ml}^{-1}$ ERAP1 in the presence of $0\text{--}70 \mu\text{M}$ L(N-Me)VAFKARAF or LVAFKARAF. The hydrolysis of L-AMC was followed for 5 min with 380 nm excitation and 460 nm emission. WEVYEKC^{dnp}ALK ($10 \mu\text{M}$) was incubated at 25°C with $6.9 \mu\text{g ml}^{-1}$ ERAP1 in the presence of $0\text{--}90 \mu\text{M}$ L(N-Me)VAFKARAF with removal of amino-terminal tryptophan followed using excitation at 295 nm and emission at 350 nm, as described previously⁴¹.

Enzyme Kinetics

WRVYEKC^{dnp}ALK, L-AMC and L-pNA (Sigma) kinetic studies were performed as described above except using $1.5 \mu\text{g ml}^{-1}$ ERAP1 and at 27°C , with L-pNA reactions in 50 mM Hepes, pH 7.5, 0.1 M NaCl, and L-AMC reactions with 1% (v/v) DMSO. Initial velocities of WRVYEKC^{dnp}ALK, L-AMC, and L-pNA hydrolysis were measured at concentrations of $0\text{--}70 \mu\text{M}$, $0\text{--}2 \text{ mM}$, and $0\text{--}5 \text{ mM}$ respectively. Equation (1) was used to fit initial rate vs. substrate concentrations profiles. For the simple Michaelis-Menten model the Hill coefficient, h , was fixed at one. For allosteric models h was allowed to vary.

$$v = \frac{V m_{app} [S]^h}{(K m_{app})^h + [S]^h} \quad (1)$$

At high L-AMC concentrations the fluorescence signal from liberated AMC was attenuated. A term describing an inner filter correction was included, and equation (2) was used to fit the L-AMC data.

$$v = \log^{-1} (Abs_{corr} [S]) * \frac{V m_{app} [S]^h}{(K m_{app})^h + [S]^h} \quad (2)$$

All data from a single experiment were fit globally using a single Abs_{corr} value, which matched the values obtained in control experiments using L-AMC and AMC only ($450\text{--}500 \text{ M}^{-1}$). Several control experiments were performed to ensure that the observed sigmoidal behavior was not a result of experiment artifact. Free AMC in assay conditions exhibited linear fluorescence intensity with concentration from 0.4 to $3.3 \mu\text{M}$. Initial rates of L-AMC ($100 \mu\text{M}$) hydrolysis were linear with ERAP1 concentrations $8.6\text{--}45.7 \text{ nM}$ ($0.9\text{--}4.8 \mu\text{g/ml}$). AMC alone, ERAP1 alone, and L-AMC alone each in reaction buffer exhibited no appreciable change of fluorescence during the assay time. Preincubated L-AMC solutions in assay wells that were transferred to other wells for measurements exhibited no reduction in L-AMC hydrolysis.

Hydrodynamic studies

Size exclusion experiments were performed using a Superdex S200 column (GE Healthcare) in PBS, with fractions analyzed for L-AMC hydrolysis as described above. Dynamic light scattering measurements were performed on a Protein Solutions DynaPro instrument at 6°C .

ERAP1 (1 $\mu\text{g ml}^{-1}$ in PBS) was filtered through 0.22 μm Spin-X filters (Corning) before analysis.

