




# Conformational dynamics linked to domain closure and substrate binding explain the ERAP1 allosteric regulation mechanism

Zachary Maben<sup>1,3</sup>, Richa Arya<sup>1,3</sup>, Dimitris Georgiadis<sup>2</sup>, Efstratios Stratikos<sup>2</sup> <sup>2</sup> & Lawrence J. Stern<sup>1</sup> <sup>1</sup> 

The endoplasmic-reticulum aminopeptidase ERAP1 processes antigenic peptides for loading on MHC-I proteins and recognition by CD8 T cells as they survey the body for infection and malignancy. Crystal structures have revealed ERAP1 in either open or closed conformations, but whether these occur in solution and are involved in catalysis is not clear. Here, we assess ERAP1 conformational states in solution in the presence of substrates, allosteric activators, and inhibitors by small-angle X-ray scattering. We also characterize changes in protein conformation by X-ray crystallography, and we localize alternate C-terminal binding sites by chemical crosslinking. Structural and enzymatic data suggest that the structural reconfigurations of ERAP1 active site are physically linked to domain closure and are promoted by binding of long peptide substrates. These results clarify steps required for ERAP1 catalysis, demonstrate the importance of conformational dynamics within the catalytic cycle, and provide a mechanism for the observed allosteric regulation and Lys/Arg528 polymorphism disease association.

<sup>1</sup>Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA. <sup>2</sup>Department of Chemistry, National and Kapodistrian University of Athens, Athens, Greece. <sup>3</sup>These authors contributed equally: Zachary Maben, Richa Arya. ✉email: [Lawrence.Stern@umassmed.edu](mailto:Lawrence.Stern@umassmed.edu)

Endoplasmic-reticulum aminopeptidase 1 (ERAP1) is a key determinant of antigen processing and presentation. In this role, ERAP1 activity digests peptides in class I major histocompatibility complex (MHC-I) antigen-processing pathway, in a manner that can create or destroy peptide epitopes for presentation to CD8 T cells<sup>1</sup>. Polymorphisms within ERAP1 correlate with predisposition to autoimmune diseases, pre-eclampsia, and cancer<sup>2,3</sup>. The ERAP1 single-nucleotide polymorphism (SNP) rs30187, which encodes lysine or arginine at position 528, has been associated with autoimmune diseases in epistasis with specific MHC alleles<sup>4,5</sup> and alters the set of peptides presented by MHC<sup>6,7</sup>. In biochemical assays, lysine/arginine 528 substitution affects ERAP1's peptide hydrolysis activity<sup>8</sup>. How this polymorphism 25 Å away from the active site zinc affects enzymatic activity and specificity is unknown.

ERAP1 is a member of the M1 aminopeptidase family, which is found in all domains of life and plays numerous functional roles<sup>9</sup>. These proteins contain a thermolysin-like domain<sup>10</sup> (domain II), which holds the active site. The M1 aminopeptidase family has a conserved HEAT-motif-repeat domain (domain IV) forming a bowl-shaped C-terminal structure<sup>9</sup>. In most crystal structures of this family, domain IV is closely associated with domain II and forms an internal pocket that separates the active site from the bulk solvent. This interdomain interaction sterically blocks substrate binding and product release, and it was hypothesized that aminopeptidases in this family must change structure during the catalytic cycle in order to accommodate these steps<sup>9</sup>.

The first crystal structures of ERAP1 provided some structural underpinning to this picture. One crystal structure matched the closed conformation that had been observed for other family members<sup>11</sup>. Two other crystal structures captured an open conformation, in which domain IV had undergone a rigid-body translocation away from domain II, exposing the internal cavity to bulk solvent<sup>11,12</sup>. Along with the domain IV reorientation, the active site had reorganized with tyrosine 438 rotating away in a position not optimal for catalysis. This suggested a connection between the conformational state and catalytic activity. A similar motion of the homologous catalytic tyrosine 549 was later observed in crystal structures of insulin-regulated aminopeptidase (IRAP), in this case linked to the presence of a transition-state analog in the catalytic center<sup>13–15</sup>.

ERAP1 has an unusual substrate-length dependence, efficiently trimming peptides 10–15 residues long but sparing shorter peptides<sup>16,17</sup>, well suited to its role in processing peptides for loading onto MHC-I proteins, which typically bind peptides of ~9 residues. A connection between the active site and a hypothesized distal regulatory site within ERAP1 has been suggested previously to explain ERAP1 substrate-length dependence and allosteric activation<sup>12</sup>. ERAP1 hydrolysis of activated amino-acid substrates such as leucine-*p*-nitroanilide (Leu-*p*NA) shows deviations from Michaelis–Menten behavior suggestive of cooperativity within the catalytically active monomer<sup>12</sup>. In addition, catalysis of short substrates is increased by the presence of short peptides 4–14 residues in length<sup>12,18</sup>. These data were interpreted in terms of a mechanism whereby the allosteric modulators bind within the overall substrate-binding cavity but distal to the active site, inducing conversion to a more active form<sup>12</sup>. These findings, along with data showing preferential hydrolysis of substrates longer than eight amino acids<sup>16</sup>, support a model that an optimal length substrate peptide simultaneously binds the active site and a distal regulatory site that activates catalysis<sup>12,17</sup>. Binding sites within the overall substrate-binding cavity have been identified for two small-molecule ERAP1 allosteric modulators<sup>19,20</sup>, one of which overlaps with the binding site identified for the C-terminus of peptides with a hydrophobic terminal residue<sup>21,22</sup>, but how occupancy of these sites induces

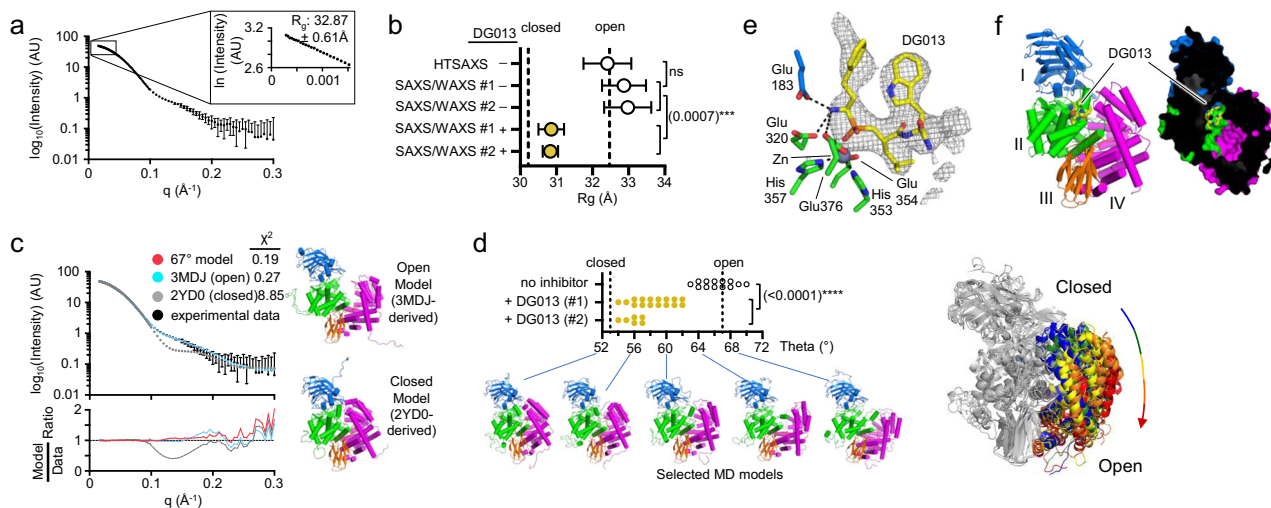
activation, and whether activation is related to domain reorganization, remains unknown.

Here, we examine conformational states of ERAP1 using small-angle X-ray scattering (SAXS) and observe that ERAP1 adopts a largely open conformation in solution, but closes in the presence of substrates, inhibitors, or an allosteric activator that binds far from the active site at an interdomain junction. The disease-related polymorphic position 528 (lysine/arginine) alters the stability of the closed conformation and modulates an interdomain electrostatic interaction with glutamate 913, providing a mechanistic explanation for altered catalytic efficiency. The crystal structure of a substrate-mimic inhibitor bound to ERAP1 in the open conformation shows how helix 4a folds in the presence of substrate, coupling active-site alterations to large-scale domain motions. The binding site for peptide substrate C-terminal hydrophobic amino-acid residues identified previously in crystal structures<sup>21,22</sup>, along with alternate sites identified here by chemical cross-linking, are used to model the effect of peptide binding on domain closure energetics. These results suggest a model for regulation of ERAP1 activity that connects substrate binding and active-site reorganization through a large-scale domain closure motion promoted by binding long peptide substrates.

## Results

**ERAP1 adopts an open conformation in solution and closes with active site occupancy.** ERAP1 has been observed to adopt open and closed conformations in X-ray crystal structures<sup>11,12</sup>, but whether these states interconvert in solution, and the relevance of the conformational change to catalysis, has not been determined. In order to investigate conformational transitions of ERAP1 in solution, we used small-angle X-ray scattering (SAXS). In initial experiments, we used ERAP1 allele IV isoform 2 (Supplementary Fig. S1a), as previously used for structural characterization<sup>12</sup>, in the presence or absence of the broad-range aminopeptidase inhibitors bestatin<sup>23</sup> or leucine-thiol (LeuSH)<sup>24</sup>, which target the substrate S1 and S1' subsites and active site zinc, respectively<sup>11,12,23–25</sup>. In the presence of either of these inhibitors, ERAP1 becomes more compact, with the radius of gyration,  $R_g$ , decreasing by ~2 Å (Supplementary Fig. S2a). To investigate the effect of a more physiological ligand on ERAP1 conformation, we used the octamer peptide SIINFEKL. This peptide is generated efficiently from longer substrates by ERAP1 in vitro<sup>17</sup> and in vivo<sup>26</sup>, and itself is a suboptimal ERAP1 substrate<sup>16</sup>. In the presence of a saturating amount of SIINFEKL, ERAP1 adopted a closed conformation (Supplementary Fig. S2a).

To facilitate detailed SAXS analysis, we used a different ERAP1 construct (allele II isoform 1, Supplementary Fig. S1), which lacks a nine-residue unstructured C-terminal tail and recombinant *myc* tag not visualized in the X-ray structure. Removing these unstructured regions simplified fitting the experimental scattering data to molecular models. We collected solution X-ray scattering data including wide-angle WAXS data to allow for direct measurement of the water scattering profile (maximum intensity at  $q = 2 \text{ \AA}^{-1}$ ), producing high-quality background-subtracted scattering curves<sup>27</sup>.  $R_g$  values as determined by Guiner fitting matched the  $R_g$  value calculated for the open conformer observed by X-ray crystallography, as previously reported<sup>28</sup> (Fig. 1a). We observed no differences in  $R_g$  for samples of this isoform collected by either conventional SAXS or SAXS/WAXS (Fig. 1b, open circles). The SAXS/WAXS buffer subtraction method improves data quality at higher  $q$  values, allowing higher-resolution analysis in which the full range of scattering data can be fit to experimental or computational atomic models. The full-profile dataset fit better to a model based on an open-crystal structure



**Fig. 1** ERAP1 adopts an open conformation in solution and closes when bound to the substrate and substrate-mimic inhibitors. **a** Representative ERAP1 SAXS curve collected using SAXS/WAXS data collection. Data are for ERAP1 reference sequence (isoform I) in the absence of inhibitor or substrate ( $n = 1$  sample was assessed in an independent experiment). Error bars represent uncertainty in log intensity values from radial averaging of scattering profiles. Inset, Guinier plot showing  $R_g$  determination using a linear fit of low-resolution scattering data ( $n = 1$ ,  $q < 0.04 \text{ \AA}^{-1}$ ); fit value for slope and associated uncertainty are shown. **b** DG013 induces closed ERAP1 conformation as measured by  $R_g$ . Each data point represents an independent experiment ( $n = 1$ ) processed individually, with the error of linear fit for each dataset shown as error bars. Model for open and closed conformations are based on previously determined ERAP1 crystal structures (2YD0 and 3MDJ chain B) shown for reference are right, with calculated  $R_g$  values indicated with dotted vertical lines. An unpaired two-tailed  $t$  test was used to assess the significance of the difference between SAXS/WAXS  $R_g$  values for ERAP1 in the presence or absence of DG013 from two independent experiments. **c** ERAP1 experimental SAXS curve ( $n = 1$ ) from panel **a** overlaid with calculated SAXS curves generated from three structural models. A largely open model derived from molecular dynamics simulation with  $q = 67^\circ$  fits best. Fit residuals for these three models are shown below. **d** DG013 induces closed ERAP1 conformation as measured by SAXS curve fitting to structural models from molecular dynamics. Models that fit experimental SAXS data with  $\chi^2 < 1$  are shown as points on the graph from an independent experiment, forming a distribution of models that each fit scattering data within experimental uncertainty. An unpaired two-tailed  $t$  test was used to assess significance. Selected ERAP1 models from MD simulation that demonstrate the relationship between angle theta and structural conformation are shown below and on the right is an overlay of selected MD models showing the hinge-like motion of domain IV. **e** DG013 binds to ERAP1 active site. Polder map (gray mesh) contoured at  $3.5\sigma$  shows DG013 electron density at ERAP1 active site. **f** DG013-bound ERAP1 crystal structure is in the closed conformation. Shown are cartoon and surface representations colored by domain as indicated. Surface cutaway reveals DG013 bound at the active site. Source data are provided as a Source Data file.

