

# Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides

Efthalia ZERVOUDI\*<sup>1</sup>, Athanasios PAPAKYRIAKOU†<sup>1</sup>, Dimitra GEORGIADOU\*, Irini EVNOUCHIDOU\*, Anna GAJDA‡, Marcin POREBA‡, Guy S. SALVESEN§, Marcin DRAG‡§, Akira HATTORI||, Luc SWEVER§¶, Dionisios VOURLIOMIS† and Efstratios STRATIKOS\*<sup>2</sup>

\*Protein Chemistry Laboratory, IRRP, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece, †Chemical Biology Laboratory, Institute of Physical Chemistry, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece, ‡Division of Medicinal Chemistry and Microbiology, Wrocław University of Technology, 50-370 Wrocław, Poland, §Program in Apoptosis and Cell Death Research, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, U.S.A., ||Graduate School of Pharmaceutical Sciences, Kyoto University, Department of System Chemotherapy and Molecular Sciences, Sakyo, Kyoto 606-8501, Japan, and ¶Insect Molecular Genetics and Biotechnology Group, Institute of Biology, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece

ERAP1 (endoplasmic reticulum aminopeptidase 1), ERAP2 and IRAP (insulin-regulated aminopeptidase) are three homologous enzymes that play critical roles in the generation of antigenic peptides. These aminopeptidases excise amino acids from N-terminally extended precursors of antigenic peptides in order to generate the correct length epitopes for binding on to MHC class I molecules. The specificity of these peptidases can affect antigenic peptide selection, but has not yet been investigated in detail. In the present study we utilized a collection of 82 fluorogenic substrates to define a detailed selectivity profile for each of the three enzymes and to probe structural and functional features of the S1 (primary specificity) pocket. Molecular modelling of the three S1 pockets reveals substrate–enzyme interactions that

are critical determinants for specificity. The substrate selectivity profiles suggest that IRAP largely combines the S1 specificity of ERAP1 and ERAP2, consistent with its proposed biological function. IRAP, however, does not achieve this dual specificity by simply combining structural features of ERAP1 and ERAP2, but rather by a unique amino acid change at position 541. The results of the present study provide insights on antigenic peptide selection and may prove valuable in designing selective inhibitors or activity markers for this class of enzymes.

**Key words:** aminopeptidase, antigen, kinetics, peptide specificity, site-directed mutagenesis, substrate library.

## INTRODUCTION

### Antigen presentation and processing

Cytotoxic T-lymphocytes identify infected or transformed cells by recognizing small antigenic peptides bound on to cell-surface receptors of MHC I. These antigenic peptides are derived from the proteolysis of intra- or extra-cellular proteins and constitute an indicator of the health status of the cell. Aberrant generation of antigenic peptides can lead either to immune evasion or autoimmunity. Antigenic peptides are generated intracellularly by complex proteolytic pathways [1]. A key component of these pathways is the proteasome, a large intracellular multi-subunit protease that generates fragments from intracellular or endocytosed proteins. Peptides generated by the proteasome are transported into the ER (endoplasmic reticulum) by a specialized ATP-dependent peptide transporter, TAP (transporter associated with antigen processing) [2]. A similar but distinct pathway, the cross-presentation pathway, operates in specialized intracellular vesicles that contain endocytosed extracellular proteins [3]. The proteasome-generated peptides usually have the correct C-terminus as the final antigenic peptides, but also have N-terminal extensions that make them too large to bind on to MHC class I molecules that have stringent length requirements with a general preference for nonamers. Although these extensions vary from one to six amino acids long, the most common extension is one amino acid [4]. Inside the ER, resident aminopeptidases trim these antigenic peptide precursors to generate the mature antigenic peptides that can then bind on to nascent MHC class I molecules [5,6].

ERAP1 (ER aminopeptidase 1) and ERAP2 are two specialized aminopeptidases that reside in the ER and have been demonstrated to trim antigenic peptide precursors to generate mature antigenic peptides [7–10]. Recently, a homologous aminopeptidase, IRAP (insulin-regulated aminopeptidase; or PLAP) has been demonstrated to perform a similar function in cross-presentation vesicles [11]. These three aminopeptidases share approximately 50% sequence identity. As a result of their shared homology and function, it has been proposed that they constitute a distinct sub-family of aminopeptidases within the metalloprotease classification M1 [12]. ERAP1 is the best characterized of the three and has been shown to affect antigen presentation *in vivo*, shaping the pool of antigenic peptides and influencing immunodominance [13–17]. Inhibition of ERAP1 by the broad-spectrum metalloprotease inhibitor Leucinethiol was sufficient to replicate gene knock-down experiments in cells and to induce alterations in the repertoire of the antigenic peptides [14]. ERAP1 has unusual, for an aminopeptidase, enzymatic properties, preferring to trim longer peptides down to a length of 8–10 amino acids, the appropriate length for MHC class I binding [18]. It has relatively broad substrate specificity showing preferences for side chains throughout the peptide-substrate sequence [19]. IRAP shares some of the molecular properties of ERAP1 in generating mature antigenic epitopes, although recent findings suggest that it does so in distinct patterns, suggesting differences in specificity [20].

The trimming specificity of the N-terminal amino acid from antigenic peptide precursors by aminopeptidases is a strong determinant for the generation of mature antigenic peptides and the determination of the antigenic peptide repertoire. A large

Abbreviations used: ACC, 7-amino-4-carbamoylmethylcoumarin; AMC, 7-amino-4-methylcoumarin; ER, endoplasmic reticulum; ERAP, ER aminopeptidase; hArg, homo-arginine; hTyr, homo-tyrosine; IRAP, insulin-regulated aminopeptidase; R-AMC, L-arginine-AMC.

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed (email stratos@rrp.demokritos.gr or stratikos@gmail.com).

number of antigenic peptide precursors carry only a single amino acid extension, whose trimming will be largely affected by the N-terminal specificity of the aminopeptidase [4]. The *in vitro* trimming preferences of ERAP1 have been recently demonstrated to largely determine antigenic peptide presentation in cultured cells [21]. Although highly homologous, ERAP1/2 and IRAP do not have the same specificity. Using chromogenic substrates it has been reported that the preferred residue for ERAP1 is leucine, whereas for ERAP2 is arginine [22,23]. IRAP can cleave both substrates [24]. The exact role of these specificity differences in the biological function of these enzymes is not clear, nor have they been investigated in any detail. In the present study we set forth to characterize in detail the shape, size and composition of the S1 specificity pocket of each enzyme, in an effort to better understand the molecular determinants that contribute to antigenic peptide repertoire generation. By a combination of substrate library screening, molecular modelling and site-directed mutagenesis, we unravel key features of the S1 pocket of these enzymes that are consistent with their distinct biological functions and may be valuable for the rational design of selective inhibitors or activity markers.

## EXPERIMENTAL

### Protein expression and purification

Recombinant ERAP1 was produced by insect cell culture after infection with recombinant baculovirus carrying the ERAP1 coding sequence and isolated from the cell supernatant as previously described [19]. A recombinant and soluble form of IRAP was produced by 293F cells grown in suspension after transfection with a plasmid vector carrying the IRAP coding sequence as previously described [20].

For production of recombinant ERAP2, the sequence coding for full-length human ERAP2 was inserted in the pFastBac1 vector between the BssHIII and NotI restriction endonuclease sites. The final construct contained the 21-bp A-rich sequence derived from a lobster tropomyosin cDNA leader sequence adjacent to the initiation codon and a C-terminal His<sub>6</sub> tag for efficient expression and purification. The pFastBac1-ERAP2 vector was used to generate recombinant baculovirus according to the manufacturer's instructions (Invitrogen). The recombinant baculovirus was used to infect Hi5 cells grown in suspension in SF900II serum-free medium. At 3 days post-infection, recombinant ERAP2 was found in the cell supernatant, harvested by centrifugation and isolated by Ni-NTA (Ni<sup>2+</sup>-nitrilotriacetate) affinity chromatography as previously described for ERAP1 [19].

### Mutagenesis

Site-directed mutagenesis for the construction of the E541R mutation in human IRAP was performed using the QuikChange<sup>®</sup> II XL kit according to the manufacturer's instructions (Agilent Technologies). The primers used for the mutagenesis were 5'-TCATCTGTTTCAGTCTTCAGAACAAATTCGAGAAATGTTGATTCTCTTCC-3' (sense) and 5'-GGAAAGAGAATCAAACATTTCTCGAATTTGTTCTGAAGACTGAACAGATGA-3' (antisense). Successful mutagenesis was confirmed by DNA sequencing.