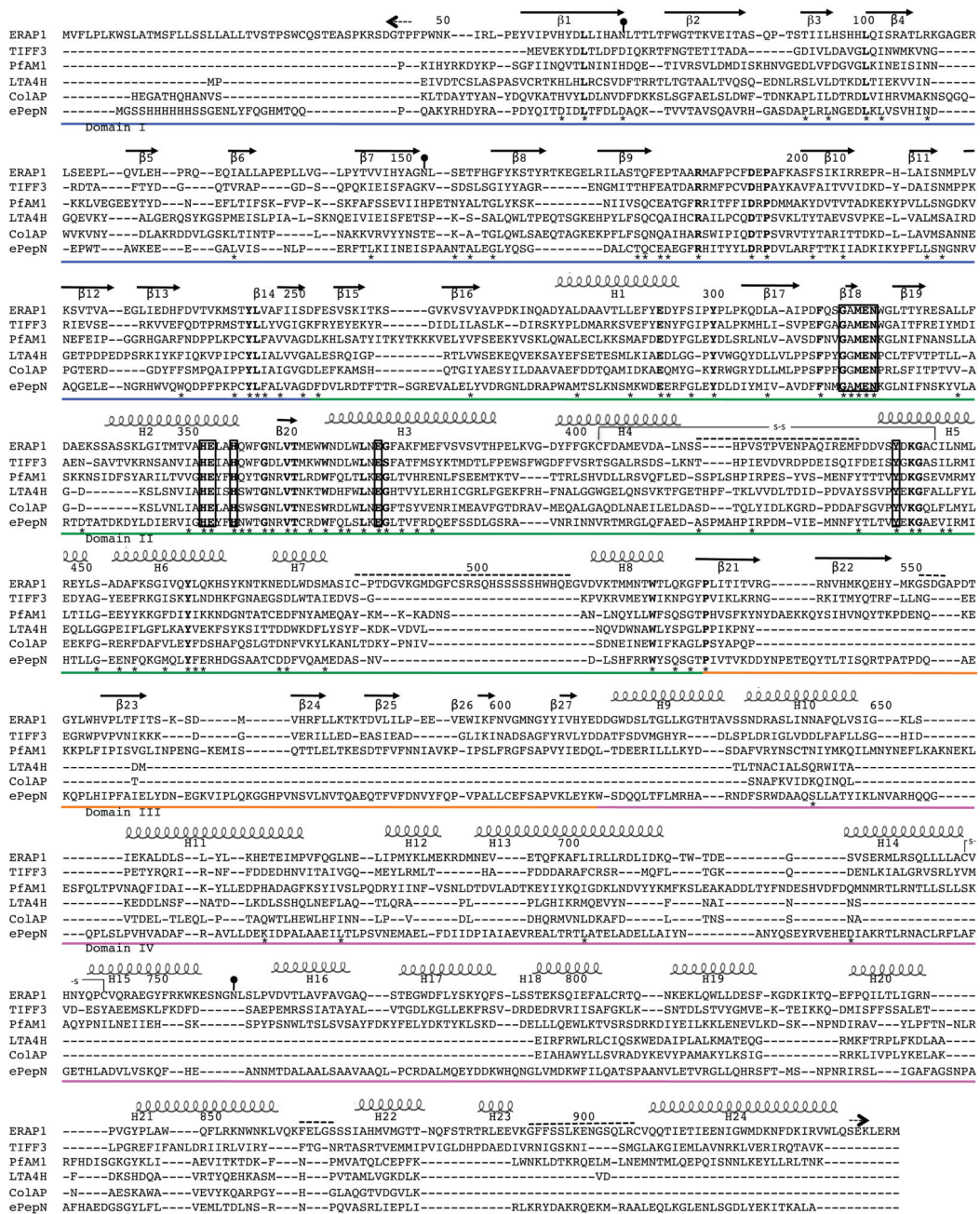
REFERENCES

1. Rock KL, York IA, Goldberg AL. Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat Immunol.* 2004; 5:670–7. [PubMed: 15224092]
2. Fruci D, Niedermann G, Butler RH, van Endert PM. Efficient MHC class I-independent amino-terminal trimming of epitope precursor peptides in the endoplasmic reticulum. *Immunity.* 2001; 15:467–76. [PubMed: 11567636]
3. Serwold T, Gaw S, Shastri N. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat Immunol.* 2001; 2:644–51. [PubMed: 11429550]
4. Saric T, et al. An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat Immunol.* 2002; 3:1169–76. [PubMed: 12436109]
5. Serwold T, Gonzalez F, Kim J, Jacob R, Shastri N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature.* 2002; 419:480–3. [PubMed: 12368856]
6. York IA, et al. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat Immunol.* 2002; 3:1177–84. [PubMed: 12436110]
7. Blanchard N, et al. Endoplasmic reticulum aminopeptidase associated with antigen processing defines the composition and structure of MHC class I Peptide repertoire in normal and virus-infected cells. *J Immunol.* 2010; 184:3033–42. [PubMed: 20173027]
8. York IA, Brehm MA, Zendzian S, Towne CF, Rock KL. Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance. *Proc Natl Acad Sci U S A.* 2006; 103:9202–7. [PubMed: 16754858]
9. Yan J, et al. In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules. *J Exp Med.* 2006; 203:647–59. [PubMed: 16505142]
10. Firat E, et al. The role of endoplasmic reticulum-associated aminopeptidase 1 in immunity to infection and in cross-presentation. *J Immunol.* 2007; 178:2241–8. [PubMed: 17277129]
11. Blanchard N, et al. Immunodominant, protective response to the parasite *Toxoplasma gondii* requires antigen processing in the endoplasmic reticulum. *Nat Immunol.* 2008; 9:937–44. [PubMed: 18587399]
12. Brown MA. Genetics of ankylosing spondylitis. *Curr Opin Rheumatol.* 2010; 22:126–32. [PubMed: 20084006]
13. Fung EY, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun.* 2009; 10:188–91. [PubMed: 19110536]
14. Mehta AM, et al. Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. *Genes Chromosomes Cancer.* 2009; 48:410–8. [PubMed: 19202550]
15. Yamamoto N, et al. Identification of 33 polymorphisms in the adipocyte-derived leucine aminopeptidase (ALAP) gene and possible association with hypertension. *Hum Mutat.* 2002; 19:251–7. [PubMed: 11857741]
16. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proc Natl Acad Sci U S A.* 2005; 102:17107–12. [PubMed: 16286653]
17. Kisselev AF, Akopian TN, Woo KM, Goldberg AL. The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J Biol Chem.* 1999; 274:3363–71. [PubMed: 9920878]
18. Momburg F, Roelse J, Hammerling GJ, Neefjes JJ. Peptide size selection by the major histocompatibility complex-encoded peptide transporter. *J Exp Med.* 1994; 179:1613–23. [PubMed: 8163941]
19. Schumacher TN, et al. Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. *J Exp Med.* 1994; 179:533–40. [PubMed: 8294864]

20. van Endert PM, et al. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity*. 1994; 1:491–500. [PubMed: 7895159]
21. Kanaseki T, Blanchard N, Hammer GE, Gonzalez F, Shastri N. ERAAP synergizes with MHC class I molecules to make the final cut in the antigenic peptide precursors in the endoplasmic reticulum. *Immunity*. 2006; 25:795–806. [PubMed: 17088086]
22. Falk K, Rotzschke O, Rammensee HG. Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature*. 1990; 348:248–51. [PubMed: 2234092]
23. Infantes S, et al. H-2Ld class I molecule protects an HIV N-extended epitope from in vitro trimming by endoplasmic reticulum aminopeptidase associated with antigen processing. *J Immunol*. 2010; 184:3351–5. [PubMed: 20200278]
24. Hattori A, Matsumoto H, Mizutani S, Tsujimoto M. Molecular cloning of adipocyte-derived leucine aminopeptidase highly related to placental leucine aminopeptidase/oxytocinase. *J Biochem*. 1999; 125:931–8. [PubMed: 10220586]
25. Evnouchidou I, et al. The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS One*. 2008; 3:e3658. [PubMed: 18987748]
26. Hearn A, York IA, Rock KL. The specificity of trimming of MHC class I- presented peptides in the endoplasmic reticulum. *J Immunol*. 2009; 183:5526–36. [PubMed: 19828632]
27. Saveanu L, et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol*. 2005; 6:689–97. [PubMed: 15908954]
28. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. *Nucleic Acids Res*. 2009; 38:D227–33. [PubMed: 19892822]
29. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. Comparison of ARM and HEAT protein repeats. *J Mol Biol*. 2001; 309:1–18. [PubMed: 11491282]
30. Suhre K, Sanejouand YH. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*. 2004; 32:W610–4. [PubMed: 15215461]
31. Thompson MW, Archer ED, Romer CE, Seipelt RL. A conserved tyrosine residue of *Saccharomyces cerevisiae* leukotriene A4 hydrolase stabilizes the transition state of the peptidase activity. *Peptides*. 2006; 27:1701–9. [PubMed: 16597475]
32. Addlagatta A, Gay L, Matthews BW. Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc Natl Acad Sci U S A*. 2006; 103:13339–44. [PubMed: 16938892]
33. Thunnissen MM, Nordlund P, Haeggstrom JZ. Crystal structure of human leukotriene A(4) hydrolase, a bifunctional enzyme in inflammation. *Nat Struct Biol*. 2001; 8:131–5. [PubMed: 11175901]
34. Ito K, et al. Crystal structure and mechanism of tripeptidyl activity of prolyl tripeptidyl aminopeptidase from *Porphyromonas gingivalis*. *J Mol Biol*. 2006; 362:228–40. [PubMed: 16914159]
35. Tholander F, et al. Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chem Biol*. 2008; 15:920–9. [PubMed: 18804029]
36. Fournie-Zaluski MC, et al. Structure of aminopeptidase N from *Escherichia coli* complexed with the transition-state analogue aminophosphinic inhibitor PL250. *Acta Crystallogr D Biol Crystallogr*. 2009; 65:814–22. [PubMed: 19622865]
37. Tsujimoto M, Hattori A. The oxytocinase subfamily of M1 aminopeptidases. *Biochim Biophys Acta*. 2005; 1751:9–18. [PubMed: 16054015]
38. Tanioka T, et al. Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. *J Biol Chem*. 2003; 278:32275–83. [PubMed: 12799365]
39. Goto Y, Tanji H, Hattori A, Tsujimoto M. Glutamine-181 is crucial in the enzymatic activity and substrate specificity of human endoplasmic-reticulum aminopeptidase-1. *Biochem J*. 2008; 416:109–16. [PubMed: 18593381]
40. Monecke T, et al. Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science*. 2009; 324:1087–91. [PubMed: 19389996]