(Fig. 1c, cyan symbols) than to one based on the closed crystal structure (Fig. 1c, gray symbols). For the open-crystal structure model, residuals for the fit of experimental to calculated scattering intensities were small and evenly distributed throughout the resolution range (Fig. 1c, bottom pane), and were smaller than expected experimental uncertainties based on radial averaging of observed scattering patterns (Fig. 1c, black bars). To explore other conformers, we used a set of 45 structural models selected from a molecular dynamics simulation<sup>28</sup> that sample a range of opening angles with opening angle theta<sup>28</sup> ranging from 54 to 74° (Fig. 1d). Of these, the best fitting model had theta = 67° (Fig. 1c, red symbols), similar to that observed for the open-crystal structure, but with a lower  $\chi^2$  value. To account for uncertainties in the experimental data, we considered all models with  $\chi^2 < 1$ . This provided a cluster of models fitting the experimental ERAP1 data, with theta angles bracketing the value calculated for the open-crystal structure (Fig. 1d, open symbols).

To examine the effect of active-site occupancy on ERAP1 conformation in solution, we used a tight-binding tripeptide analog DG013 ( $IC_{50} = 33 \text{ nM}$ <sup>29</sup>), which has a phosphinic group expected to mimic the tetrahedral intermediate that would form at the scissile peptide bond during catalysis (Supplementary Fig. S1c). We determined the X-ray crystal structure of ERAP1 bound to DG013 (Table 1, Supplementary Table 1, Supplementary Figs. 3 and 4, see “Methods” for details) and confirmed that DG013 binds as expected to the ERAP1 active site (Fig. 1e, f and Supplementary Table S1a), with the enzyme adopting the closed conformation in the crystal (Fig. 1f and Supplementary Fig. S3a–f). The binding pose is similar to that of another

phosphinic peptide, DG046, bound to ERAP1<sup>30</sup>, but distinct from DG013 bound to ERAP2, most likely due to lack of an aromatic P2'-binding site.

In the presence of DG013, ERAP1  $R_g$  decreases by  $\sim 2 \text{ \AA}$ , an effect observed in multiple measurements and for two independent protein preparations (Fig. 1b, yellow circles). Comparison of this experimental finding with  $R_g$  calculated for structural models derived from open and closed crystal structures (PDB ID 3MDJ and 2YD0, respectively, dotted vertical lines in Fig. 1b) shows that the  $\sim 2 \text{ \AA}$   $R_g$  shift upon ERAP1 binding DG013 is consistent with a shift of the conformational equilibrium from open toward closed. Fitting the ERAP1 + DG013 scattering profiles to molecular dynamics structures revealed a cluster of models consistent with the experimental data ( $\chi^2 < 1$ ) with theta between 54 and 62° (Fig. 1d, yellow symbols). This cluster is distinct and nonoverlapping with that observed for ERAP1 in the absence of DG013, and consistent with a conformation similar to but slightly more open than observed for ERAP1 in closed-conformation crystal structures.

Overall, the SAXS analysis shows that ERAP1 adopts an open conformation in solution and closes upon binding substrate or substrate-like inhibitors.

**The crystal structure of a transition-state peptide analog bound to the open conformer reveals conformational changes induced by substrate binding.** To begin to address how domain closure is coupled to catalysis, and to dissect structural features related to substrate binding from those related to the open-closed

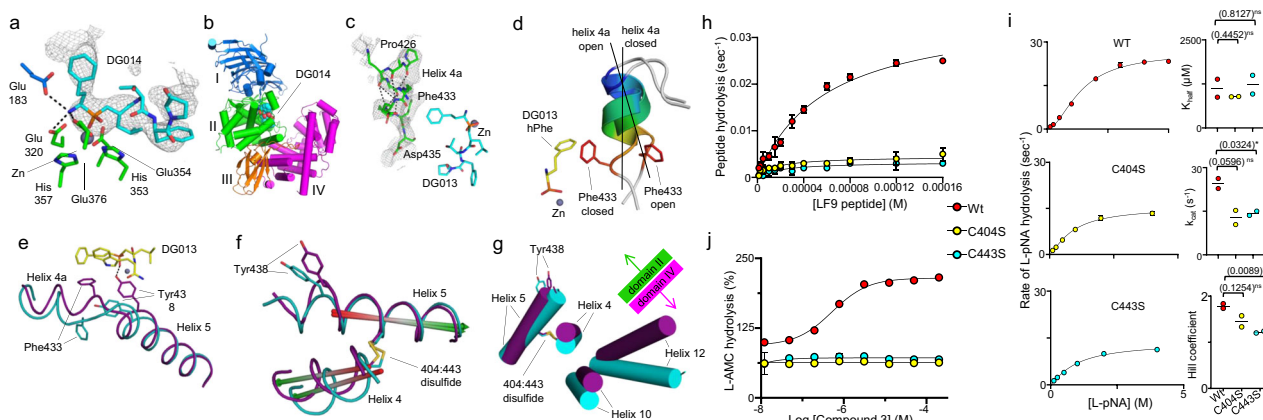
**Table 1** Data collection and refinement statistics for ERAP1 bound to phosphinic inhibitors.

|                                | ERAP1 in complex with phosphinic pseudotripeptide inhibitor DG013 (PDB ID 6M8P) | ERAP1 in complex with phosphinic pseudodecapeptide inhibitor DG014 (PDB ID 6MGQ) |
|--------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Wavelength (Å)                 | 0.979                                                                           | 1.110                                                                            |
| Resolution range (last shell)  | 41.7–3.3 (3.4–3.3)                                                              | 56.1–2.9 (3.0–2.9)                                                               |
| Space group                    | P2 21 21                                                                        | P1 21 1                                                                          |
| Unit cell a, b, c              | 125.79, 548.68, 589.06                                                          | 56.14, 234.7, 132.28                                                             |
| α, β, γ                        | 90, 90, 90                                                                      | 90, 93.12, 90                                                                    |
| Total reflections              | 6,551,362 (1,046,163)                                                           | 734,382 (80,646)                                                                 |
| Unique reflections             | 601,903 (19,094)                                                                | 73,472 (1842)                                                                    |
| Multiplicity                   | 10.9 (10.5)                                                                     | 10.0 (8.8)                                                                       |
| Completeness (%)               | 74.7 (19.2)                                                                     | 81.7 (20.1)                                                                      |
| Completeness, elliptical (%)   | 87.6 (51.9)                                                                     | 99.2 (98.9)                                                                      |
| Mean I/σ(I)                    | 2.68 (0.54)                                                                     | 20.11 (0.56)                                                                     |
| Wilson B-factor                | 29.6                                                                            | 59.5                                                                             |
| R-merge                        | 1.121 (4.858)                                                                   | 0.657 (4.604)                                                                    |
| R-meas                         | 1.177 (5.105)                                                                   | 0.692 (4.896)                                                                    |
| R-pim                          | 0.353 (1.55)                                                                    | 0.215 (1.631)                                                                    |
| CC1/2                          | 0.952 (0.145)                                                                   | 0.861 (0.234)                                                                    |
| CC*                            | 0.988 (0.503)                                                                   | 0.962 (0.616)                                                                    |
| Reflections used in refinement | 449,952 (19,086)                                                                | 60,060 (1841)                                                                    |
| Reflections used for R-free    | 2123 (83)                                                                       | 2987 (92)                                                                        |
| R-work                         | 0.285 (0.330)                                                                   | 0.198 (0.309)                                                                    |
| R-free                         | 0.295 (0.361)                                                                   | 0.256 (0.371)                                                                    |
| CC (work)                      | 0.725 (0.426)                                                                   | 0.865 (0.564)                                                                    |
| CC (free)                      | 0.758 (0.280)                                                                   | 0.844 (0.448)                                                                    |
| Protein molecules per ASU      | 22                                                                              | 3                                                                                |
| Number of non-hydrogen atoms   | 155,927                                                                         | 20,671                                                                           |
| Number of macromolecules       | 151,544                                                                         | 20,279                                                                           |
| Number of ligands              | 4383                                                                            | 138                                                                              |
| Number of solvent              | NA                                                                              | 39                                                                               |
| Protein residues               | 18,942                                                                          | 2562                                                                             |
| RMS (bonds)                    | 0.004                                                                           | 0.005                                                                            |
| RMS (angles)                   | 0.64                                                                            | 0.79                                                                             |
| Ramachandran favored (%)       | 95.94                                                                           | 94.13                                                                            |
| Ramachandran allowed (%)       | 4.06                                                                            | 5.52                                                                             |
| Ramachandran outliers (%)      | 0                                                                               | 0.35                                                                             |
| Rotamer outliers (%)           | 0.42                                                                            | 1.7                                                                              |
| Clashscore                     | 6.27                                                                            | 27.32                                                                            |
| Average B-factor               | 39.5                                                                            | 70.7                                                                             |
| Average macromolecules         | 39.0                                                                            | 70.7                                                                             |
| Average ligands                | 55.8                                                                            | 59.0                                                                             |
| Average solvent                | NA                                                                              | 32.77                                                                            |
| Number of TLS groups           | 1                                                                               | 14                                                                               |

transition, we crystallized ERAP1 bound to the decamer phosphinic inhibitor, DG014, and determined the three-dimensional structure by X-ray crystallography (Table 1, Supplementary Fig. S3g–j, and Supplementary Table S1b) DG014 shares DG013's phosphinic group and N-terminal residues but is seven amino-acid residues longer, and similarly inhibits ERAP1 with high potency (Supplementary Fig. S1c, S5). Like DG013, DG014 induced closing as assessed by SAXS/WAXS  $R_g$  and theta analysis (Supplementary Fig. S2). Serendipitously, we were able to obtain crystals of the ERAP1–DG014 complex in the open conformer, and we determined its structure by X-ray crystallography (Table 1, Supplementary Table 1, see “Methods” for details). The crystal structure shows that DG014 binds as expected with the phosphinic group at the active site, but with only the first five residues sufficiently structured to allow confident modeling (Fig. 2a, b). Ordered electron density of sufficient volume to contain ~2 amino-acid residues was observed 22 Å away from the active site, but we could not confidently ascertain whether this corresponded to buffer components or disconnected peptide C-terminal density, both of which have been observed at similar locations in domain IV in other crystal structures<sup>19,21,22,30</sup>.

Relative to previously determined structures of the open conformation for ERAP1 bound to bestatin<sup>12</sup>, conformational changes induced by peptide binding are apparent. A region between helices 4 and 5 (residues 426–433) that was not visualized in previous open conformation structures<sup>11,12</sup> is now ordered and forms a short helix 4a (Fig. 2c). Helix 4a is in an intermediate state in the open DG014 structure relative to closed-form structures, with a shift required to adopt the configuration present in the closed conformer (Fig. 2d). In the closed state, phenylalanine 433 makes substantial contacts with bound ligand, as observed previously in the closed-state structure of ERAP1 bound to bestatin<sup>11</sup>, and conversion between the DG014-bound (open) and DG013-bound (closed) structures requires translation and rotation of helix 4a to bring phenylalanine 433 in contact with substrate P1 side chain (Fig. 2d).

