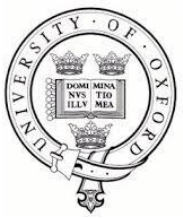
University of Oxford



ERAP1 inhibitor development

Pavel Guzanov

Linacre College

Systems Approaches to Biomedical Sciences

Industrial Doctoral Centre

A thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Michaelmas Term 2017

Abstract

ERAP1 has been associated with a set of immune-mediated diseases, including ankylosing spondylitis. Currently, there is no selective and potent ERAP1 inhibitor available that would allow in-depth research of its functions and roles in diseases. Therefore, the aim of this work is to apply fragment-based drug design approach for the development of such inhibitor.

A set of 12 structurally diverse hits was identified as a result of a screen containing 1200+ fragments using orthogonal ERAP1 activity assays. These hits can be used as a basis for development of a larger lead inhibitor, which requires structural information about their binding mode to ERAP1. Three different approaches have been tried in order to obtain such information about the binding mode. Unfortunately, ERAP1 crystallisation trials did not succeed, the potential reason being heterogeneity of the protein samples due to presence of several glycoforms. A set of 31 analogues of an IRAP inhibitor, which has been shown to inhibit ERAP1 activity, was synthesised and screened providing valuable structure-activity relationship information. Taken together, the screening campaign has resulted in a novel ERAP1 inhibitor with a single digit micromolar potency against ERAP1.

Acknowledgements

I would like to thank my supervisors Prof. Udo Oppermann and Prof. Paul Brennan for setting up this challenging project and their support throughout it. I am truly grateful to my other supervisor Dr. James Dunford for his day-to-day assistance in the laboratory and passing me the practical skills of a protein chemist. Thanks to my industrial supervisor Dr. Christian Wiesmann from Novartis AG I got an invaluable opportunity to perform part of my work at NIBR, which provided important results for the project. I also want to thank Christelle Henry from NIBR, who taught me to perform the NMR screening experiments.

Both Oppermann and Brennan research group members were extremely supportive. I would like to thank specifically Dr. Martin Philpott for his assistance with cloning and Dr. Catrine Johansson for her tremendous support with protein purifications and crystallisations.

In addition, this project would not have developed so far without assistance of the following people. Dr. Anthony Tumber taught me to use the RapidFire. Dr. Rod Chalk and Oktawia Borkowska helped with protein Mass-Spec analysis, which was tremendously important in this project. Dr. Pavel Savitsky supervised several gene cloning experiments all of which were successful, after I had been failing to do them myself for a few months. Dr. Shubhashish Mukhopadhyay assisted with fixing issues with protein expression from insect cells. Dr. Tobias Krojer assisted with protein crystallisation optimisations, obtained the X-ray data and taught to analyse it.

I also want to thank EPSRC, Novartis, Royal Commission for the Exhibition 1851 and Prof. Udo Oppermann for providing financial support during my DPhil.

Finally, but not the least important, I would like to thank my parents and my fiancée Liana for their moral support all along four years, and especially during months when this thesis was written.

Contents

List of figures

List of tables

Abbreviations

ACN – acetonitrile

AnPEP – human aminopeptidase N

BCS – basic chemical space crystallisation screen

CV – column volume

Da – daltons

DCM – dihloromethane

DMF – dimethylformamide

DMSO – dimethyl sulfoxide

DNA - deoxyribonucleic acid

DSS – 4,4-dimethyl-4-silapentane-1-sulfonic acid

Dppf – 1,1'-Bis(diphenylphosphino)ferrocene

EC_{50} – Half maximal effective concentration

ERAP1 – endoplasmic reticulum aminopeptidase 1

ERAP2 - endoplasmic reticulum aminopeptidase 2

ESI – electron spray ionisation

FBDD – fragment-based drug design

FTIR – Fourier-transform infrared spectroscopy

GF – gel filtration or size-exclusion chromatography

IC_{50} – half maximal inhibitory concentration

IMAC –immobilized metal ion affinity chromatography

IPTG - isopropyl β-D-1-thiogalactopyranosid

JCSG - Joint Center for Structural Genomics crystallisation screen

HCS – Hampton Crystal Screen

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIN – Hampton Index crystallisation screen

LB – lysogeny broth

LC-MS – liquid chromatography mass-spectrometry

LTA4H – LTA_4 hydrolase

Man – mannose

MS – mass-spectrometry

msd – mean square displacement

NAG – N-acetyl glucosamine

NMR – nuclear magnetic resonance

NTA – nitrilotriacetic acid

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

ppm – parts per million

prep-HPLC – preparatory high performance liquid chromatography

SAR – structure activity relationship

SDS – sodium dodecyl sulphate

SOC – super optimal broth

TB – terrific broth

TEV – tobacco etch virus

Tris – tris(hydroxymethyl)aminomethane

uHPLC - Ultra-High Performance Liquid Chromatography

UV - ultraviolet

1. Introduction.

1.1 The role of ERAP1 in the adaptive immune system and other reported functions

Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) is an aminopeptidase that plays an important role in the adaptive immune system. Its alternative names are Aminopeptidase Regulator of TNFR1 shedding (ARTS-1), endoplasmic reticulum aminopeptidase related with antigen processing (ERAAP), adipocyte-derived leucine aminopeptidase (A-LAP) and puromycin-insensitive leucyl-specific aminopeptidase (PILS-AP).

ERAP1 is expressed in all tissues. Higher expression has been found in the tissues, which have higher Major Histocompatibility Complex I (MHCI) expression, for example liver, lung, spleen and thymus. [1] It has been reported that ERAP1 expression is induced by IFN-γ, a proinflammatory cytokine that induces expression of several molecules that play a role in the MHC I antigen processing and presentation pathway. [2] ERAP1 has been found to be localised mainly in the endoplasmic reticulum (ER) of the cell. [3]

The role of ERAP1 in the adaptive immune system was elucidated by three independent research groups in 2002. [1], [2], [3] In the MHC I mediated immune response system cytotoxic T-lymphocytes (CTLs) are able to recognise infected or mutated cells by unusual proteins they produce and present on the surface (Figure 1-1). In the ER MHC I binds small polypeptide fragments derived from proteolytic antigen processing of intracellular proteins and this complex is transported to the cell membrane. CTLs interact with the membrane bound MHC I through its Complementarity Determining Regions (CDRs) of T-Cell Receptors (TCRs) present on the membrane. Each set of monoclonal T-lymphocytes has CDRs with specificity towards only one particular epitope – the short polypeptide characteristic of an antigen. When CTLs recognise their target epitope, a cytotoxic cascade of events takes place, which eventually leads to apoptosis of the aberrant cell. [4] MHC I has specificity for the polypeptide substrate length – depending on its type, MHC I binds solely 8, 9 or 10

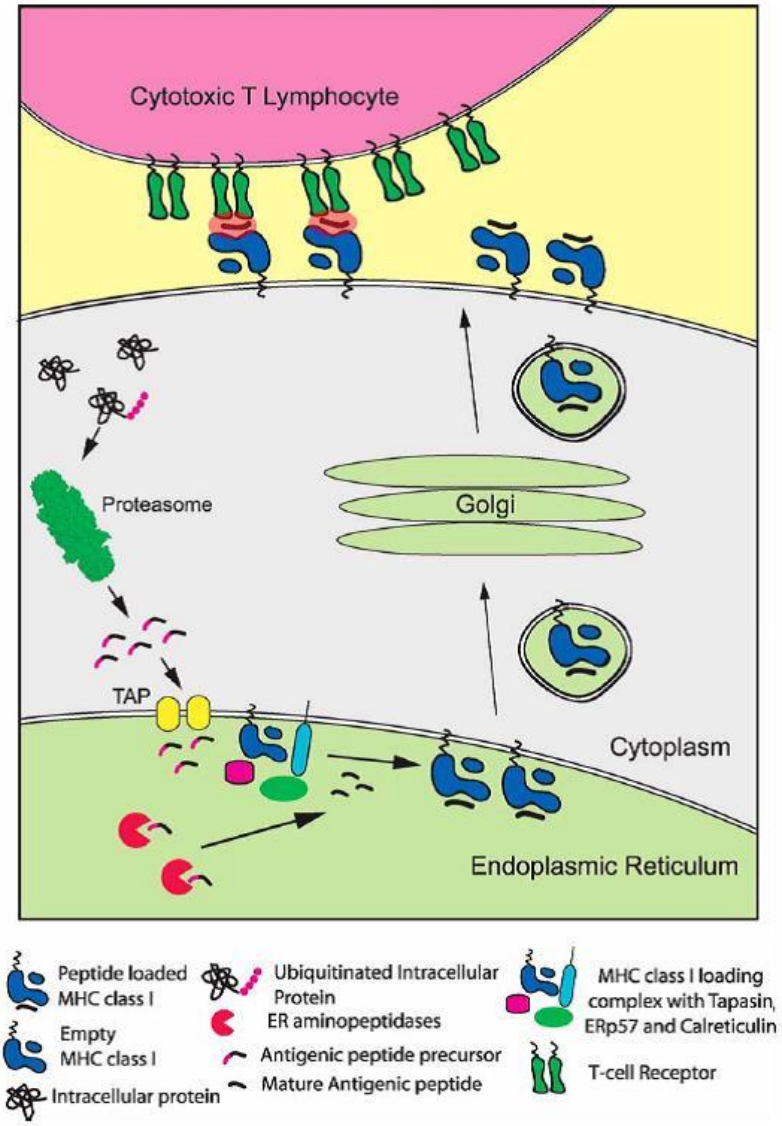


Figure 1-1. Schematic representation of the ERAP1 role in the adaptive immune system. The proteasome degrades cytosolic proteins, producing peptide chains. These antigen precursors are transported from the cytosol to ER by TAP. ERAP1 trims the precursors to the appropriate length for binding MHCI complex. Upon binding an antigen, MHCI is transported to the cell membrane, where the antigens serve for tumour or viral infection recognition by T-cells. (Adapted from I. Evnouchidou *et al.* [17]*).*

residue long protein fragments. [5] However, proteins degraded by the cytosolic proteasomes result in 3-16 residues long peptide fragments, thus creating antigenic peptide precursors with elongated N-termini. [6] These polypeptides are transported to the ER by the Transporter associated with Antigen Processing (TAP) protein. [7] The function of ERAP1 is to trim such precursors' N-termini into optimal fragments to allow their binding to MHC I. ERAP1 activity is dependent on the length of the substrate – it effectively trims peptides 8-16 residues long. [8] This substrate size specificity allows ERAP1 to perform its antigen precursor trimming function. Knock-out experiments in mice have shown that ERAP1 is not necessary for survival. [9], [10], [11] These experiments have identified that

knocking-out ERAP1 leads to the decrease of the cell surface levels of most MHCI alleles. However, the cell surface levels of some MHCI alleles do not change or even increase. Changes in the MHC II cell surface expression of the knock-out have not been observed [10], which confirms that ERAP1 is only involved in MHC I antigen presentation pathway.

MHC I molecules that are not bound to an antigen peptide are unstable, therefore their cell surface expression reduces. [12] Consequently, the range of the MHCI alleles found on the cell surface is directly affected by the changes in the repertoire of the mature antigens in the ER. Some novel MHCI-peptide complexes that were identified in the ERAP1-deficient mice caused strong T- and B-cell responses [11], which indicates that the absence of ERAP1 may lead to the increase of immunogenicity. All the above findings in the knock-out mice indicate that ERAP1 is capable of both creating mature MHC I antigens and destroying them. In this way ERAP1 changes the repertoire of the mature antigens available for MHC I binding, thus modulating the immune response.

There are a number of other functions that have been attributed to ERAP1. Its involvement in blood pressure regulation has been proposed. [13] Intracellular ERAP1 has been reported to promote angiogenesis [14], but on the contrary the secreted form has been found to supress it. [15] Finally, ERAP1 has been reported to play a role in the innate immune by enhancing the production of cytokine receptors. [16] However, its MHC I antigen processing function is the most clearly demonstrated one. Since its discovery in 2002, the overwhelming majority of ERAP1 research has been done in connection with this function. [17]

1.2 Diseases associated with ERAP1

Genome Wide Association Studies (GWAS) have discovered a relationship between single nucleotide polymorphism (SNP) of ERAP1 and Ankylosing Spondylitis. [18], [19] AS is a type of arthritis, an immune-mediated disease that causes the chronic inflammation of joints. In AS joints of the spine are affected, which may lead to the total loss of spinal cord flexibility in the most severe cases. [20] It has been found that ERAP1 SNP has association with AS only in patients bearing a specific HLA-B27

allele, HLA-B being a gene coding for MHC I in humans. [21] This suggests that the mechanism of the diseases involves alteration of the whole MHC I antigen presenting pathway.

Another immune-mediated disease that has been associated with ERAP1 by GWAS is type I diabetes. [22] This condition leads to elevated blood sugar levels due to insufficient insulin production. The disease mechanism involves autoimmune mediated destruction of the insulin-producing β-cells in the pancreas. [23]

A significant ERAP1 association has also been found between ERAP1 SNP and cervical carcinoma (CC). [24] CC is a type of cancer that is caused by Human Papilloma Virus (HPV). In CC oncogenic HPV types infect the uterine cervix and lead to cancer. [25] It has been suggested that the diseases mechanism involves changes in the antigen presenting machinery or its components' gene expression. [26]

ERAP1 has been found to be significantly downregulated in several cancer cell-lines. [27] The reduction in ERAP1 expression may lead to reduction in the amount of tumour-related antigens presented by MHCI. This allows the tumour to hamper the immune response and develop further.

The molecular mechanism explaining how ERAP1 genetic variations affect the disease predisposition has not yet been proposed. It has been hypothesised that SNPs associated with the diseases might alter the enzyme specificity and trimming ability, thus changing the repertoire of the presented epitopes. This in turn could cause a defective immune response. [17] Accordingly, ERAP1 has a crucial role in the MHCI antigen presenting pathway and reported associations with a set of immune-mediated diseases such as AS and cancers make ERAP1 a very interesting therapeutic target.

1.3 Overall structure of ERAP1

Full-length ERAP1 structure has been successfully solved in our group in both open (PDB: 2YD0) and closed (PDB: 3QNF) conformations [28] with the resolution limits of 2.7Å and 3.0Å respectively. A

3.0Å closed state ERAP1 crystal structure has also been reported by Stern *et al.* [29] The proposed ERAP1 structures from these groups are in agreement with each other, even though the latter group used a different construct.

ERAP1 is a monomeric M1 class aminopeptidase. [17] It has been found to be similar to the previously crystallised aminopeptidases TIFF3 (**1Z5H**), ePepN (**2ZXG**) and LTA4H (**1HS6, 3B7U**). ERAP1 is divided into 4 domains (Fig. 2): domain I (residues 46-254), domain II bearing the active site (residues 255-529), domain III (residues 530-614) and domain IV (residues 615-930).

Domain I consists solely of β-sheets. An 8-stranded β-sheet saddle-like motif in the centre is adjacent to a 3-stranded β-sheet on one side and a 4-stranded one on the other. Domain II contains 8 α-helices and 5 β-strands forming a thermolysin-like catalytic domain. Domain III is composed of 6 β-strands forming a β-sandwich. This domain pulls domain IV in and out thereby switching ERAP1 between the closed and the opened conformations (Figure 1-2). Finally, the bowl-shaped domain IV consists only of α-helices. The proposed substrate C-terminus binding site is located at the intersections of helices 21, 31 and 35 of this domain. In the opened conformation domain IV does not interact with domain I and II thus allowing substrate to enter and bind to the active site of ERAP1. However, in order to catalyse the peptide hydrolysis reaction ERAP1 has to switch into the closed conformation, where domain IV completely separates the substrate binding cavity from the environment. The catalytic proteolysis takes place in the closed state, when the catalytic cavity is fully structured.

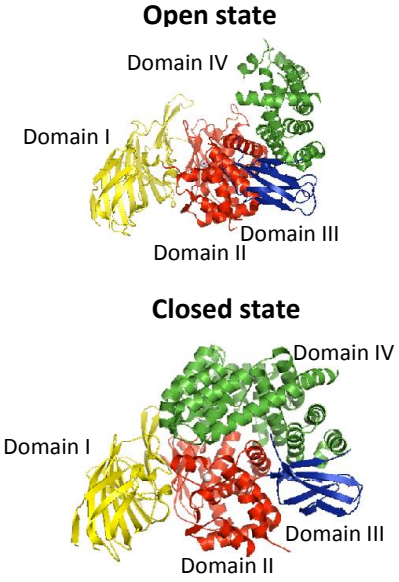


Figure 1-2. ERAP1 in open (top, PDB: 3QNF) and closed (bottom, PDB: 2YD0) conformations. Domains I-IV are coloured accordingly yellow, red, blue and green. White sphere represents Zn in the active site. Created using Pymol.

1.4 ERAP1 catalytic site and the proposed catalytic mechanism

The active site contains G^{317}AMEN and H^{353}-E-X-X-H-X_{18}-E Zn-binding motifs, both being common for M1 class exopeptidases (Figure 1-3). [30] The zinc-binding motif coordinates Zn via residues His^{353}, His^{357} and Glu^{376}. Ser^{316} and Met^{319} of the GAMEN motif, create a small hydrophobic cleft, where the substrate's N-terminus is accommodated. The size of this cleft dictates ERAP1 specificity for the N-terminal residues of substrates. Substrates with N-terminal leucine and methionine are processed the most efficiently, since these residues size optimally match the cleft. Substrates with larger and smaller N-terminal residues show slower trimming rates. [31] Glu^{181} has also been reported to be important for ERAP1 N-terminal specificity. Its mutation to aspartate leads to a switch of the substrate N-terminal residue preference from hydrophobic amino acids to the basic ones. [32]

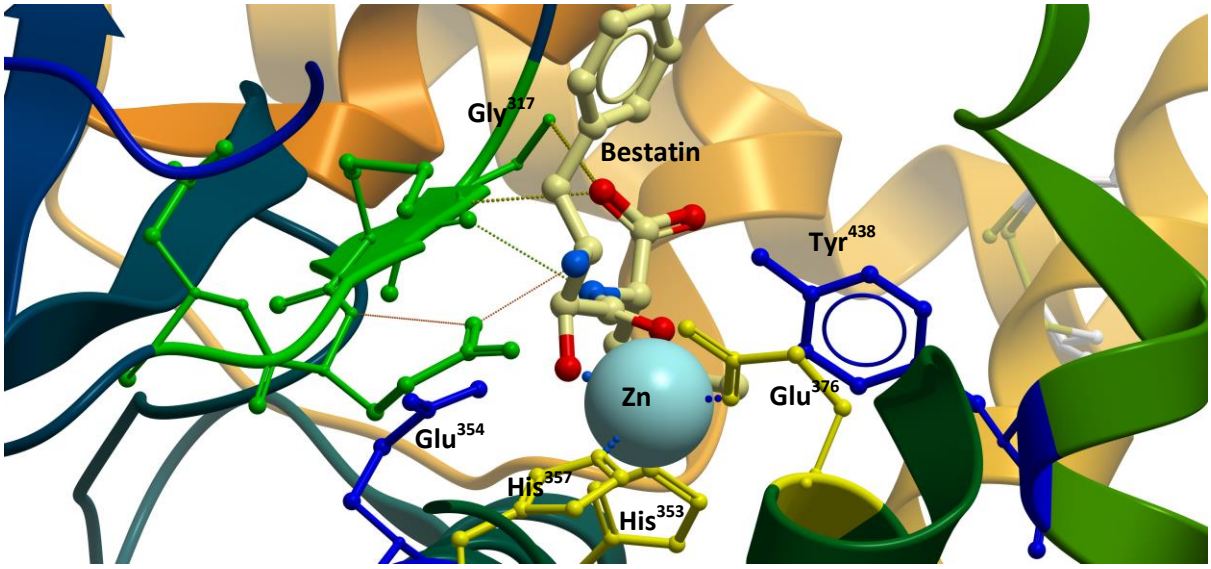


Figure 1-3. ERAP1 active site. White sphere – Zn. Zn-binding H^{353}, H^{357} and Glu^{376} are in yellow. G^{317}AMEN in green. Catalytically important Tyr^{438} and Glu^{354} are in blue. Bestatin bound to the active site is in pink, its H-bond interactions with active site residues are represented by doted lines. Created using ICM-Pro, PDB: 2YD0.

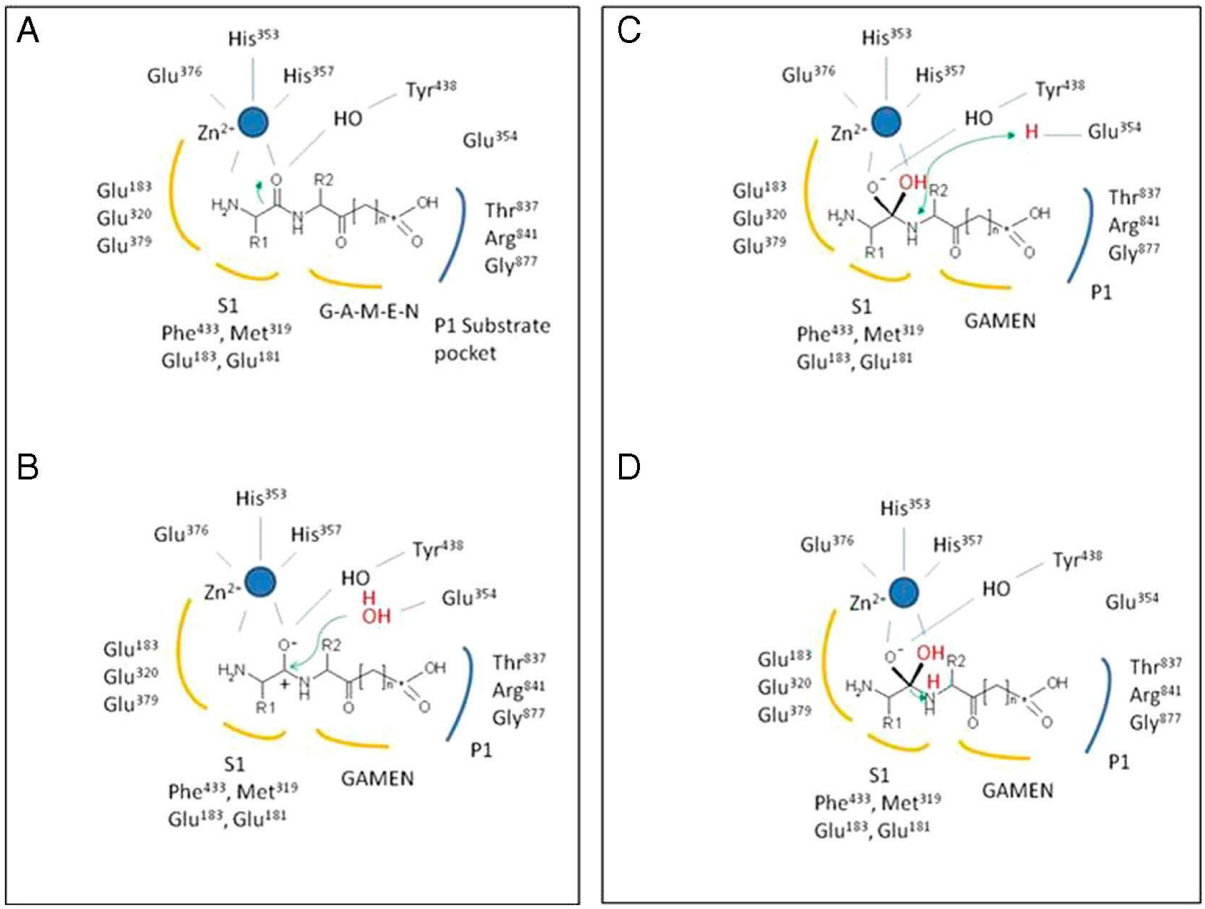


Figure 1-4. Schematic representation of the proposed ERAP1 peptide hydrolysis mechanism. Domain II active site surface – yellow, domain IV surface – blue. (Adapted from G. Kochan *et al.* [28])

1.5 ERAP1 proposed catalytic mechanism and substrate specificity

The proposed mechanism for the ERAP1 peptide bond hydrolysis was based on previously established peptide cleavage mechanism found in the orthologous LTA4 enzyme (Figure 1-4). , [33], [34], [35] The amino and carbonyl terminus are accommodated in the active site by coordination to the Zn^{2+} ion. The hydroxyl group of Tyr^{438} makes an H-bond to the carbonyl group and polarises it. A water molecule performs a nucleophilic attack facilitated by an ionised Glu^{354} on the polarized carbonyl bond. The obtained tetrahedral intermediate is stabilised by the Zn^{2+} ion and interactions with Tyr^{438} followed by proton capture by Glu^{354}. In the next step the hydrolysed amide nitrogen is protonated by Glu^{354}, which converts it into a good leaving group. It was proposed that the protonated amide may be stabilised in some way, for example *via* interactions with backbone carbonyls. [28]

As mentioned before, ERAP1 has a strong preference for 8-16 residue long peptide substrates. Two models have been proposed to account for the ERAP1 substrate length specificity. One model suggests that ERAP1 trims N-elongated antigenic peptide precursors while they are already weakly bound to MHCI, leaving the loose elongated N-terminus accessible for ERAP1. Once the peptide is trimmed to the correct length, the remaining polypeptide is "shielded" by MHCI, hence ERAP1 cannot access it and crop further. [36] The other model assumes that the specificity is an inherent ERAP1 property, *i.e.* this enzyme displays a "molecular ruler" mechanism, and MHCI binds a peptide fragment after it has been already trimmed by ERAP1. [37] Findings of several research groups [28], [38] are in favour of the "molecular ruler" mechanism. Modelling experiments resulted in the same peptide substrate length specificity as obtained in experimental studies. [28]

It has been suggested that ERAP1 contains a hydrophobic binding site for the peptide C-terminus [37]. The distance between the putative C-terminus binding site and the active site has been measured in the computational experiments based on the obtained crystal structures. From the obtained distance it has been concluded that peptides shorter than 8-9 residues are not long enough to reach the ERAP1 active site, where peptide bond cleavage occurs. [28] However, as full-length ERAP1 complex bound to a peptide has not yet been crystallised, it is not possible to confidently confirm which model describes the trimming process correctly.