### Library synthesis

Of the 82 fluorogenic substrates in the library, 61 have been described previously [25]. All new compounds [D-amino acids-ACC (where ACC is 7-amino-4-carbamoylmethylcoumarin), L-hTyr (homo-tyrosine)-ACC, L-4-guanidino-phenylalanine-ACC

and L-dehydrotryptophan-ACC] were synthesized using protocols described previously [25]. HPLC purification and post-purification analysis of all new compounds were conducted on a Waters M600 solvent delivery module equipped with a Waters M2489 Detector system using preparative Waters Spherisorb S100DS2 or analytical Waters Spherisorb S50DS2 columns. Solvent composition: system A [water/0.1% TFA (trifluoroacetic acid)] and system B [acetonitrile/water 80%:20% (v/v) with 0.1% TFA]. All substrates were at least 95% pure and were validated by ESI-MS (electrospray ionization MS) at the Mass Spectrometry Facility at the Department of Chemistry of University of Wrocław. The chemical structures for all 82 substrates are shown in Supplementary Figure S1 (at <http://www.BiochemJ.org/bj/435/bj4350411add.htm>).

### Fluorogenic assay

Trimming of the fluorogenic peptide substrates by ERAP1, ERAP2 and IRAP was followed using a TECAN infinite<sup>®</sup> M200 microplate fluorescence reader. The samples were excited at 380 nm and fluorescence was recorded at 460 nm. The reactions were followed for 5–10 min at 24 °C. In all cases the rise in fluorescence was linear with time indicating steady-state kinetics. The slope of the time-course was used to calculate the reaction rate. L-AMC [L-leucine-AMC (7-amino-4-methylcoumarin)] and R-AMC (L-arginine-AMC) substrate controls were included in every plate to allow comparison between data collected from different plates.

### Homology modelling

Multiple sequence alignment of human ERAP1 (isoform a, NP\_057526.3), ERAP2 (NP\_001123612.1) and IRAP (isoform a, NP\_005566.2) was performed using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) with the default parameters (Supplementary Figure S2 at <http://www.BiochemJ.org/bj/435/bj4350411add.htm>). The good overall sequence identity of ERAP2 and IRAP with ERAP1 (49% and 44% respectively), especially considering the higher degree of identity at the catalytic subsites of interest, provided a solid template for homology modelling. On the basis of the crystal structure of ERAP1 (PDB code 2XDT), ERAP2 and IRAP models were generated using Modeller 9v4 [26]. Residues Pro<sup>46</sup>–Arg<sup>940</sup> from the PDB 2XDT structure were used as a template for the generation of ERAP2 (Arg<sup>61</sup>–Thr<sup>960</sup>) and IRAP (Leu<sup>60</sup>–Leu<sup>1025</sup>) models (excluded residues are shown in Supplementary Figure S2). The model with the lowest objective function value was selected for further optimization using AMBER 9 [27]. Hydrogen and missing heavy atoms, including Zn(II), disulfide and metal–ligand bonds were added using XLEAP. The AMBER-based parm99SB force field was applied to all protein atoms, and parameters for the Zn(II) co-ordination sphere were taken from [28]. Subsequently, the position of hydrogen atoms and the metal ion site was optimized with energy minimization in a vacuum using a distance-dependent dielectric and a 20 Å (1 Å = 0.1 nm) cut-off for non-bonded interactions. The quality of ERAP2 and IRAP models was assessed using the Structural Analysis and Verification Server (<http://nihserver.mbi.ucla.edu/SAVS>), which exhibited a reasonable degree of quality by virtue of their sequence alignment.

### Substrate docking

The substrate library was generated starting from the SMILES (simplified molecular input line entry specification)

representation of each compound, then OMEGA [29] was used to calculate the initial three-dimensional co-ordinates and QUACPAC (Openeye Inc.) to apply AM1-BCC atomic charges [30]. Docking of the substrates was performed using AutoDock 4.2 [31] with default parameters except for the number of docking rounds, which was increased to 100. Non-polar hydrogen atoms were merged and Kollman charges were applied to the protein atoms using AutoDockTools 1.4.5. Ligands were treated as fully flexible excluding amide bonds and guanidinium groups. The docked model chosen for analysis was among the highest binding energy conformations with proper orientation for substrate binding (i.e. the scissile bond  $C=O \cdots Zn(II)$  distance  $< 2.5 \text{ \AA}$ ). No further optimization of the predicted enzyme–substrate interactions was attempted. Electrostatic potential surfaces were generated using APBS and PME electrostatics packages [32]. VMD 1.8.6 and PyMOL (<http://www.pymol.org>) was used for visual inspection and rendering of the Figures [34].

## RESULTS

### Screening strategy

We used a collection of 82 fluorogenic substrates to generate a selectivity profile for ERAP1, ERAP2 and IRAP. The experimental conditions for each screen were designed so that every substrate used would be assayed under sub- $K_m$  concentrations so that the rate of cleavage would linearly correlate with the  $k_{cat}/K_m$  value of each substrate and enzyme as previously described [25]. To ensure this, we first generated Michaelis–Menten plots for the best-known substrate for each enzyme. We used L-ACC for ERAP1 and IRAP and R-ACC for ERAP2. The  $K_m$  for ERAP1 is larger than 1 mM (estimated to be  $1150 \pm 305 \mu\text{M}$ ) and the  $K_m$  values for ERAP2 and IRAP are  $90 \pm 3 \mu\text{M}$  and  $85 \pm 30 \mu\text{M}$  respectively (results not shown). As a result we chose to do all screening assays at substrate concentrations below  $10 \mu\text{M}$ . After the preliminary screening, substrates for which no signal was measured were re-screened at a concentration of  $100 \mu\text{M}$  in an effort to quantify trimming rates for poor substrates. Using this strategy we estimated a minimum trimming rate difference between good and non-processed substrates of approximately 200-fold.

### Selectivity profiles for ERAP1, ERAP2 and IRAP

The relative trimming rates for each of the 82 substrates with each of the three enzymes are shown in Figure 1. Rates are plotted as a fraction of the best substrate for each enzyme. The three enzymes share key preferences but also display marked differences. ERAP1 efficiently trimmed approximately 16 of the 82 substrates, showing significant preference for hydrophobic and aromatic amino acids, as well as for long aliphatic side chains. Accordingly, the best performing substrate was hTyr. ERAP2 displayed a significantly different profile than ERAP1, with strong preferences for positively charged amino acids. Overall, ERAP2 efficiently trimmed approximately 10 of the 82 substrates, with several key differences from ERAP1. The best two substrates, arginine and hArg (homo-arginine), had a guanidinium group, revealing a strong preference for extended chains with positively charged ends. Shorter hydrophobic side chains were processed to a smaller degree. Similarly to ERAP1, ERAP2 appeared to prefer extended carbon side chains, but in contrast with ERAP1, ERAP2 displayed a very strong preference for a positive charge at the end of those chains. The selectivity profile of IRAP was the most permissive of the three enzymes. IRAP was able to trim at least

25 of the 82 substrates in the library. Interestingly, in almost all cases, IRAP was able to process the substrates that were trimmed by either ERAP1 or ERAP2. This finding suggests that IRAP has the combined specificity of ERAP1 and ERAP2. However, some exceptions were evident since a few substrates were not processed by ERAP1 or ERAP2 but trimmed by IRAP (cyclopentyl-glycine, Abu, Bpa) and vice versa (3-NO<sub>2</sub>-tyrosine). These observations suggest that although IRAP can generally process the sum of substrates of ERAP1 and ERAP2, it may use distinct molecular interactions to achieve this specificity.

### Non-natural side chains probe the characteristics of the S1 pockets

We employed amino acids with unnatural side chains to gain insight on structural and functional features of the S1 pockets (Figure 1B). Interestingly, we identified a much larger number of unnatural side chains as good substrates for all three enzymes. Accordingly, the best substrate for ERAP1 was hTyr, hArg was the second best substrate for ERAP2, and both of those substrates were optimal for IRAP. This observation suggests that the S1 pockets are not strictly optimized for natural amino acids but can easily accommodate more complex structures.