Reorientation of Phe433 and motion of the helix 4a is coupled to rotation and tilt of helix 5, which repositions the side chain of Tyr438 (Fig. 2e). Based on mechanistic studies of M1 family aminopeptidases<sup>31</sup> and mutagenesis studies of ERAP1<sup>12</sup>, Tyr438 is expected to play an important role in the catalytic mechanism by stabilizing the oxyanion that forms in the transition state with



**Fig. 2** Ordering and reorientation of helix 4a near the active site with substrate binding and role of Cys403-443 disulfide in coupling active-site rearrangements with interdomain interactions. **a** DG014, cyan, is shown bound to ERAP1 active site. Polder map (gray mesh) contoured at  $3.5\sigma$  shows DG014 density in chain A. **b** ERAP1 was crystallized in the open conformation with DG014. Domains I (blue), II (green), III (orange), and IV (magenta) are shown as cartoons. DG014 is shown as cyan spheres. **c** Helix 4a is ordered along the bound inhibitor. Polder map (gray mesh) contoured at  $2.8\sigma$ . Asp435, a5 the C-terminal end of helix H4a, contacts DG014 at the P2' and P4' positions. **d** Helix H4a tilts and rotates between open and closed conformations. Helix 4a is colored as N- to C-terminal blue-red rainbow. Helix 4a tilts  $35^\circ$  between open and closed conformations (angle between helix central axes, black lines). Rotation along the helix central axis is depicted using Phe433 side chain for reference. DG013 homophenylalanine-phosphinyl group is shown bound at active site zinc. Phe433 in closed-conformation contacts S1' homophenylalanine (hPhe) of DG013 in a T-shaped  $\pi$ - $\pi$  interaction. **e** Phe433 motion occurs in conjunction with Tyr438 repositioning. Helix 4a moves as helix 5 twists, positioning Tyr438 for catalysis. Interaction with DG013 shown as dashed line. **f** Helix 5 rotation aligns Tyr438 for catalysis<sup>11</sup>. Helix 5 twist is coupled to reorientation of helix 4 through the 404:443 disulfide. **g** Helices 10 and 12 from domain 4 are adjacent to helix 4, coupling the domain interface to the substrate-binding site. **h** Mutation of residues participating in the Cys404-Cys443 disulfide bond by replacement with serine reduces LF9 peptide hydrolysis and (i) L-pNA enzymatic activity. The data are fit into Michaelis-Menten equation for peptide substrate (**h**) and allosteric sigmoidal for L-pNA with fit values for  $K_{\text{half}}$ ,  $k_{\text{cat}}$ , and Hill coefficient are shown in (i). An unpaired two-tailed *t* test was used to assess the significance of the difference between  $K_{\text{half}}$ ,  $k_{\text{cat}}$ , and hill coefficient for ERAP1 and Cys mutants from two independent experiments ( $n = 3$ ). **j** Cys404-Ser and Cys443-Ser mutations block allosteric activation by compound 3. Activity assays in (**h**, **j**) were performed in three independent experiments with two replicates ( $n = 2$ ) for (**h**) and (**j**) in each independent assay. Error bars on (**h**-**j**) represent the variation between the samples in an independent experiment. Source data are provided as a Source Data file.

water attack on the scissile bonds of a peptide substrate. It was noted previously that Tyr438 near the end of helix 5 is oriented away from the substrate in the open conformation, but moves towards the active site in the closed conformation<sup>11,12</sup>. In the ERAP1-DG014 (open) complex, (Fig. 2e, cyan) as in previous open ERAP1 crystal structures, Tyr438 is oriented away from the active site and moves to interact with the peptide bond transition-state mimic phosphinic group in the ERAP1-DG013 (closed) complex (Fig. 2e, magenta), in a position analogous to the previously determined bestatin-bound ERAP1 closed structure<sup>11</sup>. Although this motion of tyrosine 438 into the active site previously has been associated with domain closure<sup>12,13</sup>, the coupling mechanism has been obscure. We suggest that the region between helices 4 and 5 (residues 426–433) acts as a sensor for substrate binding, ordering to form helix 4a in the presence of sufficiently long peptide-like substrates, and providing a site by which domain closure can be coupled to active-site reorganizations required for catalysis.

We note that helix 5 is linked to helix 4 through a disulfide bond between residues Cys443 and Cys404 (Fig. 2f), and that helix 4 forms the edge of domain II, where it lies along helices 10 and 12 at the edge of domain IV (Fig. 2g). The disulfide bond between helix 4 and helix 5 is unique to ERAP1 and ERAP2 among the M1 family of zinc aminopeptidases and is highly evolutionarily conserved among ERAP1 orthologs (Supplementary Fig. S6). For IRAP and CD13, two other M1 family members that have been crystallized in both open and closed conformations, helix 4 moves between open and closed states in conjunction with domain IV as is found in ERAP1. However, in these proteins, which lack a disulfide corresponding to Cys443-Cys404 in ERAP1, helix 4 motion does not couple with motion of the N-terminus of helix 5 (Supplementary Fig. S7). To test the

role of this disulfide in ERAP1 in coupling domain closure with enzymatic activation, we introduced cysteine-to-serine mutations C404S and C443S to disrupt disulfide bond formation and measured the effect on enzymatic activity. The C404S and C443S mutations resulted in a significant decrease in peptide hydrolysis activity (Fig. 2h). To explore if this reduction was due to reduced coupling of domain closure and enzymatic activation, we examined the cooperativity of Leu-pNA hydrolysis. Fitting L-pNA hydrolysis data to an allosteric Michaelis-Menton model<sup>12</sup> (see “Methods” for details), showed that  $K_{\text{half}}$  was not affected, indicating that the mutations did not affect substrate binding, but  $k_{\text{cat}}$  was significantly reduced, indicating an effect on the enzymatic processes subsequent to substrate binding (Fig. 2i). The Hill coefficient was reduced by disulfide bond disruption, as expected if it helped to couple allosteric site binding to active-site rearrangement (Fig. 2i). To critically test the role of the C404-C443 disulfide in coupling domain closure to enzymatic activation, we tested the effect of C404S and C443S mutation on allosteric activation induced by compound 3 (Supplementary Fig. S1), a previously characterized non-peptidomimetic small-molecule allosteric modulator<sup>20</sup>. Compound 3 activates ERAP1 for Leu-AMC hydrolysis; the activation has been hypothesized to act by stabilizing the closed conformer<sup>20</sup>, and this is confirmed below. The C404S and C443S mutations abrogated the allosteric activation effect of compound 3 (Fig. 2j). These results support a role for the C404-C443 disulfide in coupling binding site occupancy to enzymatic activation via domain closure.

**Peptide C-terminal binding sites localize within domain IV at sites accessible in the open conformation with additional contacts formed upon domain closure.** ERAP1 exhibits side

chain specificity at the C-terminal as well as N-terminal end of peptide substrates<sup>17,19,21</sup>, suggesting contact(s) between enzyme and substrate beyond the five residues of DG014 for which ordered peptide density was observed. A binding site for C-terminal aliphatic amino-acid residues has been identified within the interior of ERAP1's substrate-binding cavity in a crystal structure of a substrate peptide bound to ERAP1's isolated C-terminal domain<sup>22</sup>, and in a recent structure of intact ERAP1 in closed conformation bound to a long phosphinic peptide inhibitor<sup>21</sup>. Although aliphatic residues are preferred at the C-terminus of peptide substrates, peptides with other C-terminal amino acids also can be processed efficiently<sup>8,17,32</sup>. We used a photo-cross-linking approach to identify additional C-terminal binding site(s) for bound peptide substrates. We designed an 11-mer cross-linking probe, designated DG023, with peptide sequence based on a known ERAP1 substrate<sup>8</sup>, the non-hydrolyzable phosphinic group replacing the first peptide bond, and the unnatural amino-acid *p*-benzoyl-L-phenylalanine (BPA)-amide at the C-terminus<sup>33</sup> (Fig. 3a). UV irradiation (350 nm) of BPA generates a carbene that can form a covalent bond with nearby molecules<sup>33</sup>. We performed cross-linking reactions in this manner using DG023 and purified ERAP1 and identified three cross-linking sites by LC-MS/MS (Fig. 3b, Supplementary Fig. S8a, b, see "Methods" for details). Two of the sites, Leu677 and Leu838, lie in domain IV close to the C-terminal aliphatic amino-acid residue binding site identified X-ray crystallography<sup>11,22,34</sup> (Fig. 3b, c). The third site, Leu686, borders domains II, III, and IV at the *in silico* docking site identified for compound 3 (Fig. 3b, c); the homologous site is accessed by long peptides bound to mammalian aminopeptidase N (CD13), a structural homolog of ERAP1 involved in neuropeptide processing<sup>35</sup>. We were able to build conceptual models with good geometry for DG023 peptides extending from the active site Zn to each of the cross-linking sites, using the DG013 and DG014 structures as guides (Fig. 3d). Recent crystal structures of 10-mer and 15-mer phosphinic peptides reveal binding sites in the same general region in the interior of ERAP1's substrate-binding cavity (Fig. 3e)<sup>21</sup>.

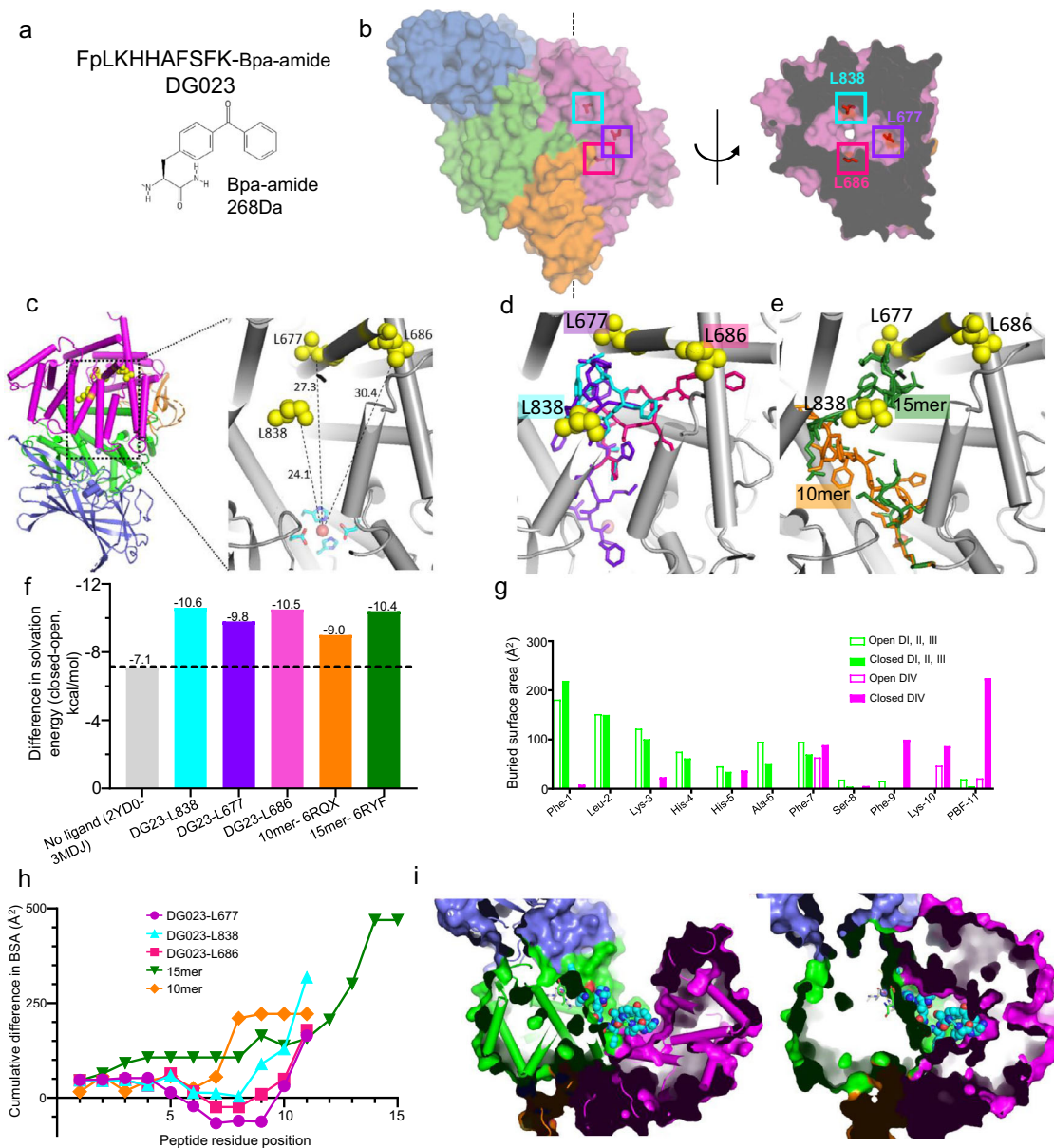
We performed buried surface area and solvation energy calculations for each of the structural models, shown in Fig. 3d, e, to help understand the impact of binding long peptide substrates on domain closure. Comparisons of open and closed conformers for ERAP1 bound to short inhibitors or peptides show that the closed structure has many additional interdomain contacts, mostly between domains II and IV, involving 77 residues (Supplementary Tables 2 and 3), and resulting in ~2800 Å<sup>2</sup> of protein surface area accessible to solvent in the open conformation but excluded from solvent in the closed conformation. Such "buried" surface area correlates with greater energetic stability by decreasing the energetic requirement of solvation<sup>36</sup>. For ERAP1, domain closure results in a reduction of solvation energy between open and closed conformers estimated to be ~7 kcal/mole, calculated for ERAP1 crystal structures with the short peptide-like inhibitor bestatin (Fig. 3f, gray bar). Binding of long peptides creates additional sites for interdomain interaction, corresponding to an additional 355–1019 Å<sup>2</sup> of protein surface area buried by domain closure (Supplementary Fig. S9a), and increased stabilization of the closed conformer by 1.9–3.5 kcal/mole (Fig. 3f, colored bars). ERAP1 makes contact with the bound peptides all along their lengths, but the contacts do not substantially involve domain IV, or contribute to the overall stabilization of the closed conformer, until the peptide extends beyond 6–9 residues, as shown in detail for the ERAP1–DG023–L838 model (Fig. 3g, see Supplementary Fig. S9b–e for the other complexes). A similar length dependence was observed for all the peptide models (Fig. 3h). A cross-sectional view of the

ERAP1–DG023–L838 model illustrates how long peptide binding creates many sites for stabilization of the closed conformation (Fig. 3i, see Supplementary Tables 2 and 3 for the other complexes).