The results of enzyme activity experiments have suggested that the ERAP1 trims the substrate peptides in a non-processive manner. [28] Thus, the proposed catalytic cycles consist of the following stages. First, the substrate anchors to the ERAP1 with its C- and N-termini. This binding converts ERAP1 to its closed conformation, to allow effective trimming of the N-terminal substrate residue. This is supported by the fact that short dipeptide substrates are trimmed at a much slower rate, since such substrates are not able to stabilise the ERAP1 closed conformation. But the addition of 7-mer peptide to the enzyme reaction allosterically activates the hydrolysis of such short

substrates. [39] Finally, the trimmed substrate must dissociate from ERAP1 to allow the next round of trimming.

The ability to trim long peptide substrate distinguishes ERAP1 from other aminopeptidases, which typically are able to trim 4-mer or shorter substrates. [8] ERAP1 contains a very large internal substrate binding cavity (2,920 $Å^3$) formed by the domains II and IV in the closed state, which is considerably larger compared to the similar aminopeptidases ePEpN (2,200 $Å^3$) or LTA4 (1,130 $Å^3$) making the ERAP site the largest cavity seen to date for an aminopeptidase. [29] Such a cavity allows accommodation of large substrates of variable length. Substrates that are too long (16+ residues) do not allow ERAP1 switching to the closed conformation. Hence, they cannot be effectively trimmed.

1.6 Reported ERAP1 inhibitors

There is currently little information about potent and selective ERAP1. Bestatin – an aminopeptidase inhibitor isolated from actinobacteria [40] – has been used as a non-selective inhibitor for ERAP1 studies (IC_{50} = 87 μM) and its crystallisation. It is the effective inhibitor of the broad range of aminopeptidases from various families. Its main drawback is lack of selectivity and low potency against ERAP1.

DG013A was reported by E. Stratikos *et al.* in 2013 (Figure 1-5). [41] It is a phosphinic pseudo-peptide that mimics the tetrahedral carbanion intermediate – the transition state intermediate of the peptide cleavage reaction. It had IC_{50} = 33 nM against ERAP1 in the *in vitro* assay, which makes it a much more potent ERAP1 inhibitor than bestatin. It has been successfully used in *in vitro* cell-based assays and has been shown to induce the same effect as ERAP1 silencing. [42] The inhibitor concentration used in the cell assay was 1 μM, which is high for such a potent inhibitor. But its main drawback is lack of selectivity against other M1 family aminopeptidases, having a similar potency against ERAP2 and IRAP. The poor target selectivity is typical for this class of metalloproteinase inhibitors, because

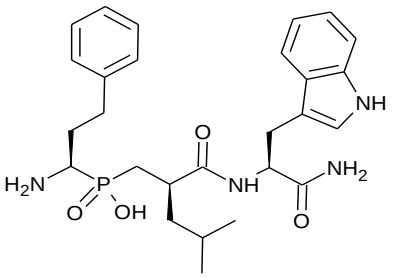


Figure 1-5. Structure of the reported ERAP1 inhibitor DG013A.

the potency is mainly dependent on the chelation strength of the phosphinic group to the enzymatic zinc ion. [43] In addition, its synthesis strategy restricts the potential structural diversity of the analogues, which further hampers the development of a more selective inhibitor.

The same group reported thiomersal – an antiseptic and antifungal agent used as a vaccine preservative – as potential ERAP1 inhibitor. [44] It has sub-micromolar potency against ERAP1 in the spectroscopic enzyme activity assay, making it a less potent inhibitor than their previously described molecule DG013A. However, thiomersal has been shown to have a very strong selectivity profile over ERAP2 and IRAP. No further exploration of the structure-activity relationship of that scaffold's analogues was reported in that paper.

Thus, at the moment the only reported potent ERAP1 is DG013A, which does not offer selectivity over homologous aminopeptidases. Hence, it cannot be used as a research tool to study ERAP1 functions and role in the diseases in detail.

1.7 Human M1 aminopeptidases.

All the aminopeptidases that belong to the M1 family contain three highly-conserved motifs. $HEXXHX_{18}E$ motif that participates in zinc-binding is a hallmark of this metalloproteinase class. In addition, a conserved GAMEN motif that is important for binding the N-termini of the substrates. Finally, all M1 aminopeptidases contain two conserved active site residues – glutamic acid, which is part of the GAMEN, and a peripheral tyrosine residue that plays an important role in the catalytic mechanism. The catalytic domain of M1 aminopeptidases is in general highly conserved.

Currently, there are eleven human aminopeptidases that belong to the M1 class apart from ERAP1 - leucyl-cystinyl aminopeptidase (LNPEP, insulin-related aminopeptidase, IRAP), endoplasmic reticulum aminopeptidase 2 (ERAP2), aminopeptidase Q (APQ, laeverin, LVRN), puromycin-sensitive aminopeptidase (NPEPPS), thyrotropin-releasing hormone-degrading ectoenzyme (TRHDE), Glutamyl aminopeptidase (aminopeptidase A, ENPEP), aminopeptidase N (AnPEP), aminopeptidase O (APO, AOPEP), leukotriene A4 hydrolase (LTA4H), arginyl aminopeptidase B (RNPEP) and arginyl

aminopeptidase B like-1 (RNPEPL1). This class remains relatively unexplored at the moment. Structural information about half of the class members – APQ, NPEPPS, TRHDE, AOPEP, RNPEP and RNPEPL1 – is not available, since they have not yet been crystallised according to the PDB database.

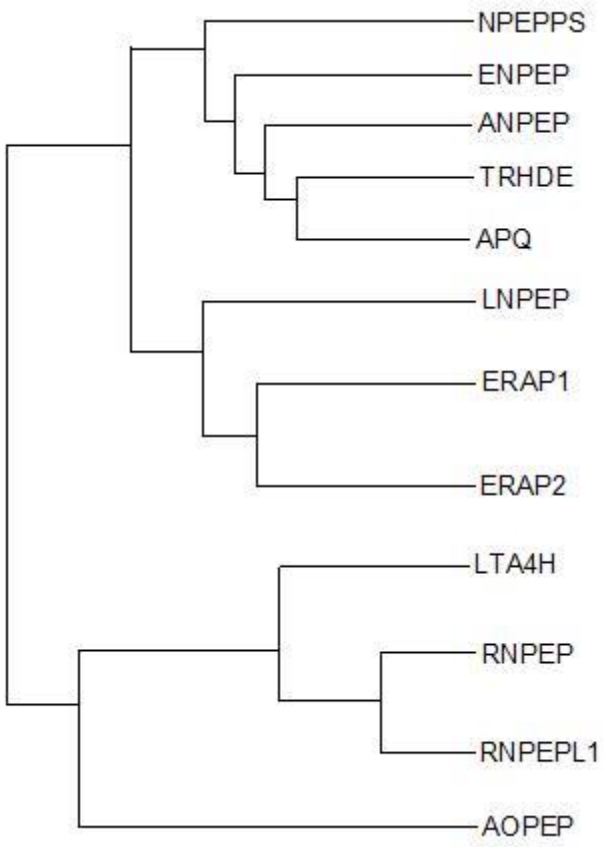


Figure 1-6. A phylogenetic tree of human M1 aminopeptidases based on the results from K. Maynard et al. [45]

Based on the phylogenetic analysis of M1 aminopeptidase thermolysin has been suggested as the common ancestor of this class of enzymes, which subsequently have evolved into three subgroups. [45] The first group contains APQ, TRHDE, ANPEP, ENPEP, NPEPPS, LNPEP, ERAP1, and ERAP2. This group can be further subdivided into two, the first subgroup containing ENPEP, NPEPPS, TRHDE, ANPEP and APQ, and the second subgroup containing ERAP1, ERAP2 and LNPEP. LTA4H, RNPEP, and RNPEPL1 belong to the second group. The third group consists of only one enzyme – AOPEP.

1.8 Selection of an off-target enzyme panel for M1 aminopeptidase inhibitors.

When the enzymes for the selectivity panel were chosen, several aspects were taken into consideration. Firstly, the enzymes should have been previously successfully crystallised. Secondly, they preferably should not be membrane proteins to facilitate purification and crystallisation. Finally, at least some of the enzymes should be close orthologs of ERAP1 to allow high selectivity of the developed inhibitor.

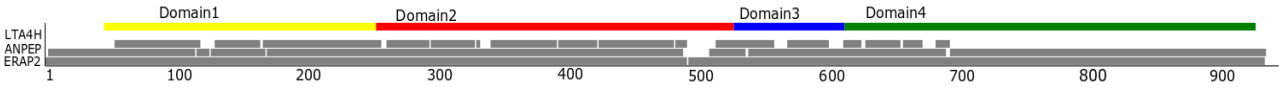


Figure 1-7. The sequence alignment of LTA4H, AnPEP and ERAP2 with ERAP1. ERAP1 domain structure is given on the top of the alignment. The alignment was performed using Saturn software (http://www.thesgc.org/saturn)

LTA4H, AnPEP and ERAP2 have been selected for the panel (Figure 1-7 and Figure 1-8). All three have been previously successfully crystallised [46], [47], [48]. Moreover, work on LTA4H and AnPEP has been done in groups that we collaborate with, hence the unpublished material was accessible. ERAP2 is the closest ERAP1 ortholog, which shares 51% sequence identity and belongs to the same M1 subgroup. AnPEP is a more distant ERAP1 ortholog. It belongs to the different subgroup to ERAP1 and shares 33% sequence identity. Finally, LTA4H belongs to the different M1 aminopeptidase group and shares 25% sequence identity with ERAP1.

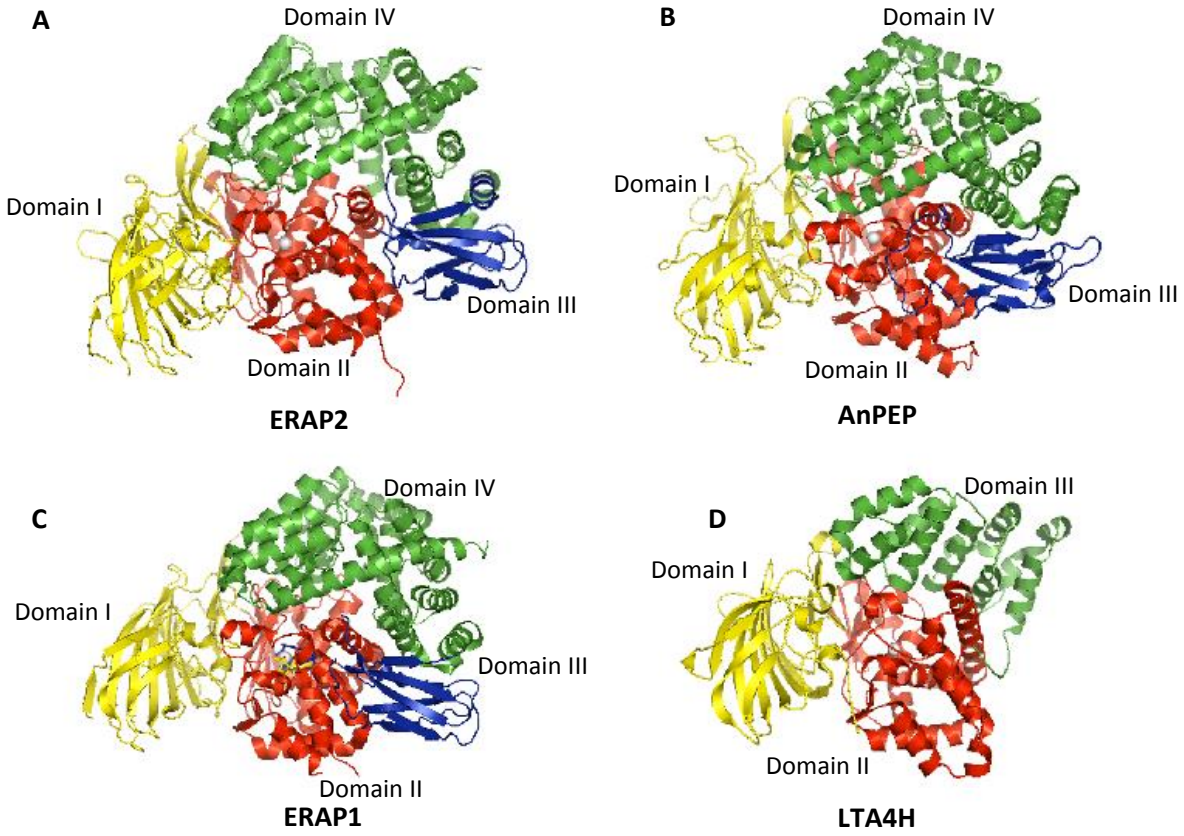


Figure 1-8. ERAP1 and selectivity panel enzymes crystal structures. ERAP2 (A, PDB: 3SE6), AnPEP (B, PDB: 4FYQ), ERAP1 (C, PDB: 2YD0) and LTA4H (D, PDB: 1HS6). Domain I is yellow, catalytic domain II – red, C-terminal domain IV – green and β-sandwich domain III blue. LTA4H is the only enzyme in this set that does not have the β-sandwich domain. Zn in the active site is represented by a white sphere. Created using Pymol.

1.8.1 LTA4H

The primary function of LTA4H is hydrolysis of leukotriene A_4 (LTA_4) to leukotriene B_4 (LTB_4). [49] LTB_4 is a potent lipid chemoattractant that is involved in a variety of diverse processes in the human body such as inflammation [50], immune response [51], lipid homeostasis [52]. LTA4H also exhibits aminopeptidase activity, however its function in the human body has not yet been elucidated.

LTA4H consists of three domains, which correspond to domains I, II and IV in ERAP1. Its sequence identity with the latter is 25%. N-terminal domain I (residues 1-207) has a β-sandwich structure typical for this class of aminopeptidases. Domain II (residues 208-450) contains the enzyme catalytic site and creates substrate binding clefts at the interface with domain III (residues 451-610). LTA4H has two distinct binding sites for leukotriene and peptide substrates. [35]

1.8.2 ERAP2

ERAP2 plays a similar role as ERAP1 in the MHCI system, trimming the antigen precursors in the ER to appropriate length for binding the MHCI molecule. It has been found to have a complementary activity to ERAP1, because it can trim the substrates that ERAP1 cannot process efficiently. [53]

The structure of ERAP2 is very similar to that of ERAP1, the enzymes sharing 51% sequence identity. It consists of four domains, which correspond to domains I, II, III and IV in ERAP1. N-terminal domain I (residues 54-271) has a β-sandwich structure typical for this class of aminopeptidases. Domain II (residues 272-546) contains the enzyme catalytic site. Domain IV (residues 648-960) interfaces with the domain II and forms the substrate binding cleft with a comparable size to ERAP1. However, the shape and the calculated electrostatic potential of the cavity are different. This explains differences in substrate specificity between ERAP1 and ERAP2. [47] In addition, ERAP2 does not have the same unusual specificity for the substrate size, as it has been shown to efficiently digest 8-mer peptide substrates that are not trimmed by ERAP1. [8]

1.8.3 AnPEP

AnPEP converts Angiotensin III (AngIII, RVYIHPF) to Angiotensin IV (AngIV, VYIHPF) by trimming the N-terminus of the former. Thus, AnPEP aminopeptidase activity plays an important role in the renin-angiotensin system, which is important for blood pressure regulation and electrolyte control. [54]

AnPEP consists of four domains, each corresponding to a corresponding domain in ERAP1. Its sequence identity with the latter is 33%. N-terminal domain I (residues 66-286) has a β-sandwich structure typical for this class of aminopeptidases. Domain II (residues 287-549) contains the enzyme catalytic site. Domain IV (residues 636-967) interfaces with domain II which forms the substrate binding cleft. No AnPEP crystal structure in the open conformation has been obtained. However, in the closed conformation the internal cavity does not have any significant channel or opening that would allow a substrate to reach it. Thus, it has been proposed that AnPEP undergoes the same

change from open to closed state upon substrate binding as ERAP1. Computational studies have supported this idea. [55]

1.9 Fragment-based drug design

Fragment-based drug design (FBDD) is becoming an increasingly popular rational drug development approach both in academia and in industry. In difference to the traditional high-throughput screening (HTS) approach, in FBDD the initial hits are generated by screening libraries of small molecules (<300 Da). The smaller compounds allow more efficient screening of chemical space, thus in a typical FBDD screening campaign libraries of hundreds or thousands of compounds are used compared to millions in a typical HTS screening (Figure 1-9). This significantly accelerates the screening and makes it less resource intensive, which makes this approach popular in academia. FBDD hits have higher ligand efficiency – binding free energy per ligand heavy atom – than a typical HTS hit. This implies that more atoms in the FBDD hit contribute to the interactions with the target and such hit is less likely to have a group interfering with binding to the target. One of the main disadvantages of FBDD is that binding affinities of the hits are typically small, in a high micromolar to millimolar range. Thus, much more sensitive screening methods are necessary compared to HTS. Therefore, X-ray crystallography and NMR are the most widely used biophysical techniques for this purpose due to their exceptional sensitivity.

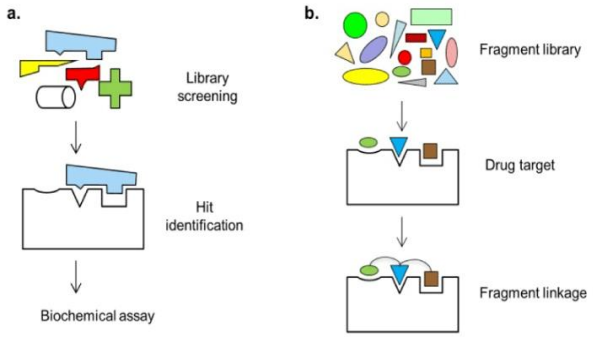


Figure 1-9 Schematic representation of HTS approach (a) and FBDD approach (b). In FBDD the identified hits usually have better ligand efficiency and fewer groups that interfere with the binding to the target. The main disadvantage of FBDD is that the identified hits are much weaker binders, thus more sensitive biophysical assays are required for screening. (Adapted from S. Park et al. [108])

The lead compound is designed by growing a hit or merging with other small fragment hits. For this step structural information about the hits' binding mode is of paramount importance. Blind modification of the hit structure parts in order to find a favourable structure-activity relationship that is normally used in HTS can be rarely applied to FBDD hits. They are simply too small to allow modifications of separate parts of the molecule.

There are a number of targets, especially metalloenzymes, where the FBDD approach has been successful for generating potent inhibitors. The examples of such compounds include tyrosine-protein kinase (JAK2 and JAK3) inhibitor AT928 [56], cycline-dependant kinase 2 (CDK2) inhibitor AT7519 [57], matrix metalloproteinase-2,3 (MMP-2 and MMP-3) inhibitor ABT-518 [58] and LTA4H inhibitor DG051 [59]. All these compounds have progressed into clinical trials. There are examples of FBDD developed inhibitors that have reached the market, e.g. serine/threonine-protein kinase B-raf (B-RAF) inhibitor Zelboraf®, currently used for treatment of late-stage melanoma.

1.10 Project goal and thesis layout

There is a clear need for a selective and potent ERAP1 inhibitor. Such a research tool would allow more detailed investigation of its functions and roles in the associated diseases. In this work a fragment-based drug design approach is implemented for the first time for the development of an ERAP1 inhibitor. The goal of this work was to identify SAR for the ERAP1 hits. Ideally, a potent and selective ERAP1 inhibitor ought to be designed

The initial project plan involved three main stages. The first stage involved the fragment libraries screening against ERAP1 using two orthogonal enzyme activity assays. Then, binding mode of the identified hits had to be elucidated using X-ray crystallography. Using this structural information a larger lead molecule was supposed to be designed by merging several hits together. The final stage involved the lead compound optimisation in terms of its potency and selectivity against the selectivity panel enzymes.

There is an informal consensus about chemical probe properties regarding potency and selectivity. Typically, < 100 nM potency in the *in vitro* assay and < 1μM in cellular assays are desired for a compound to be classed as a probe molecule. The selectivity against the same sub-family proteins is supposed to be > 30-fold. [60], [61] These requirements have been set as the target properties for the inhibitor in this work.

This thesis is divided into four chapters. In the first chapter the work on enzyme cloning, expression and purification is described. In chapter 2 the results from the fragment screening campaign are reported and discussed. The following chapter describes the work on crystallisation of ERAP1. Finally, in chapter 4 the chemical synthesis that has been done for this project is described.

2. Enzyme cloning, expression and purification

2.1 Introduction

In this project purified ERAP1 and a panel of enzymes for selectivity studies were required for the activity assays and crystallisation experiments. Therefore, cloning, expression and purification of these enzymes has been performed independently.

The full-length human ERAP1 has been previously successfully cloned, expressed and purified by the former members of our research group. [28] This has been done using the Baculovirus Expression Vector System (BEVS), which is described in more detail below.

Human LTA4H, AnPEP and ERAP2 have been chosen as the enzymes for the selectivity panel. Such panel would allow the development of an inhibitor with good selectivity properties that could be used for the in-depth research of ERAP1 functions and role in the associated diseases. The following three factors have influenced the choice of the proteins for the panel. Firstly, these enzymes belong to the same M1 aminopeptidase family, having a significant degree of homology. LTA4H has 25% sequence identity with ERAP1, AnPEP has 33% and ERAP2 51%, the latter being the most similar ERAP1 ortholog in the M1 family. Secondly, these aminopeptidases have been previously successfully expressed, purified and crystallised. Finally, all three are soluble proteins and are not attached to cell membranes (a secretion AnPEP construct has been used for enzyme expression, purification and crystallisation), which facilitates purification and crystallisation of the proteins

Full-length human LTA4H [35] has been expressed using the BL21 bacterial expression system, whereas full-length human AnPEP [48] and ERAP2 have been expressed using the same BEVS system used for ERAP1. [47]

The bacterial expression system is one of the most widely used for exogenous protein expression. Its advantages are fast culture growth – bacteria doubling time is approximately 20 minutes [62] , high

cell density in the culture, quick and facile exogenous DNA transformation, and cost effectiveness relative to the alternative expression. [63] BL21 is the most widely used *E. coli* strain for this purpose. [63]

However, this system is usually not suitable for proteins larger than 60 kDa [63] and eukaryotic proteins where post-translational modifications such as glycosylation are important. [64] In these cases the insect cell expression system is typically applied instead.

The advantage of the insect cell expression system is that it allows producing exogenous proteins with natural fold and post-translational modifications. Many protein processing events that occur in mammalian cells also take place in the insect ones, for example N-glycosylation. [65] Considering that the insect cell expression system is much easier, cheaper and biologically safer to use than the mammalian one, the former is widely used for the expression of human proteins, which cannot be expressed in bacteria.

Human ERAP1, ERAP2 and AnPEP have been expressed using the Bac-to-Bac® BEVS. In this system a recombinant *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) containing the gene of interest is used to infect Sf9 cells and makes them express the required protein. These cells are a clonal isolate of Sf21 cell line, which was developed from the pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda*. The expression of the target gene is controlled by the viral polyhedrin promoter, which triggers the cloned gene overexpression in the very late (also known as the viral occlusion protein stage) stage – 20-36h after the infection.

2.1.1 Description of the Bac-to-Bac® BEVS

The preparation of the active recombinant baculovirus involves three stages. First, a recombinant pFastBac (pFb) plasmid is prepared, which is a special vector that was designed for the Bac-to-Bac® system. The SGC developed pFb-CT10HF-LIC and pFb-Sec plasmids were used in this work, both conferring ampicillin resistance (Fig. 2-1).

Once the recombinant plasmid is obtained, it is used for production of a bacmid – a recombinant viral DNA. For this purpose DH10α *E. coli* strain is used that contains an empty bacmid - bMON14272 (136 kb) with kanamycin resistance and a helper plasmid pMON7124 (13.2 kb). The latter encodes a transposase, which ensures transposition of the desired gene from the donor plasmid into the bacmid (Figure 2-1). In addition, the helper plasmid provides tetracycline resistance.

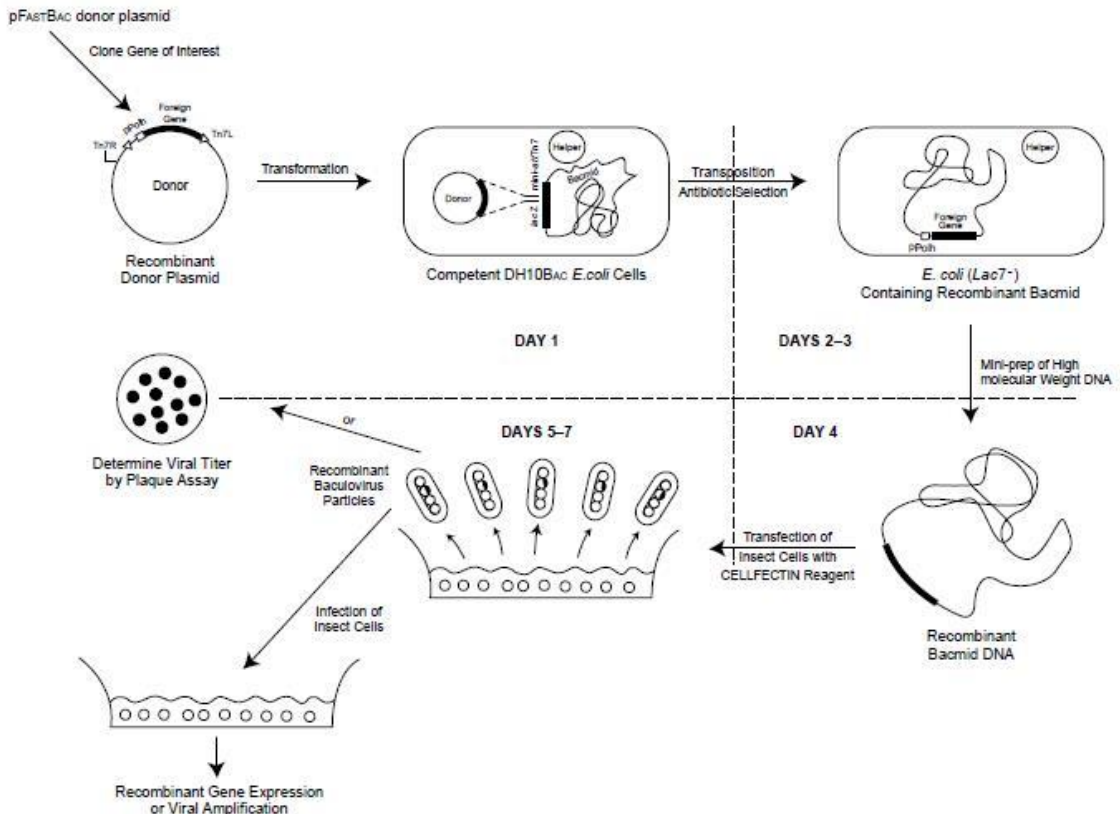


Figure 2-1. Bac-to-Bac is used system to generate a recombinant baculovirus. The image adapted from the Invitrogen Instruction Manual, Guide to BEVS and Insect Cell Culture techniques.