D-amino-acid-based substrates were poorly processed by all three enzymes, suggesting that the L-configuration is a prerequisite for binding and/or catalysis. Under typical experimental conditions only ERAP2 was found to be able to trim D-arginine, albeit  $\sim 50$ -fold slower than L-arginine. Michaelis–Menten analysis of these two substrates suggested that the lower trimming rate was due to both changes on the  $K_m$  ( $90 \pm 3 \mu\text{M}$  for L-Arg and  $1053 \pm 304 \mu\text{M}$  for D-Arg) and  $k_{cat}$  parameters ( $0.177 \pm 0.003 \text{ s}^{-1}$  for L-Arg and  $0.038 \pm 0.018 \text{ s}^{-1}$  for D-Arg) (see Supplementary Figure S3 at <http://www.BiochemJ.org/bj/435/bj4350411add.htm>). These findings suggest that the L-configuration is crucial for both binding and catalysis for this family of enzymes.

Similarly to human aminopeptidase N (CD13) [25], the enzymes analysed in the present paper had a very strong preference for amino acids with an amino group in the  $\alpha$  position and were completely inactive toward substrates with a hydroxy group in the  $\alpha$  position, such as Apns, or amino acids with no amino group present in this position, such as 6-Ahx or  $\beta$ -Ala. This finding is consistent with the important role for the peptide N-terminus in substrate recognition [35].

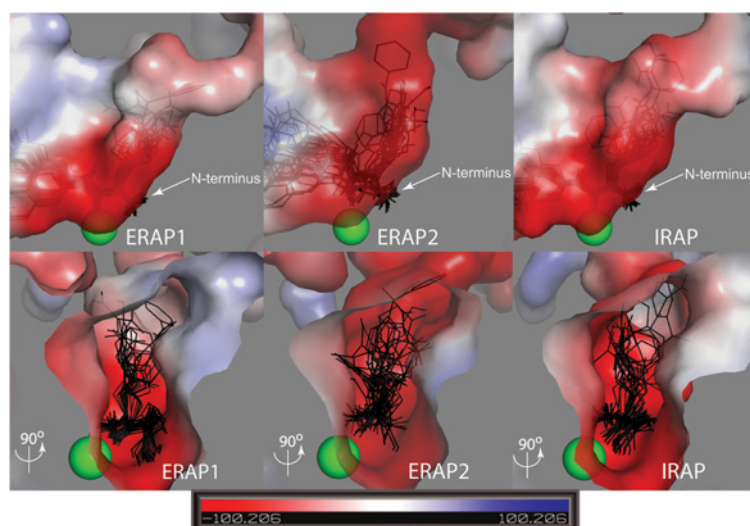
Some of the substrates in the library have side chains of substantial size and would be expected to fit only in large S1 pockets. The S1 pocket of IRAP in particular appears to be able to accommodate the most bulky and hydrophobic substrates in the library (Bpa and Igl), whereas ERAP1 and ERAP2 processed them poorly. This result indicates that the S1 pocket of IRAP may be the largest of the three. Finally, lack of processing of conformationally restricted substrates such as 1-Nal, 2-Nal or Bip by ERAP1, ERAP2 and IRAP suggests that although the pocket is large, it is well-defined and rigid so as to exclude side-chain structures that are not flexible enough to adopt appropriate configurations. Overall, these observations suggest that it may be possible to optimize S1 recognition for each enzyme by incorporating bulky non-natural side chains in the substrate.

### ERAP1 and ERAP2 mixture behaves similarly to IRAP

Saveanu et al. [10] have previously suggested that ERAP1 and ERAP2 operate in a concerted manner in the ER. In contrast, IRAP has been suggested to operate on a separate pathway of cross-presentation, distal from compartmentalized ERAP1 and







**Figure 3** Surface representation of the S1 pocket for ERAP1, ERAP2 and IRAP coloured by electrostatic potential

The best substrates C for each enzyme are shown as stick models in the predicted conformations.

N-terminal docking site and by additional negatively charged side chains in the S1 pocket. ERAP2 has the most negatively charged residues of the three enzymes (Glu<sup>177</sup>, Asp<sup>198</sup> and Asp<sup>888</sup>), whereas IRAP has two (Glu<sup>426</sup> and Glu<sup>541</sup>) and ERAP1 only one (Glu<sup>865</sup>). The only basic residue within the three S1 pockets belongs to ERAP1 (Arg<sup>430</sup>), leading to altered electrostatic potential near the top of the S1 pocket (Figure 3, blue coloured region), an observation that supports the poorer ability of ERAP1 to trim positively charged P1 substrates.

### Key residues that control specificity

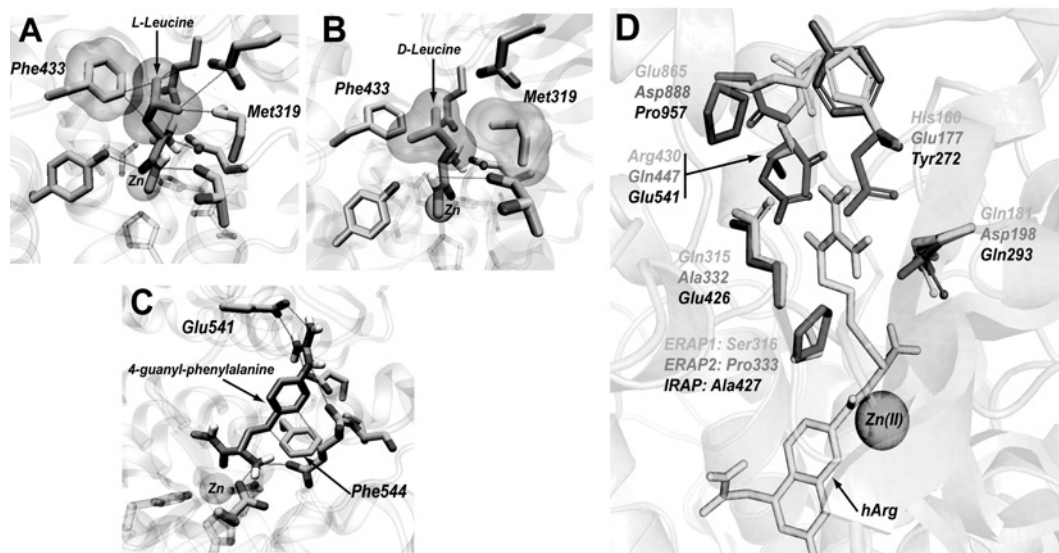
By analysing the interactions between docked substrates and protein side chains, we were able to define the residues that line the S1 pocket for the three enzymes. These residues are listed in Table 1 and indicated in the alignment in Supplementary Figure S2. Half of these residues are conserved between the three enzymes (positions 184, 314, 319, 433, 864 and 868 in ERAP1 numbering) and presumably contribute to the common general characteristics of the pocket. Five of these conserved residues are non-polar (Pro<sup>184</sup>, Phe<sup>314</sup>, Met<sup>319</sup>, Phe<sup>433</sup> and Phe<sup>864</sup> in ERAP1) and may support the preference of all three enzymes for non-polar P1 substrates. Interestingly, Phe<sup>433</sup>/Phe<sup>450</sup>/Phe<sup>544</sup> comes in close proximity to the  $\beta$ -carbon of the substrate backbone, leading to unfavourable steric hindrance with any substrates with  $\beta$ -branched side chains such as valine, threonine or isoleucine (Figure 4A). Instead, the Phe<sup>433</sup>/Phe<sup>450</sup>/Phe<sup>544</sup> residues are predicted to provide favourable aromatic  $\pi$  interactions with the guanidinium groups of arginine and hArg P1-bearing substrates as well as favorable CH- $\pi$  interactions with linear aliphatic chains (Figure 4C). Accordingly, the longer hTyr, hLeu (homo-leucine) and norleucine are the same or better substrates in comparison with tyrosine, leucine and isoleucine. On the opposite side of the pocket, Met<sup>319</sup>/Met<sup>336</sup>/Met<sup>430</sup> is predicted to make contacts with the C $\beta$  atom of the substrates, leading to steric hindrance for substrates with D-configuration and decreased binding affinity (Figure 4B). D-Substrates are predicted to bind with a relatively different configuration of their scissile peptide bond compared with L-substrates (compare Figure 4A with 4B), consistent with the reduced catalytic efficiency we observed (lower  $k_{cat}$ ).

**Table 1** Residues of ERAP1, ERAP2 and IRAP that are predicted to provide key interactions with substrates and help form the S1 specificity pocket

Conserved residues between the three enzymes are in bold.