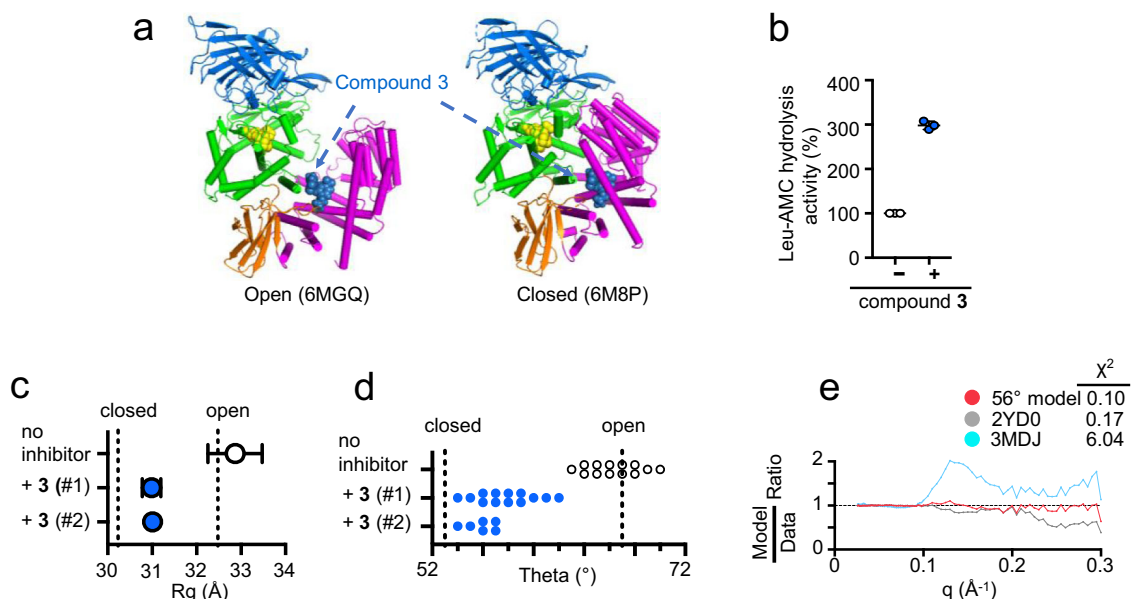
**An allosteric small-molecule activator induces ERAP1 domain closure.** To investigate the relationship between conformational states and allosteric modulation of ERAP1 enzymatic activity, we characterized ERAP1 bound to the non-peptidomimetic small-molecule allosteric modulator compound 3 mentioned above. Compound 3 docks in a pocket at the junctions of domains II, III, and IV formed by the closed conformation of ERAP1 (Fig. 4a), and acts as an allosteric activator of ERAP1 by increasing hydrolysis activity towards the dipeptide substrate Leu-AMC<sup>20</sup> (Fig. 4b). Compound 3 induced ERAP1 to adopt a closed conformation in solution, as shown by SAXS/WAXS  $R_g$  analysis (Fig. 4c) and by model fitting full scattering curves to those calculated for MD models as just described (Fig. 4d, e).  $R_g$  values and theta opening values were similar to those observed for ERAP1 bound to DG013 and the other small-molecule inhibitors. These results indicate that an allosteric activator can induce a similar closing motion as substrate binding and associate the closed conformation with the 'active conformer' hypothesized to explain the allosteric activation in other contexts.

**Position 528 Lys/Arg polymorphism alters domain closure energetics.** Numerous genome-wide association studies (GWAS) have linked a common ERAP1 polymorphism at position 528 with susceptibility to inflammatory conditions of autoimmune etiology such as psoriasis, Behçet's disease, and most strongly, with ankylosing spondylitis<sup>5,37,38</sup>. The disease-correlated variant codes for lysine at this position, compared to arginine in the protective ERAP1 allele. This seemingly conservative substitution nevertheless has an effect on immune function, presumably by altering the presentation of peptides that stimulate an autoinflammatory response<sup>7</sup>. Proposed mechanisms for how the polymorphism influences ERAP1 activity include alterations in protein expression<sup>39</sup> or altered protein conformational flexibility<sup>18</sup>. As reported previously<sup>8,18</sup>, we find that Arg528 distinctly alters hydrolysis activity of the dipeptide substrate Leu-pNA compared to Lys528, slightly increasing the enzyme catalytic rate  $k_{cat}$  but substantially decreasing the apparent substrate affinity as measured by increasing  $K_{half}$  (Supplementary Fig. S12).

Examination of the ERAP1 structure and electrostatic calculations shows that a positively charged Arg or Lys at position 528 forms a long-range electrostatic interaction with Glu913 in domain IV (Fig. 5a, long dashes). For both Arg528 and Lys528, the interaction would contribute preferentially to the closed conformation, as the pair of amino acids moves far apart in the open conformation (Fig. 5a, top panel). Calculation of the electrostatic potential between positions 528 and 913 using DelPhi shows that Lys528 has a stronger energetic contribution to the closed conformation than does Arg528, suggesting that it stabilizes the closed relative to the open conformation of ERAP1. The energetic difference as estimated by DelPhiForce is ~4.5 kJ/mol. Besides the Glu913 electrostatic interaction, Arg528 and Lys528 also interact with Asn414 and His548, in domains II and III, respectively<sup>40</sup>. Similarly, to the Glu913 interaction, the Arg/Lys528 interactions with Asn414 and His548 contribute preferentially to the closed conformation, with more favorable geometry for Lys528 as compared to Arg528 (Fig. 5a). Thus, through the long-range electrostatic interaction with Glu913 and interactions with Asn414 and His548, Lys528 would appear to stabilize the closed conformation relative to Arg528.



**Fig. 3** Binding of substrates and inhibitors increase interactions between domains II and IV. **a** Crosslinker peptide sequence is based on the sequence of characterized ERAP1 substrate<sup>8</sup>. First two amino acids are phenylalanine and leucine analogs connected by a phosphinic group “p”. Peptide C-terminus bears a photocrosslinker *p*-benzoylphenylalanine (BPA). After cross-linking, trypsin cleaves probe peptide leaving BPA-amide as a 268 Da adduct at the crosslink site. **b** Cross-linked ERAP1 residues reside in domain IV (magenta). Orthogonal views shown of closed ERAP1 structure (PDB ID 2YD0). Cross-linked amino acids shown as sticks with colored boxes. On right, a cutaway view reveals cross-linked sites in domain IV. **c** Close views of BPA-cross-linked Leu838, 677 and 686 localized in domain IV and their distances from the active site Zn is 24.1, 27.3 and 30.4 Å, respectively. **d** Close views of the paths of modeled DG023 peptide to each cross-linked site. **e** Paths of 10-mer (orange) and 15-mer (green) phosphinic peptides bound to ERAP1<sup>21</sup> include L838 and L677 cross-linking sites. **f** Solvation energy of the closed conformer increases on binding long peptide substrates as compared to open conformer. 2YD0 (closed) and 3MDJ (open) were used as controls because of the absence of long peptides in the structures. DG023 peptide models and crystal structures of ERAP1 bound with 10-mer and 15-mer were used to calculate the solvation energy of open and closed conformers using PISA. **g** Buried surface area (BSA) of peptide residues in domain I, II, III (green), and domain IV (magenta) in open (open bars) and closed conformer (closed bars) along the length of the peptide. N-terminal peptide residues are more buried in domains I, II, III in both open and closed conformers while C-terminal peptide residues make more contacts in domain IV on its closure. BSA of the DG023-L838 peptide model is plotted here. **h** Cumulative difference in BSA along the length of the peptide between closed and open conformer increases for the C-terminal residues for peptide substrates. **i** Cross-sectional view of ERAP1 with DG023-L838 peptide model bound to open (left) and closed (right) conformation shows an increased number of contacts for the peptide C-terminal with domain IV in closed conformer (right). Source data are provided as a Source Data file.



**Fig. 4 Allosteric small-molecule activator compound 3 induces ERAP1 closure.** **a** Preferred binding pose for compound **3** at an interdomain interface distal to the active site<sup>20</sup>. **b** Compound **3** activates hydrolysis of dipeptide substrate leucine-amidomethylcoumarin (Leu-AMC). **c–e** ERAP1 adopts closed conformation in the presence of compound **3**, as measured by **(c)**  $R_g$  analysis, and **(d)** fitting to MD models. **e** Residual plot of best structural model fit to SAXS data in the presence of compound **3**, with crystal structure model fits shown for reference. Each data point in **(c)** represents a single independent experiment with  $n = 1$  and the error of linear fit for each independent experiment is shown as error bars. Source data are provided as a Source Data file.

To investigate the effect of the 528 Arg/Lys polymorphism on the ability of ERAP1 to adopt the closed conformation, we measured the concentration dependence of ERAP1 modulation by two compounds that induce ERAP1 closure, the pseudopeptide active-site inhibitor DG013 (Fig. 5d) and the allosteric activator compound **3** (Fig. 5b, c), using a Leu-AMC hydrolysis assay. We used two allelic variants that differ only at position 528 (Supplementary Fig. S1). We reasoned that if the predicted electrostatic stabilization of the closed conformer of Lys528 ERAP1 over Arg528 ERAP1 were correct, compounds binding to the closed conformer should show preferential binding to Lys528 ERAP1. That is what was observed. Both compounds exhibited a higher potency for Lys528 ERAP1 (circles) than for Arg528 ERAP1 (squares); with lower  $IC_{50}$  for the inhibitor DG013 (Fig. 5b), and lower  $AC_{50}$  for the activator compound **3** (Fig. 5c). A similar polymorphism-dependent difference in inhibitory concentration has been reported previously for bestatin<sup>41</sup>, which also stabilizes the closed conformation of ERAP1 (Supplementary Fig. S2a). This behavior can be contrasted with that observed for compound **2** (Supplementary Fig. S1), a small-molecule inhibitor of ERAP1 that preferentially docks to the open conformer<sup>20</sup>, and does not induce ERAP1 domain closure, even when present at saturating concentrations (Fig. 5d, f, purple symbols). Unlike compound **3** and DG013, compound **2** exhibits equivalent potency for the two ERAP1 lysine/arginine 528 variants (Fig. 5d).