41. Evnouchidou I, Berardi MJ, Stratikos E. A continuous fluorogenic assay for the measurement of the activity of endoplasmic reticulum aminopeptidase 1: Competition kinetics as a tool for enzyme specificity investigation. *Anal Biochem.* 2009; 395:33–40. [PubMed: 19638272]
42. Segel, I. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems.* Vol. 984. Wiley-Interscience; New York: 1993.
43. Moussaoui M, Guasch A, Boix E, Cuchillo C, Nogues M. The role of non- catalytic binding subsites in the endonuclease activity of bovine pancreatic ribonuclease A. *J Biol Chem.* 1996; 271:4687–92. [PubMed: 8617733]
44. Birrell GB, Zaikova TO, Rukavishnikov AV, Keana JF, Griffith OH. Allosteric interactions within subsites of a monomeric enzyme: kinetics of fluorogenic substrates of PI-specific phospholipase C. *Biophys J.* 2003; 84:3264–75. [PubMed: 12719256]
45. Kyrieleis OJ, Goettig P, Kiefersauer R, Huber R, Brandstetter H. Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. *J Mol Biol.* 2005; 349:787–800. [PubMed: 15893768]
46. Towler P, et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J Biol Chem.* 2004; 279:17996–8007. [PubMed: 14754895]
47. Comellas-Bigler M, Lang R, Bode W, Maskos K. Crystal structure of the *E. coli* dipeptidyl carboxypeptidase Dcp: further indication of a ligand-dependent hinge movement mechanism. *J Mol Biol.* 2005; 349:99–112. [PubMed: 15876371]
48. Holland DR, et al. Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry.* 1992; 31:11310–6. [PubMed: 1445869]
49. Grams F, et al. Structure of astacin with a transition-state analogue inhibitor. *Nat Struct Biol.* 1996; 3:671–5. [PubMed: 8756323]
50. Matsuura Y, Stewart M. Structural basis for the assembly of a nuclear export complex. *Nature.* 2004; 432:872–7. [PubMed: 15602554]
51. Shen Y, Joachimiak A, Rosner MR, Tang WJ. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature.* 2006; 443:870–4. [PubMed: 17051221]
52. Malito E, et al. Molecular bases for the recognition of short peptide substrates and cysteine-directed modifications of human insulin-degrading enzyme. *Biochemistry.* 2008; 47:12822–34. [PubMed: 18986166]
53. Georgiadou D, et al. Placental leucine aminopeptidase efficiently generates mature antigenic peptides in vitro but in patterns distinct from endoplasmic reticulum aminopeptidase 1. *J Immunol.* 2010; 185:1584–92. [PubMed: 20592285]
54. Burton PR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet.* 2007; 39:1329–37. [PubMed: 17952073]
55. Harvey D, et al. Investigating the genetic association between ERAP1 and ankylosing spondylitis. *Hum Mol Genet.* 2009; 18:4204–12. [PubMed: 19692350]
56. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology.* 1997; 276:307–326.
57. McCoy AJ, et al. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
58. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr.* 2000; 56:965–72. [PubMed: 10944333]
59. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2000; 66:213–21. [PubMed: 20124702]
60. DeLano, WL. *The PyMOL Molecular Graphics System.* DeLano Scientific LLC; Palo Alto, CA, USA: 2008. <http://www.pymol.org>