The transposition of the target gene disrupts *lacZα* gene in the bacmid, which complements the deleted *lacZα* on the DH10α chromosome. Hence, the colonies with successful transposition can be identified using the blue white screening method. Bacteria in such colonies do not express functional β-galactosidase, so these colonies do not turn blue when grown in presence of Blue-gal.

Finally, the recombinant bacmid is transformed into the active viral stock that is used for further protein expressions. The bacmid infects Sf9 cells with help of a transfecting reagent, such as

GeneJet® or GeneJuice® used in this work. The reagent allows the recombinant bacmid DNA to pass the insect cells wall. The infected cells produce the active virus, which does not need adjuvants for the cell infection.

2.1.2 Chapter overview

In this chapter the attempts to reproduce the cloning, expression and purification of ERAP1 using the previously optimised protocols are described. Then, the work on the cloning, expression and purification of ERAP2, AnPEP and LTA4H for the selectivity panel is reported.

2.2 ERAP1

Two ERAP1 constructs have been used in this work – ERAP1-c200 and ERAP1-c201. ERAP-c200 is the full-length human $_{1\text{-}941}$ERAP1 construct designed for the BEVS expression system. It contains C-terminal TEV-cleavable His and FLAG tags. This construct was previously used in our group for ERAP1 expression and subsequent crystallisation. [28] It is supposed to have an internal secretion signal, since the protein was secreted by Sf9 cells when cloned into a non-secretory pFb-CT10HF-LIC plasmid (Figure 2-2).

ERAP1-c201 is a shorter human $_{37\text{-}941}$ERAP1 construct with N-terminal TEV-cleavable His-tag. It was used in an attempt to improve ERAP1 crystallisability by reducing the number of labile residues (see Table 4-1). ERAP1-c201 recombinant pFb-Sec-NH plasmid was available from the previous work on ERAP1 in our group. The plasmid allowed the protein to be secreted by the insect cells in the absence of the internal secretion signal.

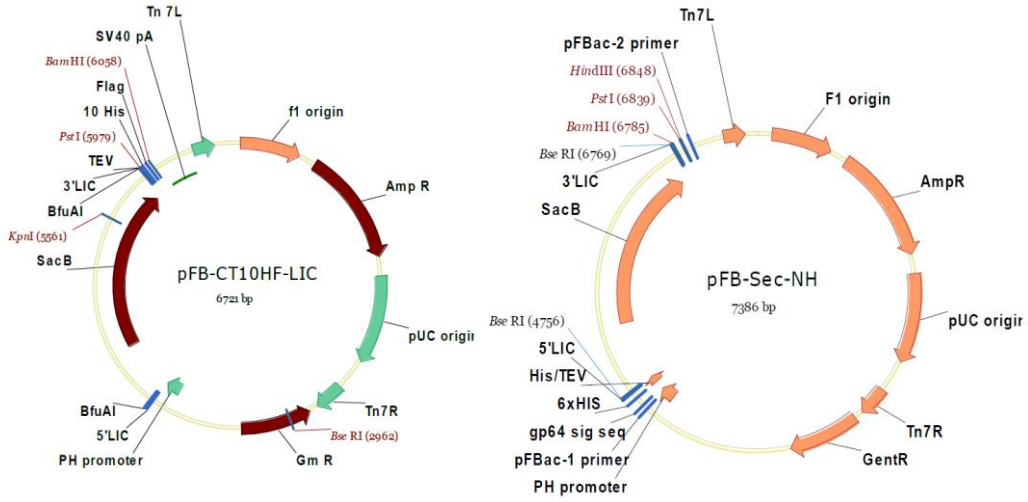


Figure 2-2. pFb-CT10HF-LIC plasmid (left) used for cloning of ERAPc-c200, ERAP2 and AnPEP-c001 and pFb-Sec-NH plasmid (right) used for cloning of AnPEP-c002.

2.2.1 Cloning and viral stock preparation

ERAP1-c200 was obtained by PCR from the full-length human ERAP1 entry clone and the relevant primers. The PCR product was successfully cloned into the pFb-CT10HF-LIC plasmid by Dr. Martin Philpott and confirmed by sequencing. This plasmid was used for the preparation of the recombinant bacmid *via* transformation in DH10α *E. coli* cells. The successful transposition of the insert of the correct size was confirmed by the PCR of the purified bacmid samples (Figure 2-3).

Numerous attempts were made to transform the obtained batches of the ERAP1-c200 recombinant bacmid into the active viral stock using either GeneJet® or GeneJuice® transfection reagents. Different volumes of the obtained P0 virus – 60 µl, 120 µl or 320 µl per 2 ml of the cell culture – were tried. All the test-infections using the obtained P0 virus stock batches were unsuccessful. The test-purifications were performed either 72 or 96 hours after the infection with the bacmid (Figure 2-3). The Sf9 cells looked infected under the microscope, but no ERAP1 was isolated upon test purifications. The protein band in the supernatant flow-through was attributed to Sf9 cells secreted protein and the same band was observed in the test expressions of other enzymes (Figure 2-18). Test-infection of 10 ml cell culture using the 0.5 ml P1 virus was attempted once, but the result was the same. Therefore, no further amplification of the obtained virus batches was performed.

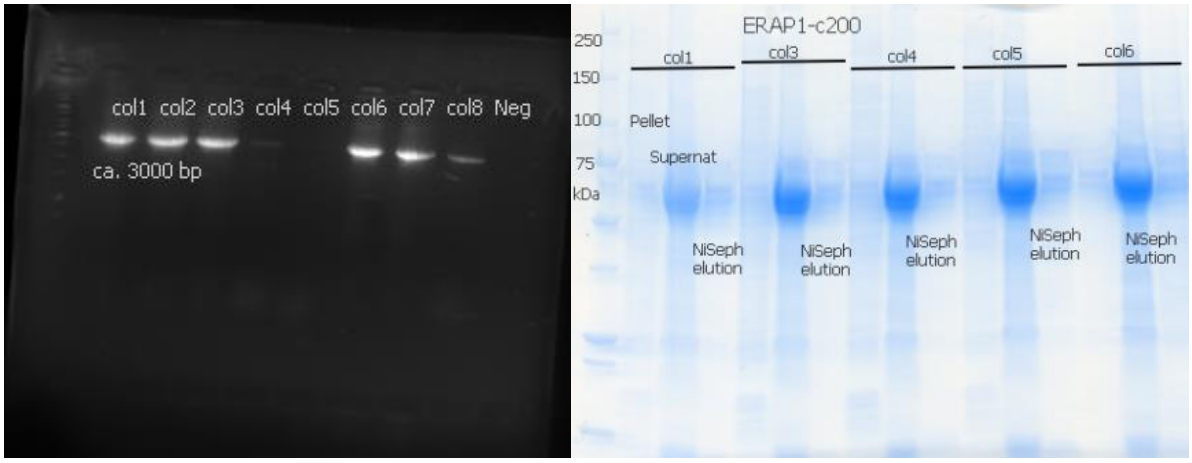


Figure 2-3. ERAP1-c200 bacmids PCR (left). Colonies 1, 2, 3, 6 and 7 have strong bands at around 3000 bp, indicating a successful gene insertion. $_{1\text{-}941}$ERAP1 gene length is 2826 base-pairs. SDS-PAGE analysis of the fractions from ERAP1-c200 test-purification (right). Test-expression from the Sf9 cultures infected with ERAP1-c200 recombinant bacmids obtained from five different DH10α colonies. Purification performed using the standard IMAC test-purification protocol. Both the cell pellet and the supernatant were analysed for protein expression. ERAP1-c200 expected mass is 110 kDa.

The recombinant ERAP1-c201 bacmid was successfully obtained using the available recombinant plasmid and transformed into the active viral stock using the standard protocol (Figure 2-4).

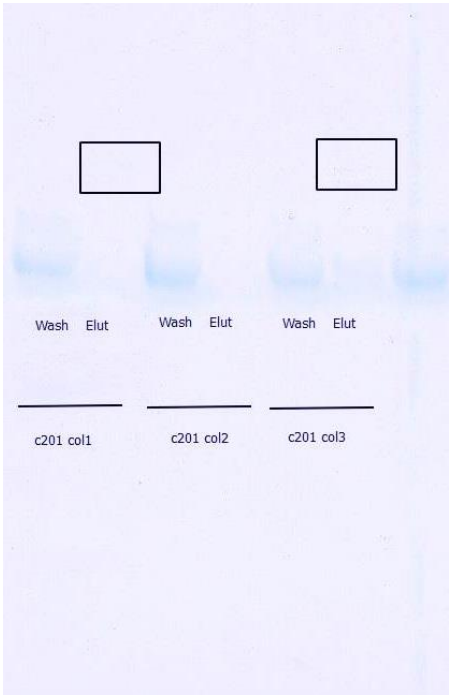


Figure 2-4. Test expression from the Sf9 cultures infected with ERAP1-c201 recombinant bacmids obtained from several DH10α colonies. ERAP1-c201 expected mass is 108 kDa. Since the protein was supposed to be secreted, only culture media was analysed. Purification performed using the standard IMAC protocol. The bands are very faint because the gel was washed for too long. The identity of the protein was confirmed by the MS-MS analysis of the tryptic digest products of the gel-band.

2.2.2 Protein expression and purification

ERAP1-c200

Initially, ERAP1-c200 was expressed using the remaining P3 viral stock from the previous work on ERAP1 in our group. By the beginning of the final year of the project this stock had depleted and all our attempts to prepare fresh viral stock were unsuccessful. Thus, it was not possible to produce more protein for further crystallisation attempts and other experiments on ERAP1. A small purified ERAP1-c200 enzyme stock was kept for the enzyme activity assays.

ERAP1-c200 was purified from Sf9 cells media after the cells infection with the virus in two steps. The first step was Ni-affinity chromatography. Initially, HisTrap HP column was used for this purpose. But since the protein eluted in many fractions (> 6 CV), Ni-sepharose beads were used instead in later purifications (Figure 2-5). The second purification step was gel-filtration (Figure 2-6).

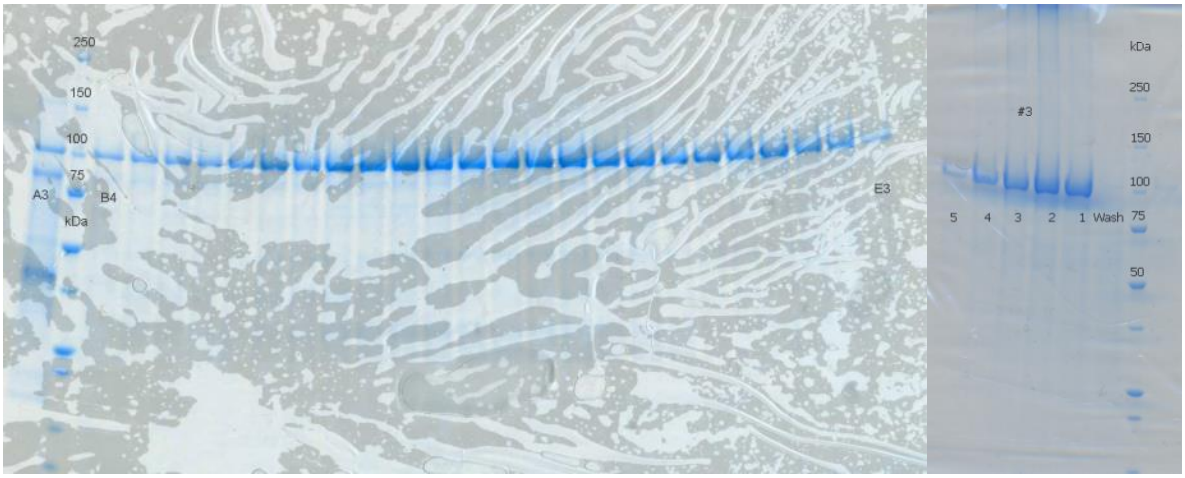


Figure 2-5. ERAP1-c200 elution from IMAC using either 5 ml HisTrap column (left) or Ni-sepharose beads in batch binding (right). Each fraction on the left gel is 0.4 column volume. Each fraction on the right gel is 1.2 column volume.

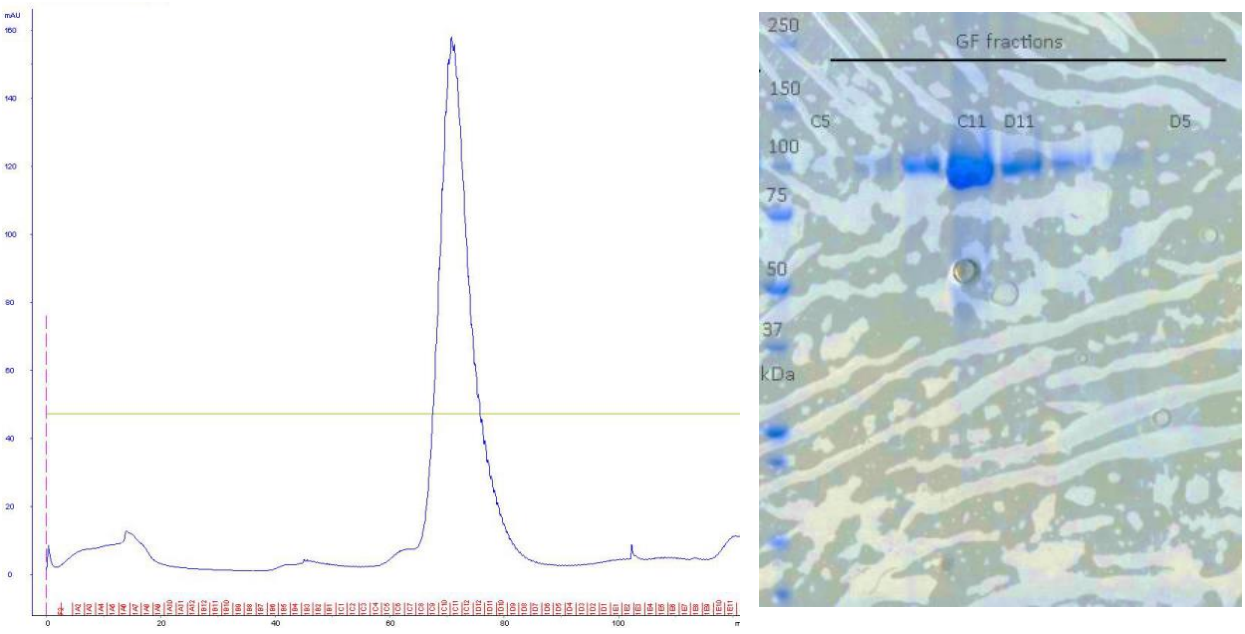


Figure 2-6. Example of a typical GF-chromatogram from the second step of ERAP1-c200 purification (left). A sharp peak indicated a monomeric protein that was well-folded. SDS-PAGE analysis of the protein-containing fraction (right) showed that the sample was not contaminated by other proteins.

The main issues with the purifications of this construct were its reproducibility, the quality and amount of the obtained protein, which varied from one batch to another:

- The obtained yields of the purified protein varied from 0.2 mg to 4.7 mg per L of the cell culture, despite using the same amount of P3 virus for the cell culture infection. This was most likely due to the variability in the insect cells used.
- Tailing (3-10 column volumes) was observed upon protein elution from the Ni-beads (Figure 2-5). Raising imidazole concentration of the elution buffer from the standard 250 mM to 500 mM did not lead to any noticeable improvement. Ni-NTA and cobalt TALON® beads were tried as an alternative. ERAP1 eluted from Ni-NTA with the same tailing effect, whereas the protein did not bind to the TALON® beads.
- Occasionally the protein partly went through the membrane with 30kDa cut-off upon concentration using Sartorius Vivaspin Turbo centrifugal filter. The flow-through ERAP1 had the same enzymatic activity as the other samples. Native MS suggested that such sample was contaminated by an unknown polymer, but still found most of the protein in a folded state (Figure 2-10). Later experiments revealed that this was likely due to the leakage of the

concentrators and the polymer detected may have been from the concentrator membrane filter.

2.2.3 Protein quality assessment

Due to the highlighted issues and continued failed attempts to crystallise ERAP1 (see Chapter 4) the quality of the obtained protein samples was assessed in more detail. Firstly, SDS-PAGE gels of the obtained samples were compared to those that had been successfully crystallised. No noticeable differences between the two were identified (Figure 2-7).

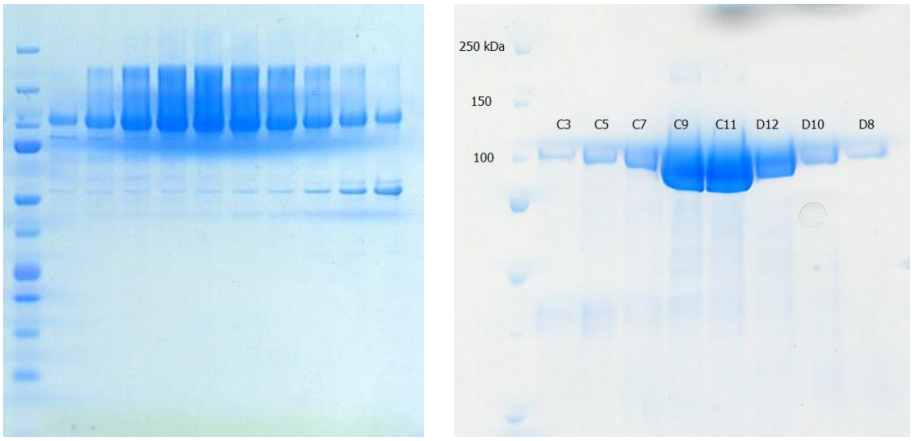


Figure 2-7. A typical SDS-PAGE analysis of the ERAP1 containing fraction after size-exclusion chromatography purification step from the earlier work in our group (left) and now (right). The same protein ladder was used on both gels.

The purified ERAP1-c200 was also analysed by intact MS (Figure 2-8). The results confirmed that the enzyme was glycosylated upon expression by the Sf9 cells. The spectra showed that all samples contained the same ERAP1 glycoforms – each glycoform having a glycan chain of a particular length. However, the distribution of the glycoforms in samples from different expressions varied. This was expected, because the glycosylations were performed by the Sf9 cells and the exact extent and pattern of the expressed protein glycosylation is not closely regulated. Such heterogeneity might be the potential reason for the repeated failures of the crystallisation experiments.

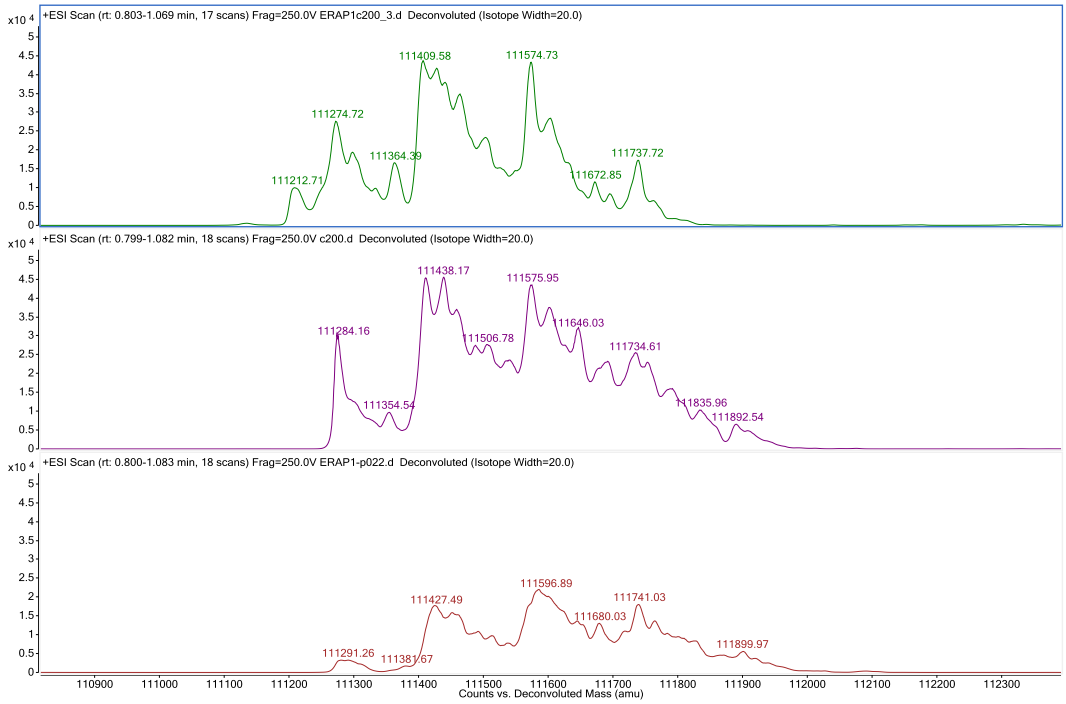


Figure 2-8. Deconvoluted intact MS spectra of three purified ERAP1-c200 samples from the different expression batches. The main peaks, each representing a particular glycoform, were the same for all samples. The distribution of the glycoforms in each protein sample varied from one expression batch to another

Considering the aforementioned issues during ERAP1 purification, the folding of the obtained samples was investigated. For this purpose native MS measurements and data analysis were performed by Dr. Rod Chalk and Oktawia Borkowska. These experiments showed that most samples had no issues with folding. The spectrum of the sample that went through the 30 kDa membrane upon concentration contained a slightly higher amount of unfolded protein than normal and an unidentified polymer contamination (Figure 2-10).

It was hypothesised that the purification issues might be due to low stability of ERAP1 in one of the buffers. Therefore a T_m-shift assay was performed on ERAP1-c200 sample in a set of the buffers used in the purification or crystalisation (Figure 2-9). The bell-shaped curves indicated the protein in the sample was properly folded. The peaks of the curves are all in the narrow range of 52-55^{o}C. This suggested that none of the buffers used significantly destabilised ERAP1and may have led to the problems with the purification.

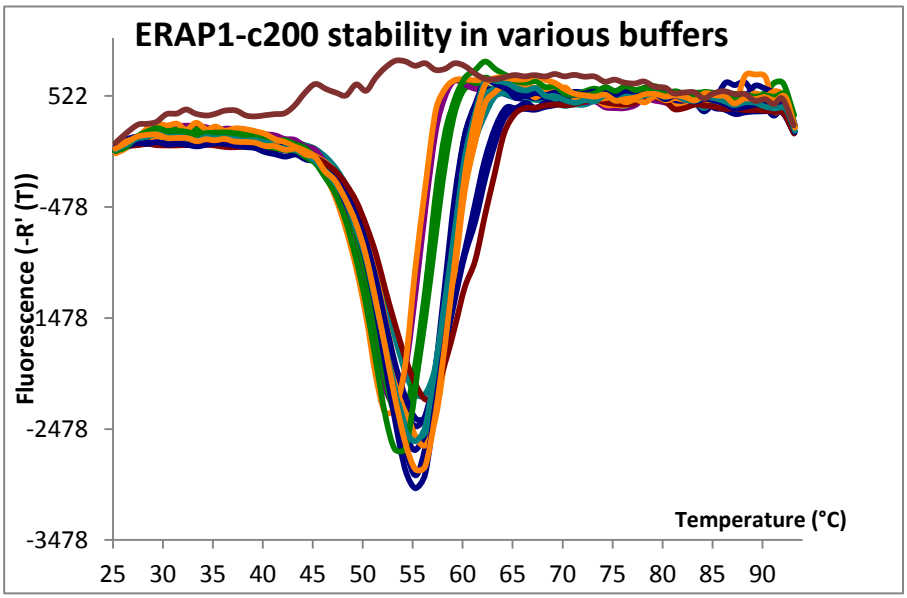


Figure 2-9. Result from the T_M-shift fluorescent measurements of ERAP1-c200 stability in various buffers used in this project. No buffer was found to significantly affect the ERAP1 stability.

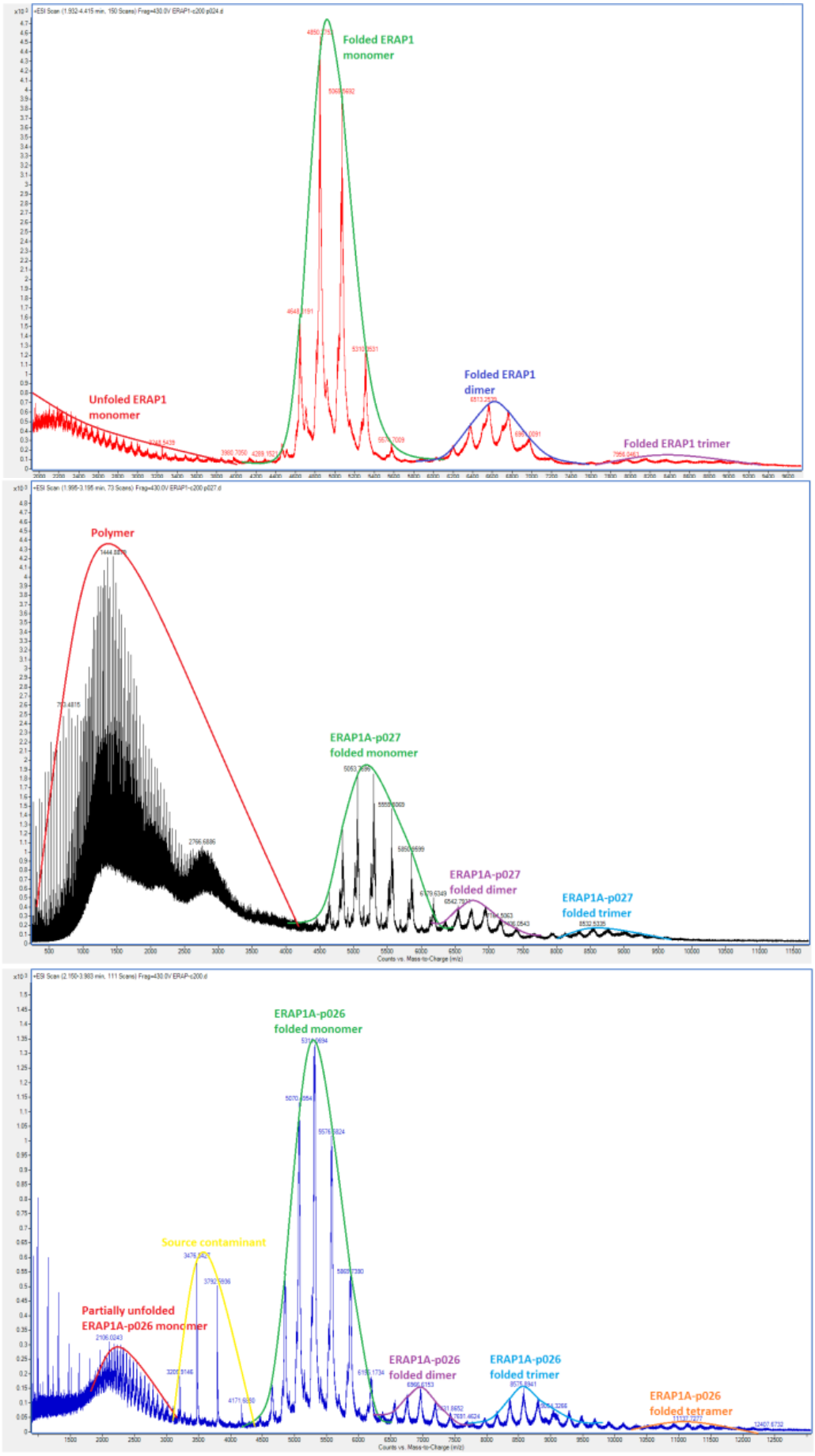


Figure 2-10. Native MS showing a well-folded ERAP1-c200 sample (top). The concentrated protein sample (bottom) and the ERAP1 sample that went through the 30 kDa membrane (middle) both indicated presence of the folded enzyme. The sample that went through contained an unidentified polymer contaminant.