ERAP1	ERAP2	IRAP
His <sup>160</sup>	Glu <sup>177</sup>	Tyr <sup>272</sup>
Gln <sup>181</sup>	Asp <sup>198</sup>	Gln <sup>293</sup>
<b>Pro<sup>184</sup></b>	<b>Pro<sup>201</sup></b>	<b>Pro<sup>296</sup></b>
<b>Phe<sup>314</sup></b>	<b>Phe<sup>331</sup></b>	<b>Phe<sup>425</sup></b>
Gln <sup>315</sup>	Ala <sup>332</sup>	Glu <sup>426</sup>
Ser <sup>316</sup>	Pro <sup>333</sup>	Ala <sup>427</sup>
<b>Met<sup>319</sup></b>	<b>Met<sup>336</sup></b>	<b>Met<sup>430</sup></b>
Arg <sup>430</sup>	Gln <sup>447</sup>	Glu <sup>541</sup>
<b>Phe<sup>433</sup></b>	<b>Phe<sup>450</sup></b>	<b>Phe<sup>544</sup></b>
<b>Phe<sup>864</sup></b>	<b>Phe<sup>887</sup></b>	<b>Phe<sup>956</sup></b>
Glu <sup>865</sup>	Asp <sup>888</sup>	Pro <sup>957</sup>
<b>Ser<sup>868</sup></b>	<b>Ser<sup>891</sup></b>	<b>Ser<sup>960</sup></b>

Six of the 12 residues that define the S1 pockets vary between the enzymes and contribute to differences between the three S1 pockets that underlie changes in specificity (Figure 4D). Of these residues, two were found to be of particular importance for interactions that appear critical for the differences in specificity between the three enzymes. The polar residue at position 181/198/293 (ERAP1/ERAP2/IRAP numbering) is a glutamine in ERAP1 and IRAP, but is an aspartate in ERAP2. Its positioning adjacent to the GAMEN motif makes it appropriate for interactions with positively charged side chains and has been shown to be important for the selectivity of ERAP2 by site-directed mutagenesis [36]. Interestingly, although IRAP, similarly to ERAP2, is able to process substrates with positively charged side chains, it does not contain an aspartate residue at position 181/198/293 but resembles ERAP1 by having a glutamine residue. This observation raises the question on how IRAP is able to recognize positive charges in the S1 pocket. Docking of positively charged substrates in IRAP suggests that at least a non-conserved, negatively charged amino acid in IRAP, Glu<sup>541</sup>, may be a candidate residue for providing salt-bridge interactions to stabilize positively charged substrates in the S1 cavity of IRAP.



**Figure 4** Key residues that define the S1 pocket

(A) Phe<sup>433</sup> in ERAP1 is stacked closely with a leucine side chain of the substrate. (B) Met<sup>319</sup> makes unfavourable steric interactions with the C,β of D-leucine, leading to an altered binding conformation of the scissile peptide bond. (C) Simulated interactions between a 4-guanyl-phenylalanine side chain with Phe<sup>544</sup> and Glu<sup>541</sup> in IRAP. (D) The six non-conserved amino acids that define the S1 pocket of each enzyme; the predicted conformation of hArg is indicated by an arrow.

(Figure 4C). ERAP1 has an arginine residue at the equivalent position (Arg<sup>430</sup>) and ERAP2 has a glutamine residue (Gln<sup>447</sup>).

#### Mutagenesis confirms that Glu<sup>541</sup> in IRAP is important for positively charged substrate recognition

To test the prediction that Glu<sup>541</sup> in IRAP is important for the enzyme's preference for positively charged side chains, we used site-directed mutagenesis to change the Glu<sup>541</sup> in IRAP to an arginine, the equivalent residue in ERAP1. The IRAP E541R variant was expressed in recombinant form and purified to homogeneity (see Supplementary Figure S4 at <http://www.BiochemJ.org/bj/435/bj4350411add.htm>). We probed the substrate selectivity of the mutant IRAP using the L-substrate library and compared it with the wild-type protein (Figure 5). As predicted, the mutant IRAP had an altered selectivity profile and was much less potent in trimming positively charged residues (noted by arrows in Figure 5). In this context the selectivity profile of E541R IRAP was similar to the profile of ERAP1. Michaelis–Menten analysis using the fluorogenic substrate R-AMC indicated that the difference in specificity for IRAP E541R was primarily due to loss in affinity ( $K_m$ ) (Figure 6). We concluded that Glu<sup>541</sup> in IRAP is largely responsible for allowing IRAP to mimic the substrate preferences of ERAP2, without losing the preferences of ERAP1.

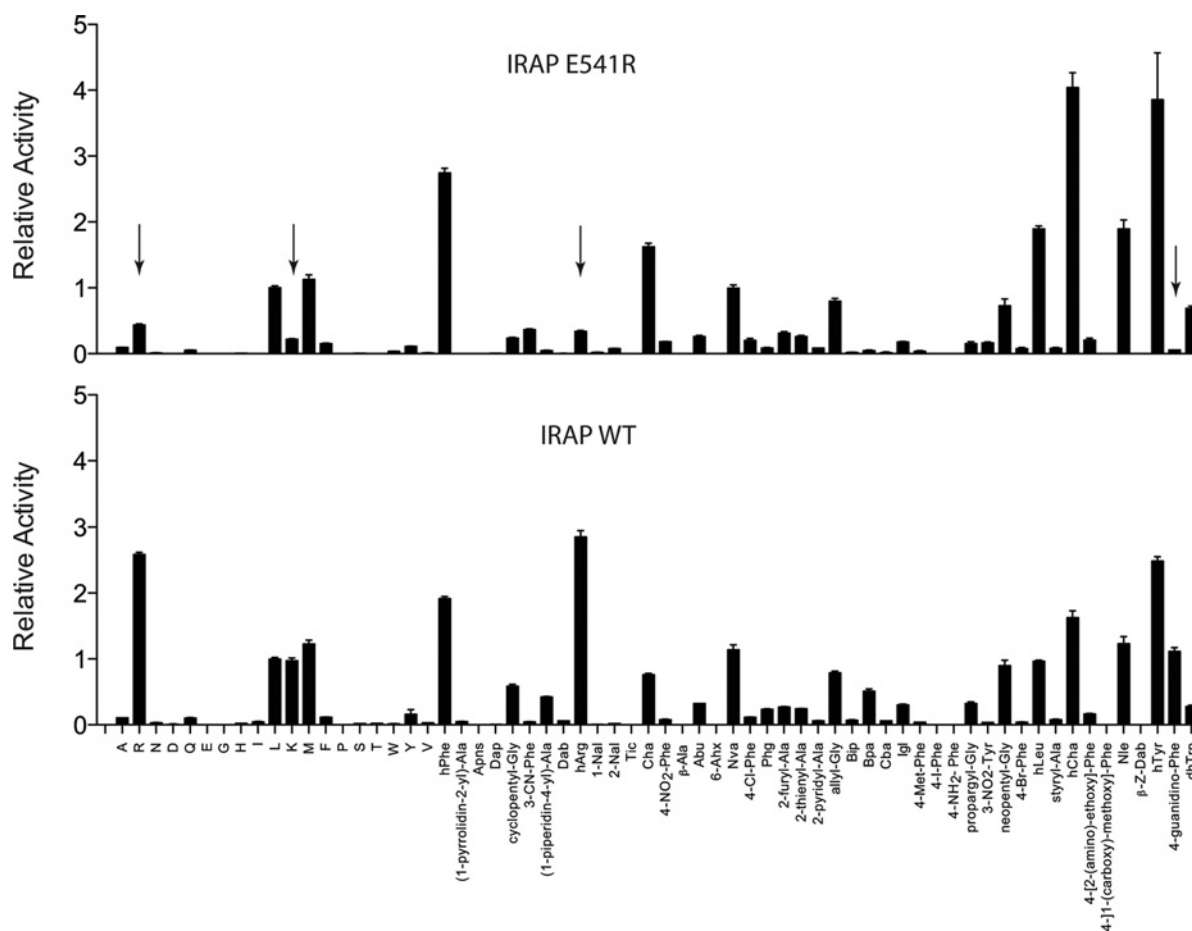
#### DISCUSSION

The importance of antigenic peptide precursor trimming by aminopeptidases has emerged the last few years as both a necessary step for antigenic peptide generation and also as a novel paradigm of regulation of the adaptive immune response [17,37]. However, the discovery that three distinct aminopeptidases participate in antigen processing has raised important questions regarding the regulation of antigenic peptide generation that are far from answered. We hypothesized that the necessity for multiple aminopeptidases performing what is seemingly an identical role

lies in key differences between the specificity of these enzymes. Towards testing this hypothesis, we systematically characterized the S1 specificity of ERAP1, ERAP2 and IRAP. We discovered that these three enzymes share many features between their S1-binding pockets, but at the same time have key differences that may help explain their distinct biological functions.