a



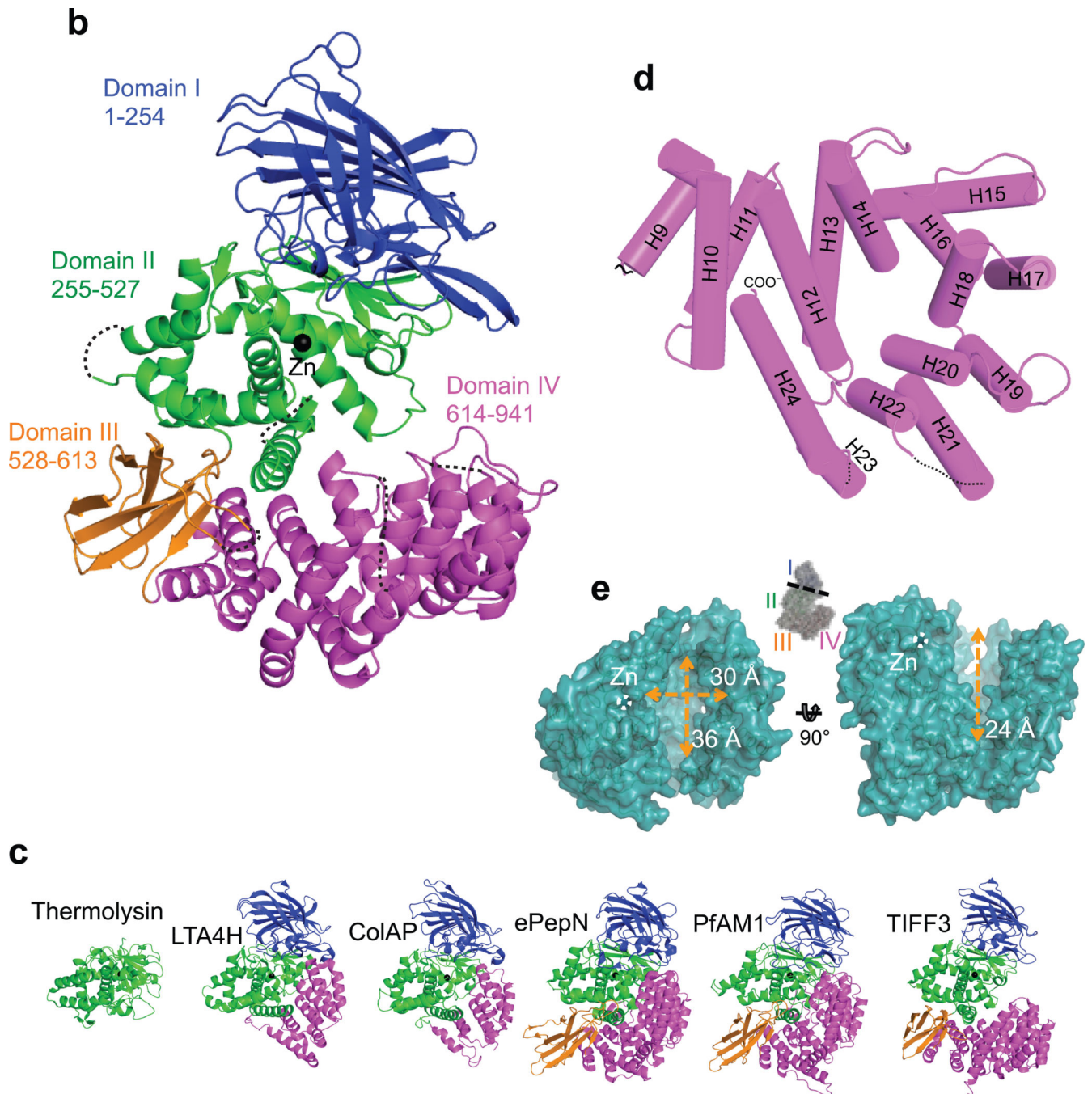


Figure 1. ERAP1 structure and domain arrangement. (a) Structure based alignment of ERAP1 with M1 aminopeptidase family members human leukotriene A4 hydrolase (LTA4H), *T. acidophilum* tricorn interacting factor F3 (TIFF3), *E. coli* aminopeptidase N (ePepN), *P. falciparum* aminopeptidase M1 (PfAM1), and *C. psychrerythraea* cold active aminopeptidase (ColAP). Numbering corresponds to human ERAP1, colored lines indicate domains in ERAP1; coils and strands represent helices and strands respectively, and dashes represent loops not observed in crystal structure. Asterisk marked residues are homologous

among the shown aminopeptidases; boxed residues are highly conserved M1 aminopeptidase motifs HExxEx₁₈E, GAMEN, and Tyr438. Positions of disulfide bonds are indicated and observed glycosylation sites are marked by balloon icons. **(b)** Overall shape of ERAP1 represented as a ribbon diagram and colored according to domains: blue is domain I, green is domain II, orange is domain III, and pink is domain IV. Dotted lines represent disordered loops. **(c)** Structures of M1 aminopeptidases family members and thermolysin, colored by domain: *B. thermoproteolyticus* thermolysin (PDB 3FV4), LTA4H (PDB 3FUH), ColAP (PDB 3CIA), ePepN (PDB 2HPT), PfAM1 (PDB 3EBG), and TIFF3 (PDB 1Z5H). **(d)** Schematic diagram of C-terminal domain of ERAP1 with helices represented as cylinders showing the ARM/HEAT motif spiral forming a large cavity lined by the even numbered helices. **(e)** Large cavity formed by the catalytic domain and C-terminal domain shown in a surface representation. Inset shows how the N-terminal domain was removed above the dashed black line to provide a clear view of the cavity. Orange dashed lines represent estimated distances across the cavity.

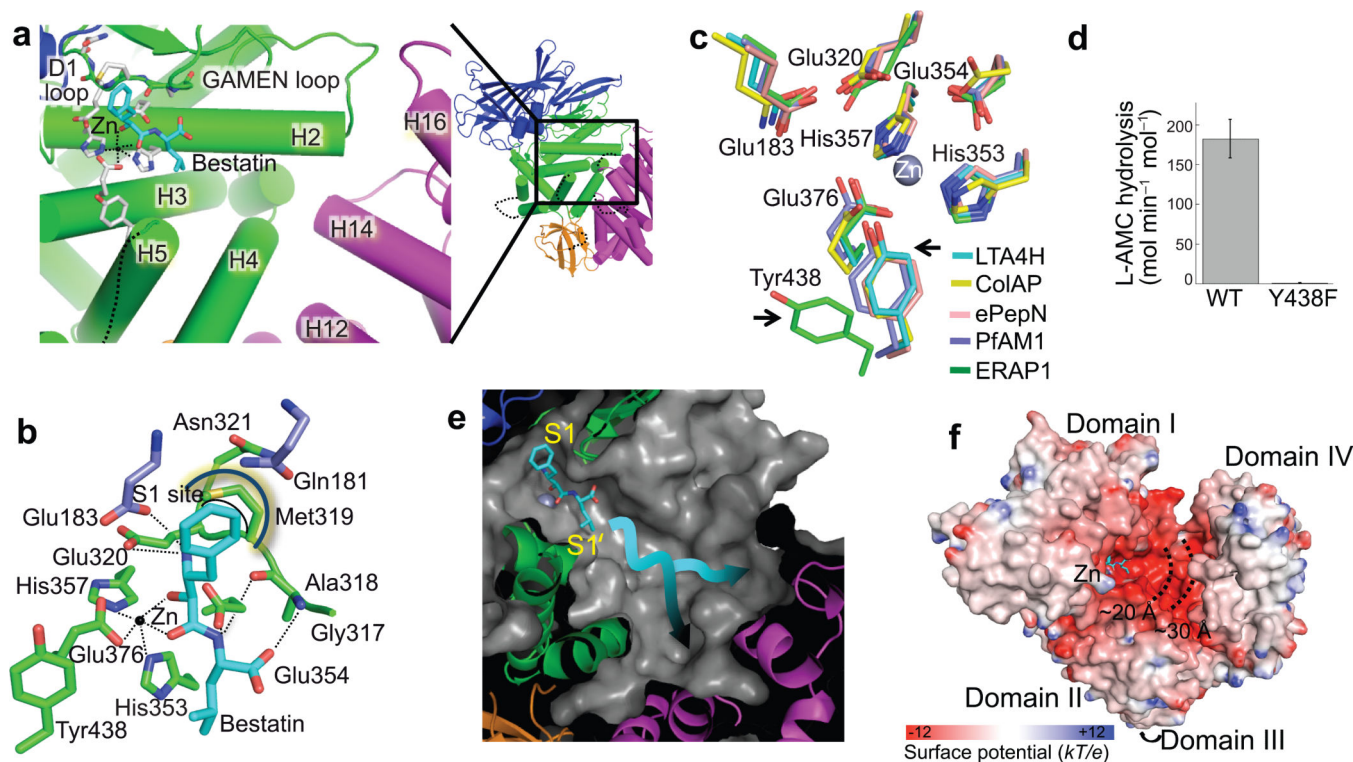


Figure 2. ERAP1 active site. **(a)** Structural elements forming the ERAP1 active site. Loops and helices, shown as cylinders, are colored as in Fig. 1b. Bestatin is shown with cyan bonds, zinc as a black sphere. ERAP1 residues in the active site are shown with white bonds. Inset at right shows the entire protein for orientation; in the view shown helices 19 and 21 that occlude the active site were removed for clarity. **(b)** ERAP1 interactions with bound aminopeptidase inhibitor bestatin. Side chains of conserved catalytic residues of ERAP1 are shown carbon atoms colored by domain. Black arc indicates S1 site. **(c)** Conservation of catalytic residues among M1 family aminopeptidase structures, shown as an overlay of the active sites of LTA4H (cyan), COLAP (yellow), ePepN (pink), PfAM1 (purple), and ERAP1 (green). ERAP1 has an altered orientation of Tyr438 relative to the other aminopeptidases (arrows). Residue numbers correspond to ERAP1. **(d)** ERAP1 enzymatic activity is dependent on the side-chain hydroxyl of Tyr438. L-AMC hydrolysis activity of wild-type and Y438F ERAP1. Error bars show standard deviation of 3 measurements. **(e)** The substrate binding cavity extends from the active site into domain IV. Surface representation of ERAP1 near the active site, with the catalytic zinc shown as a dark gray sphere, and bestatin with cyan bonds. Expected positions of peptide side chain binding pockets S1 and S1' indicated. Arrows represent possible paths of peptides extending from the active site into the cavity. **(f)** Electrostatic surface map of ERAP1. Surface colored from red to blue representing negatively charged to positively charged regions. Dotted lines represent indicated distances from zinc ion (grey sphere). Bestatin colored with cyan bonds.

