High-resolution crystal structure of ERAP1 with bound phosphinic transition-state analogue inhibitor

Petros Giastas[1], Margarete Neu[2,] Paul Rowland[2], and Efstratios Stratikos[1]

1 National Center for Scientific Research Demokritos, Agia Paraskevi, Athens 15310, Greece

2 Medicinal Science and Technology, GlaxoSmithKline, Stevenage, Hertfordshire SG1 2NY, UK

Materials and Methods

Protein construct

The protein sequence used is shown below and includes a 25 amino acid substitution of the exon 10 loop by the linker GSG (in bold) as well as C-terminal 10-His purification tag (in bold) next to a TEV recognition site (underlined, TEV cleavage bond is indicated by a dash). The final construct has a length of 922 amino acids and a MW of 105171 kDa. The codon optimised ERAP1 gene was synthesised and sub-cloned into a pFASTBAC1 vector using BamH1 and XhoI to generate pFASTBAC1-ERAP1 (1-941) Δ486-513 GSG insert, 528K-Tev-10His.

MVFLPLKWSLATMSFLLSSLLALLTVSTPSWCQSTEASPKRSDGTPFPWNKIRLPEY
VIPVHYDLLIHANLTTLTFWGTTKVEITASQPTSTIILHSHHLQISRATLRKGAGER
LSEEPLQVLEHPRQEQIALLAPEPLLVGLPYTVVIHYAGNLSETFHGFYKSTYRTKE
GELRILASTQFEPTAARMAFPCFDEPAFKASFSIKIRREPRHLAISNMPLVKSVTVA
EGLIEDHFDVTVKMSTYLVAFIISDFESVSKITKSGVKVSVYAVPDKINQADYALDA
AVTLLEFYEDYFSIPYPLPKQDLAAIPDFQSGAMENWGLTTYRESALLFDAEKSSAS
SKLGITMTVAHELAHQWFGNLVTMEWWNDLWLNEGFAKFMEFVSVSVTHPELKVGDY
FFGKCFDAMEVDALNSSHPVSTPVENPAQIREMFDDVSYDKGACILNMLREYLSADA
FKSGIVQYLQKHSYKNTKNEDLWDSMASI**GSG**GVDVKTMMNTWTLQKGFPLITITVR
GRNVHMKQEHYMKGSDGAPDTGYLWHVPLTFITSKSDMVHRFLLKTKTDVLILPEEV
EWIKFNVGMNGYYIVHYEDDGWDSLTGLLKGTHTAVSSNDRASLINNAFQLVSIGKL
SIEKALDLSLYLKHETEIMPVFQGLNELIPMYKLMEKRDMNEVETQFKAFLIRLLRD
LIDKQTWTDEGSVSERMLRSQLLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVD
VTLAVFAVGAQSTEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQWLLDESFK
GDKIKTQEFPQILTLIGRNPVGYPLAWQFLRKNWNKLVQKFELGSSSIAHMVMGTTN
QFSTRTRLEEVKGFFSSLKENGSQLRCVQQTIETIEENIGWMDKNFDKIRVWLQSEK
LERM<u>ENLYFQ</u>-*G**HHHHHHHHHH***

Protein expression and purification

ERAP1 was isolated from the supernatant of 10L Hi5 cells after infection with recombinant baculovirus carrying the ERAP1 gene. The cell medium was harvested by centrifugation (4500rm, 4°C, 45 min) and concentrated to 2L using a Sartorius centrifugal ultrafiltration device with a 30 kDa MWCO Hydrosart membrane. The concentrate was adjusted to contain 20mM Tris pH 7.5, 300mM NaCl and 1mM $NiSO_4$. 5ml Ni-NTA resin was equilibrated with 50mM HEPES, 500mM NaCl, 5% glycerol, pH 7.5, 0.5mM TCEP and added to the concentrated supernatant. The slurry was incubated for 2hrs at 4°C and then collected in a gravity column, washed with 50mM HEPES, 500mM NaCl, 5% glycerol, pH 7.5,

0.5mM TCEP and eluted with 50mM HEPES, 500mM NaCl, 5% glycerol, pH 7.5, 500mM Imidazole, 0.5mM TCEP. The ERAP1 containing eluate was dialysed (MWCO 3500 Da) overnight at 4°C against 1L of 10mM Tris, 150mM NaCl, 5% Glycerol, pH 8.0, 0.5mM TCEP in the presence of TEV protease with a molar ratio of 1:20 (TEV:ERAP1). After digestion, the dialysate was concentrated to 3.5 ml and further purified by size exclusion chromatography on a Superdex S200pg column (26x900mm) equilibrated with 10mM HEPES buffer pH 7.0 and 100mM NaCl. The ERAP1 containing fractions were pooled and concentrated to 12.7 mg/ml.