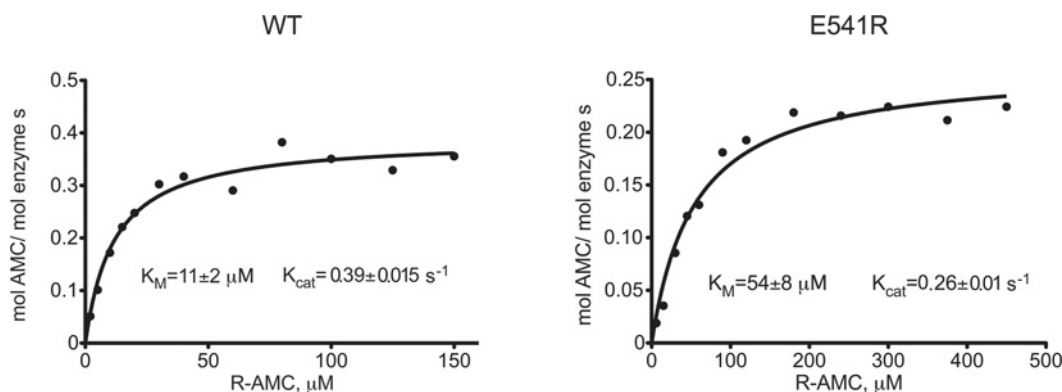
Our analysis suggests that to a large extent IRAP combines the N-terminal specificity of ERAP1 and ERAP2. This is consistent with the recently proposed function of IRAP in a distinct processing compartment inside the cell [11]. ERAP1 and ERAP2 have been proposed to function in tandem inside the ER, with ERAP2 behaving as an accessory protease, assisting ERAP1 in trimming sequences that would otherwise be poorly processed. IRAP, however, appears to act alone inside cross-presentation compartments and as a result it needs to be able to process both ERAP2 and ERAP1 substrates. ERAP2 gains the ability to process positively charged amino acids by a key change at position 181/198/293 [36]. However, this particular change reduces its affinity for hydrophobic chains, specializing it for positively charged amino acids. IRAP, however, cannot afford this option; it needs to be able to trim both ERAP1 and ERAP2 substrates. IRAP achieves this not by altering position 181/198/293 but by altering position 430/447/541 instead, allowing it to combine both specificities. This elegant solution to this apparent specificity problem is indicative on how key amino acid changes inside a specificity pocket can guide selectivity in this family of aminopeptidases. Generally it has proven difficult to alter the primary specificity of proteases by single amino acid replacements because the S1 pocket is influenced by a large number of interactions and may even be intrinsically disordered, as seen for example in the trypsin/chymotrypsin family [38]. However, certain scaffolds tolerate specificity switching by single residue substitutions, for example the chymase/granzyme family [39].

Regardless of the differences between their specificity, the three enzymes share some striking similarities for the side chains they fail to recognize. Neither enzyme can process negatively charged side chains presumably due to the strong negative electrostatic potential of the general region of the S1 pocket. Furthermore,



**Figure 5** Selectivity profile of IRAP E541R mutation compared with wild-type protein

Data have been normalized for leucine. Arrows highlight the most significant changes brought about by the mutation. WT, wild-type.



**Figure 6** Michaelis–Menten kinetics for hydrolysis of the substrate R-AMC by E541R IRAP as well as the wild-type enzyme

The enzymatic parameters  $K_m$  and  $k_{cat}$  are depicted for the wild-type (WT) and mutant (E541R) IRAP enzyme.

all three enzymes fail to process substrates with side chains that carry  $\beta$ -carbon or oxygen branching (such as valine, isoleucine or threonine) due to the limited space in the S1 pocket and the stringent stereochemical requirements for the recognition of the N-terminus of the substrate by the conserved GAMEN motif. Phe<sup>544</sup> plays a key role in this phenomenon and is critical for enzyme activity [40]. These findings, however, raise a crucial

question. If all the aminopeptidases that perform antigenic peptide processing before MHC class I loading cannot efficiently process common amino acids such as valine, isoleucine, threonine, glutamate or aspartate, how do such antigenic peptide precursors get processed? Inspection of the SYFPEITHI antigen database reveals that many antigenic peptides may be derived from precursors that would require the excision of such amino acids



(<http://www.syfpeithi.de/>). One notable example is the antigenic peptide from human ovalbumin, SIINFEKL, that can be processed by ERAP1 efficiently, although a common precursor sequence contains a glutamate residue (ESIINFEKL) [9]. A possible answer to this question lies on the specificity of ERAP1 for amino acids distal to the N-terminus of the substrate [19], a property that may be shared by ERAP2 and IRAP. An alternative explanation would include the participation of a currently unidentified accessory aminopeptidase.

The important role for these aminopeptidases in antigen presentation in combination with their distinct selectivity profiles suggests that selective inhibition of a single one may lead to subtle alterations of the antigenic peptide repertoire that can be used to modulate the immune response. Recently, polymorphic variation in ERAP1 and ERAP2 has been linked with predisposition to autoimmunity, viral infection and cancer, suggesting that manipulation of the activity of these aminopeptidases may have an important therapeutic potential [41]. Indeed, use of the non-selective general metalloproteinase inhibitor leucine-thiol in cultured cells has been demonstrated to alter antigen presentation [17]. Furthermore, we recently demonstrated that polymorphic variation in ERAP1 can affect antigen processing *in vitro* [42]. Therefore the development of highly potent and selective inhibitors for this class of enzymes may constitute a useful approach towards the modulation of the adaptive immune response. In addition, the development of highly selective substrates can be useful for investigating established pathogenic links and developing diagnostic and prognostic markers. Our results suggest that although these three enzymes are highly homologous, they still carry key differences in their S1 pockets that can be exploited for the design of selective inhibitors or specific activity markers that can be used to follow antigen processing *in vivo* or *ex vivo*.

In summary, we have performed a detailed analysis of the S1 specificity of ERAP1, ERAP2 and IRAP, three enzymes that process antigenic peptide precursors and are crucial to the functioning of the adaptive immune system. By combining small-substrate library screening, molecular modelling and mutagenesis we revealed key differences and similarities between the three enzymes that underlie their biological function. Furthermore, our analysis can facilitate efforts towards the rational design of small-molecular-mass selective inhibitors and activity markers that can be used to manipulate and characterize the adaptive immune response.

## AUTHOR CONTRIBUTION

Efthalia Zervoudi performed and analysed the enzymatic experiments. Athanasios Papakyriakou and Dionisios Vourloumis performed and helped interpret the molecular modelling. Dimitra Georgiadou, Irini Evnouchidou, Akira Hattori and Luc Swevers designed the ERAP2 and IRAP expression systems, purified the recombinant enzymes and helped interpret the specificity data. Anna Gajda, Marcin Poreba and Marcin Drag designed and synthesized the substrate library and helped interpret the specificity data. Marcin Drag, Athanasios Papakyriakou and Guy Salvesen helped design the study, interpret the data and write the paper. Efstratios Stratikos conceived and supervised the project, helped analyse the data and wrote the paper. All authors read and approved the final manuscript.

## ACKNOWLEDGEMENTS

We thank Dr Florentia Fostira for help with confirming mutagenesis reactions by DNA sequencing and Paulina Platek for help with synthesis of fluorogenic substrates.

## FUNDING

This work was funded in part by a NCSR Demokritos internal grant (to E.S.) and by the Foundation for Polish Science and the State for Scientific Research Grant [grant number N N401 042838 (to M.D.)]. I.E. and D.G. acknowledge financial support by the National Centre for Scientific Research "Demokritos" graduate fellowship programme.