To investigate the effect of the position 528 Arg/Lys polymorphism on the ERAP1 open and closed conformations, we measured SAXS/WAXS scattering profiles. In the absence of ligand, the two ERAP1 variants are not distinguishable by Guinier  $R_g$  analysis (Fig. 5e, open circle and open square for Lys528 and Arg528, respectively) or by fitting scattering profiles to MD simulation-derived structures with different theta values (Fig. 5f, same symbols as 5e). In the presence of saturating concentrations of the inhibitor DG013 or the allosteric modulator compound **3**, both ERAP1 variants adopted the closed conformation (Fig. 5e, f).  $R_g$  and theta values were identical for the two variants. Thus, the 528 Arg/Lys polymorphism does not appreciably alter the

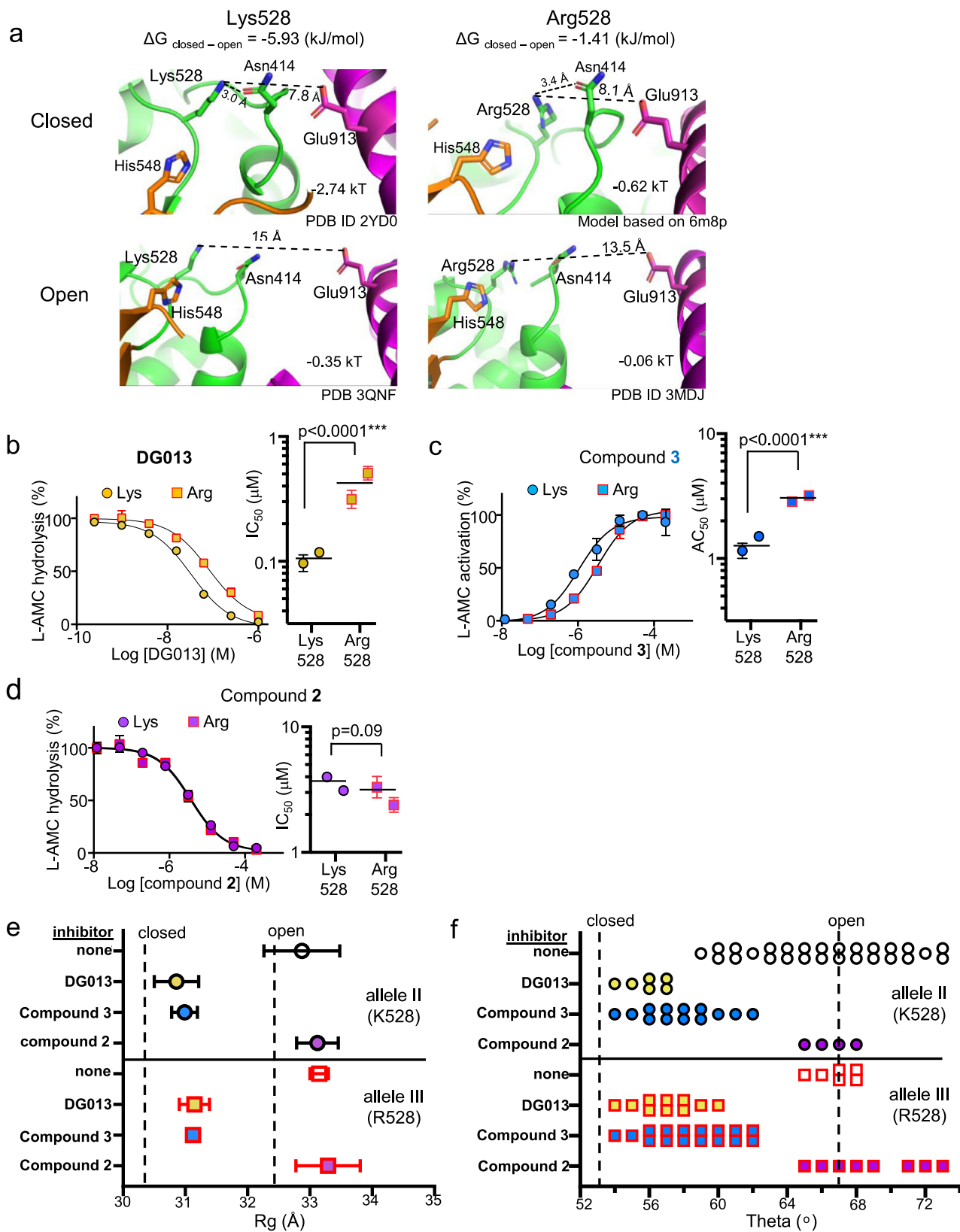
conformation of ERAP1 in either the open or closed forms, despite its apparent effect in regulating the conversion between them.

Position 528 is neither located near the ERAP1 active site nor the docking sites for compounds **2** or **3**. Thus the greater effects of lysine/arginine 528 substitution on the binding of compounds that stabilize the closed over the open conformer are consistent with a model in which ERAP1 domain closure is altered by polymorphic variation at position 528, with arginine at position 528 impairing ERAP1's ability to dynamically adapt the closed form. Interestingly, similar effects on enzymatic parameters and DG013 binding were observed previously when ERAP1 conformational equilibrium was disrupted by mutations that destabilize ERAP1 interdomain salt bridges present in the closed conformation<sup>28</sup>.

## Discussion

These results, taken together, provide a comprehensive view of large-scale domain closure motions in ERAP1 and their role in catalytic function. ERAP1 in solution adopts a largely open conformation in the resting state, as shown by small-angle X-ray scattering (Fig. 1). Under these conditions, the opening angle theta<sup>28</sup>, which describes the orientation of domain IV relative to domains I, II, and III, is similar to that of the open conformer observed in crystal structures. In this conformation, the substrate-binding site is accessible to solution, but it is not configured for efficient processing, with interhelical loops disordered and a key active site tyrosine not optimally oriented for catalysis<sup>11,12</sup>. Binding of a variety of ligands including allosteric activators and peptide-like inhibitors promotes domain closure, in each case stabilizing a form with theta angle similar to that observed crystallographically for the closed conformer (Figs. 1 and 4). A complex of ERAP1 bound to the long peptidomimetic inhibitor DG014 trapped crystallographically in the open conformation exhibited some of the conformational attributes previously observed in closed-state structures, in particular ordering of helix





4a, suggesting that these changes were coupled to substrate binding (Fig. 2). Other changes associated with the closed conformation appeared to require domain closure, including the key rotation of helix 5 that brings Tyr438 into position for catalysis. Peptide models extending from the catalytic zinc to C-terminal binding sites identified by chemical cross-linking, as well as two

recent crystal structures of ERAP1 bound to 10-mer and 15-mer peptides<sup>21</sup>, showed that bound peptides provide many sites for interdomain interaction, with residues beyond positions 7–9 contributing preferentially to stabilization of the closed conformer (Fig. 3). This observation suggests that domain closure will be more efficient for longer peptides because of the larger

**Fig. 5 K/R polymorphism at position 528 modulates a conformationally dependent interdomain electrostatic interaction and regulates binding of small-molecule ERAP1 modulators that induce closing.** **a** Lys or Arg at position 528 makes a long-distance electrostatic interaction with Glu913, in addition to interactions with Asn 414 and His 448. Open and closed conformations for proteins with Lys528 or Arg528 are shown. The closed Arg528 conformation is modeled based on 6M8P, all others are from crystal structures as indicated. Closest interatomic distances between Lys/Arg528 and Glu913 are 7.8 Å (closed) and 13.5 Å (open), respectively. Lys528 stabilizes the closed ERAP1 conformation more than does Arg528. Pairwise electrostatic potentials between residues 528 and 913 shown in kT units for each structure, calculated using DelPhiForce. **b** DG013 exhibits greater potency for Lys528 ERAP1 than Arg528 ERAP1. Leucine-AMC hydrolysis rate was measured and normalized to control condition without DG013. Data were fit to a sigmoidal curve with constrained top and bottom = 100 and 0% activity respectively. Significance calculated using two-tailed ANOVA (\* $P < 0.0001$ ). **c** Compound **3** exhibits greater potency for Lys528 ERAP1 than Arg528 ERAP1. L-AMC hydrolysis rate was measured and normalized to control condition without compound **3**. Data were fit to a sigmoidal curve with constrained bottom = 100% activity. Significance calculated using two-tailed ANOVA (\*\*\* $P < 0.0001$ ). **d** Compound **2** does not show ERAP1 variant-specific potency. Leucine-AMC hydrolysis rate was measured and normalized to control condition without compound **2**. Data were fit as in (b). For binding studies with all the three small molecules (b–d), two independent experiments were performed with triplicate ( $n = 3$ ) samples in each experiment and variation between the replicates are represented by error bars. For both Lys528 (allele II) and Arg528 (allele III) variants, ERAP1 conformation remains open in the absence of ligand and in presence of compound **2**, as shown by  $R_g$  (e) or SAXS/WAXS model fitting (f). Both Arg/Lys528 variants of ERAP1 remain closed in the presence of compound **3** and DG013, as shown in the same analysis in (e) and (f). A single independent experiment ( $n = 3$ ) was performed for  $R_g$  analysis on Lys528 and Arg528 variants of ERAP1 and error of linear fit for each independent experiment is shown as error bars (e). Source data are provided as a Source Data file.

number of sites available for interdomain interaction and that this effect can provide the basis for ERAP1's striking length-dependent "molecular ruler" cleavage activity<sup>16,17</sup>.

Early ideas about structural determinants for ERAP1 conformational change and allosteric regulation focused on a single regulatory site postulated to be accessible to peptides of eight residues or longer<sup>12</sup>. Occupancy of this site, for example by the C-terminus of sufficiently long peptides, or by allosteric activators, was hypothesized to trigger a conformational change to the closed conformation<sup>12</sup>. Subsequent structural work identified the binding sites for a peptide substrate<sup>22</sup> and a small-molecule allosteric modulator (compound **1**<sup>19,20</sup>), and the location of C-terminal residues for 10-mer and 15-mer phosphinic peptide inhibitors<sup>21</sup> (Supplementary Fig. S10). Subsequent docking studies revealed a distinct location for a different allosteric activator (compound **3**<sup>20</sup>), identified also as the binding site for a buffer component from the crystallographic mother liquor<sup>30</sup>, and along the path for a long peptide exiting the substrate-binding cavity of the homologous protein human aminopeptidase N<sup>35</sup> (Supplementary Fig. S10). The cross-linking studies reported here reveal that the C-terminus of bound DG023 peptide can access both of these regions in solution (Supplementary Fig. S10). ERAP1 has wide specificity for peptides of different sequences, with different peptide sequences exhibiting different internal and to a lesser extent C-terminal sequence preferences<sup>17,42</sup>, and this would appear to require degenerate binding modes. The surface area and energetic calculations reported here indicate that bound peptides in a variety of conformations can stabilize the active, closed conformation of ERAP1. Based on these results, we suggest that the large interdomain interaction surface comprising >70 residues provides many opportunities for stabilization of the closed form by allosteric activators and by long peptides that can bridge between domains II and IV, regardless of the exact path of the bound peptide (Supplementary Tables 2 and 3).

A recent analysis of ERAP1 molecular dynamics trajectories calculated from high-resolution closed-conformer crystal structures revealed a path of correlated dynamic motions linking Tyr438 in the active site with residues near the C-terminus of the bound 15-mer peptide<sup>21</sup>. The network path was shorter in the presence of an N-terminally bound tetrapeptide, suggesting that substrate binding results in stronger dynamic coupling between active and regulatory sites<sup>21</sup>. Our results help to confirm and extend this model. The C404–C433 disulfide, required for efficient allosteric activation, lies along the proposed path (Supplementary Fig. S11), and provides a mechanism for at least some of the proposed conformational coupling. Perhaps more

importantly, we note that in the open conformation the path of correlated motions is broken because of the movement of domain IV helices 10 and 12 relative to domain II helices 4 and 5 (Supplementary Fig. S11). Thus, domain closure would appear to be required for efficient conformational coupling of regulatory and active sites via the proposed network path.

Electrostatic stabilization of the closed conformer of ERAP1 provides a mechanism for the observed effects of the key Lys/Arg528 polymorphism. Of the nine common ERAP1 single amino-acid polymorphisms<sup>43</sup>, K528R is the most strongly associated with autoimmune disease<sup>4,5,43</sup>, with substantial effects on enzymatic activity toward both single-residue and peptide substrates<sup>8</sup>. The K528R amino-acid substitution is conservative and located away from the catalytic center and substrate binding cavity, and thus the mechanism by which Lys/Arg528 impacts ERAP1 activity has been an outstanding question in the field. Some other common polymorphic sites are located within the interior compartment, in position to interact with a bound peptide<sup>12</sup>, or along the path of coupled motions<sup>21</sup>, but position 528 is on the surface of the protein at the junction of domains II, III, and IV). Crystal structures of ERAP1 show that lysine or arginine at this position can participate in a long-range electrostatic interaction with Glu913 in domain IV when the protein is in the closed but not the open conformation (Fig. 5a). Poisson–Boltzmann electrostatic calculations suggest that this interaction stabilizes the closed form. In addition, an intradomain interaction involving Lys/Arg528, Asn414, and His548 and specific to the closed conformation is more favorable in the Lys528 variant, as previously noted<sup>40</sup>, also tending to stabilize the closed conformer. The energetic effects of these changes are predicted to be relatively small, and we did not observe any substantial differences in the SAXS profiles of K528 and R528 that would have been indicative of substantial alterations in the equilibrium mixture of open and closed forms, or a change in the extent of interdomain motion in the open form. However, we did observe significant differences in the potency of two inhibitors that bind at two different sites and induce domain closure (Figs. 1 and 4). Both compounds were more potent for the K528 variant, consistent with the proposed electrostatic stabilization of the closed form for this variant and preferential binding of the inhibitors to binding to the closed form. Previous molecular dynamics simulations also suggested a role for the K/R528 polymorphism in regulating ERAP1 conformational dynamics and domain motions, with K528 shifting the conformational ensemble towards the closed conformation<sup>18</sup>. As ERAP1 inhibitors are increasingly being considered as potential therapeutics for cancer

immunotherapy and autoimmunity, the effect of common ERAP1 polymorphisms on inhibitor potency must be considered. With regard to understanding the role ERAP1 polymorphism in autoimmune etiology, stabilization of the closed form (as observed for K528) would tend to promote increased activity under conditions where substrate binding and/or domain closure were limiting, but for some substrates turnover might be limited by product release or other aspects of the enzymatic cycle involving the open conformer. In these cases, the greater relative stability of the closed conformation might be detrimental to the overall activity. Thus, the effects of the ERAP1 polymorphism and domain closure on processing particular candidate autoimmune epitopes<sup>6,7</sup> might be difficult to predict.