Enzymatic assays

The activity of ERAP1 was measured by a fluorigenic assay as described in (Stamogiannos, 2017) on a TECAN infiniteM200 microplate fluorescence reader. Determination of inhibitor IC_{50} was performed using the same assay as described (Stamogiannos, 2017).

Crystallization and data collection

Crystals of ERAP1 in complex with DG046 were grown by the sitting-drop vapor-diffusion method in 100 mM DL-malic acid, Bis-tris propane, pH 7.0, 25% PEG 1500 with a protein concentration of 12 mg ml^{-1} and a molar ratio of protein to DG46 equal to 1:5. Cryoprotection of the crystals was achieved by immersing them in a solution containing the precipitant and 20% ethylene glycol for a few seconds and subsequently, crystals were flash frozen in liquid N2. Data were collected at 100 K at a wavelength of 1.0 Å on beamline P13 of PETRA III at EMBL, Hamburg, Germany. The reflections were integrated with XDS (Kabsch, 2010)), the space group was determined with POINTLESS (Evans, 2011), and the data merging was carried out with SCALA (Evans, 2006) of the CCP4 (Winn et al., 2011) suite.

Structure determination and refinement

The structure was solved by molecular replacement with PHASER (McCoy et al., 2007) using as a search model the closed structure of ERAP1 (PDB ID 2YD0) (Kochan et al., 2011)). Inspection of the initial electron density maps showed clearly the presence of a large and continuous electron density in the binding site region, which was subsequently attributed to DG46. Structure refinement was carried out with PHENIX (Afonine et al., 2012) using restrained refinement and at the final stages was implemented TLS refinement. Model building and real-space refinement were performed in COOT (Emsley et al, 2010). The high-resolution limit was determined with the CC1/2 and mean(I/σI) criteria (Karplus and Diederichs, 2012), using as cutoff the values of 50 % and 1.5, respectively. The overall geometry of the refined model was very good, with 99.5 % of the residues lying in Ramachandran favored or allowed regions, no peptide omega angle outliers and 0.6 % of the residues being in unexpected rotameric state, however, justifiable by the electron density map. The electron density for the regions 110-113 and 553-557, belonging to loops connecting secondary structure elements as well as the density for the Gly-Ser-Gly loop that substituted the exon 10 loop, could not be determined and therefore the corresponding residues were not built in the

model. The program PyMOL (http://www.pymol.org/) was used for structure visualization and for the generation of figures.

References

- Kabsch, W., *Acta Cryst.* D66, 125-132 (2010). DOI: 10.1107/S0907444909047374
- Evans, P.R., *Acta Cryst.* D67, 282-292 (2011). DOI: 10.1107/S090744491003982X
- Evans, P.R., *Acta Cryst.* D62, 72-82 (2006). DOI: 10.1107/S0907444905036693
- Winn, M.D., et al. *Acta. Cryst.* D67, 235-242 (2011). DOI: 10.1107/S0907444910045749
- McCoy, A.J., et al. *J. Appl. Cryst.* 40, 658-674 (2007). DOI: 10.1107/S0021889807021206
- Kochan, G., et al., *Proc Natl Acad Sci* USA.108(19) 7745-50 (2011). DOI: 10.1073/pnas.1101262108
- Afonine, P.V., et al., *Acta Cryst.* D68, 352-67 (2012). DOI: 10.1107/S0907444912001308
- Emsley, P., et al., *Acta Cryst.* D66, 486-501 (2010). DOI: 10.1107/S0907444910007493
- Karplus, P.A. and Diederichs, K., *Science*. 336(6084), 1030–1033 (2012). DOI: 10.1126/science.1218231
- Stamogiannos, A. et al. *Biochemistry 2017 Mar 14;56(10):1546-1558. DOI: 10.1021/acs.biochem.6b01170.*

Supplemental Table 1: Data collection and refinement statistics.