## REFERENCES

- Rock, K. L. and Goldberg, A. L. (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **17**, 739–779
- Lauvau, G., Kakimi, K., Niedermann, G., Ostankovitch, M., Yotnda, P., Firat, H., Chisari, F. V. and van Ender, P. M. (1999) Human transporters associated with antigen processing (TAPs) select epitope precursor peptides for processing in the endoplasmic reticulum and presentation to T cells. *J. Exp. Med.* **190**, 1227–1240
- Rock, K. L. and Shen, L. (2005) Cross-presentation: underlying mechanisms and role in immune surveillance. *Immunol. Rev.* **207**, 166–183
- Cascio, P., Hilton, C., Kisselev, A. F., Rock, K. L. and Goldberg, A. L. (2001) 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J.* **20**, 2357–2366
- Roelse, J., Gromme, M., Momburg, F., Hammerling, G. and Neefjes, J. (1994) Trimming of TAP-translocated peptides in the endoplasmic reticulum and in the cytosol during recycling. *J. Exp. Med.* **180**, 1591–1597
- Serwold, T., Gaw, S. and Shastri, N. (2001) ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat. Immunol.* **2**, 644–651
- Serwold, T., Gonzalez, F., Kim, J., Jacob, R. and Shastri, N. (2002) ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419**, 480–483
- Saric, T., Chang, S. C., Hattori, A., York, I. A., Markant, S., Rock, K. L., Tsujimoto, M. and Goldberg, A. L. (2002) An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat. Immunol.* **3**, 1169–1176
- York, I. A., Chang, S. C., Saric, T., Keys, J. A., Favreau, J. M., Goldberg, A. L. and Rock, K. L. (2002) The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8–9 residues. *Nat. Immunol.* **3**, 1177–1184
- Saveanu, L., Carroll, O., Lindo, V., Del Val, M., Lopez, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G. and van Ender, P. M. (2005) Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* **6**, 689–697
- Saveanu, L., Carroll, O., Weimershaus, M., Guernonprez, P., Firat, E., Lindo, V., Greer, F., Davoust, J., Kratzer, R., Keller, S. R. et al. (2009) IRAP identifies an endosomal compartment required for MHC class I cross-presentation. *Science* **325**, 213–217
- Tsujimoto, M. and Hattori, A. (2005) The oxtocinase subfamily of M1 aminopeptidases. *Biochim. Biophys. Acta* **1751**, 9–18
- York, I. A., Brehm, M. A., Zendzian, S., Towne, C. F. and Rock, K. L. (2006) Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides *in vivo* and plays an important role in immunodominance. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9202–9207
- Hammer, G. E., Gonzalez, F., Champsaur, M., Cado, D. and Shastri, N. (2006) The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules. *Nat. Immunol.* **7**, 103–112
- Yan, J., Parekh, V. V., Mendez-Fernandez, Y., Olivares-Villagomez, D., Dragovic, S., Hill, T., Roopenian, D. C., Joyce, S. and Van Kaer, L. (2006) *In vivo* role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules. *J. Exp. Med.* **203**, 647–659
- Firat, E., Saveanu, L., Aichele, P., Staeheli, P., Huai, J., Gaedicke, S., Nil, A., Besin, G., Kanzler, B., van Ender, P. and Niedermann, G. (2007) The role of endoplasmic reticulum-associated aminopeptidase 1 in immunity to infection and in cross-presentation. *J. Immunol.* **178**, 2241–2248
- Hammer, G. E., Gonzalez, F., James, E., Nolla, H. and Shastri, N. (2007) In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. *Nat. Immunol.* **8**, 101–108
- Chang, S. C., Momburg, F., Bhutani, N. and Goldberg, A. L. (2005) The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17107–17112
- Evnouchidou, I., Momburg, F., Papakyriakou, A., Chroni, A., Leondiadis, L., Chang, S. C., Goldberg, A. L. and Stratikos, E. (2008) The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS ONE* **3**, e3658
- Georgiadou, D., Hearn, A., Evnouchidou, I., Chroni, A., Leondiadis, L., York, I. A., Rock, K. L. and Stratikos, E. (2010) Placental leucine aminopeptidase efficiently generates mature antigenic peptides *in vitro* but in patterns distinct from endoplasmic reticulum aminopeptidase 1. *J. Immunol.* **185**, 1584–1592
- Hearn, A., York, I. A. and Rock, K. L. (2009) The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J. Immunol.* **183**, 5526–5536
- Hattori, A., Kitatani, K., Matsumoto, H., Miyazawa, S., Rogi, T., Tsuruoka, N., Mizutani, S., Natori, Y. and Tsujimoto, M. (2000) Characterization of recombinant human adipocyte-derived leucine aminopeptidase expressed in Chinese hamster ovary cells. *J. Biochem.* **128**, 755–762

- 23 Tanioka, T., Hattori, A., Masuda, S., Nomura, Y., Nakayama, H., Mizutani, S. and Tsujimoto, M. (2003) Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. *J. Biol. Chem.* **278**, 32275–32283
- 24 Matsumoto, H., Rogi, T., Yamashiro, K., Kodama, S., Tsuruoka, N., Hattori, A., Takio, K., Mizutani, S. and Tsujimoto, M. (2000) Characterization of a recombinant soluble form of human placental leucine aminopeptidase/oxytocinase expressed in Chinese hamster ovary cells. *Eur. J. Biochem.* **267**, 46–52
- 25 Drag, M., Bogyo, M., Ellman, J. A. and Salvesen, G.S. (2010) Aminopeptidase fingerprints, an integrated approach for identification of good substrates and optimal inhibitors. *J. Biol. Chem.* **285**, 3310–3318
- 26 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815
- 27 Case, D. A., Cheatham, 3rd, T. E., Darden, T., Gohlke, H., Luo, R., Merz, Jr, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688
- 28 Papakyriakou, A., Spyroulias, G. A., Sturrock, E. D., Manessi-Zoupa, E. and Cordopatis, P. (2007) Simulated interactions between angiotensin-converting enzyme and substrate gonadotropin-releasing hormone: novel insights into domain selectivity. *Biochemistry* **46**, 8753–8765
- 29 Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A. and Stahl, M. T. (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **50**, 572–584
- 30 Jakalian, A., Jack, D. B. and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641
- 31 Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791
- 32 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041
- 33 Reference deleted
- 34 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38
- 35 Tsujimoto, M., Goto, Y., Maruyama, M. and Hattori, A. (2008) Biochemical and enzymatic properties of the M1 family of aminopeptidases involved in the regulation of blood pressure. *Heart Fail. Rev.* **13**, 285–291
- 36 Goto, Y., Tanji, H., Hattori, A. and Tsujimoto, M. (2008) Glutamine-181 is crucial in the enzymatic activity and substrate specificity of human endoplasmic-reticulum aminopeptidase-1. *Biochem. J.* **416**, 109–116
- 37 Georgiadou, D. and Stratikos, E. (2009) Cellular mechanisms that edit the immunopeptidome. *Curr. Proteomics* **6**, 13–24
- 38 Hung, S. H. and Hedstrom, L. (1998) Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity. *Protein Eng.* **11**, 669–673
- 39 Caputo, A., James, M. N., Powers, J. C., Hudig, D. and Bleackley, R. C. (1994) Conversion of the substrate specificity of mouse proteinase granzyme B. *Nat. Struct. Biol.* **1**, 364–367
- 40 Albiston, A. L., Pham, V., Ye, S., Ng, L., Lew, R. A., Thompson, P. E., Holien, J. K., Morton, C. J., Parker, M. W. and Chai, S. Y. (2010) Phenylalanine-544 plays a key role in substrate and inhibitor binding by providing a hydrophobic packing point at the active site of insulin-regulated aminopeptidase. *Mol. Pharmacol.* **78**, 600–607
- 41 Haroon, N. and Inman, R. D. (2010) Endoplasmic reticulum aminopeptidases: biology and pathogenic potential. *Nat. Rev. Rheumatol.* **6**, 461–467
- 42 Evnouchidou, I., Kamal, R. P., Seregin, S. S., Goto, Y., Tsujimoto, M., Hattori, A., Voulgari, P. V., Drosos, A. A., Amalfitano, A., York, I. A. and Stratikos, E. (2011) Cutting edge: coding single nucleotide polymorphisms of endoplasmic reticulum aminopeptidase 1 can affect antigenic peptide generation *in vitro* by influencing basic enzymatic properties of the enzyme. *J. Immunol.* **186**, 1909–1913

Received 13 December 2010/7 February 2011; accepted 11 February 2011  
Published as BJ Immediate Publication 11 February 2011, doi:10.1042/BJ20102049