As indicated by the intact MS, the protein was N-glycosylated, hence the next step was the glycosylation pattern comparison of the obtained samples with that of the crystallised protein. ERAP1 has five potential N-glycosylation sites – Asn^{70}, Asn^{154}, Asn^{414}, Asn^{760} and Asn^{901}. According to the crystal structure (PDB: 2YD0), ERAP1 had three N-glycosylated residues – Asn^{70}, Asn^{154} and Asn^{414}. The observed mass of the glycosylated ERAP1-c200 was 111284 Da (Fig. 8), which was 2544 Da higher than that of the deglycosylated protein (Figure 2-12), this being the total mass of all N-glycans. In the insect cells a usual N-glycan had the core structure of NAG-NAG-Man-$(Man)_2$ (MW = 910 Da), which also may be fucosylated at the first sugar. [66] It was concluded that the ERAP1 glycans were not fucosylated, since the protein was successfully deglycosylated, and PNGase F is not capable of the cleavage of fucosylated N-glycans. So it was proposed that the purified ERAP-c200 most likely also had three N-glycosylations, because the calculated total glycan mass could accommodate 2.8 proposed glycan cores. Comparing the tryptic digest products' of the glycosylated and deglycosylated enzyme samples one glycosylation site was identified - Asn^{414} (Appendix III). Typically, if a protein was expressed using the same insect cell line the N-glycosylation pattern was expected to remain the same, but the length of the glycans varied from one expression to another.

The intact MS spectrum of the deglycosyltated protein indicated that the deglycosylated ERAP1 was much more homogenous – there was only one major protein entity in the sample rather than a mixture of different glycoforms. The observed mass was 108740 Da, while the expected mass of the full-length $_{1\text{-}941}$ERAP1 was 110555 Da (Figure 2-12). The identity of the protein was confirmed by the MS-MS of sample tryptic digest products (Appendix II). In addition, the clearly observed Zn-adducts on the intact MS also indicated that it was ERAP1, since the protein contained a zinc binding motif. Thus, the intact MS results were consistent with the deglycosylated protein being a truncated ERAP1-c200.

The difference between the expected full-length mass and the observed mass of the protein was consistent with the truncation of approximately 16 residues. From the MS-MS analysis of the

sample's tryptic digest products (Appendix II) it was possible to conclude that the truncated protein contained residues 40-940. This was in agreement with the published crystal structure (PDB: 2YD0), where residues 46-940 were modelled. It was not possible to identify exactly the residues that were truncated, which was to be expected for such a large protein with many potential post-translational modifications.

Due to a much higher homogeneity of the deglycosylated ERAP1-c200 sample, it was decided to try preparing a large ERAP1-c200 deglycosylated sample for the crystallisation. The large scale deglycosylation was successful, according to the intact MS (Fig. 2-12). It was not possible to remove PNGase from the ERAP1 sample neither by IMAC nor by concentrating using a centrifugal filter with a 50 kDa cut-off (Figure 2-11). Surprisingly, the native MS identified only glycosylated ERAP1. Supposedly, the deglycosylated protein is less stable and precipitates from low salt buffers, which were used to prepare the sample for native MS.

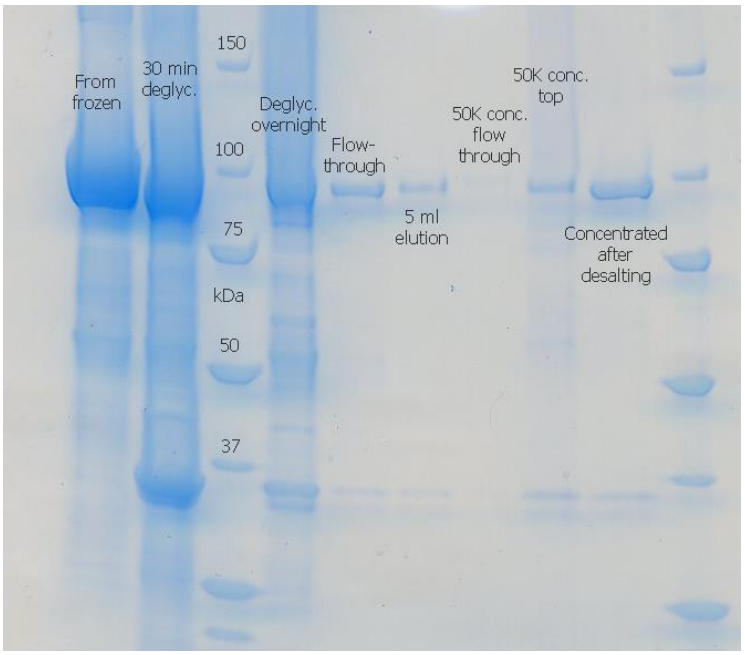


Figure 2-11. SDS-PAGE analysis of samples from the large scale ERAP-c200 deglycosylation. PNGase F molecular weight is ca. 36 kDa. It was not removed neither by IMAC chromatography nor by concentration using a centrifugal filter with a large cut-off.

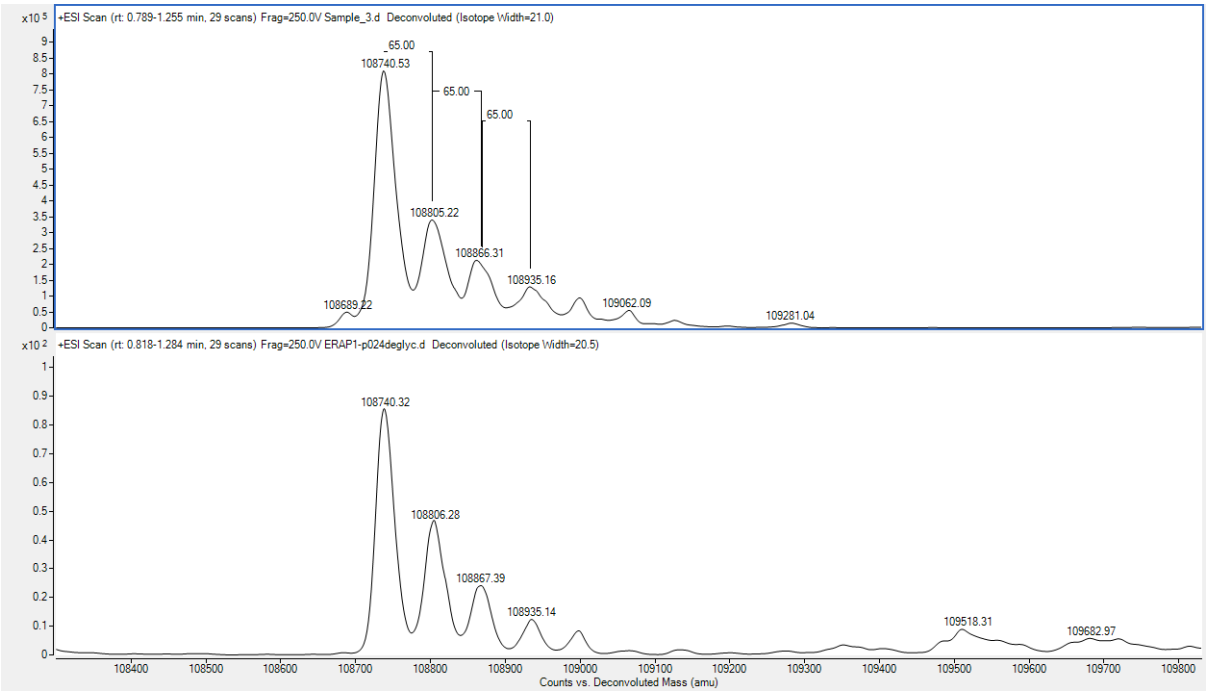


Figure 2-12. Deconvoluted intact MS spectra of the small (top) and large (bottom) scale delgycosylations of ERAP1-c200 using the PNGase F. The minor peaks were attributed to formation Zn-adducts (+ 65 Da), as ERAP1 contained Zn-binding motif. The MS signal of the bottom sample is much weaker, indicating a much lower protein concentration.

The analysis of the ERAP1 crystal structure (PDB: 2YD0) showed that the glycans might be potentially important for the protein-protein interactions in the crystal due to their proximity to the neighbouring protein molecule's residues (Figure 2-13). Hence, it was decided not to optimise the purification and crystallisation of the deglycosylated ERAP1-c200. Instead, the attempts to crystallise the shorter ERAP-c201 construct were made (see Chapter 4.2).

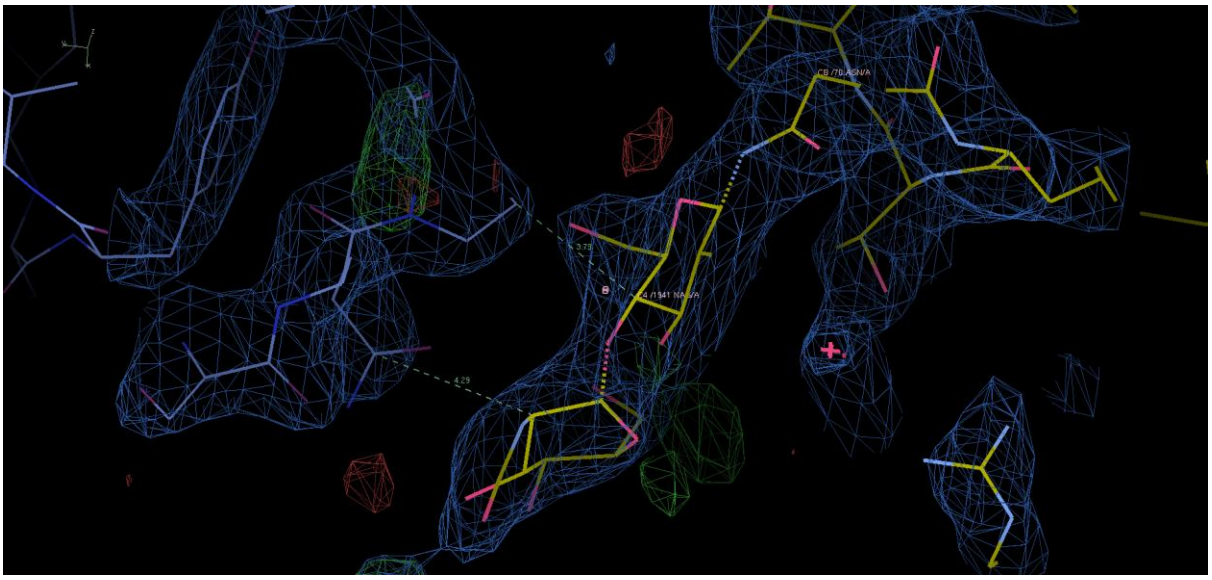


Figure 2-13. Asn^{70} glycan NAGs are in close proximity to the Gln^{88} and Pro^{89} of the neighbouring molecule. A typical distance for symmetry-related molecules in the crystal is considered to be approximately 4Å. [111] The image was created in Coot using 2YD0 crystal structure at 1.22 msd.

ERAP1-c201

ERAP1-c201 was successfully expressed and purified from the Sf9 cells media after the cells infection using the freshly obtained viral stock (Figure 2-14). The first purification step was Ni-affinity chromatography using Ni-sepharose beads for batch binding followed by overnight TEV-cleavage. The second step was the gel-filtration of the TEV-cleaved protein sample. The IMAC and GF purification protocols were the same as for ERAP1-c200.

The same problems were encountered upon the protein purification as with the full-length protein. In an attempt to improve the purification, ion exchange chromatography was tried as an alternative method to the gel filtration. The test purification worked successfully, but when this method was applied to the full-scale purification no protein bound to the ion-exchange column.

The analysis of the native MS of a TEV-cleaved purified ERAP1-c201 sample was performed with help by Dr. Rod Chalk and Oktawia Borkowska (Figure 2-15). It showed that the predominant protein glycoform had all its five sites N-glycosylated. In addition, glycoforms with N-glycosylation states 4 and 0 were observed. Thus the protein from this construct was even more heterogeneous than the

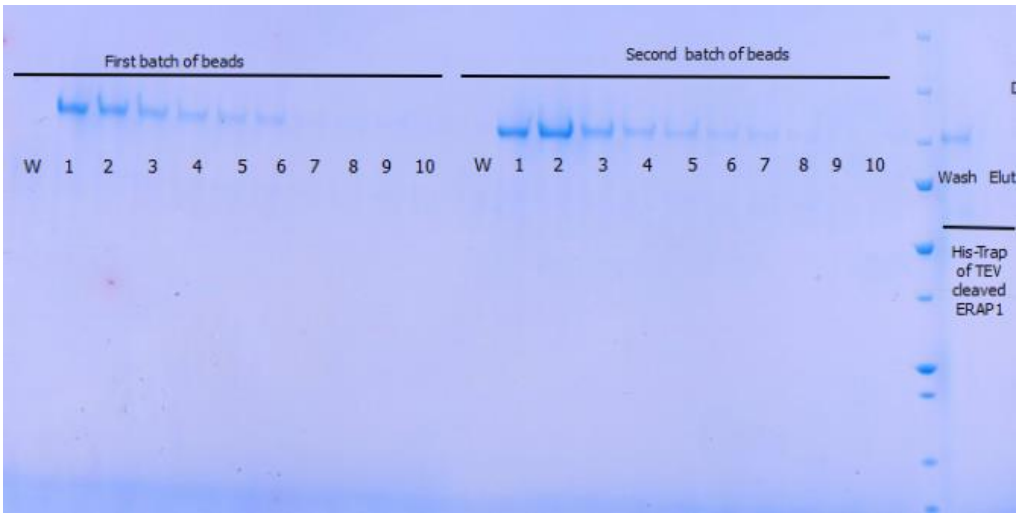


Figure 2-14. ERAP1-c201 elution from Ni-sepharose beads. Each elution fraction is 1.2 column volumes. The fractions from the purification of the TEV-cleaved ERAP-c201 are on the last two lanes on the gel. The TEV-cleavage was successful, since the protein did not bind to the column (it was in the wash fraction, not the elution). After the Sf9 media was incubated with the first nickel beads batch, there was still some ERAP1 left in the media. Thus, another batch of nickel beads was used to bind the remaining protein.

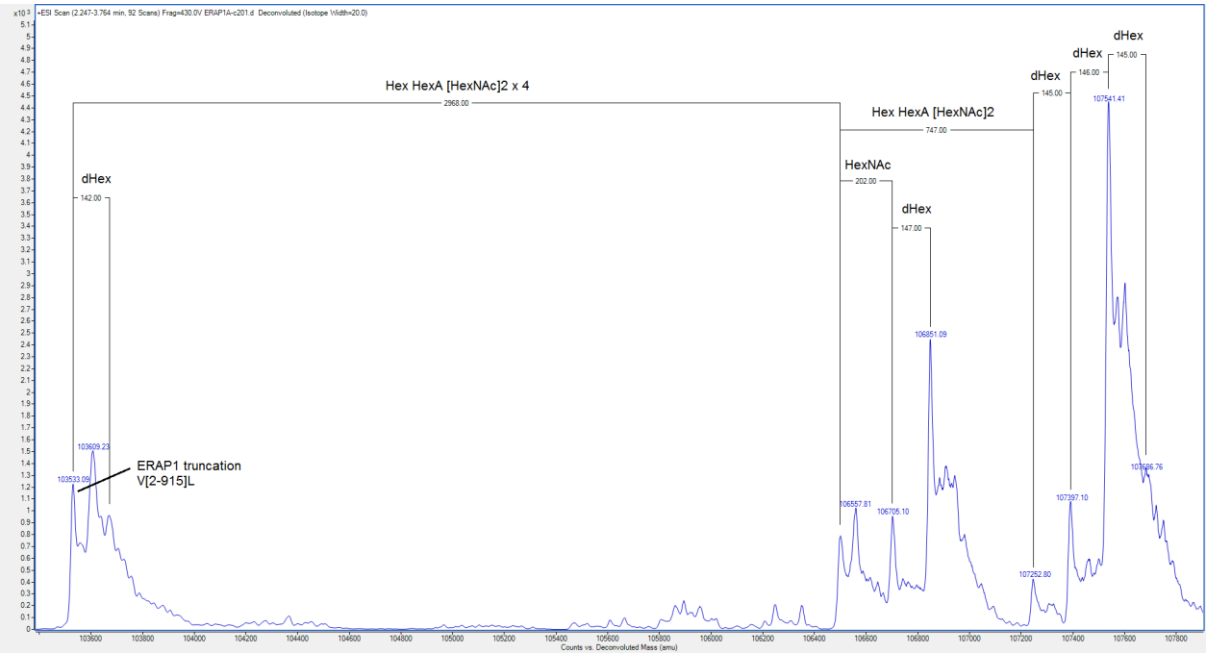


Figure 2-15. Analysis of the deconvoluted native MS of ERAP1-c201 TEV-cleaved purified sample with proposed glycan structures.

ERAP-c200, which may explain why the attempts to crystallise this protein failed (see Chapter 4.2).

2.3 ERAP2

Previously, there was an attempt in our group to clone, express and purify full-length human $_{1\text{-}960}$ERAP2 construct with C-terminal His-tag using the pFb-CT10HF-LIC vector and the same Bac-to-Bac® approach as for ERAP1-c200 above. The recombinant bacmid was successfully obtained, but no ERAP2 expression was identified by test purification. However, another group reported its successful cloning, expression and purification [67] using "21-bp A-rich sequence derived from a lobster tropomyosin cDNA leader sequence adjacent to the initiation codon of ERAP2", which was supposed to enhance the gene expression. [68] They used the same pFb plasmid as a vector. Therefore, it was attempted to reproduce their results and used a construct with the same leader sequence ($_{1\text{-}960}$ERAP2-c002).

2.3.1 Cloning and viral stock preparation

ERAP2-c002 was obtained by PCR from the full-length human $_{1\text{-}960}$ERAP2 entry clone and the relevant primers. The PCR product was successfully cloned into the pFb-CT10HF-LIC plasmid (Figure 2-16).

The recombinant plasmid formation was confirmed by its sequencing and it was used for preparation of the recombinant bacmid *via* transformation in DH10α *E. coli* cells. The gene insertion was verified by PCR of the purified bacmid samples. Unfortunately, none of the numerous attempts

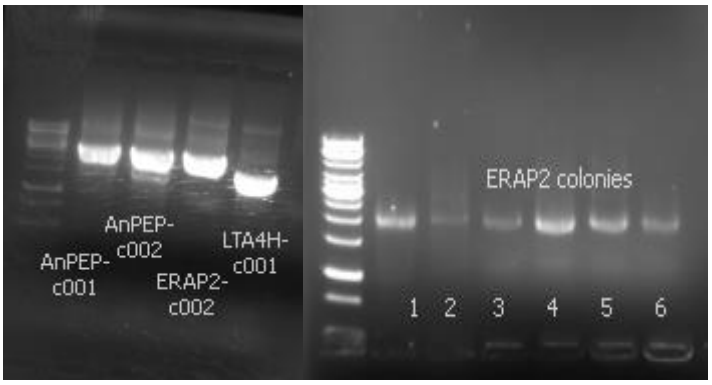


Figure 2-16. PCR product of the selectivity panel enzymes constructs (left). $_{33\text{-}976}$AnPEP gene is 2832 base-pairs, $_{1\text{-}960}$ERAP2 gene is 2883 base-pairs and $_{1\text{-}611}$LTA4H gene is 1836 base-pairs. PCR of the plasmids purified from Mach1 colonies using ERAP2-c002 specific primers (right). The purified plasmids were sequenced to confirm cloning.

to prepare the recombinant bacmid proved to be successful according to the PCR of the purified bacmids.

2.4 AnPEP

The full-length human $_{1\text{-}976}$AnPEP contains a membrane binding anchor [69], but to facilitate the protein purification two constructs encompassing only the soluble part, residues 59 to 976 were cloned. AnPEP-c001 contained a C-terminal TEV-cleavable His-tag and AnPEP-c002 contained a N-terminal TEV-cleavable His-tag.

2.4.1 Cloning and viral stock preparation

AnPEP-c001 and AnPEP-c002 were obtained by PCR from full-length human $_{1\text{-}976}$AnPEP entry clone using the relevant primers (Figure 2-16). The PCR products were successfully cloned into the pFb-CT10HF-LIC and the pFb-Sec-NH plasmids respectively (Figure 2-17). The recombinant plasmid formation was confirmed by DNA sequencing and these plasmids were used for preparation of AnPEP-c001 and AnPEP-c002 recombinant bacmids as described previously. The successful cloning of

the correct size insert was confirmed by PCR of the purified bacmid samples (Figure 2-17). Numerous attempts were made to transform various batches of the obtained recombinant bacmids into the active viral stock. All of them were unsuccessful. The Sf9 cells looked infected under the microscope, but no AnPEP protein was isolated upon test purification (Figure 2-18). The protein band in the

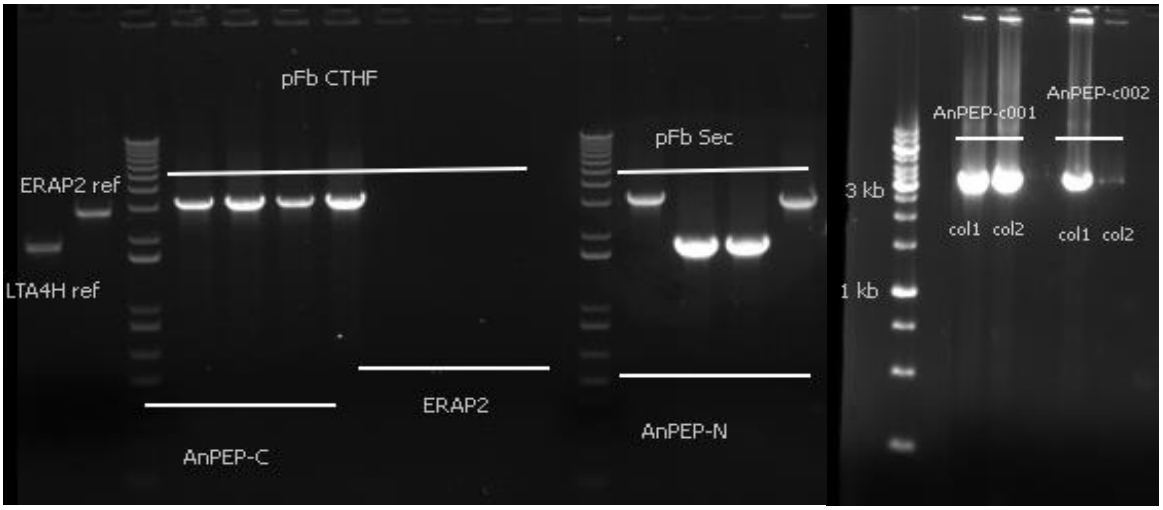


Figure 2-17. PCR of the plasmids purified from Mach1 colonies (left) using AnPEP-c001 and AnPEP-c002 specific primers. The purified plasmids were sequenced to confirm cloning. ERAP2 and LTA4H PCR products used as a DNA size reference. AnPEP-c001 and AnPEP-c002 bacmids PCR (right). Bacmids from both DH10α colonies transformed with AnPEP-c001 plasmid strong bands at around 3000 bp, indicating a successful gene insertion. Bacmid from the DH10α colony 1 transformed with AnPEP-c002 plasmid strong bands at around 3000 bp, indicating a successful gene insertion.

supernatant flow-through was attributed to an unknown Sf9 cell secreted protein. The same band was observed in test expressions of ERAP1-c200 (Figure 2-3).

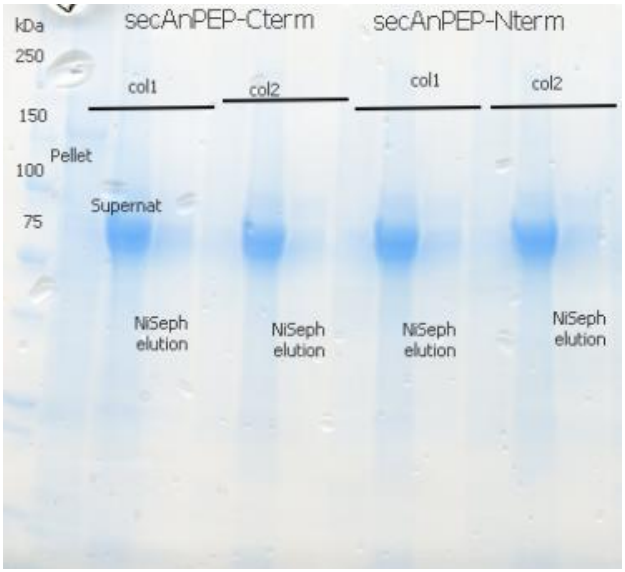


Figure 2-18. Test-expression from the Sf9 cultures infected with AnPEP-c001 and AnPEP-c002 recombinant bacmids obtained from several DH10α colonies. Purification was performed using the standard IMAC protocol. Both the cell pellet and the supernatant were checked for protein expression. AnPEP-c001 and AnPEP-c002 expected mass is 103 kDa.