Wavelength (Å)	0.9763
Resolution range (Å)	91.45 - 1.6 (1.657 - 1.6)
Space group	P 2 21 21
Unit cell	57.68 116.67 147.27 90 90 90
Total reflections	727238 (73802)
Unique reflections	130971 (12873)
Multiplicity	5.6 (5.7)
Completeness (%)	99.49 (98.94)
Mean I/sigma(I)	9.95 (1.33)
Wilson B-factor (Å^2)	21.00
R-merge	0.1172 (1.315)
R-meas	0.1295 (1.484)
R-pim	0.05421 (0.5728)
CC1/2	0.998 (0.501)
CC*	0.999 (0.8)
Reflections used in refinement	130934 (12855)
Reflections used for R-free	6495 (630)
R-work	0.1804 (0.3138)
R-free	0.2133 (0.3446)
CC(work)	0.967 (0.734)
CC(free)	0.948 (0.664)
Number of non-hydrogen atoms	8229
macromolecules	7104
ligands	311
solvent	814
Protein residues	862
RMS(bonds) (Å)	0.006
RMS(angles) (°)	0.83
Ramachandran favored (%)	97.04
Ramachandran allowed (%)	2.60
Ramachandran outliers (%)	0.35
Rotamer outliers (%)	1.73
Clashscore	8.39
Average B-factor (Å^2)	27.03
macromolecules	25.30
ligands	40.86
solvent	36.89
Number of TLS groups	3

Statistics for the highest-resolution shell are shown in parentheses.

Supplemental Figures

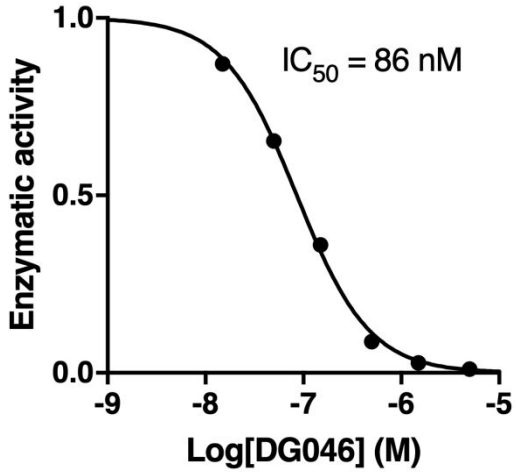

Figure S1: Titration of compound DG046 leads to full inhibition of ERAP1 with an IC_{50} of 86 nm.

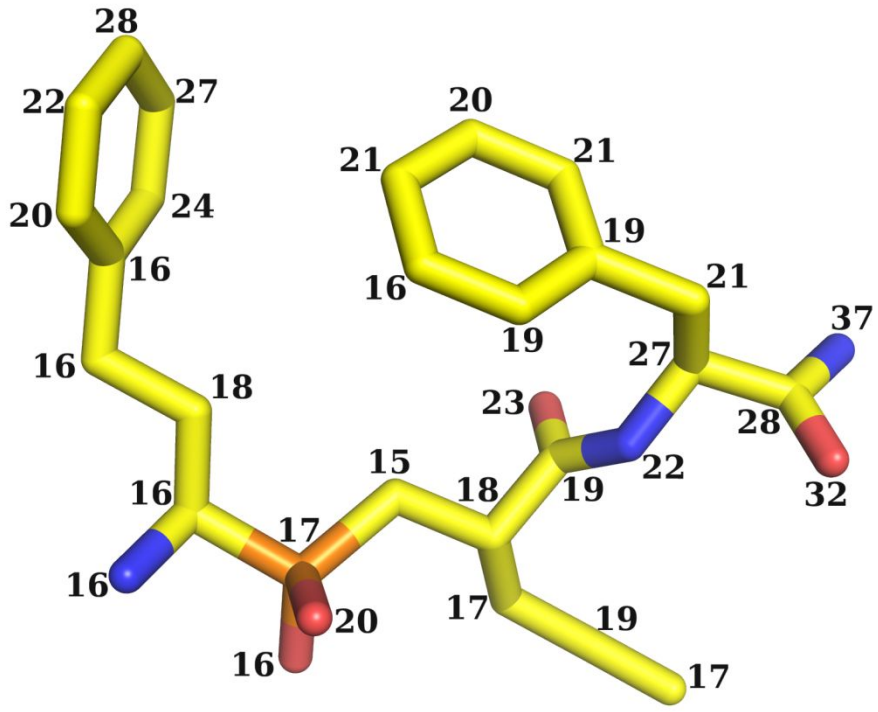

Figure S2: Model of the inhibitor DG046 in the active site of ERAP1. Carbon atoms are in yellow, oxygen in red, nitrogen in blue and phosphorus in orange. Numbers indicate the calculated B-factors for each atom (in $Å^2$).