## SUPPLEMENTARY ONLINE DATA

# Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides

Efthalia ZERVOUDI\*<sup>1</sup>, Athanasios PAPAKYRIAKOU†<sup>1</sup>, Dimitra GEORGIADOU\*, Irini EVNOUCHIDOU\*, Anna GAJDA‡, Marcin POREBA‡, Guy S. SALVESEN§, Marcin DRAG‡§, Akira HATTORI||, Luc SWEVERS¶, Dionisios VOURLOUMIS† and Efstratios STRATIKOS\*<sup>2</sup>

\*Protein Chemistry Laboratory, IRRP, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece, †Chemical Biology Laboratory, Institute of Physical Chemistry, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece, ‡Division of Medicinal Chemistry and Microbiology, Wroclaw University of Technology, 50-370 Wroclaw, Poland, §Program in Apoptosis and Cell Death Research, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, U.S.A., ||Graduate School of Pharmaceutical Sciences, Kyoto University, Department of System Chemotherapy and Molecular Sciences, Sakyo, Kyoto 606-8501, Japan, and ¶Insect Molecular Genetics and Biotechnology Group, Institute of Biology, National Centre for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece

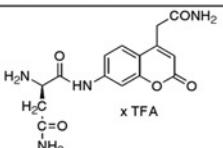
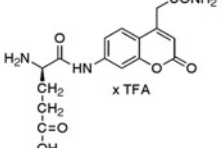
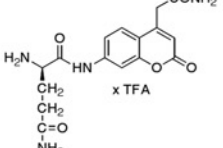
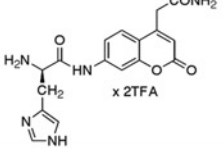
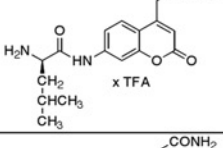
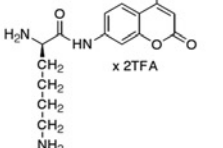
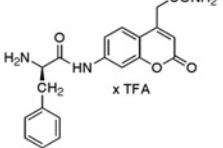
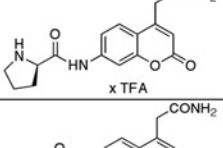
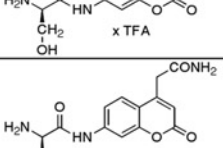
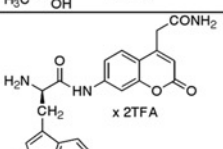
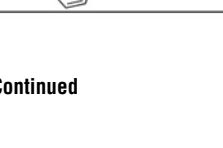
See the pages that follow for Supplementary Figures S1–S4

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed (email stratos@rrp.demokritos.gr or stratikos@gmail.com).

Entry	Structure	Name	M.W.
1		Ala	289.1
2		Arg	376.2
3		Asn	332.1
4		Asp	333.1
5		Gln	346.2
6		Glu	347.1
7		Gly	275.1
8		His	355.2
9		Ile	331.2
10		Leu	331.2
11		Lys	346.2
12		Met	349.1
13		Phe	365.2
14		Pro	315.1
15		Ser	305.1
16		Thr	319.1
17		Trp	404.2
18		Tyr	381.2
19		Val	317.2
20		D-Ala	289.1
21		D-Arg	376.2
22		D-Asp	333.1

Figure S1 Chemical structures of substrates used in the screening

Entry	Structure	Name	M.W.
23		D-Asn	332.1
24		D-Glu	347.1
25		D-Gln	346.2
26		D-His	355.2
27		D-Leu	331.2
28		D-Lys	346.2
29		D-Phe	365.2
30		D-Pro	315.1
31		D-Ser	305.1
32		D-Thr	319.1
33		D-Trp	404.2

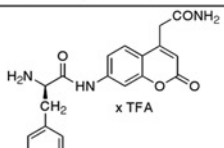
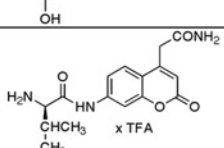
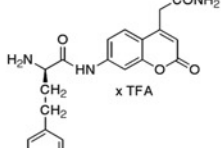
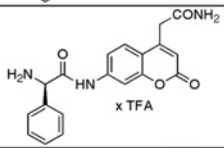
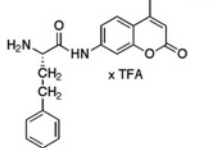
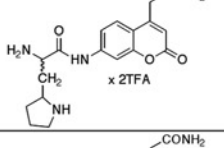
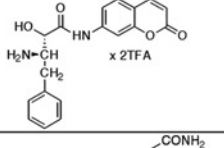
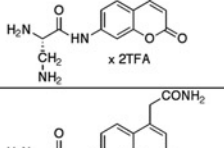
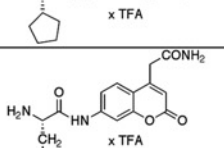
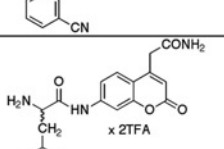
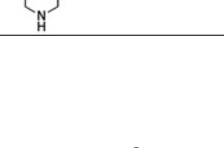
Entry	Structure	Name	M.W.
34		D-Tyr	381.2
35		D-Val	317.2
36		D-hPhe	379.2
37		D-Phg	351.2
38		hPhe	379.2
39		(1-pyrrolidin-2-yl)-Ala	358.2
40		Apns (2S,3S)	395.2
41		Dap	304.1
42		cyclopentyl-Gly	343.2
43		3-CN-Phe	390.2
44		(1-piperidin-4-yl)-Ala	372.2

Figure S1 Continued

Entry	Structure	Name	M.W.	Entry	Structure	Name	M.W.
45		Dab	318.2	55		Nva	317.2
46		hArg	388.2	56		4-Cl-Phe	399.1
47		1-Nal	415.2	57		Phg	351.1
48		2-Nal	415.2	58		2-furyl-Ala	355.1
49		Tic	377.2	59		2-thienyl-Ala	371.1
50		Cha	371.2	60		2-pyridyl-Ala	366.1
51		4-NO <sub>2</sub> -Phe	410.1	61		allyl-Gly	315.1
52		β-Ala	289.1	62		Bip	441.2
53		Abu	303.1	63		Bpa	469.2
54		6-Ahx	331.2				

Figure S1 Continued

Entry	Structure	Name	M.W.	Entry	Structure	Name	M.W.
64		Cba	328.1	73		hLeu	345.2
65		Igl	391.2	74		styryl-Ala	391.2
66		4-Met-Phe	379.2	75		hCha	385.2
67		4-I-Phe	491.0	76		4-[2-(amino)ethoxy]-Phe	424.2
68		4-NH2-Phe	380.1	77		4-[1-(carboxy)methoxy]-Phe	439.2
69		propargyl-Gly	313.1	78		Nle	331.2
70		3-NO2-Tyr	426.1	79		$\beta$ -Z-Dab	452.2
71		neopentyl-Gly	345.2	80		hTyr	381.13
72		4-Br-Phe	443.1				