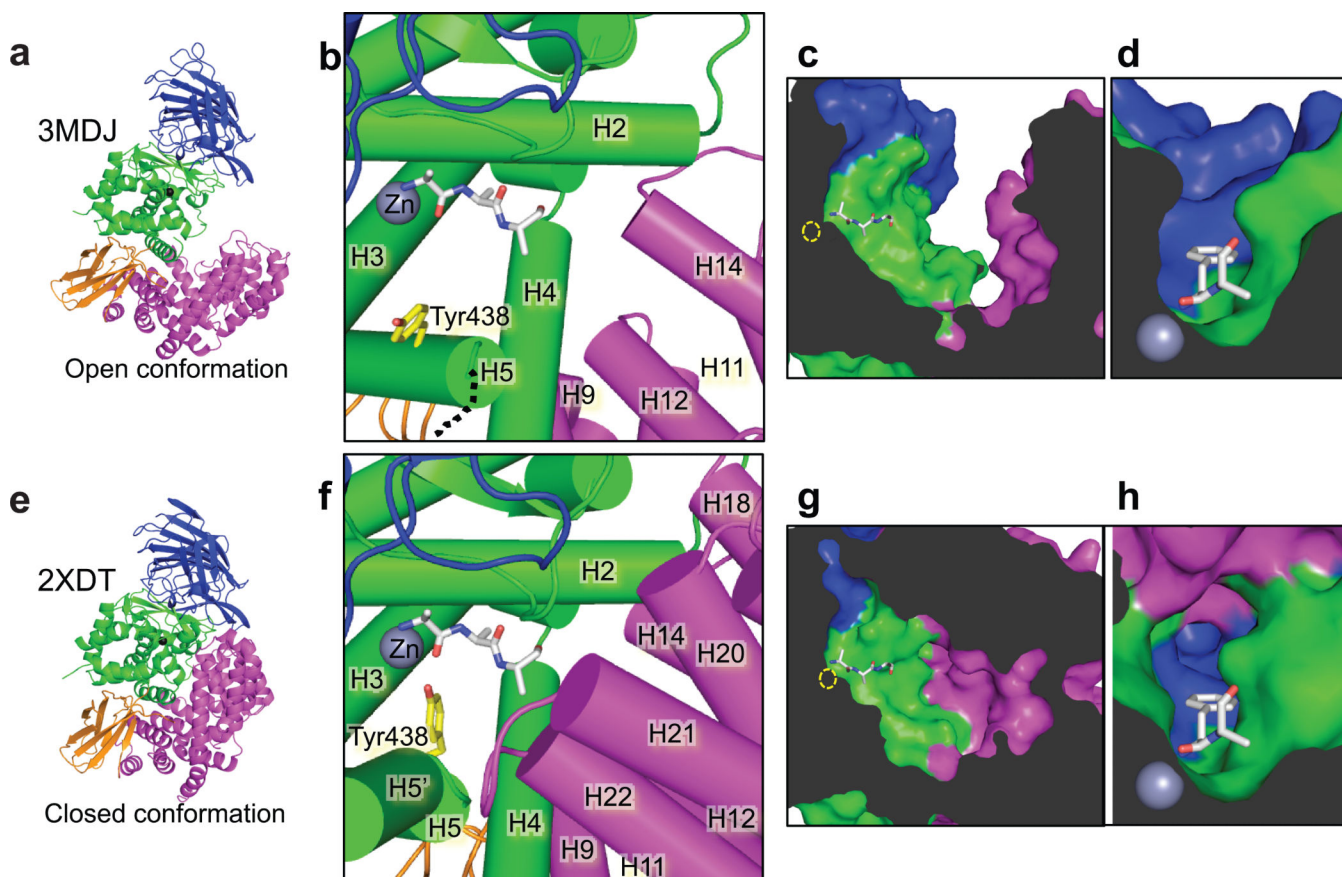


Figure 3. Open and closed conformations of ERAP1. **(a–d)** ERAP1-bestatin in an open conformation as reported here (PDB 2MDJ). **(e–h)** ERAP1 in closed conformation as reported in a recent crystal structure (PDB 2XDT). **(a,e)** Ribbon diagrams, aligned as in Figure 1c, showing domain reorientation. **(b,f)** View of active site, with zinc atom shown in gray, model tripeptide substrate with white bonds, and Tyr438 with yellow bonds. ERAP1 helices shown as cylinders and colored according to domain. Motion of helices H18–H22, reorientation of H5, and folding of an additional region H5' N-terminal to H5 are apparent in the closed conformation. **(c,g)** Cutaway view of peptide binding cavity, with surface of cavity colored according to domain, model peptide substrate with white bonds, and position of zinc atom outlined in yellow. Although the cavity is isolated from solvent in the closed conformation, its length is not substantially changed. **(d,h)** Closeup view of S1 site, looking from the cavity toward towards the S1 site. A phenylalanine side chain has been modeled at the P1 position.