Several questions and uncertainties about ERAP1 enzymatic mechanism and regulation remain. Solution-scattering studies were highly informative in relating ERAP1 conformational states in solution to those observed in crystal structures, and in revealing that binding of substrates, inhibitors, and allosteric activators all induce domain closure from a resting open conformation. However, an important caveat of using SAXS to study conformational states in this system is the difficulty in distinguishing mixtures of states from single conformations. Although in principle information about the distribution of conformational states is available from SAXS studies, we observed similar residuals in fits of experimental scattering curves to single conformations and to mixtures of conformations. For each experimental condition, we identified an ensemble of individual conformations consistent with the experimental data, but whether the individual conformations represent actual species or weighted averages of open and closed forms is not clear. Moreover, we only considered conformations observed in crystal structures or accessible by standard or accelerated molecular dynamics simulations starting from the crystal structures<sup>18,28</sup>. Model-independent approaches to SAXS data analysis such as low-resolution envelope or bead-model generation<sup>44</sup> potentially could identify novel conformations present in the solution ensemble. We expect that ERAP1 molecules transition throughout an energy landscape between open and closed low-energy states, as suggested by molecular dynamics simulations, and with the ensemble of molecules in solution populating these states and any intermediates therein in a dynamic equilibrium. Several lines of evidence support the dynamic nature of the conformational equilibrium, including the crystallographic trapping of ERAP1 in the open state with inhibitors like DG014 and bestatin despite predominance of the closed form in solution, and the allosteric promotion of enzymatic cycling by allosteric activators that stabilize the closed form. Another question relates to the relationship of domain closure to the catalytic cycle. Further studies will be required to elucidate the kinetics of domain closure, and whether these match the catalytic turnover rate, which is  $\sim 1\text{--}10\text{ s}^{-1}$  for optimal substrates under  $V_{\max}$  conditions<sup>8,20</sup>. ERAP1 can process peptides as long as 25 residues<sup>17,32</sup>, albeit inefficiently, and these would appear to be too long to fit completely in the substrate-binding chamber, particularly if they need to access any of the currently identified C-terminal binding sites. Potentially, such peptides could be accommodated with their C-termini outside of the site, exiting the chamber in the vicinity of Leu686 where helix 4 contacts helix 12, at the domain II/IV junction, as does substance P bound to the ERAP1 structural homolog APN (Supplementary Fig. 10). However, this binding mode would preclude complete domain closure without substantial reorientation of helix 4 and/or helix 12. Whether partial closure as observed for APN would be sufficient to induce active-site rearrangements allowing catalysis remains to be determined. A final caveat to the model presented here relates to the steep increase in ERAP1 processing rate observed as the peptide length

increases beyond 8 or 9 residues<sup>12,17</sup>. The stabilization of the closed conformation due to the accumulation of interdomain contacts mediated by bound peptide substrate also increases with peptide length beyond  $\sim 7$  residues (Fig. 3h), but the increase is more gradual. This would suggest that additional slow steps associated with domain closure remain to be identified, such as the ordering of interhelical loops disordered in the open form<sup>11,12</sup>. These would help to stabilize to closed conformer upon domain closure and synergize with C-terminal binding to promote the network of correlated residue motions<sup>21</sup> leading to the formation of the catalytically competent active-site configuration.

## Methods

**Mutagenesis and baculovirus production.** Sequence variants of ERAP1 were generated (inserted in pFastBac plasmid) using Agilent QuikChange II XL site-directed mutagenesis. Primers used for site-directed mutagenesis of ERAP1 are listed in Supplementary Table 4. All ERAP1 variants were validated by sequencing the entire open reading frame in the expression plasmid to confirm identity and lack of unexpected mutations. Positive clones were used to make bacmids and baculoviral stocks were prepared using SF9 insect cell line (CRL-1711, ATCC) following Bac-to-Bac baculoviral system protocol.

**Protein expression and purification.** High Five cells (BTI-Tn-5B1-4, CRL-10859, ATCC) were infected with baculoviral stocks and cultured for three days at 27 °C. Cells were then pelleted, and the supernatant was concentrated to  $\sim 100$  mL and buffer exchanged  $>100$ -fold into 50 mM Tris pH 8, 300 mM sodium chloride, 10 mM imidazole. Samples were bound to Ni-NTA-agarose resin, washed with 10 mM imidazole buffer, then washed with 20 mM imidazole buffer, then eluted with 100 mM imidazole buffer. Samples were then purified by anion-exchange chromatography and size-exclusion chromatography and stored at  $-80$  °C until use. The purity and integrity of all purified proteins were analyzed by SDS-PAGE with Coomassie Blue R250 staining (Supplementary Fig. S1b). For purification of C404S and C443S ERAP1 mutants, 1 mM DTT was kept through all the steps to prevent oxidation of the free cysteine. For activity assays of these mutants and wild-type ERAP1, a comparable amount of DTT was added to each sample.

## Chemical synthesis

**Synthesis of phosphinic peptide DG013.** The synthesis of phosphinic pseudopeptide DG013 has been described<sup>29</sup>.

**Synthesis of phosphinic peptide DG014.** Phosphinic pseudopeptide DG014 was synthesized by applying standard solid-phase peptide synthesis, on trityl alcohol lanterns (15  $\mu\text{mol}/\text{pin}$ ) using a Fmoc chemical protocol. A solution of acetyl chloride in dry dichloromethane (1:10 v/v) at room temperature was used to afford the trityl chloride lanterns. Attachment of the first amino-acid Fmoc-Tyr (tBu)-OH (30  $\mu\text{mol}/\text{pin}$ ) was performed by using *N,N*-diisopropylethylamine (18  $\mu\text{l}/\text{pin}$ ) in dry dichloromethane (0.4 mL/pin) at room temperature for 12 h<sup>45</sup>. The loading amount of Fmoc-Tyr (tBu)OH was evaluated to be 12  $\mu\text{mol}/\text{pin}$ , after cleavage from the polymer-support with 0.5% trifluoroacetic acid (TFA)/dichloromethane (room temperature, 1 h). Fmoc deprotection was performed with a solution of 20% piperidine in *N,N*-dimethylformamide over 1 h for each cycle of the synthesis. Fmoc protected amino acids (45  $\mu\text{mol}/\text{pin}$ ), 1-hydroxybenzotriazole (45  $\mu\text{mol}/\text{pin}$ ), and diisopropylcarbodiimide (45  $\mu\text{mol}/\text{pin}$ ) in dichloromethane/*N,N*-dimethylformamide (6/1) (0.4 mL/pin), were used for the coupling steps and each coupling reaction was allowed to proceed for 5 h. Coupling of the building block Boc-(R) hPhe[PO(OAd)CH<sub>2</sub>](R,S)LeuOH (23  $\mu\text{mol}/\text{pin}$ )<sup>29</sup> was performed using the coupling conditions described above (36  $\mu\text{mol}/\text{pin}$  of each reagent). Deprotection and removal of the final pseudodecapeptide from the solid support was accomplished by using a solution of TFA/dichloromethane/triisopropylsilane/H<sub>2</sub>O 39/58/2/1 for 2 h at room temperature. After concentration in vacuo, the crude product was precipitated in cold diethyl ether. DG014 was obtained after purification by analytical RP-HPLC and characterized by mass spectroscopy [ESMS *m/z* (*z* = 1): calcd for [C<sub>62</sub>H<sub>90</sub>N<sub>9</sub>O<sub>18</sub>P + H]<sup>+</sup> 1281.4; found: 1281.5].

**Synthesis of phosphinic peptide DG023.** The phosphinic pseudoundecapeptide DG023 was prepared by conventional solid-phase peptide synthesis, on Rink amide lanterns (8  $\mu\text{mol}/\text{pin}$ ), using the Fmoc strategy. Fmoc deprotection and amino-acid coupling steps were performed as described above for DG014. After the introduction of Fmoc-Bpa-OH all synthetic steps were performed in light-protected conditions. For the introduction of phosphinic pseudodipeptidic sequence, the building block Boc-(R)Phe[PO(OAd)CH<sub>2</sub>](S)LeuOH was synthesized in three steps starting from the R-stereoisomer of the Boc-protected aminophosphinic analog of phenylalanine<sup>46,47</sup>. The phosphinic pseudodipeptide Boc (R)Phe[PO(OH)CH<sub>2</sub>](S)LeuOEt was prepared as previously described<sup>48</sup>, and obtained in a stereochemically pure form after two recrystallizations with AcOEt. Subsequent adamantylation of the phosphinic group and saponification of the C-terminal ethyl

ester group afforded the final building block Boc (R)Phe[PO (OAd)-CH<sub>2</sub>] (S) LeuOH<sup>48</sup>, which was incorporated in the last step of the solid-phase synthesis. For the coupling of the aforementioned building block, 16 μmol/pin were used by using standard coupling conditions. The final pseudoundecapeptide was cleaved and deprotected from the solid support in presence of a solution of trifluoroacetic acid (TFA)/H<sub>2</sub>O/trisopropylsilane 95/2.5/2.5 over 2 h at room temperature. The solution of deprotected peptide was concentrated in vacuo and the residue was treated with cold dry diethyl ether. DG023 was obtained after purification by analytical RP-HPLC and characterized by mass spectroscopy [ESMS *m/z* (*z* = 1): calculated for [C<sub>79</sub>H<sub>104</sub>N<sub>17</sub>O<sub>14</sub>P + H]<sup>+</sup> 1546.8; found: 1546.9].

**Small-angle X-ray scattering.** Homogeneity of all ERAP1 samples used for SAXS studies was assessed the day before analysis by size-exclusion chromatography and SDS-PAGE. Samples were then stored at 4 °C until data collection. Samples used for SAXS studies were stored after data collection, and purity and integrity were confirmed a day after the data collection by SDS-PAGE. Just prior to analysis, purified protein samples were mixed with inhibitor, if indicated, and concentrated in rinsed Centricon 10 kDa MWCO 0.5 mL centrifugal concentrators equilibrated in 50 mM HEPES pH 7.5, 200 mM NaCl, 0.02% (w/v) NaN<sub>3</sub>. Concentrator retentate and flow through were stored and used as sample and buffer, respectively, during SAXS data collection. SAXS data were collected for each sample at three concentrations (generally 4 mg/mL, 2 mg/mL, and 1 mg/mL). Small-molecule additives were included at the following concentrations: 2 mM bestatin, 30 μM leucinethiol with 30 μM dithiothreitol, 120 μM SIINFEKL peptide, 100 μM DG013, 200 μM compound 3, 200 μM compound 2. Scattering curves were buffer subtracted using matched buffer-scattering curves. The concentration series was then compared, scaled, and merged manually using PRIMUS (in ATSAS v2.5.2) and SCATTER (version 3.0.g). Merged SAXS curve *R<sub>g</sub>* was calculated using AutoRG. Minimal ensemble search was performed using the FOXS webserver<sup>49</sup> with data fit to single-structural models of ERAP1.

**ERAP1 structural modeling.** Structural models were generated by two methods. For models derived from crystal structures (PDB ID 2YD0, and 3MDJ chains A/B/C), missing portions of polypeptide and complete high mannose N-glycans were added by rounds of simulation using the AllosMod webserver<sup>50</sup>. Alternately, structural models were sampled from a prior molecular dynamics simulation of ERAP1<sup>28</sup>.

**Crystallization.** Purified protein samples were prepared in 10 mM Tris or 10 mM HEPES buffer. Crystal trials were set up and hits were optimized. For DG013-bound ERAP1 cocrystal, final crystallization conditions were: 1.3 M ammonium sulfate, 100 mM MES pH5.5, 15 mg/mL ERAP1 preincubated with saturating DG013, grown at 25 °C, dehydrated by changing well solution to 2 M ammonium sulfate one day prior to looping, cryoprotected with lithium sulfate. For DG014-bound ERAP1 cocrystal, final crystallization conditions were: 15% PEG8000, 100 mM Tris pH 8.60, 3% D-sorbitol, 7.5 mg/mL ERAP1 preincubated with saturating DG014, grown at 4 °C, cryoprotected with ethylene glycol crystals were mounted on loops and frozen by plunging in liquid nitrogen in preparation for data collection.