Figure S1 Continued

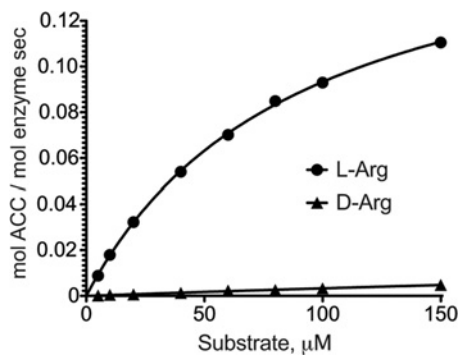




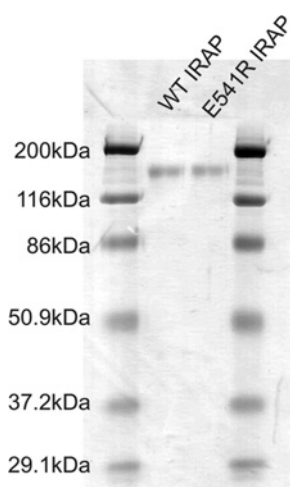
IRAP	MEPFTNDRLQLPRNMIENSMFEEEPDVVDLAKEPCLHPLEPDEVEYEPGRSRLLVRLGLGE	60
ERAP1		MVFLPKWSLAT 12
ERAP2		MFHSSAMVNSHRKPMFNIHRGFYC 24
IRAP	HEMEDEEDYESSAKLLGMSFMNRSSGLRNSATGYRQSPDGACSVPSARTMVVCAFVIV	120
ERAP1	MSFLLSLLALLTVSTPS--WCQSTEA---SPKRS DGT PFPWNKIRLPEYVIPVHYDLL	66
ERAP2	LTAILPQICICSFVSPSSYHFTEDPGA---FPVATNGERFPWQELRLPSVVIPLHYDLF	81
IRAP	AVSVIMVIYLLPRCTFTKEGCHKKNQSIGLIQPFATNGKLFPPWAQIRLPTAVVPLRYELS	180
		*** :*** :*** :***
ERAP1	IHANLTLTFWGTTKVEITASQPTSTIILHSHHLQISRATLRKGA GERLSE--EPLQVLE	124
ERAP2	VHPNLTSLDFVASEKIEVLVSNATQFIILHSDKLEITNATLQSEEDSRMYKPGKELKVL	141
IRAP	LHPNLTSMTFRGVSITVSQLQVTWNIILHSTGHNISRVTFMSA---VSSQEKQAEILE	236
		:. :***: * : :. :. :. : * ***** :*:. : * : : : *
ERAP1	HPRQEIALLAPEPLLVLGYPYVVIHYAGNLSEFHGFYKSTYRTKEGELRI LASTQFEP	184
ERAP2	YPAHEQIALLVPEKLTPLHKYVAMDFQAKLGDGFE GFYKSTYRTLGGETRILAVTDFEP	201
IRAP	YAYHGQIAIVAPEALLAGHNYTLKIEYSANISSSYGFYGFSTYDESNEKKYFAATQFEP	296
		:. : ***: *
ERAP1	TAARMAFCFDEPAFKASFISIKIRREPRHLAISNMPLVKSVTVAEGLIEDHFDVTVKMST	244
ERAP2	TQARMAFCFDEPLFKANFSIKIRRESRHLAISNMPLVKSVTVAEGLIEDHFDVTVKMST	261
IRAP	LAARSAFCFDEPAFKATFIKIIRDEQYALSMPKSSVLDLDDLVQDEFSESVMKST	356
		** ***** *
ERAP1	YLVAFIISDFESVKITKSGVKVSVYAVPKINQADYALDAAVTLLEFYEDYFSIPYPL	304
ERAP2	YLVAYIVCDFHLSGFTSSGVKVIYASPKRNQTHYALQASLKLDFYEKYFDIYYPL	321
IRAP	YLVAFIVGEMKNSQDVN-GTLVSIYAVPEKIGQVHYALETTVKLEFFQNYFEIQYPLK	415
		****: * : :. : *
ERAP1	KQDLAAIPDFQSGAMENWGLTTRYRESALLFDAEKSSASSKLGITMTVAHELAHQWFGNLV	364
ERAP2	KLDLIAIPDFAPGAMENWGLITYRETSLLFPDKTSSASDKLWTRVIAHELAHQWFGNLV	381
IRAP	KLDLVAIPDFEAGAMENWGLTFREETLLYDSNTSSMADRKLVTIKIAHELAHQWFGNLV	475
		** *
ERAP1	TMEWNDLWLNDFGFAKMEFVSVVTHPELVKVDYFFGKCFDAMEVDALNSSHPVSTPVE	424
ERAP2	TMEWNDIWLKEGFAKYMELIAVNATYPELQDDYFLNVCFEVITKDSLNSRPIKPAE	441
IRAP	TMKWNDLWLNDFGFAKMEFVSVVTHPELVKVDYFFGKCFDAMEVDALNSSHPVSTPVE	535
		** *
ERAP1	NPAQIREMFDVSYDKGACILNMLREYLSADAFKSGIVQYLKQHSYKNTKNEDLWDSMAS	484
ERAP2	TPTQIQEMFDEVSYNKGACILNMLKDFLGEKFKQGIQYLLKFSYRNASKNDLWSSLSN	501
IRAP	SSEQIEMFDSL SYFGSSLLMLKTYLSEDFVQHAVVLYLHNHSYASIQSDDLWDSFNE	595
		.. *
ERAP1	ICP----TDGVKMDGFCRSRQSSSSSHHQEGVDVKTMMNTWTLQKGFPLITITVRG	539
ERAP2	SCLESDFTSGGVCHSDPKMNSMFLAFLG----ENA EVKEMTTWTLQKGIPLLVKQDG	556
IRAP	VTN-----QTL DVKRMKMTWTLQKGFPLVTVQKKG	625
		: *
ERAP1	RNVHMKQEHYMKSGD-----APDTGYLWHVPLTFITSK---SDMVHFRLLKTKTDVLI	590
ERAP2	CSLRQQERFLQGVFQEDPEWRALQERYLWHIPLTYSTSS---SNVIHRHLKSKTDLTD	613
IRAP	KELFIQERFFLN---MKPEIQPSDTSYLWHIPLSYVTEGRNYSKYQSVSLDDKSGVIN	682
		: : *
ERAP1	LPEEVEWIKFNVMNGYYIVHYEDDGDWDLTGLLKGHTAVSSNDRASLINNAFQLVLSIG	650
ERAP2	LPEKTSWVKFNVDNSNGYYIVHYEGHGDQLITLQNLNHTLLRPKDRVGLIHDVFQLVGAG	673
IRAP	LTEEVLWVKVNIINMNGYYIVHYADDWEALIHQLKINPYVLSDKDRANLNNIFELAGLG	742
		* *
ERAP1	KLSIEKALDLSLYLKHETEIMPVFQGLNELIPMYKLMKDRMNEVETQKFAFLIRLLRDL	710
ERAP2	RLTLDKALDMTYLQHETSSPALLEGLSYLESFYHMDRRNISDIENLKRYLLQYFKVP	733
IRAP	KVPLKRAFDLINYLGNENHTAPITEALFQTDLIYNLLEKLGMDLASRLVTRVFKLLQNG	802
		: : *
ERAP1	IDKQWTWDEGSVSEMRLSQLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLA	770
ERAP2	IDRQSWSDKGSVDRMLRSALLKACDLNHAPCIQKAAELFSQWMESSGKLNIPDVLKI	793
IRAP	IQQQWTWDEGTPSMRELSALLEFACTHNLGNCSTTAMKLFDDWMSNGTQSLPTDVMTT	862
		* *
ERAP1	VFAVGAQSTEGWDFLYSKYQFSLSTEKSQIEFALCRTQNKELQWLLDESFGDKIKITQ	830
ERAP2	VYSVGAQTAGWNYLLEQYELSMSSAEQNKILYALSTSKHQEKLKLIELGMEGVIKIQ	853
IRAP	VFKVGAKTDKGWSFLGKYISIGSEAEKNKILEALASSEDVRKLYWLMKSSLNGDNFRQT	922
		* : *
ERAP1	EFPQILTIGRNPVGYPLAWQFLRKNWNKLVQKFE LGS SSI AHMVMGTTNQFSTRTRLEE	890
ERAP2	NLAALLHAIARRPKGQLAWDFVRENWTHLLKFDLGSYDIRMIISGTTAHFSSKDKLQE	913
IRAP	KLSFIIRTVGRHFGHLLAWDFVKNWNLVQKFE LGS YTIQINIVAGSTYLFSTKTHLSE	982
		: : : *
ERAP1	VKGFSSLKENGSLRCVQQT IETIEENIGWMDKNFDKIRVWLQSEKLERM	941
ERAP2	VKLFESLEAQSHLDFQTVLETITKNIKWLEKNLPTLRWLMVNT	960
IRAP	VQAFENQSEATFRLRCVQEALEVIQLNIQWMEKNLKSLTWML	1025
		* : *

**Figure S2** Sequence alignment of ERAP1, ERAP2 and IRAP using the ClustalW2 algorithm

The Zn(II)-binding motif HELAH as well as the N-terminus recognition motif GAMEN are highlighted in green. The residues that are postulated to participate in the S1 pocket for each enzyme are highlighted in yellow. Greyed residues were not included in the homology modelling since they are not present in the crystallographic structure of ERAP1 (PDB code 2XDT).



**Figure S3** Michaelis–Menten analysis of trimming of L-Arg-ACC and D-Arg-ACC by ERAP2



**Figure S4** SDS/PAGE of wild-type and E541R recombinant IRAP produced by 293F cells

The molecular mass in kDa is indicated on the left-hand side. WT, wild-type.