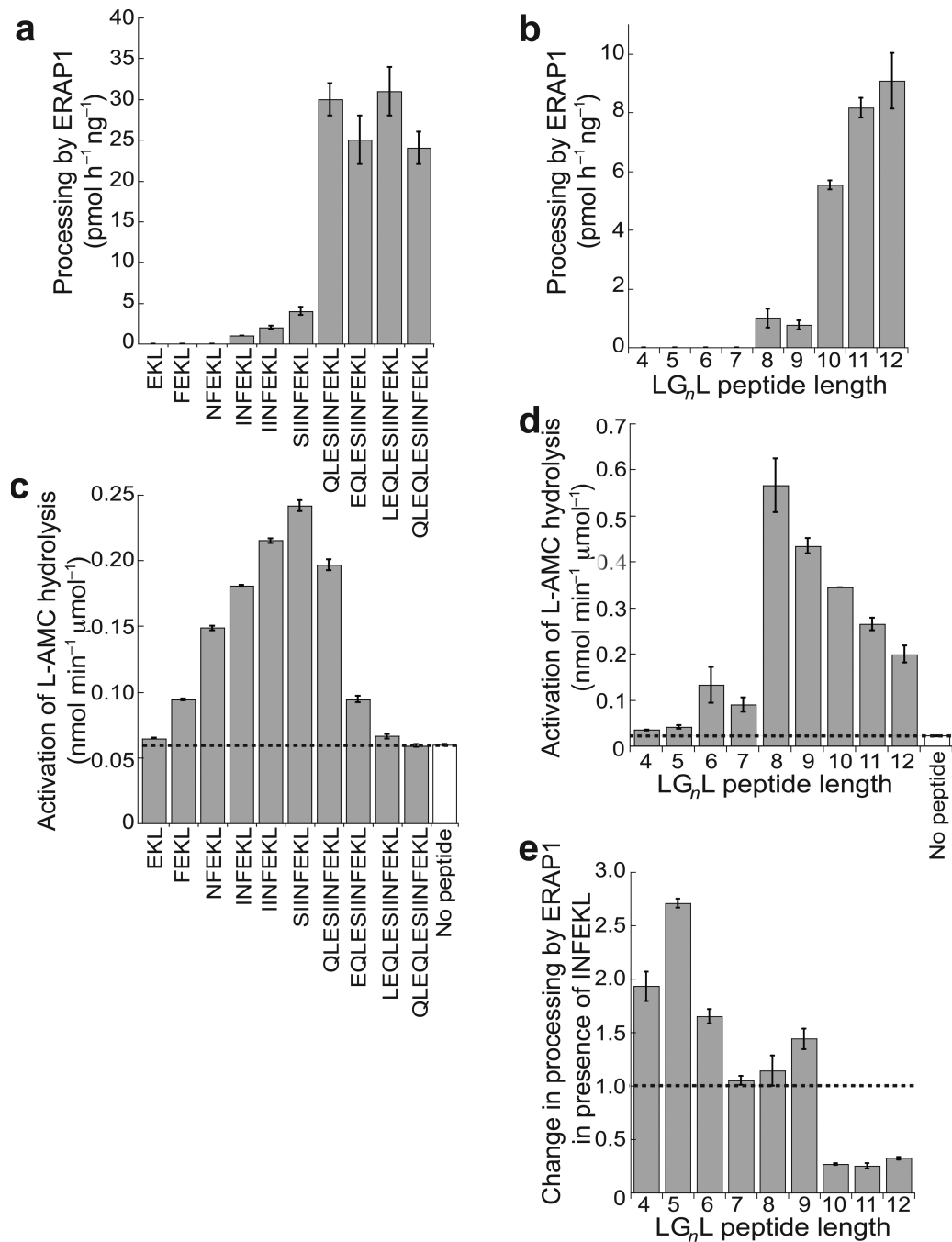


Figure 4. Length dependent peptide cleavage and allosteric activation by ERAP1. **(a)** ERAP1 processing of a series of truncated and extended SIINFEKL variants. **(b)** ERAP1 processing of a series of LG_nL peptides. **(c)** Activation of L-AMC hydrolysis by the same SIINFEKL length variants as in (a). L-AMC hydrolysis in the absence of added peptide is indicated by the open bar and the dotted line. **(d)** Activation of L-AMC hydrolysis by LG_nL peptides. **(e)** ERAP1 processing of LG_nL peptides as in panel c, but in the presence of 50 μM INFEKL peptide. The change in processing by ERAP1 is the ratio of processing with the presence of

INFEKL over the processing by ERAP1 without INFEKL. Standard deviations are indicated by error bars.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

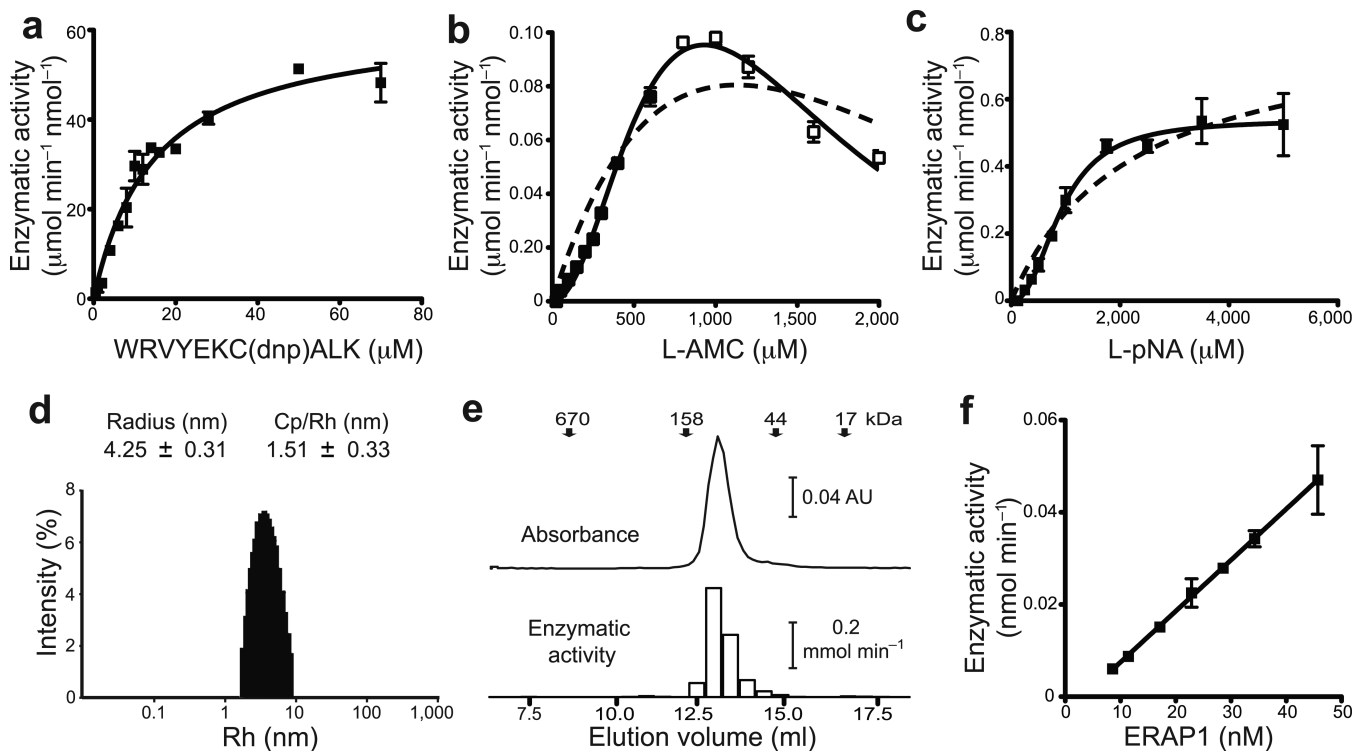


Figure 5.