**X-ray diffraction data collection and structure determination.** Crystal diffraction data were collected in a 180° arc about a static crystal (PDB ID 6MGQ) or a 360° arc about a crystal transiting along a defined vector (PDB ID 6M8P). X-ray wavelengths used were 1.110 Å (PDB ID 6MGQ) and 0.979 Å (PDB ID 6M8P). Diffraction data were collected at 100 K. X-ray diffraction data were integrated using XDS version June 1, 2017<sup>51</sup> with anisotropy assessed using STARANISO<sup>52</sup>. STARANISO assigns an anisotropic resolution cutoff, to account for different diffraction intensities in different directions in reciprocal space. For ERAP1-DG014 (PDB ID 6MGQ), anisotropy was minimal, and resolution limits in the best (2.9 Å along 0.40 a\* + 0.977 b\* + 0.211 c\*) and worst (3.5 Å along 0.427 a\* + 0.885 b\* + 0.183 c\*) directions were similar. For ERAP1-DG013 (PDB ID 6M8P), anisotropy was substantial, with resolution limits in the best (3.3 Å along 0.262 a\* + 0.965 b\*) and worst (7.4 Å along 0.998 a\* + 0.062 c\*) directions substantially different. Table 1 reports overall and last-shell completeness for both conventional spherical cutoffs and also for the elliptical cutoffs used for refinement and map interpretation. Space groups were tentatively assigned using observed point group symmetry and patterns of systemic absences and then confirmed by molecular replacement trials in all possible space groups consistent with lattice symmetry using Phaser<sup>53</sup>, and by refinement tests in all potential subgroups and supergroups using Zanuda<sup>54</sup>. Resolution limits were assigned using a CC\* cutoff of ~0.5<sup>55</sup>. Ligand occupancy was confirmed by unconstrained occupancy refinement for each molecule in the asymmetric unit, and by examination of omit maps<sup>56</sup>.

ERAP1-DG014 crystallographic space group and unit cell parameters were similar to a previously determined crystal structure for ERAP1 in the open conformation bound to bestatin (PDB ID 3MDJ)<sup>12</sup>, and the structure was determined by molecular replacement using those coordinates, searching separately for domains I + I + III and domain IV. The protein formed a trimer in the crystallographic asymmetric unit (Supplementary Fig. S3g, h), with domain organization slightly more open than for 3MDJ. The phosphinic peptide inhibitor

DG014 chemical structure was generated using the PRODRG webserver<sup>57</sup>. Restraints were then generated using eLBOW<sup>58</sup>, and models were built through cycles of refinement using Phenix<sup>53</sup> and manual rebuilding using Coot<sup>59</sup>. Omit maps were generated using Polder<sup>56</sup>. Electron density for the bound DG014 phosphinic peptide was very weak beyond the proline at residue 5. Additional density potentially corresponding to approximately two additional residues was observed 14 Å away in the vicinity of ERAP1 residues F405, D406, F674, Q675, N678, and M726. Full-length 10-residue models for DG014 with reasonable geometry would extend into this region, but the observed electron density did not support confident placement and refinement of a full-length peptide model. In the deposited PDB file, an unknown ligand record (UNL) was used to mark this location.

ERAP1-DG013 adopted a very large crystallographic unit cell, with Matthews coefficient consistent with up to 41 molecules of ERAP1 per asymmetric unit. Molecular replacement trials using ERAP1 open, closed, and isolated domain search models revealed strong solutions for 22 molecules of ERAP1 in the closed conformation, in an unusual arrangement of two stacked 11-membered rings (Supplementary Fig. S3b–d). *Z* scores from these solutions ranged from 27 (log-likelihood gain LLG = 632) for the 1st solution to 104 (LLG = 1876) for the second solution; the mean for all 22 solutions was 48 ± 19 (s.d.), with a final LLG of 92875. No other significant solutions were observed. The stacked rings aligned with their 11-fold axes of symmetry parallel to the unit cell *a* axis, and formed an open lattice (Supplementary Fig. S3). Only 16 of the 22 molecules in the asymmetric unit (8 from each stacked ring) participated in contacts in the *b*–*c* plane (Supplementary Fig. S3). The high solvent content (73%) and relatively low degree of intermolecular contacts resulting from this packing can in part help to explain the very weak diffraction observed for this crystal (overall *I*σ(*I*) = 2.68, cf. *I*σ(*I*) = 20.11 for ERAP1-DG014). Despite the weak diffraction, clear omit map electron density was observed for each of 22 ERAP1 monomers (Supplementary Fig. S4). The 22-fold averaged map was used for the initial building of DG013 and rebuilding of the ERAP1 active site. Twenty-two copies of this complex were then placed appropriately in the asymmetric unit for cycles of refinement and rebuilding. Crystallographic refinements with 22-fold non-crystallographic restraints and conservative geometry resulted in a final refined model with working and free *R*-factors, RMS deviations from ideal geometry, and main-chain and side-chain torsions and clash scores (Table 1) well within the range of PDB entries with the comparable resolution, as assessed by MolProbity<sup>60</sup>. Final refined occupancies for each of the bound DG013 monomers were near unity (Supplementary Fig. S4). Refined atomic *B*-factors were relatively low (<*B* > 39.5 Å<sup>2</sup>), perhaps surprising for a crystal with such a low-resolution limit and a high degree of anisotropy, but were consistent with the low observed Wilson *B*-factor (29.6 Å<sup>2</sup>), and demonstrate the utility of TLS refinement<sup>61</sup> in modeling anisotropic diffraction. The resolution limit of 3.3 Å was confirmed using a paired refinement test<sup>55,62</sup>; however, the effective resolution will be somewhat lower because of the anisotropic diffraction and loss of data beyond the elliptical resolution cutoff. Because of the relatively low resolution and weak diffraction observed for this crystal, we limited interpretation to localization of the bound DG013 ligand to the active site.

**Activity assays.** ERAP1 activity was measured consistently in 96-well format, 100 μL reactions containing 20 mM Tris pH 7.5, 100 mM NaCl, 0.02% sodium azide, with 0.1% (w/v) bovine specific antigen (BSA) present to block ERAP1 binding to plastic (observed even for “non-binding” polypropylene plates). For peptide hydrolysis, ERAP1 (0.2 ng/μL) was mixed with inhibitor (25 μM) and then incubated with substrate peptide (10 μM) for 10 min. Reactions were stopped by addition of trifluoroacetic acid, 0.4% (v/v) final concentration. Plates were frozen and shipped to PureHoney Technologies for quantification of substrate and product by LC-MS and summing area-under-curve of respective peaks. For leucine-amidomethylcoumarin (Leu-AMC) hydrolysis, ERAP1 (2 ng/μL) was mixed with inhibitor if indicated and reactions were begun by addition of Leu-AMC substrate (100 μM). Product formation (7-amino-4-methylcoumarin) was quantified by measuring the change in fluorescence (380 nm excitation/460 nm emission) over 20 min at using BMG POLARstar OPTIMA. For measurements of peptide hydrolysis using coupled enzyme assay, 50 μL ERAP1 (400 ng/well) was mixed with the 25 μL substrate LF9 peptide (LVAFKARKP)<sup>42</sup> at a series of concentrations along with 25 μL of coupling reagent to start the reaction as described before<sup>20</sup>. After starting the reaction, absorbance at 405 nm was measured once every 2.5 min for 20 min. Reaction slopes were quantified by calculating the absorbance change over the time course, which was linear under these conditions. Michaelis–Menten analysis was performed by fitting the slopes as a function of substrate concentration in standard Michaelis–Menten equation using GraphPad Prism.

**Photocrosslinking and mass spectrometry.** We performed several optimization and control experiments to establish appropriate conditions for photo-cross-linking. ERAP1 DG023-inhibition studies were used to estimate the saturating concentrations for the cross-linking assay. IC<sub>50</sub> for DG023 inhibition of Leu-AMC hydrolysis by ERAP1 was observed at 15 nM, similar to the other phosphinic peptides DG013 and DG014 (Supplementary Fig. S5). For the cross-linking experiment, 100 μM DG023 was used with 20 μM ERAP1; under these conditions, 99.4% inhibition of ERAP1 should be obtained. To test if BPA at the C-terminus of the peptide substrate interferes with the binding to ERAP1, we followed the

hydrolysis of a non-phosphinic peptide substrate based on DG023 (LLKHHAFSFK-BPA). ERAAP hydrolysis was similar to a known peptide substrate, LF9 (LVAFKARKE)<sup>42</sup>, (Supplementary Fig. S8d), indicating that BPA at the C-terminus did not interfere with ERAAP binding. In preliminary experiments, different irradiation times (5, 10, 15, 20, and 30 min) were tested, with optimum cross-linking observed at 30 min. To evaluate any effects of UV exposure on the structure or activity of ERAAP, L-pNA hydrolysis activity was tested before and after 30 min of irradiation, and no difference was noted (Supplementary Fig. S8c). A cross-linking experiment performed with BPA alone instead of DG023 did not identify any adducts, indicating the absence of substantial non-specific modification by this cross-linking reagent. Purified ERAAP was incubated in a V-bottom 96-well plate in the absence of any ligand, in the presence of BPA alone, or in the presence of DG023 cross-linking peptide inhibitor. Samples were irradiated for 30 min on ice using a long-wavelength UV lamp (Blak Ray B100AP/R, UVP), and then denatured, deglycosylated enzymatically, and separated by SDS-PAGE. Excised gel bands were trypsin-digested and analyzed by LC-MS/MS at the UMass Proteomic Core Facility. Crosslinks were detected by searching the SwissProt human proteome database (Uniprot\_human proteome [<https://www.uniprot.org/proteomes/UP000005640>]) using Sequest (IseNode from Proteome Discoverer 2.1.1.21, fragment ion mass tolerance of 0.050 Da, parent ion tolerance of 10.0 PPM) for a + 268 Da modification on leucine, serine, tryptophan, or methionine. One peptide identified using these criteria with *m/z* consistent with a + 268 modification of an ERAAP tryptic peptide was excluded from analysis as it appeared in the dataset generated in the absence of DG023 or BPA. Cross-linked peptides were identified with 75% peptide confidence or greater, and 2.74% propeptide FDR (spectra). Data analysis and figure preparation were performed using Scaffold 4<sup>63</sup>.

**Electrostatic potential calculation.** Structural models of ERAAP domains II–IV (lacking domain I due to limits on the number of atoms allowed) were prepared for analysis using PDB2PQR, which calculated atomic charges using AMBER force field<sup>64</sup>. The output of PDB2PQR was then input to the DelPhiForce webserver<sup>65</sup> to calculate the pairwise electrostatic potential. Settings used were: pairwise interactions mode, Gaussian dielectric distribution method, 3.0 grids/Å, 0.2 M salt concentration. To test the effect of substitution at position 528, computational mutagenesis was performed on existing structural models, using the respective residue rotamers found in the alternate ERAAP crystal structure.

**Structural modeling and analysis.** Previously determined crystal structures of ERAAP bound to 10-mer and 15-mer phosphinic peptides were observed in the closed conformation (6RQX and 6RYF). DG023 peptides with C-terminal BPA crosslinker bound to L677, L686, and L838 were modeled into the closed ERAAP conformation (2YD0) using Coot<sup>59</sup>. For solvent-accessible surface area calculations, open-conformer models with bound peptides were generated by superimposing peptide bound closed ERAAP with open ERAAP (3MDJ) using Pymol<sup>66</sup> using the pair fit function, aligning the  $\alpha$  carbons of ERAAP domains. Long peptide substrates and Zn were considered to be part of domains I, II, III. Glycans and all other ligands and buffer components were removed from the crystal structures. The PISA server<sup>67</sup> was used for surface area and solvation energy calculations, and for the identification of interdomain contacts. ERAAP was divided into two different chains, where domains I, II, and III were considered to be one chain and domain IV the other chain. Residues identified in interdomain contacts for the closed (2YD0) but not open (3MDJ) conformation are shown in Supplementary Table II. Additional residues involved in interdomain contacts in the various long peptide structures and models are shown in Supplementary Table III.

**Statistics.** For SAXS analysis, radii-of-gyration are presented as means of each curve fit with the respective fit error plotted. Fitting of SAXS data with structural models used chi-square to quantify goodness-of-fit as well as model/data ratios to qualitatively demonstrate residual deviation. Unpaired *t* tests were used to compare replicate experiments. For statistical analysis of EC<sub>50</sub>, sigmoidal curves were fit to eight concentrations of inhibitor, in triplicate. LogEC<sub>50</sub>, standard error of measurement, and the number of points (24) from two experiments were analyzed for statistical significance using two-tailed ANOVA.

**Figure design.** Structural figures were prepared using Pymol<sup>66</sup>. Graphical data were prepared with GraphPad Prism 7.0c<sup>68</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The crystallographic diffraction data and atomic coordinates generated in this study have been deposited in the world-wide Protein Data Bank (wwPDB) under accession codes 6MGQ and 6M8P. The raw crystallographic datasets have been deposited in the SBGRID repository under accession codes 605 and 606. The other X-ray crystal structure coordinates used in this study are available from in wwPDB under accession codes 3MDJ, 3QNF, 2YD0, 6RQX, and 6RYF. The human proteome sequences used for mass

spectrometry data analysis are available from the Uniprot database under accession code 5640. Source data are provided with this paper.