2.5 LTA4H

2.5.1 Cloning

The LTA4H-c001 construct was obtained by PCR from the full-length human $_{1\text{-}611}$LTA4H entry clone using the relevant primers. The PCR product was successfully cloned into the pNic28 plasmid containing N-terminal TEV-cleavable His-tag to facilitate its purification (Figure 2-19). The recombinant plasmid formation was confirmed by sequencing, before it was transformed into *E.coli* BL21 competent cells, which were used for LTA4H expression.

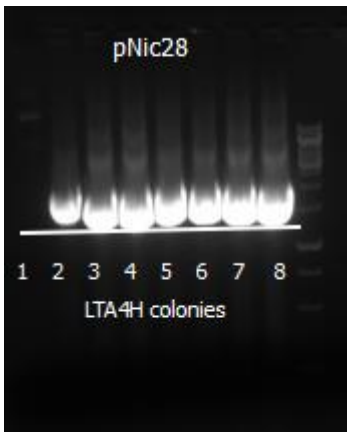


Figure 2-19. PCR of the plasmids purified from pNic28 colonies using LTA4-c001 specific primers. The purified plasmids were sequenced to confirm that the right gene was cloned.

2 Enzyme cloning, expression and purification

2.5.2 Protein expression and purification

LB, TB and PowerBroth were tried as a potential media for cell growth. The highest protein yield was obtained for the TB-cultured protein sample according to the protein band intensity. Therefore, this media was further used for large scale protein expression (Figure 2-20).

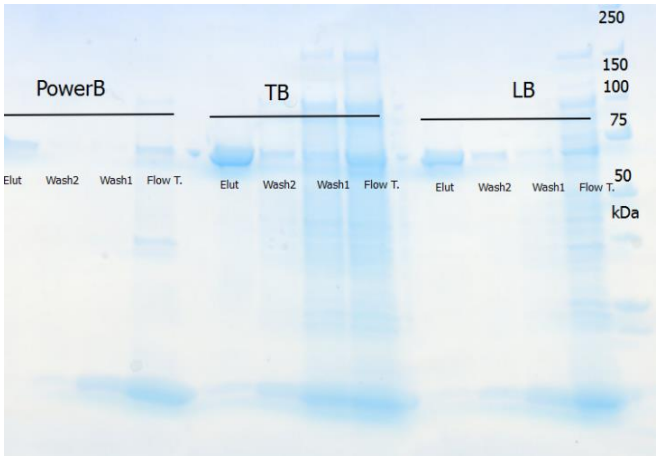


Figure 2-20. Comparison of the LTA4H expression using different growth media on the same 5 ml scale. Test-purification performed using the standard IMAC protocol. The thickest LTA4H band was obtained from the TB culture. The identity of the protein was confirmed by MS-MS analysis of the tryptic digest product of the gel-band.

The purification of LTA4H from the lysed bacterial cells was performed in two stages. The first step was Ni-affinity chromatography using Ni-sepharose beads, followed by gel-filtration (Figure 2-21). The TEV-cleavage was performed in between the two purification steps overnight, to obtain TEV-cleaved enzyme samples.

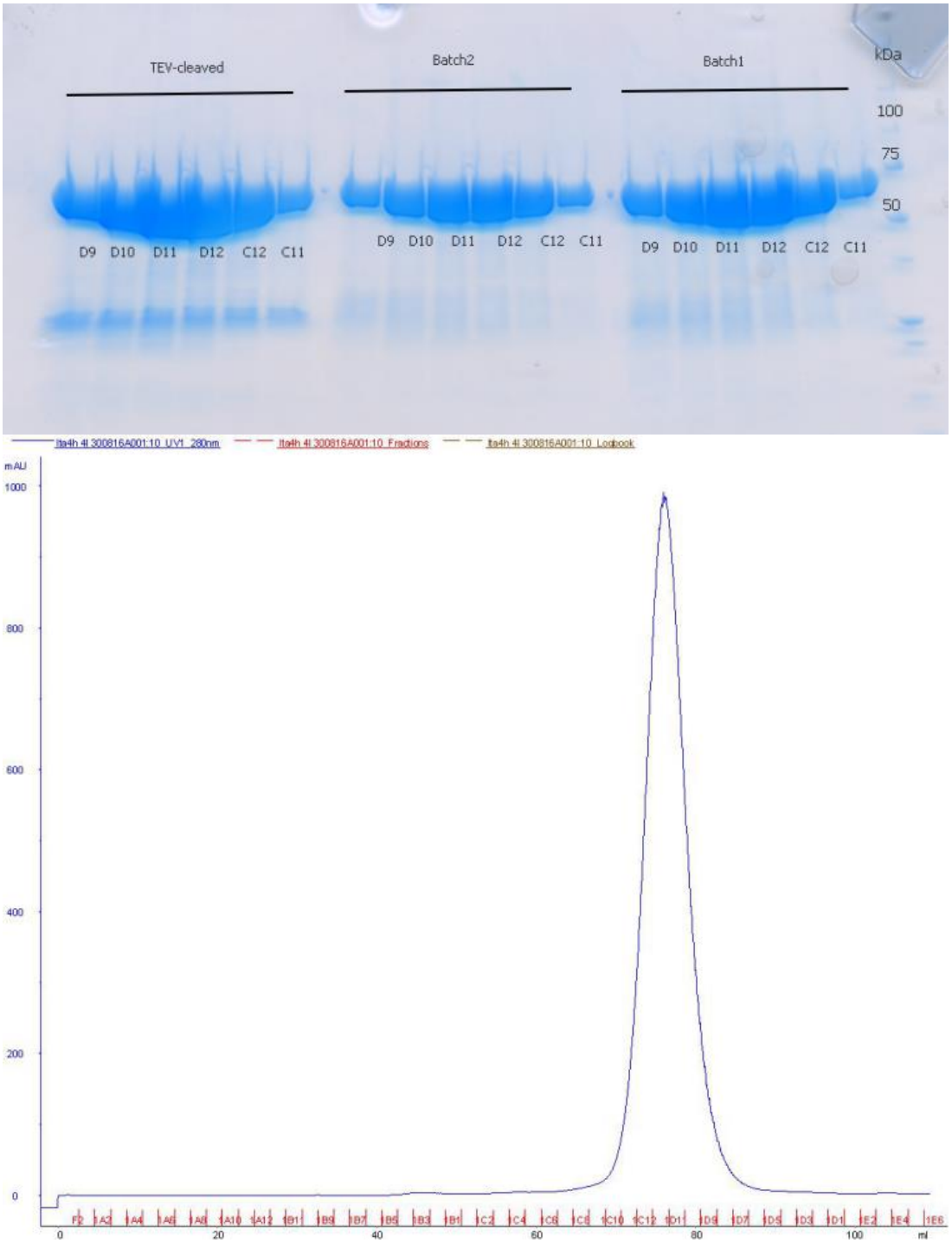


Figure 2-21. SDS-PAGE analysis (top) of the fraction obtained after S200 GF-purification of LTA4H-c001. Due to the high yield and limited capacity of the GF column, the protein was purified in several batches. The TEV-cleaved LTA4H was purified separately. GF chromatogram (bottom) of the first batch purification.

The obtained LTA4H samples were analysed by the intact MS (Fig. 23). The observed mass was 71833 Da and matched the expected mass of 71838 Da with the mass error of -67 ppm. The protein identity was confirmed by the MS-MS of sample tryptic digest products (Appendix II). The spectrum showed partial loss of N-terminal methionine, which was expected for bacterial expression. [70] TEV-cleavage was confirmed by the intact MS (Figure 2-22).The observed mass of the TEV-cleaved LTA4H was 69368 Da and matched the expected mass of 69372 Da with the mass error of 60 ppm.

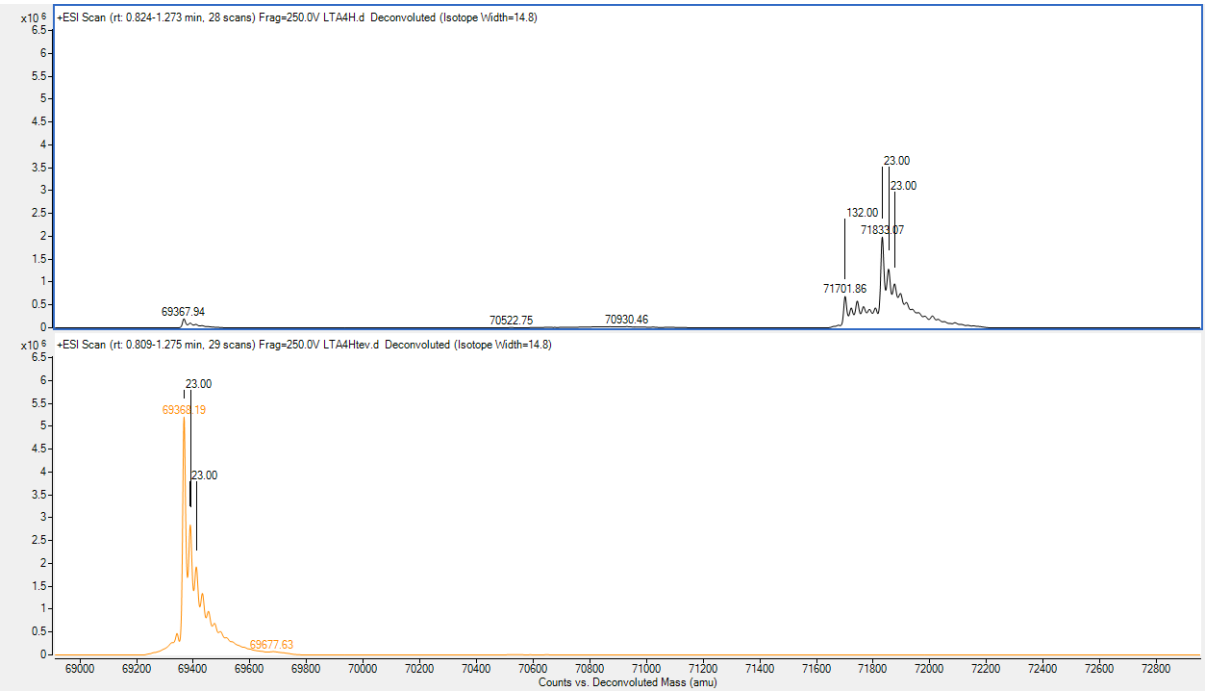


Figure 2-22. Intact MS spectra of the full-length LTA4H sample (top) and the TEV-cleaved sample (bottom). Sodium adducts (+ 23 Da) were observed on both spectra. Partial loss of Met1 (-132 Da) was observed for the full-length LTA4H

Conclusion

ERAP1-c200 and ERAP1-c201 have been successfully expressed and purified. The quality of the samples has been investigated using a set of analytical techniques and extensive analysis of the enzyme N-glycosylation has been performed. It has been concluded that the obtained ERAP1-c200 samples had similar glycosylation pattern as in the crystallised 2YD0 structure. The enzyme samples obtained for the both ERAP1 constructs have been found to be heterogeneous due to presence of several protein glycoforms, which is the probable reason why the attempts to crystallise ERAP1 have failed.

One out of three selectivity panel enzymes –LTA4H – has been successfully cloned, expressed using the BL21 expression system and purified. Recombinant bacmids carrying the *AnPEP* gene have been prepared, but no protein has been isolated upon test-expressions. A recombinant ERAP2 plasmid has been obtained and confirmed by sequencing. Unfortunately, we have never managed to transform this plasmid into the recombinant bacmid for reasons unknown. Thus, ERAP2 and AnPEP samples for the enzyme selectivity panel were not obtained.

2.6 Methods

Primer sequences used in this work

Construct	Target plasmid	Primer	Sequence
ERAP1-c200	pFb-CT10HF-Lic	Forward	TTAAGAAGGAGATATACTA TGGTGTTTCTGCCCCTCAAAT
		Reverse	GATTGGAAGTAGAGGTTCTCTGC CATACGTTCAAGCTTTTCAC
ERAP2-c002	pFb-CT10HF-Lic	Forward	TTAAGAAGGAGATATACTA CTCCTAAAAAACCGCCACCATGTTCCATTCTTCTGC
		Reverse	GATTGGAAGTAGAGGTTCTCTGC AGTATTAACCATTAGCCAAG
AnPEP-c001	pFb-CT10HF-Lic	Forward	TTAAGAAGGAGATATACTA TGCCCGCCTCGGCCACC
		Reverse	GATTGGAAGTAGAGGTTCTCTGC TTTGCTGTTTTCTGTGAACC
AnPEP-c002	pFb-Sec-NH	Forward	TTAAGAAGGAGATATACTA TGCCCGCCTCGGCCACC
		Reverse	TATCCACCTTTACTGCTACTATTTGCTGTTTTCTGTGAACC
LTA4H-c001	pNic28	Forward	TACTTCCAATCCATGCCCGAGATAGTGGATACC
		Reverse	TATCCACCTTTACTGTTAATCCACTTTTAAGTC

PCR product preparation

ERAP1-c200, ERAP2-c002, AnPEP-c001, AnPEP-c002 and LTA4H-c001 PCR products were obtained using the same PCR protocol from the appropriate entry clone using the relevant set of primers. A reaction contained 10 µl Herculase II 10x buffer, 1.25 µl dNTP, 35.75 µl DNAse-free water, 2 µl of 10 uM primer mixtures and 1 µl of 10 ng/µl DNA template. PCR cycle: 95^{o}C, 2 minutes; (95^{o}C, 45 seconds; 55 ^{o}C, 30 seconds, 72 ^{o}C, 4 minutes) x 45 cycles; 72 ^{o}C, 7 minutes. The PCR products were purified using the ThermoScientific PCR purification kit.

Protocol to remove SacB from a plasmid

5 µg of the plasmid (pFb-CT10HF-Lic or pFb-Sec-NH) were mixed with 20 µl NEBbuffer3, 20 µl of BfuAI restriction enzyme (80 units) and supplemented with DNAse-free water to make the final volume to 200 µl. The mixture was incubated at 50^{o}C overnight. The plasmid mixture was used for the next step without purification

2 Enzyme cloning, expression and purification

Ligase Independent cloning protocol

The BfuAI-digested plasmid was T4 treated using the standard SGC protocol. 50 ul of the plasmid was mixed with 13.5 µl DNAse-free water, 20 µl 5x NEB2 buffer, 10 µl 25 mM dCTP, 1 µl BSA, 0.5 µl 1M DTT and 5 µl Novagen T4 DNA polymerase. The mixture was incubated for 30 minutes at room temperature, then at 75 ^{o}C for 20 minutes to quench the reaction.

The PCR products were T4 treated using the standard protocol obtained from SGC. 5 µl DNA were mixed with 0.9 µl DNAse-free water, 2 µl 5x NEB2 buffer, 1 µl 25 mM dGTP, 0.1 µl BSA, 0.5 µl 0.1M DTT and 0.5 µl Novagen T4 DNA polymerase. The mixture was incubated for 30 minutes at room temperature, then at 75^{o}C for 20 minutes to quench the reaction.

Then 1µl of the T4-treated plasmid was mixed with 2 µl of the T4 treated insert and incubated for 15 minutes at room temperature. Then the mixture was transferred on ice, supplemented with 50 µl competent Mach1 cells and incubated for 30 minutes. The cells were heat-shocked for 45 seconds at 42^{o}C, then supplemented with 100 µl SOC and incubated for 90 minutes at 37 ^{o}C. Mach1 cells mixed with pNic28 plasmid were cultured at 37 ^{o}C on LB plates with 50 µg/ml kanamycin and 5% sucrose. Mach1 cells mixed with pFb plasmids were cultured at 37 ^{o}C on LB plates with 50 µg/ml ampicillin and 5% sucrose. The bacterial colonies emerged the next day. They were amplified in 5 ml LB cultures with the appropriate antibiotic overnight at 37 ^{o}C. The recombinant plasmids were purified the bacterial culture using the ThermoScientific plasmid mini-prep purification kit. The cloning was confirmed by sequencing of the obtained plasmids.

ERAP1-c200, ERAP2-c002, AnPEP-c001 and AnPEP-c002 bacmids preparation using the obtained plasmids.

2 µl of a recombinant plasmid were supplemented with 60 µl of defrosted DH10α competent cells and incubated for 15 minutes on ice. The cells were heat-shocked for 45 seconds at 42^{o}C, then supplemented with 500 µl SOC and incubated for 4 hours at 37 ^{o}C. Then the cells were cultured at 37

^{o}C on LB plates with 50 μg/ml kanamycin, 7 μg/ml gentamycin, 10 μg/ml tetracycline, 40 μg/ml IPTG and 5% sucrose. The plates were supplemented with 100 μg/ml Blue-gal for blue white colony screening. After 2 day the white colonies were picked and cultured on LB plates with the same conditions. In another 2 days the white colonies were amplified in 5 ml LB cultures with the same mixture of antibiotics and IPTG overnight at 37^{o}C. The bacmids were purified using the ThermoScientific plasmid mini-prep purification kit. The recombinant bacmid formation was confirmed by PCR.

Bacmid PCR protocol

1 μl of the purified bacmid was mixed with 2 μl 10x Tac buffer, 1 μl 10μM FBac1/M13 primer mixture, 1.6 μl 2.5 mM dNTP mixture, 15 μl DNAase-free water and 0.1 μl DreamTac polymerase. PCR cycle: 94^{o}C, 5 minutes; (94^{o}C, 45 seconds; 55^{o}C, 45 seconds; 72^{o}C, 5 minutes) x30 cycles; 72^{o}C, 7 minutes.

Bacmid transformation into the viral stock

GeneJuice®

The transfection was performed using the manufacturer's protocol (Novagen). 1 ml of Sf9 cells (concentration: 0.2 x 10^{6} cells/ml) were left in a plate well for 1 hour to allow cells to stick to the bottom. 2 μg were diluted with Sf-900™ II SFM media (Gibco®) to give the total volume of 60 μl. 12 μl of GeneJuice® were diluted with 48 μl Sf-900™ II SFM media (Gibco®). The DNA solution was added drop-wise to the transfection reagent, stirred mildly and incubated on the bench for 15 minutes. Then this mixture was added to the aspirated Sf9 cells. The insect cells were supplemented with Sf-900™ II SFM media (Gibco®) to make the total well volume 0.33 ml. The cells were incubated for three days at 27^{o}C. The cells were spun down by centrifugation and the obtained supernatant was kept as P0 viral stock.

GeneJet®

The transfection was performed using the manufacturer's protocol (SignaGen). 1 ml of Sf9 cells (concentration: 0.2 x 10^6 cells/ml) were left in a plate well for 1 hour to allow cells to stick to the bottom. 50 µl of the Sf-900™ II SFM media (Gibco®) were supplemented with 5 µl of GeneJet® and the bacmid (1 µg per 1 µl of GeneJet®). The mixture was incubated for 30 minutes on the bench, and then added to the aspirated Sf9 cells. After 4 hours, the cells were gently supplemented with 1 ml of Sf-900™ II SFM media (Gibco®). The cells were incubated for six days at 27^{o}C. The cells were spun down by centrifugation and the obtained supernatant was kept as P0 viral stock.

Test expression and purification protocol of the Sf9 expressed enzymes

3 ml of the Sf9 cells (concentration 1 x 10^6 cells / ml) were infected with 120 µl of P0 viral stock. Following 72 hours or 96 hours post infection incubation at 27 ^{o}C Sf9 cells were spun down by centrifugation. The obtained supernatant was used as protein source and was supplemented with Tris pH = 8.0, NaCl and $NiSO_4$ to give the final concentrations of 50 mM, 300 mM and 1 mM respectively. 200 µl per 3 ml of the protein-containing solution 50% Ni-NTA resin pre-equilibrated with the 50 mM HEPES pH 8.0 were added to the supernatant and stirred for 30 minutes. Then the suspension was loaded onto a sintered gravity column and washed with the 10 CV of washing buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10mM imidazole, 5% glycerol) and eluted with 1.5 column volumes elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 250mM imidazole). The eluted samples were analysed by SDS-PAGE for the presence of the protein of interest.

ERAP1 expression and purification

Spodoptera frugiperda Sf9 cell line was used for protein expression. Infections were performed according to the manufacturer instructions (Invitrogen). 1 ml of P3 viral stock was used per 1 l of Sf9 cells culture (concentration 1 x 10^6 cells / ml). Following 72 h post infection incubation at 27 ^{o}C cells

were removed by centrifugation. The obtained supernatant was used as protein source. It was supplemented with Tris pH = 8.0, NaCl and $NiSO_4$ to give a final concentration of 50 mM, 300 mM and 1 mM respectively. The mixture was incubated at 4^{o}C for one hour and the precipitate was removed by centrifugation. 3 ml per 1 l 50% Ni-NTA resin pre-equilibrated with the 50 mM HEPES pH 8.0 were added to the clear solution and stirred for 30 minutes. The suspension was loaded onto a sintered gravity column, washed with 20 column-volumes of washing buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10mM imidazole, 5% glycerol), and eluted with the buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 250mM imidazole. Fractions containing the protein were combined, concentrated to 2.2 ml using a Sartorius Vivaspin Turbo 30 kDa concentrator. For ERAP1-c201 samples, TEV-cleavage was performed overnight at 4^{o}C using 80 µg TEV-protease per 1 mg of the cleaved protein. The obtained protein samples was purified further by size-exclusion chromatography (Hiload 16/60 Superdex 200 column GE healthcare equilibrated with the buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol). Protein containing fractions were pooled together and concentrated to the relevant concentration using a Sartorius Vivaspin Turbo 30 kDa concentrator. The purified protein samples were stored at -80 ^{o}C for further use.

For the ion exchange purification, the IMAC purified non-concentrated protein-containing fractions were dialysed into buffer with 50 mM Tris pH 8.0 (buffer A) overnight at 4 ^{o}C. The sample was applied onto 5 ml Sepharose Q HiTrap HP column GE Healthcare pre-equilibrated with buffer A and eluted with 20 CV buffer B (50 mM Tris pH 7.5, 1 M NaCl).

LTA4H test-expression and purification

BL21 cells were grown in 10 ml TB, PowerB or LB at 37^{o}C all containing 50 µg/ml kanamycin. The protein expression was induced with 1 µl 1M IPTG when OD600 reached 2.5 for TB and PowerB, and 0.8 for LB. The bacteria were cultured overnight at 18^{o}C. Three 1ml samples of each culture were supplemented with glycerol to give final concentration of 18% and frozen at -80^{o}C as stock cultures. These samples were used for large scale expression. For each culture, the cells were spun down

(3200g, 30 min, 4°C), the supernatant discarded and the pellets resuspended in 3 ml lysis buffer (50 mM HEPES, 0.5M NaCl, 10 mM imidazole, 5% glycerol). The suspension was supplemented with 10 mg lysozyme and lysed by ultrasound. Cell debris was removed by centrifugation (17000g, 30 min, 4°C). The supernatants were applied onto gravity columns with 0.5 ml 50% Ni-sepharose pre-equilibrated with the washing buffer (the same buffer as that used for lysis), then washed with 2.5 ml washing buffer, followed by 2.5 ml washing buffer containing 20 mM imidazole and finally eluted with 1 ml elution buffer (50 mM HEPES, 0.5M NaCl, 250 mM imidazole, 5% glycerol). Samples were analysed by SDS-PAGE.

LTA4H expression and purification

LTA4H was expressed using the BL21(DE3)-R3-pRARE2 *E.coli* competent a phage-resistant derivative of BL21(DE3) carrying a pRARE2 plasmid to enhance expression of eukaryotic proteins that contain codons rarely used in *E. coli*. Cells were grown in TB at 37°C in presence of 50 µg / ml kanamycin until OD_{600} reached 2.5, then the protein expression was induced with 0.5 mM IPTG, followed by overnight incubation at 18°C. Cells were collected by centrifugation (4000 rcf, 30 min, 4°C), and the pellets were resuspended in 50 ml per 1 l culture lysis buffer (50 mM HEPES, 0.5M NaCl, 10 mM imidazole, 5% glycerol). The mixture was lysed by ultrasound in 30 ml batches on ice. Cell debris was spun down (25000g, 30 min, 4°C). The supernatant was applied onto the gravity column with 0.6 ml per 1 l bacterial culture Ni-sepharose resin pre-equilibrated with the washing buffer (the same as lysis buffer), washed with 20 CV washing buffer, followed by 20 CV the same buffer containing 20 mM imidazole and finally eluted with 4 CV elution buffer (50 mM HEPES, 0.5M NaCl, 250 mM imidazole, 5% glycerol). Fractions containing the protein were combined, concentrated to 2.2 ml using a Sartorius Vivaspin Turbo 30 kDa concentrator. For TEV-cleaved samples, the cleavage of the concentrated sample was performed overnight at 4°C using 80 µg TEV-protease per 1 mg of the LTA4H. Finally, the protein was purified by gel-filtration (Hiload 16/60 Superdex 200 GE healthcare pre-equilibrated in the buffer containing 50 mM Tris, 0.5M NaCl, 5% glycerol). Protein containing fractions were pooled together and concentrated to the relevant concentration using a Sartorius

Vivaspin Turbo 30 kDa concentrator. The purified protein samples were stored at -80 ^{o}C for further use.

Deglycosylation protocol

The deglycosylation was performed using 5:1 mass ratio to 4.5 mg/ml PNGase volume ration on ice for 30 minutes.

SDS-PAGE

4-20% Bis-Tris Criterion™ TGX precast gels (BIO-RAD) were used. 15 µl protein sample was supplemented with 5 µl SDS sample buffer 4X, the mixture was incubated at 95^{o}C for 15 minutes, then loaded onto the gel. Precision Plus Protein™ Unstained Standards (Bio-Rad) ladder was used. The gel was run in SDS buffer (0.25 M Tris pH 7.5, 2 M glycine, 35 mM SDS) at 150 volts for 45 minutes. The gel was stained using BP 3620-1 dye (Fisher Scientific).

DNA agarose gel electrophoresis

The gel was cast using 0.45 g SB Molecular Biology Grade Agarose (Severn Biotech) in 60 ml SB buffer (5 mM sodium borate, pH 8.0) supplemented with 5 µl 10mg/ml ethidium bromide (Sigma-Aldrich). 5 µl of the DNA sample was mixed with 3 µl 6X DNA Loading Dye (Fermentas) and 10 µl DNA-ase free water, and loaded onto the gel. GeneRuler™ 1kb DNA Ladder (Fermentas) was used. The gel was run in SB buffer at 150 volts for 30 minutes. DNA bands were visualised using UV.