Kinetic analysis. **(a)** Initial rate of removal of the amino-terminal residue of an internally quenched WRVYEKC^{dnp}ALK peptide, as a function of peptide concentration. Solid line shows best fit to an equation describing simple Michealis-Menten behavior with $K_m = 14.8 \pm 1.7 \mu\text{M}$ and $V_{\text{max}} = 62.3 \pm 2.9 \mu\text{mol min}^{-1} \text{nmol}^{-1}$. **(b)** Initial rate of hydrolysis of the fluorogenic L-AMC substrate as a function of concentration. Lines show best fit to simple Michaelis-Menten (dashed line) and allosteric activation (solid line) equations, which include a correction for inner filter effects and reduced apparent activity at high concentrations (see Methods for details). **(c)** Initial rate of hydrolysis of colorimetric L-pNA substrate as a function of concentration shows sigmoidal kinetics behavior. Lines show best fit to simple Michaelis-Menten (dashed line) and allosteric activation (solid line) equations. **(d)** Dynamic light scattering analysis of ERAP1 in PBS. Hydrodynamic radius estimated for a 105 kDa protein using a spherical model is 4.35 nm. **(e)** Gel filtration analysis shows enzymatically active monomeric population. Top. Elution profile for ERAP1 in PBS followed using absorbance at 280 nm, with elution position of molecular weight standards shown above plot. Bottom. Fractions were tested for L-AMC hydrolysis as a measurement of ERAP1 activity. **(f)** L-AMC hydrolysis is linearly dependent on ERAP1 concentration. Standard deviations are indicated by error bars.

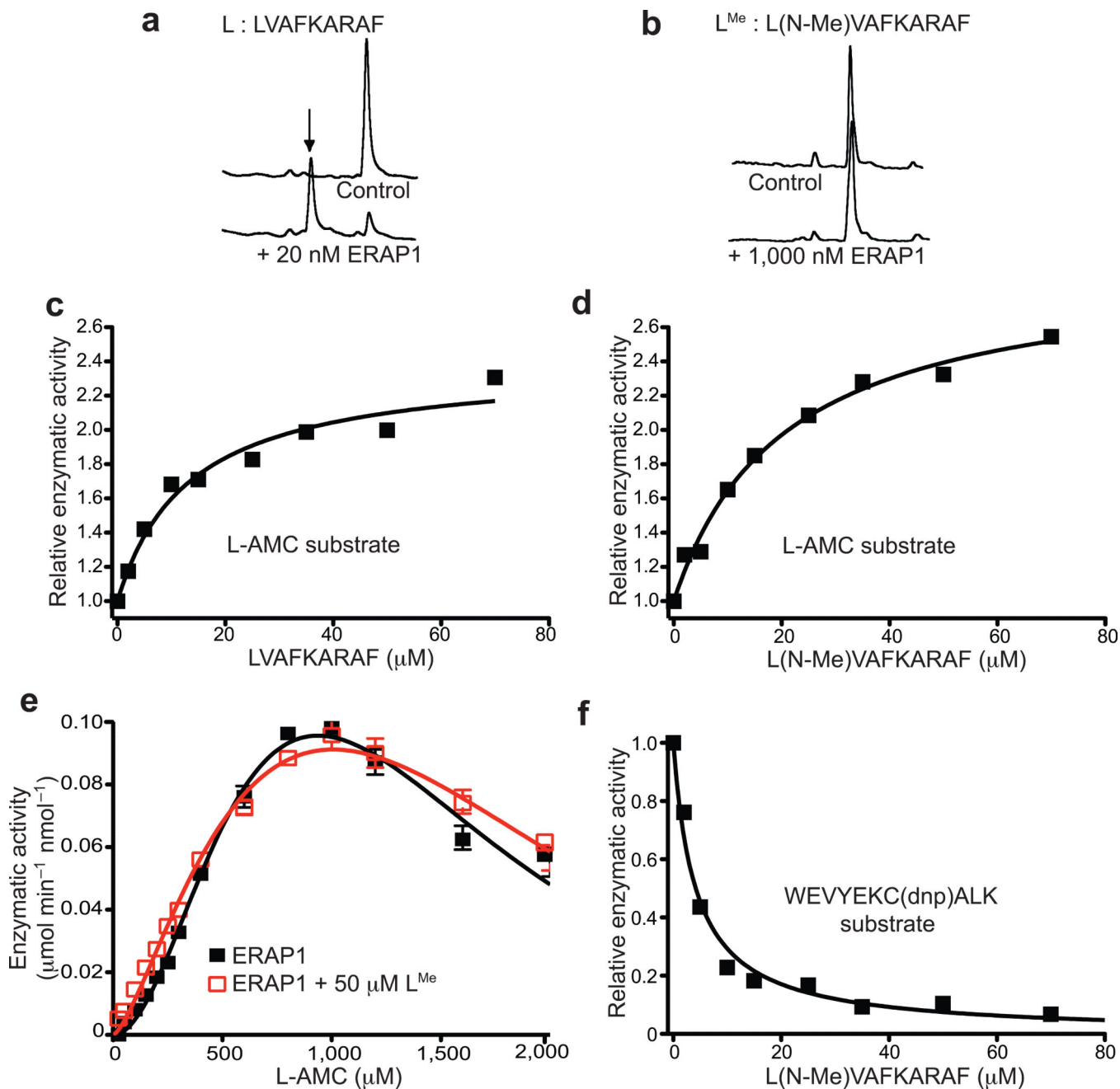


Figure 6. Non-hydrolyzable peptide activates L-AMC hydrolysis but inhibits full-length peptide hydrolysis. **(a)** An optimized ERAP1 substrate, peptide L (LVAFKARAF), is processed efficiently by ERAP1. Reverse-phase chromatograms of peptides incubated in the presence or absence of ERAP1 are shown. Arrow indicates processed product VAFKARAF. **(b)** N-methylation of the first amide bond in L^{Me} peptide prevents degradation by ERAP1. **(c)** ERAP1 hydrolysis of L-AMC is activated by increasing concentrations of peptide L. Line shows fit to an equation describing simple hyperbolic activation. **(d)** ERAP1 hydrolysis of L-AMC is activated by increasing concentrations of L^{Me} , fit as in panel c. **(e)** Concentration

dependence of initial rate of L-AMC hydrolysis by ERAP1 in the absence (filled black symbols) or presence (open red symbols) of 50 μM L^{Me} . Lines show fits to allosteric activation equation as in Fig. 5b. **(f)** ERAP1 hydrolysis of a fluorescent peptide substrate is inhibited by increasing concentrations of L^{Me} . Line shows fit to a competitive binding equation. Standard deviations are indicated by error bars.