Received: 16 December 2020; Accepted: 10 August 2021;

Published online: 06 September 2021

## References

- Nagarajan, N. A. et al. ERAAP shapes the peptidome associated with classical and nonclassical MHC class I molecules. *J. Immunol.* **197**, 1035–1043 (2016).
- Mehta, A. M. et al. Single nucleotide polymorphisms in antigen processing machinery component ERAAP significantly associate with clinical outcome in cervical carcinoma. *Genes, Chromosomes Cancer* **48**, 410–418 (2009).
- Roberts, A. R. et al. ERAAP association with ankylosing spondylitis is attributable to common genotypes rather than rare haplotype combinations. *Proc. Natl Acad. Sci. USA* **114**, 558–561 (2017).
- Zhang, Z. et al. Association of HLA-B27 and ERAAP with ankylosing spondylitis susceptibility in Beijing Han Chinese. *Tissue Antigens* **83**, 324–329 (2014).
- Chen, B., Li, D. & Xu, W. Association of ankylosing spondylitis with HLA-B27 and ERAAP: pathogenic role of antigenic peptide. *Med. Hypotheses* **80**, 36–38 (2013).
- Sanz-Bravo, A. et al. Ranking the contribution of ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 1 (ERAAP) polymorphisms to shaping the HLA-B\*27 peptidome. *Mol. Cell. Proteom.: MCP* **17**, 1308–1323 (2018).
- Sanz-Bravo, A., Campos, J., Mazariegos, M. S. & Lopez de Castro, J. A. Dominant role of the ERAAP polymorphism R528K in shaping the HLA-B27 Peptidome through differential processing determined by multiple peptide residues. *Arthritis Rheumatol.* **67**, 692–701 (2015).
- Evnouchidou, I. et al. 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 (2011).
- Drinkwater, N., Lee, J., Yang, W., Malcolm, T. R. & McGowan, S. M1 aminopeptidases as drug targets: broad applications or therapeutic niche? *FEBS J.* **284**, 1473–1488 (2017).
- Matthews, B. W. et al. Structure of thermolysin. *Nat.: N. Biol.* **238**, 41–43 (1972).
- Kochan, G. et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAAP) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl Acad. Sci. USA* **108**, 7745–7750 (2011).
- Nguyen, T. T. et al. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAAP. *Nat. Struct. Mol. Biol.* **18**, 604–613 (2011).
- Mpakali, A. et al. Crystal structure of insulin-regulated aminopeptidase with bound substrate analogue provides insight on antigenic epitope precursor recognition and processing. *J. Immunol.* **195**, 2842–2851 (2015).
- Hermans, S. J. et al. Crystal structure of human insulin-regulated aminopeptidase with specificity for cyclic peptides. *Protein Sci.: a Publ. Protein Soc.* **24**, 190–199 (2015).
- Mpakali, A. et al. Ligand-induced conformational change of insulin-regulated aminopeptidase: insights on catalytic mechanism and active site plasticity. *J. Medicinal Chem.* **60**, 2963–2972 (2017).
- York, I. A. et al. The ER aminopeptidase ERAAP enhances or limits antigen presentation by trimming epitopes to 8–9 residues. *Nat. Immunol.* **3**, 1177–1184 (2002).
- Chang, S. C., Momburg, F., Bhutani, N. & Goldberg, A. L. The ER aminopeptidase, ERAAP, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proc. Natl Acad. Sci. USA* **102**, 17107–17112 (2005).
- Stamogiannos, A., Koumantou, D., Papakyriakou, A. & Stratikos, E. Effects of polymorphic variation on the mechanism of endoplasmic reticulum aminopeptidase 1. *Mol. Immunol.* **67**, 426–435 (2015).
- Liddle, J. et al. Targeting the regulatory site of ER aminopeptidase 1 leads to the discovery of a natural product modulator of antigen presentation. *J. Med. Chem.* **63**, 3348–3358 (2020).
- Maben, Z. et al. Discovery of selective inhibitors of endoplasmic reticulum aminopeptidase 1. *J. Med. Chem.* **63**, 103–121 (2020).
- Giastas, P. et al. Mechanism for antigenic peptide selection by endoplasmic reticulum aminopeptidase 1. *Proc. Natl Acad. Sci. USA* **116**, 26709–26716 (2019).
- Sui, L., Gandhi, A. & Guo, H. C. Crystal structure of a polypeptide’s C-terminus in complex with the regulatory domain of ER aminopeptidase 1. *Mol. Immunol.* **80**, 41–49 (2016).

23. Umezawa, H., Aoyagi, T., Suda, H., Hamada, M. & Takeuchi, T. Bestatin, an inhibitor of aminopeptidase B, produced by actinomycetes. *J. Antibiotics* **29**, 97–99 (1976).
24. Chan, W. W. L-leucinthiol—a potent inhibitor of leucine aminopeptidase. *Biochemical Biophysical Res. Commun.* **116**, 297–302 (1983).
25. Wilkes, S. H. & Prescott, J. M. The slow, tight binding of bestatin and amastatin to aminopeptidases. *J. Biol. Chem.* **260**, 13154–13162 (1985).
26. York, I. A., Brehm, M. A., Zendzian, S., Towne, C. F. & Rock, K. L. Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance. *Proc. Natl Acad. Sci. USA* **103**, 9202–9207 (2006).
27. Yang, L. et al. Solution scattering at the life science X-ray scattering (LiX) beamline. *J. Synchrotron Radiat.* **27**, 804–812 (2020).
28. Stamogiannos, A. et al. Critical role of interdomain interactions in the conformational change and catalytic mechanism of endoplasmic reticulum aminopeptidase 1. *Biochemistry* **56**, 1546–1558 (2017).
29. Zervoudi, E. et al. Rationally designed inhibitor targeting antigen-trimming aminopeptidases enhances antigen presentation and cytotoxic T-cell responses. *Proc. Natl Acad. Sci. USA* **110**, 19890–19895 (2013).
30. Giastas, P., Neu, M., Rowland, P. & Stratikos, E. High-resolution crystal structure of endoplasmic reticulum aminopeptidase 1 with bound phosphinic transition-state analogue inhibitor. *ACS Med. Chem. Lett.* **10**, 708–713 (2019).
31. 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.* **15**, 920–929 (2008).
32. Mavridis, G. et al. A systematic re-examination of processing of MHCI-bound antigenic peptide precursors by endoplasmic reticulum aminopeptidase 1. *J. Biol. Chem.* **295**, 7193–7210 (2020).
33. Kauer, J. C., Erickson-Viitanen, S., Wolfe, H. R. Jr. & DeGrado, W. F. p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J. Biol. Chem.* **261**, 10695–10700 (1986).
34. Gandhi, A., Lakshminarasimhan, D., Sun, Y. & Guo, H. C. Structural insights into the molecular ruler mechanism of the endoplasmic reticulum aminopeptidase ERAP1. *Sci. Rep.* **1**, 186 (2011).
35. Chen, L., Lin, Y. L., Peng, G. & Li, F. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc. Natl Acad. Sci. USA* **109**, 17966–17971 (2012).
36. Krissinel, E. & Henrick, K. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. *J. Mol. Biol.* **372**, 774–797 (2007).
37. Evans, D. M. et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* **43**, 761–767 (2011).
38. Bettencourt, B. F. et al. Protective effect of an ERAP1 haplotype in ankylosing spondylitis: investigating non-MHC genes in HLA-B27-positive individuals. *Rheumatology* **52**, 2168–2176 (2013).
39. Costantino, F. et al. ERAP1 gene expression is influenced by nonsynonymous polymorphisms associated with predisposition to spondyloarthritis. *Arthritis Rheumatol.* **67**, 1525–1534 (2015).
40. Stratikos, E. & Stern, L. J. Antigenic peptide trimming by ER aminopeptidases—insights from structural studies. *Mol. Immunol.* **55**, 212–219 (2013).
41. Goto, Y., Hattori, A., Ishii, Y. & Tsujimoto, M. Reduced activity of the hypertension-associated Lys528Arg mutant of human adipocyte-derived leucine aminopeptidase (A-LAP)/ER-aminopeptidase-1. *FEBS Lett.* **580**, 1833–1838 (2006).
42. Evnouchidou, I. et al. The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS One* **3**, e3658 (2008).
43. Ombrello, M. J., Kastner, D. L. & Remmers, E. F. Endoplasmic reticulum-associated amino-peptidase 1 and rheumatic disease: genetics. *Curr. Opin. Rheumatol.* **27**, 349–356 (2015).
44. Grawert, T. W. & Svergun, D. I. Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.* **432**, 3078–3092 (2020).
45. Doi, T., Numajiri, Y., Takahashi, T., Takagi, M. & Shin-ya, K. Solid-phase total synthesis of (-)-apratoxin A and its analogues and their biological evaluation. *Chem., Asian J.* **6**, 180–188 (2011).
46. Georgiadis, D., Matziari, M., Vassiliou, S., Dive, V. & Yiotakis, A. A convenient method to synthesize phosphinic peptides containing an aspartyl or glutamyl aminophosphinic acid. Use of the phenyl group as the carboxyl synthon. *Tetrahedron* **55**, 14635–14648 (1999).
47. Baylis, E. K., Campbell, C. D. & Dingwall, J. G. 1-Aminoalkylphosphonous acids. Part 1. Isosteres of the protein amino acids. *J. Chem. Soc., Perkin Transact.* **1**, 2845–2853 (1984).
48. Yiotakis, A., Vassiliou, S., Jiracek, J. & Dive, V. Protection of the hydroxyphosphinyl function of phosphinic dipeptides by adamantyl. Application to the solid-phase synthesis of phosphinic peptides. *J. Org. Chem.* **61**, 6601–6605 (1996).
49. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
50. Weinkam, P., Pons, J. & Sali, A. Structure-based model of allostery predicts coupling between distant sites. *Proc. Natl Acad. Sci. USA* **109**, 4875–4880 (2012).
51. Kabsch, W. Xds. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 125–132 (2010).
52. Tickle, I. J. et al. STARANISO (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>). (Cambridge, United Kingdom: Global Phasing Ltd., 2018)
53. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 213–221 (2010).
54. Lebedev, A. A. & Isupov, M. N. Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr. D, Biol. Crystallogr.* **70**, 2430–2443 (2014).
55. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
56. Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D, Struct. Biol.* **73**, 148–157 (2017).
57. Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **60**, 1355–1363 (2004).
58. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **65**, 1074–1080 (2009).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 486–501 (2010).
60. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
61. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
62. Maly, M., Diederichs, K., Dohnalek, J. & Kolenko, P. Paired refinement under the control of PAIREF. *IUCr* **7**, 681–692 (2020).
63. Scaffold Suite Application Software Version 4.8.6. (Proteome Software Inc., 2018).
64. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
65. Li, L. et al. DelPhiForce web server: electrostatic forces and energy calculations and visualization. *Bioinformatics* **33**, 3661–3663 (2017).
66. Schrodinger, L. L. C. The PyMOL Molecular Graphics System, Version 2.0. (Schrodinger, LLC, 2015).
67. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
68. Prism 7 for Mac OS X. (GraphPad Software Inc., 2017).

## Acknowledgements

Supported by NIH grant AI153828 (L.J.S.). Some SAXS data were collected at the SIBYLS beamline at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. Some SAXS data were collected at the LiX beamline, and one X-ray diffraction dataset was collected at beamline 17-ID-2 (FMX). These beamlines are both part of the Life Science Biomedical Technology Research resource, primarily supported by the National Institute of Health, National Institute of General Medical Sciences (NIGMS) under grant P41 GM111244, and by the DOE Office of Biological and Environmental Research under grant KP1605010, with additional support from NIH under grant S10 OD012331. As a National Synchrotron Light Source II facility resource at Brookhaven National Laboratory, work performed at the LSBR is supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences Program under contract number DE-SC0012704. One X-ray diffraction dataset was collected at beamline X25 at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the United States Department of Energy, Division of Materials Sciences and Division of Chemical Sciences under contract DE-AC02-98CH10886. Proteomics data were obtained at the UMass Medical School Mass Spectrometry Facility. Molecular models of ERAP1 from molecular dynamics simulations were kindly provided by Athanasios Papakyriakou. We thank Scott Shaffer and Nadia Sultana for their assistance with proteomics.

### Author contributions

Z.M. and R.A. designed, performed, and analyzed experiments, and wrote the manuscript. D.G. designed and synthesized phosphinic inhibitors DG014 and DG023. E.S. and L.J.S. designed and analyzed experiments and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25564-w>.

**Correspondence** and requests for materials should be addressed to L.J.S.

**Peer review information** *Nature Communications* thanks Robert Hettich, David Margulies and other, anonymous, reviewers for their contributions to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021