Intact MS

Reversed-phase chromatography was performed in-line prior to mass spectrometry using an Agilent 1290 uHPLC system (Agilent Technologies inc. – Palo Alto, CA, USA). Concentrated protein samples were diluted to 0.1 mg/ml in 0.1% formic acid and 50 µl was injected on to a 2.1 mm x 12.5 mm *Zorbax* 5um 300SB-C3 guard column housed in a column oven set at 40 ^{o}C. The solvent system used consisted of 0.1% formic acid in ultra-high purity water (Millipore) (solvent A) and 0.1 % formic acid in methanol (LC-MS grade, Chromasolve) (solvent B). Chromatography was performed as follows:

Initial conditions were 90 % A and 10 % B and a flow rate of 1.0 ml/min. A linear gradient from 10 % B to 80 % B was applied over 35 seconds. Elution then proceeded isocratically at 95 % B for 40 seconds followed by equilibration at initial conditions for a further 15 seconds. Protein intact mass was determined using a 6530 electrospray ionisation quadrupole time-of-flight mass spectrometer (Agilent Technologies Inc. – Palo Alto, CA, USA). The instrument was configured with the standard ESI source and operated in positive ion mode. The ion source was operated with the capillary voltage at 4000 V, nebulizer pressure at 60 psig, drying gas at 350^{o}C and drying gas flow rate at 12 L/min. The instrument ion optic voltages were as follows: fragmentor 250 V, skimmer 60 V and octopole RF 250 V. The obtained data was analysed using Agilent MassHunter Qualitative Analysis B.00.60 software.

Native MS

Native MS measurements and data analysis were performed by Dr. Rod Chalk and Oktawia Borkowska. Reversed-phase chromatography was performed in-line prior to mass spectrometry using an Agilent 1290 uHPLC system (Agilent Technologies inc. – Palo Alto, CA, USA). Concentrated protein was diluted to 0.02 mg/ml in 0.1% formic acid and 50 µl was injected on to a 2.1 mm x 12.5 mm *Zorbax* 5um 300SB-C3 guard column housed in a column oven set at 40 ^{o}C. The solvent system used consisted of 0.1% formic acid in ultra-high purity water (Millipore) (solvent A) and 0.1 % formic acid in methanol (LC-MS grade, Chromasolve) (solvent B). Chromatography was performed as follows: Initial conditions were 90 % A and 10 % B and a flow rate of 1.0 ml/min. A linear gradient from 10 % B to 80 % B was applied over 35 seconds. Elution then proceeded isocratically at 95 % B for 40 seconds followed by equilibration at initial conditions for a further 15 seconds. Protein mass was determined using a 6530 electrospray ionisation quadrupole time-of-flight mass spectrometer (Agilent Technologies Inc. – Palo Alto, CA, USA). The instrument was configured with the standard ESI source and operated in positive ion mode. The ion source was operated with the capillary voltage at 4000 V, nebulizer pressure at 60 psig, drying gas at 350^{o}C and drying gas flow rate at 12 L/min.

The instrument ion optic voltages were as follows: fragmentor 250 V, skimmer 60 V and octopole RF 250 V.

Tryptic digest

Tryptic digest and data analysis were performed by Dr. Rod Chalk and Oktawia Borkowska.

Thermal shift assay

Thermal melting experiments were carried out using an Mx3005p Real Time PCR machine (Stratagene). Protein samples were buffered in an appropriate buffer (**1** - Washing buffer: 50 mM HEPES, 0.5 M NaCl, 10 mM imidazole, 5% glycerol, pH 7.5, **2** - Elution buffer: 50 mM HEPES, 0.5 M NaCl, 250 mM imidazole, 5% glycerol, pH 7.5, **3** - Low salt washing buffer: 50 mM Tris, 50 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8, **4** - Low salt elution buffer: 50 mM Tris, 50 mM NaCl, 250 mM imidazole, 5% glycerol, pH 8, **5** - Ion-exchange buffer A: 25 mM Tris, 50 mM NaCl, pH 8, **6** - Ion-exchange buffer B: 25 mM Tris, 1 M NaCl, pH 8, **7** - Ion-exchange buffer A + glycerol: 25 mM Tris, 50 mM NaCl, pH 8, 5% glycerol, **8** - Ion-exchange buffer B + glycerol: 25 mM Tris, 1 M NaCl, pH 8, 5% glycerol, **9** - GF buffer: 10 mM HEPES, 0.5 M NaCl, 5% glycerol, pH 7.5, **10** - GF buffer with TCEP: 10 mM HEPES, 0.5 M NaCl, 5% glycerol, 1mM TCEP, pH 7.5, **11** - Dialysis buffer: 25 mM Tris, 50 mM NaCl, pH 8, **12** - Dialysis buffer + glycerol: 25 mM Tris, 50 mM NaCl, 5% glycerol, pH 8, **13** - 50 mM Tris, 50 mM NaCl, pH 7.5, **14** - 50 mM Tris, pH 7.5, 15 - 50 mM HEPES, 0.5 M NaCl, 10 mM imidazole, 5% glycerol, pH 8.6, **16** - 50 mM HEPES, 0.5 M NaCl, 10 mM imidazole, 5% glycerol, pH 5) and assayed in a 96-well plate at a final concentration of 2 µM in 20 µL volume. Compounds were added at a final concentration of 10 µM. SYPRO Orange (Molecular Probes) was added as a fluorescence probe at a dilution of 1:1000. Excitation and emission filters for the SYPRO Orange dye were set to 465 nm and 590 nm, respectively. The temperature was raised with a step of 3 °C per minute from 25 °C to 96 °C and fluorescence readings were taken every at each interval.

3. FBDD and fragment screening

3.1 Introduction

NMR and X-ray crystallography are one of the most common biophysical techniques used for screening in FBDD. [71] They measure the protein-ligand interactions directly, which reduces the number of false positive hits. In the beginning of this project there was no established ERAP1 screening method using any of these two methods. Instead, three orthogonal ERAP1 activity assays, which are described below, had been reported at that time. This allowed us quick and facile screening of the fragment libraries, while a robust ERAP1 crystallisation protocol was developed (see Chapter 4) based on the previous successful crystallisation of this enzyme in our group. [28] The three assays relied on different measurements – absorbance, fluorescence and mass-spectrometry. Therefore, the combination of all three assays reduced the number of false-positive hits generated by the screening.

3.1.1 Colorimetry based enzyme activity assay.

For the initial fragment screening the colorimetric enzyme activity assay was used. In this assay the substrate was L-leucine-p-nitroanilide, containing an internal peptide bond that was cleaved by the aminopeptidase, releasing p-nitroaniline (Figure 3-1). The reaction was tracked by measuring the increase in absorbance at 405 nm wavelength, which was directly proportional to the concentration of the *p*-nitroaniline product. This assay has been widely used in aminopeptidase research previously, for example for work on M1 aminopeptidases such as ERAP1 [44], AnPEP [72] and LTA4H [46] which would allow us to use this assay for the target ERAP1 enzyme as well as the selectivity panel enzymes – ERAP2, AnPEP and LTA4H, making the comparison of the results easier.

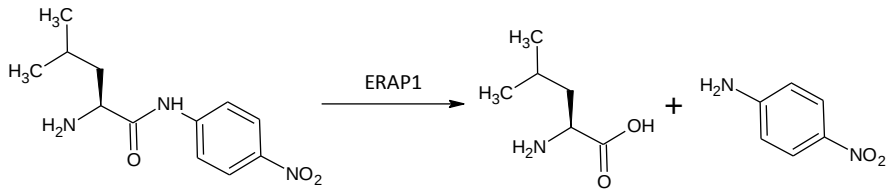


Figure 3-1. L-leucine-p-nitroanilide cleavage by ERAP1 produces p-nitroaniline, which leads to the increase in absorbance at 405 nm.

3.1.2 Fluorescence based enzyme activity assay

This assay had been developed specifically for studies of ERAP1. [73] The substrate was a 10-mer fluorogenic peptide WRVYEKC(dnp)ALK. The dinitrophenyl (dnp) moiety quenches the terminal tryptophan's fluorescence due to Förster resonance energy transfer (FRET). Trimming of the terminal tryptophan leads to an increase in fluorescence (Figure 3-2).

The colorimetric assay substrate was a dipeptide. Thus, using that assay it would be possible to identify inhibitors that bound in close vicinity of the active site and allosteric inhibitors. It was proposed that that using the fluorogenic assay with a much larger substrate it would be possible to identify inhibitors that bound further down the ERAP1 peptide binding pocket. The "molecular ruler" mechanism was a unique feature of ERAP1 and had not been described for other aminopeptidases of M1 class including the closely related ERAP2. [8] Thus, were such hits identified, they might exhibit good selectivity properties.

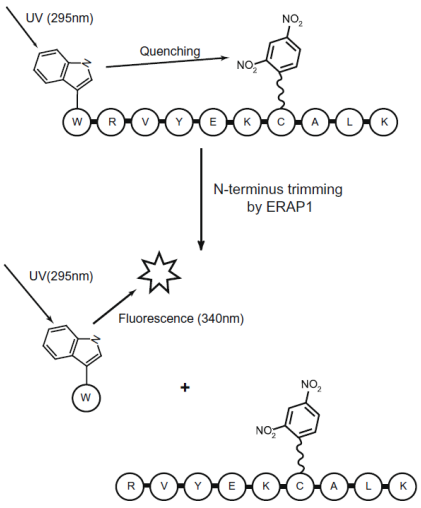


Figure 3-2. Schematic representation of the fluorogenic assay principle. When N-terminal tryptophan is cleaved from the substrate, its fluorescence is not quenched any more. Adapted from I. Evnouchidou *et al*. [73]

3.1.3 Mass-Spectrometry based enzyme activity assay.

This assay was used for the validation of the identified hits, because it relied on Mass-Spectrometry measurements rather than spectroscopic as the previous two assays. This assay had been previously used in our group for ERAP1 research. [28] The substrate was a *Chlamydia* derived peptide (protein accession number: C4PLH3)

$_{110\text{-}122}$QITANRELIQQEL, which contained HLA-B27 restricted epitope NRELIQQEL. The enzymatic reaction was tracked by measuring the concentrations of the substrate and its trimmed intermediates using QTOF MS. RapidFire® MS system was for high through-put validation of the fragment hits.

3.1.4 NMR screening

Through the collaboration with Novartis, their library of 500+ fluorinated compounds was screened against ERAP1 using ^{19}F NMR screening. In this assay the change in a compounds' ^{19}F resonance signal intensity was measured. A drop in the intensity upon addition of ERAP1 indicated that the compound was binding to the protein (Figure 3-3). Competitive binding NMR experiments with the identified hits were also performed. In the case of competitive binding, when a reference binder and a supplemented compound bound to the same site in the target enzyme, the ^{19}F signal intensity of the reference molecule increased. This technique was much more protein intensive compared to the aforementioned assays. However, only hits identified by this assay definitely bound to ERAP1, rather than being potential assay interference compounds. Thus, in the absence of structural information, this assay provided the first indication that some of the identified hits did bind to ERAP1.

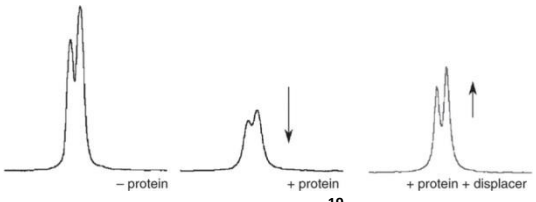


Figure 3-3. A typical change in ^{19}F resonance signal intensity of a hit upon addition of the target protein and a competitive binder. Adapted from E. Mashalidis *et al* [109]

3.1.5 Chapter overview

In this work three libraries were screened against the ERAP1 as a starting point. A library of 96 fragments with putative Zn-binding motifs based on A. Agraval *et al*. [74] A 117 fragment library with increased structural diversity through enhanced 3D characteristics of the molecules. [75] Finally, 1000 fragments (M1000) from the Maybridge Ro3 library, where fragments were selected in accordance with the "Rule of 3". [76] All fragments in the last library obey the following constraints: molecular weight < 300 Da, the number of hydrogen bond donors ≤ 3, the number of hydrogen

bonds acceptors ≤ 3 and clogP ≤ 3. The "Rule of 3" was devised through the analysis of the fragment hits of a range of targets from Astex Pharmaceuticals.

Based on the identified fragments, a follow-up library of commercially available molecules with similar structures was created. The screening results provided initial insights about the structure-activity relationship of the identified fragment hits. In addition, two reported M1 aminopeptidase inhibitors were synthesised and their ERAP1 inhibitory properties were characterised. Then a Novartis proprietary fluorinated-fragments library was screened against ERAP1 using ^{19}F NMR and the hits were characterised using the enzyme activity assays. All the identified hits were screened against LTA4H to elucidate their selectivity properties.

3.2 Screening results and discussion

3.2.1 Zn-library, 3D-library and M1000 library screening in the colorimetric and hits validation

Firstly, ERAP1 kinetic properties were established for all three activity assays. All the data was in accordance with the previously published results (Table 3-1). Bestatin was used as an inhibition control in this work. The experimental IC_{50} of bestatin was 87 µM in the colorimetric assay, which was in agreement with the reported literature [44] and the unpublished data from the previous work on ERAP1 in our group.

		Substrate K_M	v_{max}, mol s^{-1} $_{ERAP1}$mol^{-1}
Colorimetric assay	Experimental value	1.3 ± 0.55 mM	7.2 ± 0.7
	Literature value	0.9 mM [44]	2.7 [44]
MS assay	Experimental value	6.2 ± 12.5 µM	72 ± 55
	Literature value	Not reported	Not reported
Fluorogenic assay	Experimental value	7.0 ± 1.9 µM	0.7 ± 0.5
	Literature value	7.5 µM [73]	0.28 [73]

Table 3-1. Experimental and literature substrate K_M values established for the substrates from the three assays used in this work.

Since the fragments were stored as DMSO solution stocks, the effect of DMSO on the enzyme activity was assessed in the range from 0-5%. No clear trend towards a decrease in the enzyme activity at higher DMSO concentrations was observed for the colorimetric assay (Fig. 4). For consistency, all colorimetric assay reactions were performed at 2% DMSO unless stated otherwise.

	0%	0.10%	0.25%	0.50%	0.70%	1%	2%	3%	5%
ERAP1(wt)	1.00 ± 0.17	0.94 ± 0.18	0.89 ± 0.09	0.89 ± 0.11	0.90 ± 0.12	0.76 ± 0.17	1.03 ± 0.10	0.93 ± 0.06	0.83 ± 0.08

Table 3-2. ERAP1 colorimetric assay reaction velocity normalised to the positive control at different DMSO concentrations (n = 3). No clear effect of DMSO on the ERAP1 enzymatic activity was observed.

the following 9 fragments as potential hits (Table 3-3). The commercially available compounds were ordered for further characterisation. Their IC_{50} values in the colorimetric assay were established. The fragments Zn_7, and Zn_9 turned out to be false positives, since the repurchased samples showed no inhibitory activity in the concentration range 1 µM to 10 mM (Table 3-4).

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.93	1.09	1.13	1.23	0.97	1.18	0.08	1.07	0.99	0.97	0.91	1.09
B	1.35	1.16	1.02	0.09	1.00	1.07	1.12	1.00	1.10	0.45	0.96	1.08
C	1.05	1.17	1.29	1.00	1.03	0.98	1.24	1.25	1.19	1.17	1.10	1.27
D	0.20	0.17	0.82	0.76	1.01	0.74	0.04	1.09	1.11	1.04	1.03	1.53
E	1.03	1.39	0.54	1.06	1.03	0.53	0.17	1.62	1.45	1.09	0.69	0.52
F	0.76	0.90	0.36	1.04	1.01	1.04	1.02	0.93	0.93	1.04	0.74	1.05
G	1.46	0.78	1.50	1.08	0.90	0.16	1.15	1.22	1.43	1.51	1.11	1.20
H	0.10	1.24	0.65	1.35	1.19	1.23	0.73	0.73	0.89	0.87	0.84	1.11
A2	1.08	1.17	1.08	0.78	0.63	0.30	1.15	1.13	1.48	1.24	1.14	1.08
B2	1.17	0.98										

Table 3-3. Results of the Zn-binding fragments library screening (n = 3). Average colorimetric enzyme activity assay velocity (absorbance units / second) in presence of a 2 mM fragment normalised to the positive control. Fragments with >70% inhibitory effect were selected as potential hits.

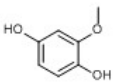
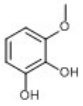
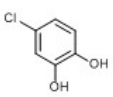
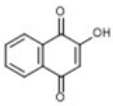
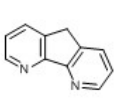
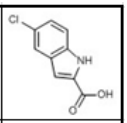
									
Plate well	A7	B4	D1	D2	D7	E7	G6	H1	Pl2 A6
Fragment ID	Zn_1	Zn_2	Zn_3	Zn_4	Zn_5	Zn_6	Zn_7	Zn_8	Zn_9
IC_{50}, mM	2.23 ± 3.01	0.24 ± 0.21	1.5 ± 1.1 µM	0.13 ± 0.14	0.20 ± 0.11	1.19 ± 2.90	Inactive	Did not order	Inactive

Table 3-4. Zn-library fragment hits and their IC_{50} determined in the colorimetric enzyme activity assay. The assay condition for the IC_{50} determination were: 5 nM ERAP1, 2 mM L-leucine-nitroanilide, 5% DMSO in 10 mM Tris pH 7.5.

Screening of the 3D-library using the colorimetric assay identified the following 15 fragments as potential hits (Table 3-5). The commercially available compounds were ordered for further characterisation. Their IC_{50} values in the colorimetric assay were established. The fragments 3D_11, and 3D_15 turned out to be false positives, since the repurchased samples showed no inhibitory activity in the concentration range 1 µM to 4 mM (Table 3-6). Fragments 3D_5 and 3D_10, and 3D_6, 3D_7, 3D_8 and 3D_9, formed two groups of structurally related hits. They potentially might have similar binding modes to ERAP1.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.56	0.80	1.06	0.97	0.09	0.48	0.73	0.92	0.73	0.12	I	1.27
B	0.94	0.40	0.54	0.90	0.60	0.82	0.73	0.25	0.52	0.63	0.45	0.63
C	0.91	0.87	1.02	0.03	0.58	0.63	0.07	I	0.00	0.89	0.82	0.97
D	0.96	0.47	0.29	0.20	0.51	I	I	0.51	0.51	0.54	0.55	1.21
E	I	0.80	0.54	0.25	0.28	0.36	0.03	0.62	0.65	0.61	0.65	1.25
F	0.36	0.47	0.28	1.53	0.48	0.87	0.83	0.35	0.27	I	0.92	1.18
G	I	0.98	0.79	0.35	0.60	0.36	0.47	0.60	0.80	0.61	0.84	1.17
H	0.55	I	0.50	0.79	0.25	1.26	0.78	0.47	1.10	0.69	0.69	I
A2	1.17	1.07	1.2	1.03	0.89	0.89	0.76	I	0.7	I	0.27	1.43
B2	0.5	0.55	0.58	0.7	I	I	0.71	0.78				

Table 3-5. Results of the 3D fragments library screening (n = 3). Average colorimetric enzyme activity assay velocity (absorbance units / second) in presence of 2 mM fragment normalised to the positive control. Fragments with >70% inhibitory effect were selected as potential hits. "I" indicate fragments that interfered with the assay.

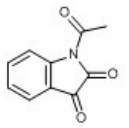
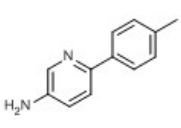
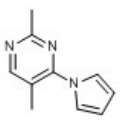
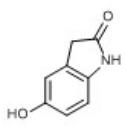
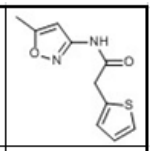
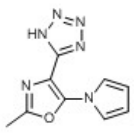
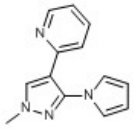
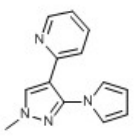
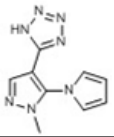
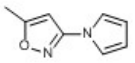
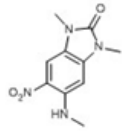
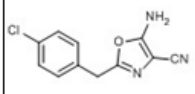
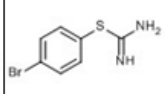
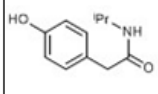
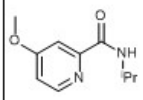
					
Plate well	**A5**	**A10**	**B8**	**C9**	**C4**
Fragment ID	**3D_1**	**3D_2**	**3D_3**	**3D_4**	**3D_5**
IC_{50}, mM	**1.19 ± 2.75**	**1.7 ± 21**	**1.18 ± 1.83**	**0.08 ± 0.05**	**0.14 ± 0.72**
					
Plate well	**D4**	**E4**	**E5**	**F3**	**D3**
Fragment ID	**3D_6**	**3D_7**	**3D_8**	**3D_9**	**3D_10**
IC_{50}, mM	**0.66 ± 0.57**	**0.50 ± 0.36**	**1.75 ± 5.75**	**0.58 ± 0.49**	**0.97 ± 1.43**
					
Plate well	**E7**	**C7**	**F9**	**H5**	**PL2_A11**
Fragment ID	**3D_11**	**3D_12**	**3D_13**	**3D_14**	**3D_15**
IC_{50}, mM	**No activity**	**ca. 0.15***	**0.75 ± 0.69**	**1.36 ± 3.03**	**No activity**

Table 3-6. 3D-library fragment hits and their IC_{50} determined in the colorimetric enzyme activity assay. The assay condition for the IC_{50} determination were: 5 nM ERAP1, 2 mM L-leucine-nitroanilide, 5% DMSO in 10 mM Tris pH 7.5. It was impossible to measure C7 fragment's IC50 precisely, because its solubility limit was 0.5 mM at the assay conditions.

Screening of the M1000-library at 1.25 mM fragments concentration using the colorimetric assay identified the following 21 fragments as potential hits (Table 3-8). The commercially available compounds were ordered for further characterisation. Their IC_{50} values in the colorimetric assay were established. The fragments M_2, M_4, M_5, M_6, M_8, M_11, M_14 and M_17 turned out to be false positives, since the repurchased samples showed no inhibitory activity in the concentration range 1 µM to 10 mM (Table 3-8). The fragment M_7 had the same hydroquinone core as the Zn_3 identified hit, so it was likely that their enzyme inhibition mechanisms were similar.

The 30 hits from all the three libraries, where inhibitory activity in the colorimetric had been confirmed by the IC_{50} evaluation, were validated using the orthogonal MS assay at 1 mM fragment concentration (Table 3-7). The hits which showed over 85% average inhibition were considered as validated. Out of 1226 fragments screened using the colorimetric assay 8% were incompatible with the assay due to low solubility or interference. 4% of the fragments were identified as hits, 43% of which were validated using the orthogonal MS enzyme activity assay.

Zn_A7	Zn_B4	Zn_D1	Zn_D2	Zn_G6	Zn_D7	3D_A5	3D_A10	3D_B8	3D_C4	3D_C7
3.49	0.06	11.14	18.91	2.09	35.23	34.83	57.83	36.42	3.42	6.99
3D_C9	3D_D3	3D_D4	3D_F3	3D_E4	3D_E5	3D_F9	3D_H5	Pos.	Neg.	Bestatin
0.08	60.14	70.11	56.64	84.09	70.70	15.19	56.46	99.43	2.67	0.01
M_4B9	M_6C7	M_6D7	M_6H7	M_9A9	M_9C11	M_9H11	M_10E3	M_10B4	M_12A2	M_1E8
9.99	29.75	3.45	84.61	33.07	51.88	78.52	0.06	14.68	1.54	70.24

Table 3-7. Validation of the hits identified by the screening of the three libraries in the ERAP1 colorimetric assay using the orthogonal MS enzyme activity assay (n = 4). Extent of the conversion of the full substrate (QITANRELIQQEL) to its trimmed intermediate (TANRELIQQEL) was used to measure the inhibitory properties of the hits at 1 mM concentration. Pos. – positive control with no fragments added. Neg – negative control with no ERAP1 added. The fragments which inhibited the conversion to < 0.15 were brought forward as hits.

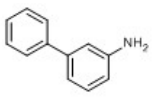
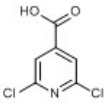
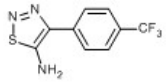
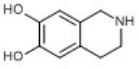
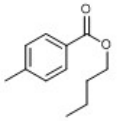
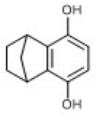
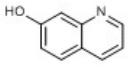
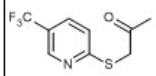
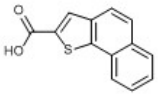
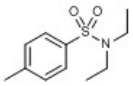
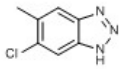
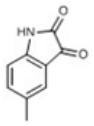
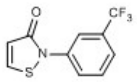
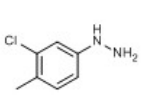
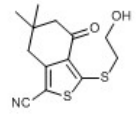
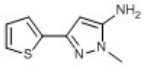
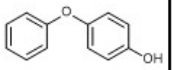
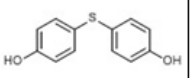
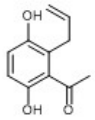
						
Plate well	1_E8	3_D8	4_B9	4_G9	5_B11	6_C7
Fragment ID	M_1	M_2	M_3	M_4	M_5	M_6
IC_{50}, mM	1.4 ± 2.7	No activity	0.41 ± 0.25	No activity	No activity	No activity
						
Plate well	6_D7	6_H7	7_C11	8_G4	8_G10	8_H4
Fragment ID	M_7	M_8	M_9	M_10	M_11	M_12
IC_{50}, mM	0.07 ± 0.09	No activity	Not available	Not available	No activity	Not available
						
Plate well	9_A9	9_B3	9_C11	9_F4	9_H11	10_E3
Fragment ID	M_13	M_14	M_15	M_16	M_17	M_18
IC_{50}, mM	0.31 ± 0.12	No activity	2.8 ± 3.2	Not available	No activity	1.42 ± 2.15
						
Plate well	10_B4	12_A2	12_H3			
Fragment ID	M_19	M_20	M_21			
IC_{50}, mM	1.65 ± 2.09	0.76 ± 0.65	Not available			

Table 3-8. M1000-library fragment hits and their IC_{50} determined in the colorimetric enzyme activity assay. The assay condition for the IC_{50} determination were: 5 nM ERAP1, 2 mM L-leucine-nitroanilide, 5% DMSO in 10 mM Tris pH 7.5.

3.2.2 3D-library and M1000 library screening in the fluorogenic assay and hit validation

As described in the introduction, the idea behind rescreening the fragment libraries using the fluorogenic enzyme activity assay was to identify hits that bound further down in the ERAP1 binding cleft. Therefore, the Zn-binding fragment library was not screened, as all its compounds contained putative Zn-binding sites and were supposed to interact with Zn in the catalytic site.