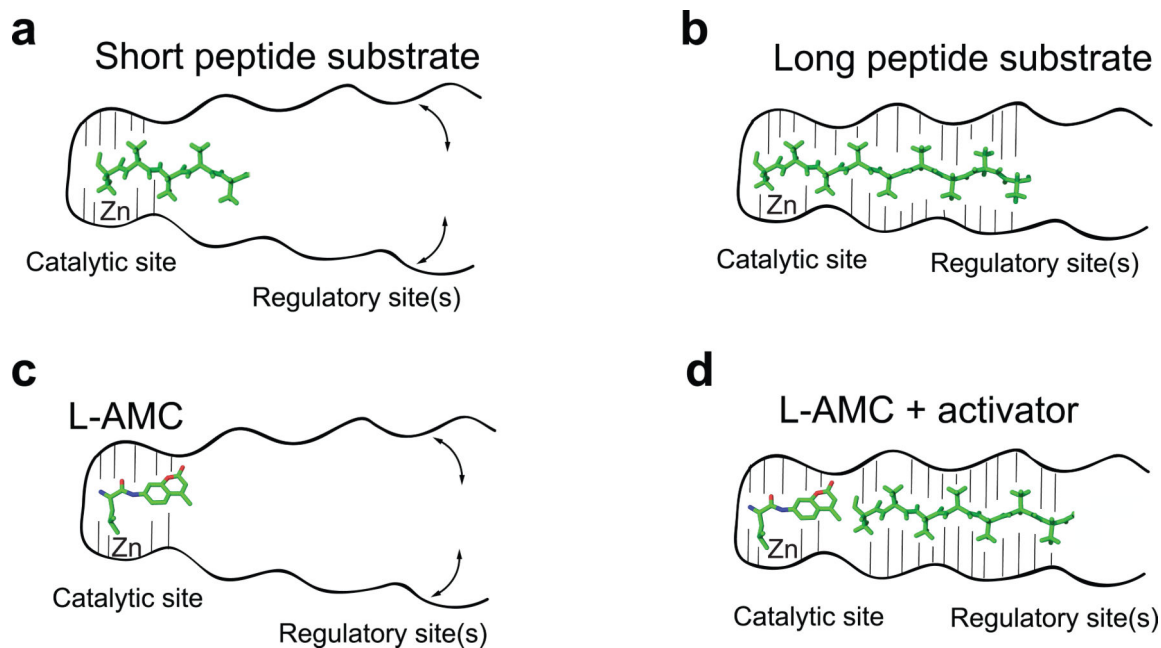


Figure 7. Model for ERAP1 length dependent cleavage activity. **(a)** A short peptide (5-mer shown) cannot reach from the catalytic site to the regulatory site. ERAP1 remains in the lower-activity open conformation and the peptide is processed inefficiently. **(b)** A long peptide (9-mer shown) can reach the regulatory site. ERAP1 adopts the higher-activity closed conformation and the peptide is processed efficiently. **(c)** A small fluorogenic substrate (L-AMC shown) cannot reach from the catalytic site to the regulatory site and is processed inefficiently. **(d)** A peptide (7-mer shown) can bind to the regulatory site together with a smaller substrate (L-AMC shown), leading to conversion to the closed conformation and increased aminopeptidase activity.

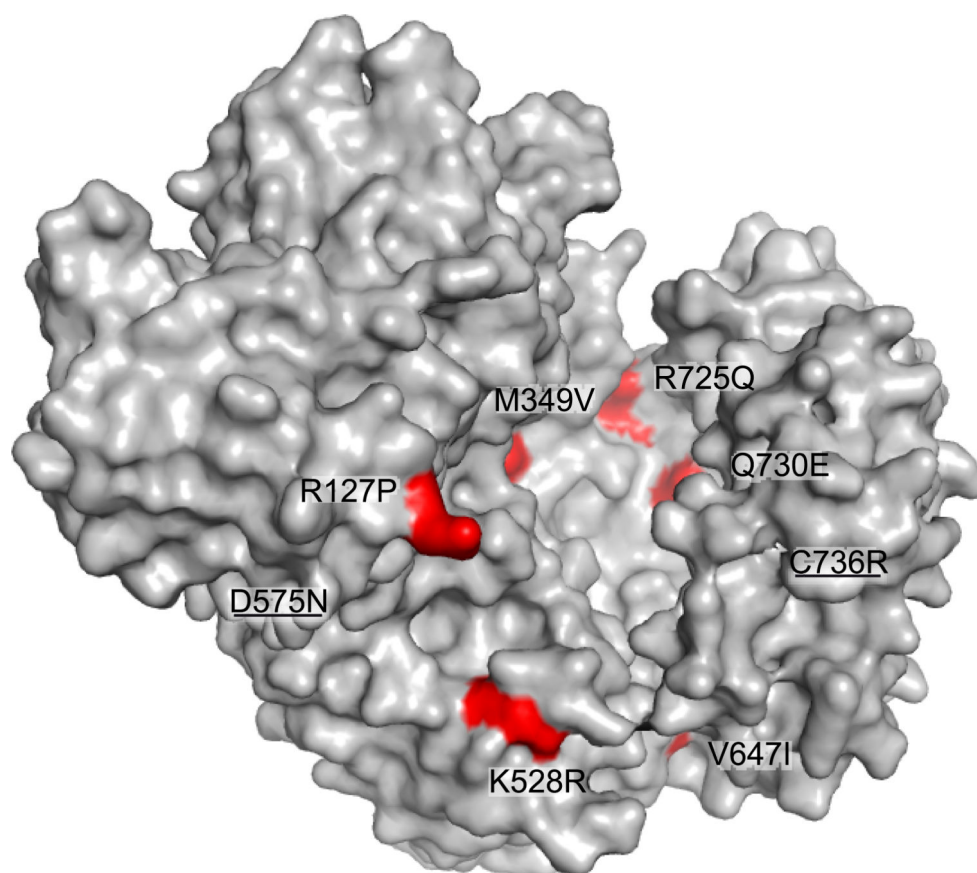


Figure 8.

Ankylosing spondylitis-associated mutations mapped on surface of ERAP1. Surface of ERAP1 in gray, with polymorphisms associated with ankylosing spondylitis shown in red. Underlined residues are hidden in this view.

Table 1

Data collection and refinement statistics

ERAP1-bestatin [†]	
Data collection	
Space group	$P2_1$
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	70.9, 234.2, 95.6
α , β , γ (°)	90.0, 103.6, 90.0
Resolution (Å)	38–2.95 (3.1–2.95)*
R_{sym}	17.8 (76.5)
$I / \sigma I$	7.5 (1.1)
Completeness (%)	99.8 (98.8)
Redundancy	3.6 (3.4)
Refinement	
Resolution (Å)	38–2.95
No. reflections	63,522
$R_{\text{work}} / R_{\text{free}}$	20 / 26
No. atoms	
Protein	20,556
Ligand / Zn ⁺ ion	66 / 3
<i>B</i> -factors	
Protein	68.0
Ligand / Zn ⁺ ion	161.9 / 41.6
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.85

* Values in parentheses are for highest-resolution shell.

[†] Dataset was from collected from one crystal.