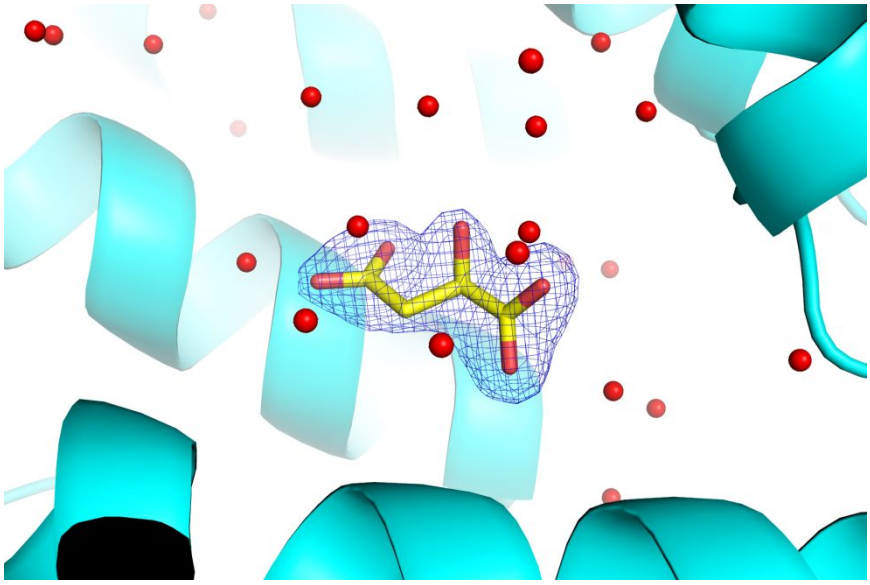

Figure S3: Model of malic acid in the cavity of ERAP1 shown in yellow sticks (oxygen atom in red), $2F_o$-F_c electron density shown as blue mesh. Nearby ordered water atoms are shown as red spheres. Contour level is 1.5 sigma.

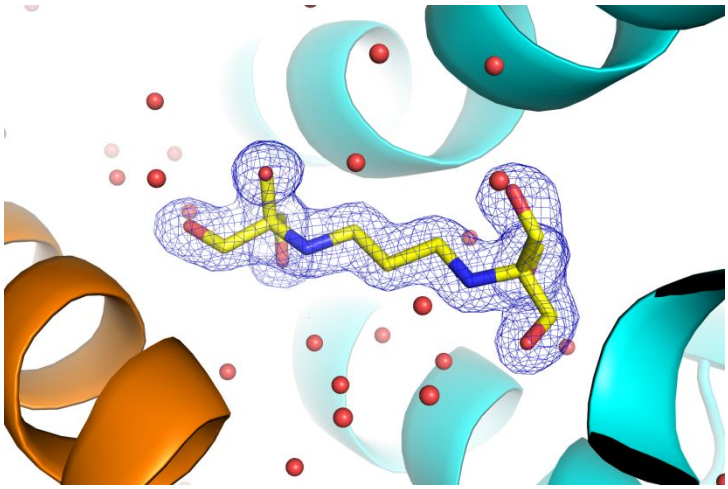

Figure S4: Bis-Tris propane molecule in the internal cavity of ERAP1 shown in stick representation (carbon=yellow, oxygen=red, nitrogen=blue). $2F_o$-F_c electron density shown as blue mesh. Nearby ordered water atoms are shown as red spheres. Contour level is 1.5 sigma.

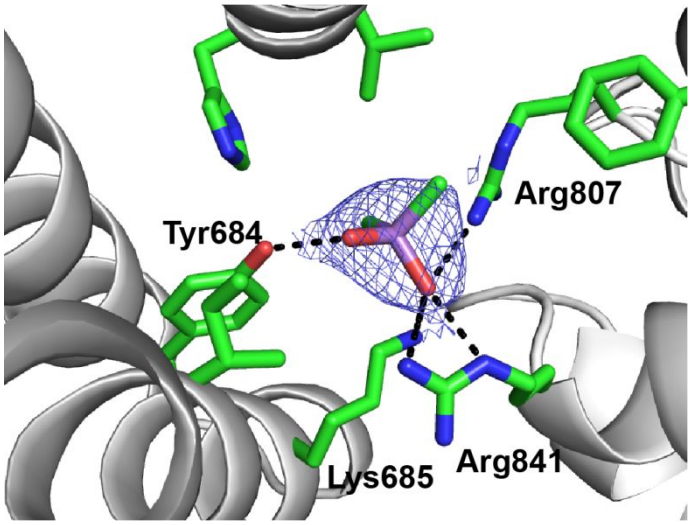

Figure S5: Cacodylate ion built in residual electron density of the ERAP1 crystal structure with PDB code 2YD0. The ion is stabilized at the exact same location as the malate and interacts with Lys685, Arg841 and Tyr684. Contour level is 1.5 sigma.