Screening of the 3D-library using the fluorogenic assay identified 10 new fragments as potential hits (Table 3-9). The commercially available compounds were ordered for further characterisation. Their IC_{50} values in the fluorogenic assay were established, D5, D9 and D12 were found to be false-positives (Table 3-10). All the hits were validated as described before using the MS assay at 1 mM fragment concentration (Table 3-11). Only one fragment, 3D_22 was confirmed to be active by the MS assay, while all the other compounds were found to be inactive. 3D_22 fragment was structurally similar to previously identified hits, for example Zn_1. In addition, it exhibited 40% inhibition, when screened in the colorimetric assay (Table 3-3). Hence, it was not clear whether it did actually bind outside the leucine-nitroanilide binding site, or it was just missed by the colorimetric assay.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.44	0.74	0.75	0.51	0.17	I	0.93	0.97	1.08	I	I	0.46
B	0.89	0.54	I	0.71	0.88	0.59	0.82	0.15	I	0.52	*0.21*	0.37
C	0.94	0.41	0.94	0.05	0.95	*0.17*	0.14	I	0.01	0.71	0.64	*0.13*
D	1.07	*0.06*	I	0.21	*0.02*	I	I	I	*0.09*	0.70	0.69	*0.18*
E	I	0.41	0.74	0.08	0.23	*0.14*	0.11	*0.11*	I	*0.14*	0.85	I
F	I	0.48	0.77	0.97	*0.10*	I	0.57	0.75	0.04	I	*0.10*	I
G	I	0.42	0.50	I	I	I	0.48	I	I	I	I	I
H	0.41	I	*0.08*	Empty	I	*0.23*	0.70	I	Empty	Empty	*0.18*	Empty
A2	0.76	0.77	0.63	0.31	0.53	0.71	0.44	I	I	0.42	0.03	*0.02*
B2	0.72	Empty	I	*0.04*	I	I	I	I				

Table 3-9. Results of the 3D fragments library screening (n = 3). Average fluorogenic enzyme activity assay velocity (fluorescence units / second) in presence of a 2 mM fragment normalised to the positive control. Fragments with >70% inhibitory effect were selected as potential hits. I indicates fragments that interfered with the assay. Values in yellow indicate new hits, which were not picked up using the colorimetric enzyme activity assay. Empty indicates that the fragment stock in the library was depleted.

B11	C6	C12	D2	E6	E8	E10	F5	F11	H3	Pos
0.66	0.69	0.83	0.67	0.62	0.75	0.33	0.80	0.77	0.75	0.75

Table 3-11. Validation of the novel 3D hits using the orthogonal MS enzyme activity assay (n = 4). Extent of the conversion of the full substrate (QITANRELIQQEL) to its trimmed intermediate (TANRELIQQEL) was used to measure the inhibitory properties of the hits at 1 mM concentration. Pos. – positive control with no fragments added. Only E10 fragment showed significant ERAP1 inhibition.

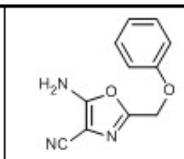
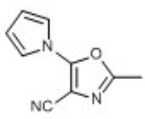
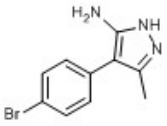
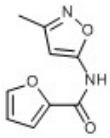
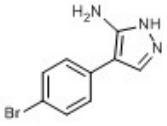
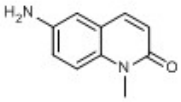
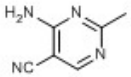
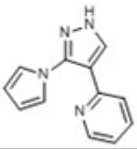
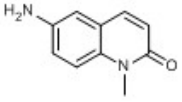
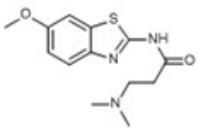
					
Plate well	**B11**	**C6**	**C12**	**D2**	**E6**
Fragment ID	**3D_16**	**3D_17**	**3D_18**	**3D_19**	**3D_20**
IC_{50}, mM	**3.0 ± 8.0**	**0.1 ± 0.2**	**0.4 ± 0.3**	**0.1 ± 0.1**	**0.4 ± 0.3**
					
Plate well	**E8**	**E10**	**F5**	**F11**	**H3**
Fragment ID	**3D_21**	**3D_22**	**3D_23**	**3D_24**	**3D_25**
IC_{50}, mM	**0.1 ± 0.1**	**0.3 ± 0.2**	**0.3 ± 0.3**	**0.1 ± 0.2**	**1.5 ± 4.5**

Table 3-10. 3D-library fragment hits and their IC_{50} determined in the fluorogenic enzyme activity assay. The assay condition for the IC_{50} determination were: 5 nM ERAP1 and 10 µM WRVYEKC(dnp)ALK substrate in 10 mM Tris (pH 7.5) buffer.

The results from 3D-library screening showed that the fluorogenic assay was not suitable for such purpose. 32% of the fragments interfered with the assay. Out of thirteen hits previously identified using the colorimetric assay twelve were picked up by the fluorogenic one, which indicated high fidelity. In addition, 17 new fragments were picked up by screening and validated using this assay. This made the total hit rate for the 3D library equal to 40%. This was clearly high, for example comparing with 11% hit rate of the colorimetric assay for the same library. The validation using the MS assay showed that 90% of the identified hits were false-positives. Due to the above reasons the fluorogenic assay was not further used for the compound screenings.

3.2.3 Follow-up library screening and characterization

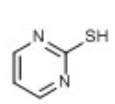
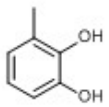
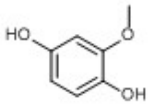
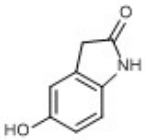
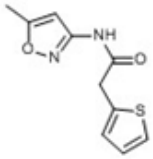
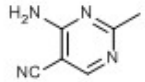
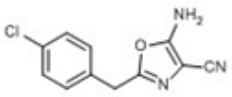
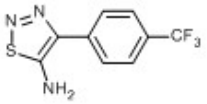
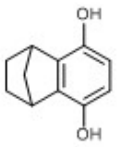
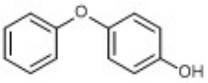
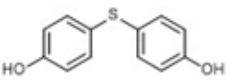
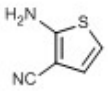
				
Fragment ID	**Zn_1**	**Zn_2**	**Zn_3**	**3D_4**
Colorimetric IC_{50}, mM	**2.23 ± 3.01**	**0.24 ± 0.21**	**0.001 ± 0.001**	**0.08 ± 0.05**
Fluorogenic IC_{50}, mM	**0.08 ± 0.07**	**Interfere**	**Interfere**	**0.08 ± 0.09**
MS, IC_{50}, mM	**0.79 ± 0.61**	**0.002 ± 0.003**	**0.90 ± 0.61**	**0.15 ± 0.21**
				
Fragment ID	**3D_5**	**3D_22**	**3D_12**	**M_3**
Colorimetric IC_{50}, mM	**0.14 ± 0.72**	**Not active**	**ca 0.15**	**0.41 ± 0.25**
Fluorogenic IC_{50}, mM	**0.05 ± 0.08**	**0.31 ± 0.22**	**0.25 ± 0.19**	**0.22 ± 0.16**
MS, IC_{50}, mM	**0.006 ± 0.013**	**0.43 ± 0.29**	**0.78 ± 3.71**	**0.92 ± 2.41**
				
Fragment ID	**M_7**	**M_18**	**M_19**	**M_20**
Colorimetric IC_{50}, mM	**0.07 ± 0.09**	**1.41 ± 2.15**	**1.64 ± 2.09**	**0.76 ± 0.65**
Fluorogenic IC_{50}, mM	**Interfere**	**0.31 ± 0.82**	**Interfere**	**0.07 ± 0.04**
MS, IC_{50}, mM	**0.1 ± 0.5 µM**	**0.85 ± 3.05**	**0.40 ± 0.53**	**0.31 ± 0.54**

Table 3-12. 12 hits identified from by the initial screening campaign were characterised using all three enzyme activity assays. In red are highlighted potential Pan Assay Interference Compounds (PAINS). [110]

Based on the identified fragment hits (Table 3-12) an extended library of 54 commercially available follow-up fragments (Ext_ prefix in the internal ID) was compiled. The fragments Ext_49 (2-aminofuran-3-carbonitrile) and Ext_50 (2-amino-1-benzothiophene-3-carbonitrile) were removed from the library as they were expensive and not paramount for the SAR. The remaining fragments were purchased and screened against ERAP1 using both the colorimetric and the MS enzyme activity assays (Table 3-13). Some fragments that did not show significant inhibitory effect in the colorimetric assay were still identified by the MS screening. Such fragments were structurally very similar to the fragments that were picked up by the colorimetric screening, for example Ext_6 and

Zn_A7. Therefore, this fact was attributed to the higher fidelity of the MS assay. It relied on the direct measurement of the enzymatic reaction product's abundance, rather than the indirect measurements of absorbance used in the colorimetric one.

Ext_1	Ext_2	Ext_3	Ext_4	Ext_5	Ext_6	Ext_7	Ext_8	Ext_9	Ext_10	Ext_11	Ext_12
1.29	0.15	I	1.00	0.84	0.60	0.56	0.22	0.24	0.75	1.20	0.28
Ext_13	Ext_14	Ext_15	Ext_16	Ext_17	Ext_18	Ext_19	Ext_20	Ext_21	Ext_22	Ext_23	Ext_24
0.92	I	0.73	0.17	0.27	I	0.22	I	1.04	0.12	1.14	0.44
Ext_25	Ext_26	Ext_27	Ext_28	Ext_29	Ext_30	Ext_31	Ext_32	Ext_33	Ext_34	Ext_35	Ext_36
1.04	0.85	0.01	0.78	0.85	0.87	0.22	1.08	0.75	0.02	0.67	1.43
Ext_37	Ext_38	Ext_39	Ext_40	Ext_41	Ext_42	Ext_43	Ext_44	Ext_45	Ext_46	Ext_47	Ext_48
I	0.79	0.73	0.06	0.03	0.44	0.19	0.58	0.59	0.28	0.94	1.00
Ext_51	Ext_52	Ext_53	Ext_54								
0.17	0.79	0.43	1.00								

Ext_1	Ext_2	Ext_3	Ext_4	Ext_5	Ext_6	Ext_7	Ext_8	Ext_9	Ext_10	Ext_11	Ext_12
1.03	0.05	0.98	0.97	0.41	0.05	0.04	0.02	0.13	0.05	0.95	0.19
Ext_13	Ext_14	Ext_15	Ext_16	Ext_17	Ext_18	Ext_19	Ext_20	Ext_21	Ext_22	Ext_23	Ext_24
1.04	0.88	0.01	0.01	0.13	0.04	0.01	0.57	0.66	0.80	0.96	0.00
Ext_25	Ext_26	Ext_27	Ext_28	Ext_29	Ext_30	Ext_31	Ext_32	Ext_33	Ext_34	Ext_35	Ext_36
0.72	0.91	0.01	1.02	1.07	0.52	0.16	0.92	0.94	0.40	0.85	0.34
Ext_37	Ext_38	Ext_39	Ext_40	Ext_41	Ext_42	Ext_43	Ext_44	Ext_45	Ext_46	Ext_47	Ext_48
0.75	0.91	0.31	0.02	0.00	0.59	0.03	0.70	0.07	0.32	0.25	0.95
Ext_51	Ext_52	Ext_53	Ext_54	Pos.							
0.05	0.29	0.75	0.88	1.02							

Table 3-13. The summarised results of the extended library screening using a colorimetric (top, n = 6) and MS (bottom, n = 3) enzyme activity assays. 2 mM fragments concentration was used for both screenings. Pos. – positive control with just ERAP1 and the substrate. Fragments that exhibited >70% inhibition were considered as potential hits. Five fragments – Ext_3 (2-(methylthio)pyrimidin-5-ol), Ext_14 (2,5-dimethoxyphenol), Ext_18 (2,3-diamonitoluene), Ext_20 (naphthalene-1,4-diol) and Ext_37 (4-[3-(trifluoromethyl)phenyl]-2,3-dihydro-1,3-thiazol-2-one) – were incompatible with the colorimetric assay due to poor solubility or interference.

Comparing the inhibitory properties of the Zn_1 and 3D_22 structural analogues (Table 3-14) the following trends were observed. The addition of a hydroxyl group to Zn_1 either in *para* or *meta* position relative to the thiol completely eliminated potency. Methylation of the thiol (Ext_1) led to the complete loss of potency. The substitution of the thiol for a hydroxyl group (Ext_10) led to a slight decrease in potency. Elimination of the methyl group (Ext_54) from 3D_22 led to total loss of the potency, while its substitution for a thiol (Ext_51) or amine (Ext_52) did not affect potency much. The most potent fragment from this series (Ext_8) contained pyrazine rather than pyrimidine core.

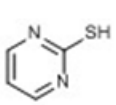
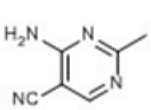
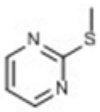
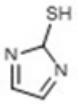
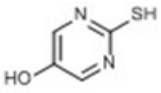
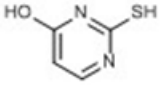
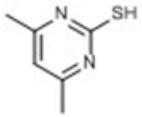
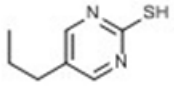
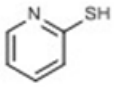
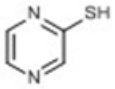
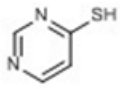
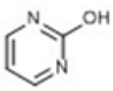
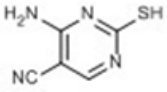
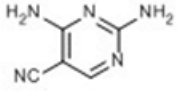
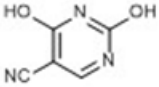
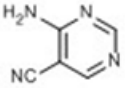
				
Fragment ID	**Zn_A1**	**3D_22**	**Ext_1**	**Ext_2**
Colorimetric IC_{50}, mM	**2.23 ± 3.01**	**Not active**	**Not active**	**0.08 ± 0.03**
MS, IC_{50}, mM	**0.79 ± 0.61**	**0.43 ± 0.29**	**Not active**	**0.62 ± 0.60**
				
Fragment ID	**Ext_3**	**Ext_4**	**Ext_5**	**Ext_6**
Colorimetric IC_{50}, mM	**Not active**	**Not active**	**Not active**	**Not active**
MS, IC_{50}, mM	**Not active**	**Not active**	**Not active**	**0.44 ± 0.28**
				
Fragment ID	**Ext_7**	**Ext_8**	**Ext_9**	**Ext_10**
Colorimetric IC_{50}, mM	**Not active**	**0.01 ± 0.01**	**0.03 ± 0.03**	**Not active**
MS, IC_{50}, mM	**0.24 ± 0.34**	**4.8 ± 3.9 µM**	**0.26 ± 0.06**	**0.72 ± 0.42**
				
Fragment ID	**Ext_51**	**Ext_52**	**Ext_53**	**Ext_54**
Colorimetric IC_{50}, mM	**0.08 ± 0.05**	**Not active**	**Not active**	**Not active**
MS, IC_{50}, mM	**0.36 ± 0.12**	**0.72 ± 4.21**	**Not active**	**Not active**

Table 3-14. The inhibitory properties of Zn_A7, 3D_E10 and the structurally related extended library fragments.

Comparing the inhibitory properties of the Zn_2, Zn_3 and M_7 structural analogues (Table 3-15) the following trends were observed. The methylation of hydroxyl groups led to total loss of potency. The hydroxyl groups could be switched for amines without a significant effect on the fragment's potency. The hydroxyl groups could be either in *ortho* or *para* positions relative one another. The most potent fragment in this series was M_7. The catechols and quinones were reported Pan Assay Interference Compounds (PAINS). [77], [78] It was not possible to rule out these hits being false-positive until the structural information on their binding mode to ERAP1 was obtained.

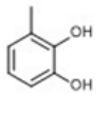
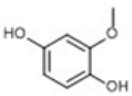
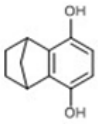
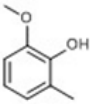
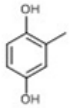
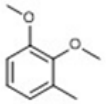
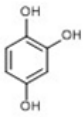
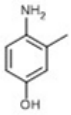
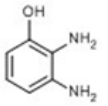
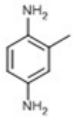
						
Fragment ID	**Zn_2**	**Zn_3**	**M_7**	**Ext_11**	**Ext_12**	**Ext_13**
Colorimetric IC_{50}, mM	**0.24 ± 0.21**	**0.001 ± 0.001**	**0.07 ± 0.09**	**Not active**	**0.03 ± 0.03**	**Not active**
MS, IC_{50}, mM	**0.002 ± 0.003**	**0.90 ± 0.61**	**0.1 ± 0.5 µM**	**Not active**	**0.22 ± 0.03**	**Not active**
						
Fragment ID	**Ext_15**	**Ext_16**	**Ext_17**	**Ext_18**	**Ext_19**	
Colorimetric IC_{50}, mM	**Not active**	**0.13 ± 0.10**	**0.27 ± 0.09**	**Interference**	**0.14 ± 0.09**	
MS IC_{50}, mM	**0.06 ± 0.19**	**0.16 ± 0.53**	**0.13 ± 0.08**	**0.10 ± 0.06**	**0.09 ± 0.05**	

Table 3-15. The inhibitory properties of Zn_2, Zn_3, M_7 and the structurally related extended library fragments.

Comparing the inhibitory properties of the 3D_5 structural analogues (Table 3-16) the following trends were observed. Thiophene was important for retaining the hit's potency. Its removal (Ext_21) or substitution with a less hydrophobic group (Ext_23) led to total loss of potency. Substitution of isoxazole with aniline also led to a complete loss of potency.

There was only one 3D_12 analogue in the library (Table 3-16). The removal of the *p*-chlorophenyl group led to total loss of potency.

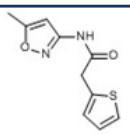
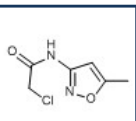
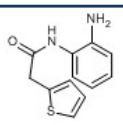
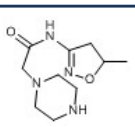
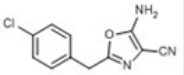
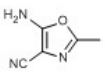
						
Fragment ID	**3D_5**	**Ext_21**	**Ext_22**	**Ext_23**	**3D_12**	**Ext_25**
Colorimetric IC_{50}, mM	**0.14 ± 0.72**	**Not active**	**1.19 ± 0.71**	**Not active**	**ca 0.15**	**Not active**
MS IC_{50}, mM	**0.006 ± 0.013**	**Not active**	**Not active**	**Not active**	**0.78 ± 3.71**	**Not active**

Table 3-16. The inhibitory properties of 3D_5 and the structurally related extended library fragments, and 3D_12 with its structurally related extended library fragment.

Comparing the inhibitory properties of the 3D_4 structural analogues (Table 3-17) the following trends were observed. Both the carbonyl (Ext_32) and the hydroxyl (Ext_29) groups were crucial for retaining potency. The hydroxyl group could be either in positions 6 or 7 (Ext_27). Methylation of the

hydroxyl group (Ext_ 26, Ext_28) led to total loss of activity. The substitution of the hydroxyl group with nitrile (Ext_31) led to significant potency loss, whilst the amine analogue (Ext_30) was completely inactive.

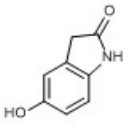
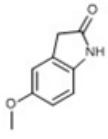
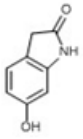
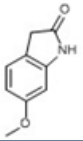
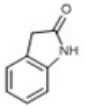
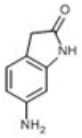
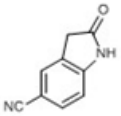
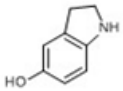
				
Fragment ID	**3D_4**	**Ext_26**	**Ext_27**	**Ext_28**
Colorimetric IC_{50}, mM	**0.08 ± 0.05**	**Not active**	**0.08 ± 0.04**	**Not active**
MS IC_{50}, mM	**0.15 ± 0.21**	**Not active**	**0.04 ± 0.02**	**Not active**
				
Fragment ID	**Ext_29**	**Ext_30**	**Ext_31**	**Ext_32**
Colorimetric IC_{50}, mM	**Not active**	**Not active**	**0.01 ± 0.02**	**Not active**
MS IC_{50}, mM	**Not active**	**Not active**	**2.90 ± 2.89**	**Not active**

Table 3-17. The inhibitory properties of 3D_4 and the structurally related extended library fragments.

Comparing the inhibitory properties of the M_3 structural analogues (Table 3-19) the following trends were observed. The removal of a trifluoromethyl moiety led to total loss of potency. All other aromatic heterocycles that were tried instead of the original thiadiazole did not show any inhibitory effects. There was only one M_20 analogue in the extended library (Table 3-19). The removal of the amine led to complete loss of compound's potency.

2-aminothiophenes and thiadiazoles were reported as potential PAINS. [78] The screening of the related fragments showed that these groups were crucial for retaining the potency in the activity assays. Thus, it was not possible to rule out these hits being false-positives without the structural information on their binding to ERAP1.

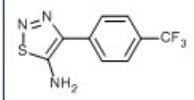
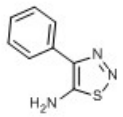
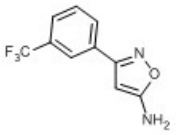
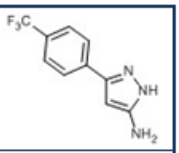
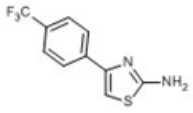
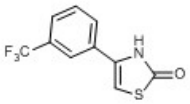
				
Fragment ID	**M_3**	**Ext_33**	**Ext_34**	**Ext_35**
Colorimetric IC_{50}, mM	**0.41 ± 0.25**	**Not active**	**0.04 ± 0.05**	**Not active**
MS IC_{50}, mM	**0.92 ± 2.41**	**Not active**	**Not active**	**Not active**
				
Fragment ID	**Ext_36**	**Ext_37**	**M_20**	**Ext_48**
Colorimetric IC_{50}, mM	**Not active**	**Not active**	**0.76 ± 0.65**	**Not active**
MS IC_{50}, mM	**Not active**	**Not active**	**0.31 ± 0.54**	**Not active**

Table 3-19. The inhibitory properties of M_3 and the structurally related extended library fragments, and M_20 with its structurally related extended library fragment.

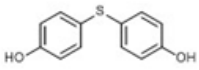
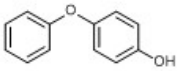
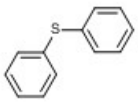
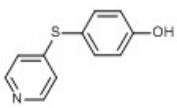
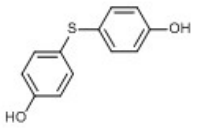
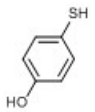
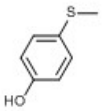
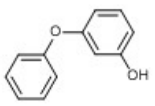
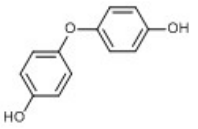
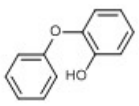
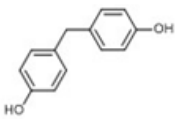
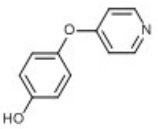
				
Fragment ID	**M_19**	**M_18**	**Ext_38**	**Ext_39**
Colorimetric IC_{50}, mM	**1.64 ± 2.09**	**1.41 ± 2.15**	**Not active**	**Not active**
MS IC_{50}, mM	**0.40 ± 0.53**	**0.85 ± 3.05**	**Not active**	**Not active**
				
Fragment ID	**Ext_40**	**Ext_41**	**Ext_42**	**Ext_43**
Colorimetric IC_{50}, mM	**0.02 ± 0.03**	**0.02 ± 0.01**	**Not active**	**ca. 1**
MS IC_{50}, mM	**0.11 ± 0.17**	**0.02 ± 0.03**	**Not active**	**1.26 ± 1.52**
				
Fragment ID	**Ext_44**	**Ext_45**	**Ext_46**	**Ext_47**
Colorimetric IC_{50}, mM	**Not active**	**Not active**	**1.39 ± 1.19**	**Not active**
MS IC_{50}, mM	**Not active**	**Not active**	**Not active**	**0.74 ± 3.09**

Table 3-18. The inhibitory properties of M_18, M_19 and the structurally related extended library fragments.

Comparing the inhibitory properties of the M_18 and M_19 structural analogues (Table 3-18) the following trends were observed. The substitution of hydroxyl groups for thiols led to a slight increase in potency (Ext_40), while their removal (Ext_38) eliminated potency completely. Hydroxylation of

the second phenyl ring in the M_18 in any position led to total loss of potency (Ext_43, Ext_ 44, Ext_45), whilst conversion of that phenyl group into pyridine did not affect potency much. Interestingly, the same pyridine substitution in M_19 made the compound (Ext_39) completely inactive. This might indicate that even though M_18 and M_19 were structurally related, they might have different binding modes to ERAP1.

3.2.4 The reported M1 aminopeptidase inhibitors screening and characterisation

The inhibitory properties of two synthesised reported M1 aminopeptidase inhibitors – DG-051-1 (a LTA4H inhibitor [59]) and IRAP-I (IRAP inhibitor, [79]) were also characterised using both the colorimetric and the MS enzyme activity assays (Table 3-20). IRAP-I was identified as a more potent inhibitor than bestatin in both assays. The calculated IC_{50} was 9 µM and was comparable to its reported IC_{50} against IRAP (2 µM) in the same colorimetric assay. [79] Hence, the potential selectivity issues for this scaffold must be taken into account, especially considering that IRAP was not part of the enzymes selectivity panel. DG-051-1 had a similar scaffold to previously identified hits M_18 and M_19. However, it did not exhibit clear inhibitory properties in any of the assays. According to the published LTA4H co-crystal structure with a DG-051-1 analogue (PDB: 3FH5) the ligand bound to the leukotriene binding pocket of LTA4H [46] (for more details on the LTA4H binding pocket see Figure 4-8). This pocket does not exist in ERAP1. Therefore, despite having a very similar core structure to the identified ERAP1 hits, DG-051-1 was not confirmed as an ERAP1 inhibitor.

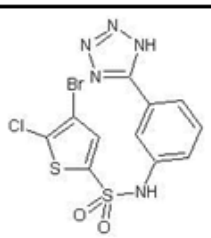
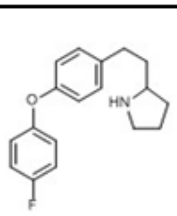
		
Fragment ID	**IRAP-I**	**DG-051**
Colorimetric IC_{50}, µM	**9 ± 6**	**Inactive**
MS IC_{50}, µM	**9 ±5**	**Inactive**

Table 3-20. The characterisation results of the reported M1 aminopeptidase inhibitors using the colorimetric assay (top value) and the MS assay (bottom value).

3.2.5 NMR screening of the fluorinated fragment library

19 Novartis proprietary fragment mixtures with 532 fragments in total were screened using ^{19}F NMR. 24 fragments showed >70% drop in ^{19}F resonance signal intensity upon addition of ERAP1 and were screened separately for validation (Table 3-21). 4 compounds were identified as false positives. Out of the remaining 20 compounds, structures of the 13 were disclosed by the company, since they were non-proprietary (Figure 3-4).

		ms	
Compound Novartis ID	19F shift, -1 x ppm	0.12	0.32
CGA215788	62.486	0.47	0.85
BPS552-NX-1	62.286	0.51	0.71
PKF234-507-NX-1	62.266	0.62	0.78
CGA157465-NX-1	74.496	0.69	0.9
NVP-AVF347-NX-1	61.176	0.92	0.97
NVP-AUV212-NX-1	68.196	0.12	0.2
CGA-262442-NX-1	61.666; 65.686; 66.956	0.51	0.75
NVP-QBL688-NX-2	60.286	0.85	0.95
CGA157116-NX-1	63.506; 63.546	0.1	0.15
NVF-AVF211-NX-1	62.156	0.43	0.61
GPA000553-NX-1	60.636	0.65	0.85
CGA180224	62.539; 62.496	0.71	0.86
NVP-AQL006-NX-1	68.116	0.62	0.81
NVP-AUV452-NX-1	60.376	0.62	0.8
CGP015134-NX-1	78.296	0.61	0.8
CGA262411-NX-1	67.036	0.5	0.68
CGA284249-NX-1	61.476; 61.516; 61.366	0.45	0.6
NVP-AKL824-NX-1	64.326; 64.366	0.45	0.73
PKF059-715-NX-1	64.326	0.3	0.49
NVP-AEC780-NX-1	61.856; 61.896	0.57	0.79
NVP-ANI961-NX-1	62.196	0.7	0.86
PKF043-479-NX-1	63.896; 63.776; 63.736	0.79	0.94
NVP-BEE064-NX-1	71.0063	0.5	0.7
NVP-BXR406-NX-1	63.446; 63.486; 63.336	0.12	0.17

Table 3-21. The results of the identified hits validation by ^{19}F NMR. The binding was measured as the ratio of a compound's fluorine resonance signal intensity without the enzyme and in presence of ERAP1. Compounds with signal intensity change < 0.4 at 120 ms relaxation time and < 0.6 at 320 ms were considered as false positives. Such compounds are marked in red. Non-proprietary compounds, which structures were available, are marked in blue.

Competitive binding experiments were performed using GPA000553-NX-1 and NVP-AEC780-NX-1 as the reference compounds. A significant increase in their ^{19}F resonance signal intensity upon addition of a confirmed NMR hit was sought. According to the results (Table 3-22) none of the added fragments displaced NVP-AEC780-NX-1. Thus none of the 18 fragments had the same binding site as this molecule. 4 molecules were found to displace GPA000553-NX-1. Structures of two of those were available – CGA-262442-NX-1 and NVP-AKL824-NX-1. All three compounds contain a Zn-binding group – either carboxylic acid or 2-hydroxy-pyridine. [74] Thus it was possible to suggest that all three coordinate to Zn in the enzyme active site. However, it was not possible to validate the above suggestion without the structural information on their binding mode to ERAP1.

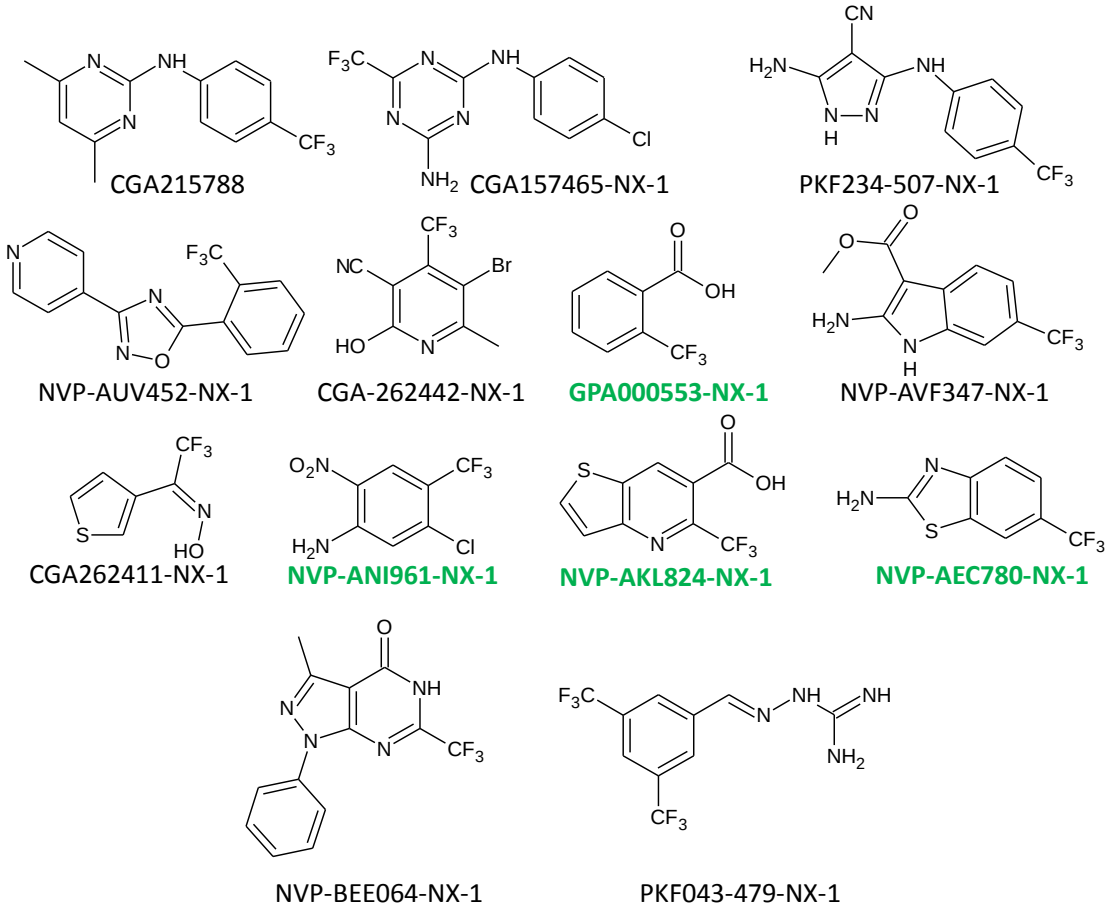


Figure 3-4. Structures of Novartis non-proprietory NMR screening hits. Commercially available compounds are indicated in green.

	120 ms		320 ms		400 ms	
Compound ID	GPA000553	AEC780-NX-1	GPA000553	AEC780-NX-1	GPA000553	AEC780-NX-1
CGA215788	0.9	1.1	0.9	1.1	0.77	1.03
BPS552-NX-1	1.04	1.07	1.05	0.96	1.02	0.83
PKF234-507-NX-1	1.18	1.18	1.1	0.74	0.9	0.74
CGA157465-NX-1	1.15	1.35	1.6	1.1	1.45	1.08
NVP-AVF347-NX-1	0.99	0.9	1.1	0.75	0.87	0.71
CGA-262442-NX-1	1.7	0.98	3.2	0.88	3.1	0.88
NVP-QBL688-NX-2	1.7	1.2	2.8	1.15	2.8	1.2
NVF-AVF211-NX-1	0.92	1.12	0.98	1.02	0.76	1.02
CGA180224	0.98	1.1	1.07	1.1	1.1	1.1
NVP-AQL006-NX-1	1.22	0.98	1.4	0.78	1.2	0.88
NVP-AUV452-NX-1	1	1	1.05	0.87	0.76	0.92
CGP015134-NX-1	2.65	0.97	6.5	0.88	7.4	0.8
CGA262411-NX-1	1.05	1.1	1.2	0.96	1.13	0.96
CGA284249-NX-1	1.1	0.23	1.34	0	1.3	0
NVP-AKL824-NX-1	2.8	1.35	6.5	1.35	6.9	1.3
NVP-ANI961-NX-1	1.14	0.43	1.45	0.2	1.23	0
PKF043-479-NX-1	0.8	0.62	0.83	0.32	0.62	0.18
NVP-BEE064-NX-1	1.15	1.25	1.34	1.14	1.26	1.2
Negative control	0.84	0.72	0.95	0.85	0.96	0.9

Table 3-22. Results from the competitive binding NMR experiments using the validated NMR fragments. GPA000553 and AEC8780-NX-1 were used as reference binders. Their ^{19}F resonance signal intensity upon addition of a competitor molecule was normalised to that prior the addition. The measurements were taken at three different relaxation times – 120 ms, 320 ms and 400ms. Commercially available fragments are in blue.

3.2.6 NMR hit validation using enzyme activity assays

The commercially available NMR hits GPA000553-NX-1 (further referred as GPA000553), NVP-ANI961-NX-1 (further referred as ANI961), NVP-AKL824_NX-1 (further referred as AKL824) and NVP-AEC780-NX-1 (further referred as AEC780) were purchased and characterized using the colorimetric and MS enzyme activity assays. Inhibitory properties of these fragments against ERAP1 were confirmed (Table 3-23).

CGA262442 was structurally related to the 3D_22 hit and its analogues from the extended library. Thus, similar hits were identified by the orthogonal screening methods. CGA262442 binding to ERAP1 was shown by NMR screening. Hence, in the absence of structural information this was the first indication the at least some hits identified by the enzyme activity assays were potentially real ERAP1 binders rather than assay interference compounds.

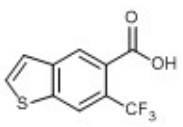
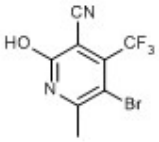
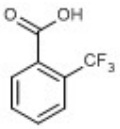
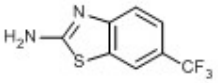
				
Fragment ID	**AKL824**	**CGA262442**	**GPA000553**	**AEC780**
Colorimetric IC_{50}, mM	**0.89 ± 1.61**	**0.66 ± 0.98**	**1.29 ± 2.37**	**4.1 ± 9.0**
MS IC_{50}, mM	**0.23 ± 0.18**	**0.33 ± 0.12**	**0.41 ± 0.14**	**0.60 ± 0.76**

Table 3-23. Results of the identified NMR hits characterisation using the colorimetric and MS enzyme activity assays.

3.2.7 The identified hit screening against LTA4H.

The hits were screened against LTA4H in the same colorimetric assay in order to assess their selectivity properties. Firstly, LTA4H kinetic properties were established for the colorimetric assay. The obtained K_M was 0.45 ± 0.05 mM and v_{max} was 0.02 ± 0.01 mol s^{-1} $_{LTA4H}$$mol^{-1}$ under the assay conditions. The reported literature K_M value for this assay was 0.26 mM in a higher salt buffer. [80] The obtained K_M under those assay conditions was 0.25 ± 0.01 mM, which was in a good agreement with the reported value.

41 compounds that had been previously identified as ERAP1 inhibitors were screened against LTA4H using the same colorimetric assay as before (Table 3-24). Only two fragments – CGA262442 and 3D_5 showed an inhibitory effect. The former had a comparable potency against both ERAP1 and LTA4H, and the latter was ten times more potent against ERAP1 than LTA4H. 30 fragments did not show any effect on the enzymatic activity of LTA4H in the colorimetric assay. Eight fragments exhibited an activating effect on the enzyme activity rather than an inhibitory one. This implied that their interaction mechanism with LTA4H was different to that with ERAP1. Fragment Ext_43 had a remarkably strong effect on LTA4H activity. At 2 mM concentration it increased the reaction speed 45-fold compared to the control. Such activating effect of the phenoxy-phenyl type molecule ARM-1 on LTA4H had been previously described. [46] The co-crystallisation of Ext_43 with LTA4H confirmed that it was binding to exactly the same site as ARM-1 (see Figure 4-10).

Thus, the screening showed that most of the identified inhibitor scaffolds had no selectivity issues with LTA4H.

IRAPi	LTA4Hi	Zn_A7	3D_E10	Ext_2	Ext_6	Ext_7	Ext_8	Ext_9	Ext_10
0.15 (x2)	Inactive	0.10 (x10)	Inactive	Inactive	Inactive	0.02 (x10)	Inactive	Inactive	Inactive
AKL824	CGA262442	Ext_51	Ext_52	Zn_B4	Zn_D1	M_6D7	GPA00055	Ext_12	Ext_15
Inactive	0.97	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
Ext_16	Ext_17	Ext_18	Ext_19	3D_C4	Ext_22	3D_C7	3D_C9	Ext_27	Ext_31
Inactive	Inactive	Inactive	Inactive	2.4	Inactive	Inactive	Inactive	Inactive	Inactive
Ext_34	M_4B9	AEC780	M_12A2	Ext_40	M_10B4	M_10E3	Ext_41	Ext_43	Ext_46
Inactive	Inactive	1.1 (x10)	Inactive	Inactive	0.005 (x4)	0.009 (x6)	Inactive	0.04 (x45)	0.20 (x5)
Ext_47									
Inactive									

Table 3-24. The results of the selectivity screening of the identified ERAP1 hits against LTA4H using the same colorimetric assay. The compounds that did not show any effect are in red. The compounds that showed inhibitory effect are in green, IC_{50} are given in mM. The compounds that showed activating effect (speed of the reaction was higher than the control) are in yellow, the value in parenthesis is the activation effect at 2 mM fragments concentration.

3.3 Conclusions

The colorimetric and the MS enzyme activity assay have been successfully used for ERAP1 inhibiting fragments screening and characterisation. The fluorogenic enzyme activity assay has been shown to be unsuitable for the above purpose because it generated a large number of false positive hits and a third of the screened fragments interfered with assay.

As the result of the screening campaign 13 structurally diverse hits have been identified and validated. The characterisation of a follow-up library containing 52 compounds structurally related to the above hits has provided some interesting insight about their structure-activity relationship. In addition, a reported inhibitor against IRAP, an aminopeptidase belonging to the same M1 family as ERAP1, has been identified as a potent ERAP1 inhibitor.

The screening by ^{19}F NMR of the Novartis proprietary library containing 532 fluorinated compounds has yielded 20 identified hits. Structures of 13 of those were disclosed by the company, only four molecules being commercially available. These compounds have been purchased and characterised using the enzyme activity assays, confirming their inhibitory properties against ERAP1. In the absence of structural information, this screening has been the first indication that some of the

identified hits did really bind to ERAP1. ANI961 NMR hit is structurally related to the 3D_22 fragment, which has been identified as a hit by the enzyme activity assays. This similarity has shown that at least some of the hits identified by the enzyme activity assay are real ERAP1 binders.

The screening of the 41 identified ERAP1 hits against LTA4H has shown that 30 of them do not affect the enzyme activity in the colorimetric assay. Only two fragments – CGA262442 and 3D_C4 – have displayed inhibitory effect on LTA4H. Eight fragments – IRAP-I, Zn_1, Ext_7, AEC_780, M_18, M_19, Ext_43 and Ext_46 – have exhibited activating effect on the LTA4H activity. Such an effect has been previously described for structurally similar compound to M_18, M_19, Ext_43 and Ext_46.

The inability to obtain structural information about the identified hits has been the main hurdle of the project. This has not allowed interpreting the observed structure-activity relationships for the identified inhibitor scaffold. Thus, it has not been possible to rationally combine several fragments scaffolds into a larger “lead” inhibitor.

3.4 Methods

All enzyme activity assay measurements were performed at room temperature.

Colorimetric assay

For kinetic studies continuous assays at different L-leucine-nitroanilide concentrations (2.67 mM, 2.3 mM, 2 mM, 1.5 mM, 1 mM, 0.7 mM, 0.5 mM, 0.25 mM) and different ERAP1 concentrations (20 nM, 10 nM, 5 nM, 2 nM, 1 nM) in 10 mM Tris (pH 7.5) buffer were recorded. Total cell volume of one assay was 300 µl. Measurements were recorded on the Spectramax Plus^{384} spectrometer. Typical run-time was 40 minutes, measurements were taken every 20 seconds. The obtained data was analysed using MatLab (see appendix I) to give the initial reaction rate. The data fitting to the Michaelis-Menten equation was done using Prism software. The substrate concentration was calculated from the measured absorbance using the Beer-Lambert-Bouguer law, assuming p-nitroaniline extinction coefficient to be $\varepsilon = 9920\ cm^{-1}\ M^{-1}$ (H. Yamaguchi et al., MIMB, 815, 187-198)

The screening continuous assay contained 5 nM ERAP1, 2 mM L-leucine-nitroanilide and 2 mM of a fragment in 10 mM Tris (pH 7.5), giving the total volume of 300 µl. Measurements were recorded on the Spectramax Plus^{384} spectrometer. Typical run-time was 40 minutes, measurements were taken every 30 seconds. The obtained data was analysed using MatLab to give the initial reaction rate.

The continuous assay for measuring fragments' IC_{50} contained 5 nM ERAP1, 2 mM L-leucine-nitroanilide and a fragment at different concentrations (1 mM, 0.75 mM, 0.5 mM, 0.25 mM, 0.1 mM, 75 µM, 50 µM, 5µM, 1µM and 0.5 µM) in 10 mM Tris (pH 7.5), the total volume of 300 µl. Measurements were recorded on the Spectramax Plus^{384} spectrometer. Typical run-time was 40 minutes, measurements were taken every 20 seconds. The obtained data was analysed using MatLab to give the initial reaction rate. The IC_{50} were calculated using the Prism software.

Mass-spectrometry assay

For kinetic studies continuous assays at different QITANRELIQQEL concentrations (15.0 µM, 10.0 µM, 7.0 µM, 4.0 µM, 2.0 µM, 1.0 µM, 0.2 µM) and 33 nM ERAP1 in 10 mM Tris (pH 7.5) buffer were recorded. Total cell volume of one assay was 500 µl. Measurements were recorded on the RapidFire QTOF Mass-spectrometer. Typical run-time was 20 minutes, measurements were taken every 130 seconds. Quantitative data was obtained by the generation and integration of Extracted Ion Chromatograms (EIC). QITANRELIQQEL and its trimming product were observed as single and double charged ions. The data fit to the Michaelis-Menten equation was done using Prism software.

The confirmation end-point assay contained 33 nM ERAP1, 6 µM QITANRELIQQEL and 1 mM of a fragment in 10 mM Tris (pH 7.5), the total volume of 50 µl. The reaction was quenched at 10 minutes by adding 1 µl of 10% formic acid and the obtained mixture was diluted by factor of ten in the same buffer as above. The diluted sample was analysed on RapidFire QTOF Mass-spectrometer. Quantitative data was obtained by the generation and integration of Extracted Ion Chromatograms (EIC) for the doubly charged ions QITANRELIQQEL (m/z 779) and TANRELIQQEL (m/z 658) of each peptide intermediate using the MassHunter Qualitative Analysis Software (Agilent). Extent of the substrate conversion was calculated as EIC area of TANRELIQQEL divided by sum of EIC areas of TANRELIQQEL and QITANRELIQQEL.

The end-point assay for measuring fragments' IC_{50} contained 33 nM ERAP1, 6 µM QITANRELIQQEL and a fragment at different concentrations (1 mM, 0.9 mM, 0.7 mM, 0.2 mM, 0.01 mM, 0.001 mM) in 10 mM Tris (pH 7.5), the total volume of 50 µl. The reaction was quenched at 10 minutes by adding 1 µl of 10% formic acid and the obtained mixture was diluted by factor of ten in the same buffer as above. The diluted sample was analysed on a RapidFire QTOF Mass-spectrometer. Quantitative data was obtained in the same way as in the confirmation assay above. The IC_{50}'s were calculated using the Prism software.

Fluorogenic assay

For kinetic studies continuous assays at different WRVYEKC(dnp)ALK concentrations (60 µM, 40 µM, 20 µM, 10 µM, 5 µM, 2 µM, 1 µM, 0.5 µM and 0.33 µM) and different ERAP1 concentrations (5 nM and 3 nM) in 10 mM Tris (pH 7.5) buffer were recorded. Total cell volume of one assay was 300 µl. Measurements were recorded on the POLARstar Omega fluorimeter. Typical run-time was 20 minutes, measurements were taken every 30 seconds. The fluorescence calibration showed that 1 µm tryptophan induced fluorescence of 50 units. The obtained data was analysed by MatLab to give the initial reaction rate. The data fit to the Michaelis-Menten equation was done using Prism software.

The screening continuous assay contained 10 µM WRVYEKC(dnp)ALK, 5 nM ERAP1 and 1 mM fragment in 10 mM Tris (pH 7.5) buffer. Total cell volume of one assay was 300 µl. Measurements were recorded on the POLARstar Omega fluorimeter. Typical run-time was 20 minutes, measurements were taken every 30 seconds. The obtained data was analysed by MatLab to give the initial reaction rate.

The continuous assay for measuring fragments IC_{50}'s contained 12 µM WRVYEKC(dnp)ALK, 6 nM ERAP1 and a fragment at different concentrations (2.4 mM, 1.2 mM, 0.6 mM, 0.15 mM, 0.015 mM, 0.0015 mM) in 10 mM Tris (pH 7.5) buffer. Total cell volume of one assay was 250 µl. Measurements were recorded on the POLARstar Omega fluorimeter. Typical run-time was 20 minutes, measurements were taken every 30 seconds. The obtained data was analysed by MatLab to give the initial reaction rate. The IC_{50} were calculated using the Prism software.

^{19}F NMR screening

The initial screening of the fluorinated compounds mixtures was performed with 1.67 mM final concentration of each fluorinated fragment in a mixture in 50 mM Tris pH 7.6 and 1 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) as a NMR standard. ERAP1-c200 was added to the NMR tube to give a final enzyme concentration of 5 µM. The measurements were performed on the AV600

Brucker machine with QNP probe-head at 120 ms and 320 ms relaxation time. Brucker Topspin Software was used for the data analysis.

The validation of the identified NMR hits was performed using 50 µM fragments in 50 mM Tris pH 7.6 and 1 mM DSS as a NMR standard. ERAP1-c200 was added to the NMR tube to give the final enzyme concentration of 5 µM. The measurements were performed on the AV600 Brucker machine with QNP probe-head at 120 ms and 320 ms relaxation time. Brucker Topspin Software was used for the data analysis.

In the competitive binding experiments 50 µM GPA000553-NX-1 and 50 µM NVP-AEC780-NX-1 were used as the reference binders. 50 mM Tris pH 7.6 was used as the buffer containing 1 mM DSS as a NMR standard. ERAP1-c200 was added to the NMR tube to give the final enzyme concentration of 5 µM. A competing fragment was added to a final concentration of 500 µM. The measurements were performed on the AV600 Brucker machine with QNP probe-head at 120 ms, 320 ms and 400 ms relaxation time. Brucker Topspin Software was used for the data analysis.

Peptide substrates

WRVYEKC(dnp)ALK and QITANRELIQQEL were purchased from Peptide Synthetics.

4. Enzyme crystallisation and X-ray data analysis

4.1 Introduction

In a typical Fragment-Based Drug Design (FBDD) campaign, once a set of fragments hits is identified, obtaining structural information about their binding mode is of paramount importance. This allows growing or merging of fragments by rational design rather than by using a combinatorial approach. In addition, this significantly decreases the number of second round compounds that have to be synthesized and screened. Thus, the availability of structural information significantly increases the overall success of the FBDD strategy. [81]

X-ray crystallography and NMR are the most widely used methods for obtaining the structural information. Their sensitivity allows detecting fragment-protein interactions, which are very weak compared to the larger ligands typically used in HTS. Historically, NMR was the primary method, due to the low throughput of crystallography. However, the advantage of the latter approach is its remarkable sensitivity and the detail of structural information it provides. [82]

The recent advances in high-throughput X-ray crystallography allow it to be routinely applied for FBDD campaigns. This progress has been achieved due to improvements in the following aspects of the crystallography pipeline. Firstly, the throughput of crystallisation trials has significantly increased as the result of the miniaturisation and automation of crystallisation trials, the introduction of standardised commercial crystallisation screens such as Hampton Research and the robotisation of crystal plate imaging [83] using systems like Minstrel HT at SGC Oxford. Secondly, brighter synchrotron beamlines, faster X-ray detectors and automation of the crystal mounting has allowed obtaining higher quality data from smaller crystals in a much shorter time. [82] Finally, the automation of data processing and refinement using software such as *CCP4* [84] *or Phenix* [85] has significantly decreased the amount of manual input required to solve each crystal structure, thus also increasing the throughput. [86] To date there are dozens of molecules in clinical trials, where

the lead was optimised from fragment hits using the structural information from X-ray crystallography, for example kinase inhibitor SNS-314 [87] or LTA4H inhibitor DG-051. [59]

The main hurdle in the application of X-ray crystallography to FBDD is the ability to reproducibly grow well-diffracting crystals (ideally at resolution limit <2.5Å) of the target protein, which tolerate soaking of the fragment libraries. Since ERAP1 was successfully crystallised in our group previously, which is described in detail in the Introduction chapter, X-ray crystallography was considered as an adequate strategy for obtaining structural information of a new chemical starting point for inhibitor design.

4.1.1 Chapter overview

First, a reproducible ERAP1 crystallisation system based on the previously optimised ERAP1 crystallisation protocol from our group has to be established in order to obtain structural information about the interactions of the identified hits (see Chapter 3) with the target protein. Considering the small size of the screened fragments, the resolution of ERAP1 co-crystals has also to be improved compared to that of the published enzyme structure 2YD0 that was solved at 2.7Å resolution. However, considering the improvements in beam lines precision and crystallographic software since then, such resolution would be achievable even using the crystals of the same quality.

In section 4.2 the attempts to find suitable crystallisation conditions for ERAP1 are described. Then a different strategy, where the fragment hits are co-crystallised with LTA4H – an aminopeptidase of the same M1 family with the active site similar to that of ERAP1 – is described in sections 4.3 and 4.4.

4.2 ERAP1 crystallisation

The initial crystallization strategy involved reproducing crystals of the full-length closed state ERAP1-c200 in complex with bestatin that had been previously obtained in our group. [28] The crystals of the closed form were obtained by mixing 100 nl of protein solution (15 mg⁄ml) and 50 nl of a precipitant consisting of 1.4 M potassium citrate; 0.1 M cacodylate pH 5.7 and 0.17 mM n-dodecyl-β-D-maltoside at 20^{o}C. The crystals appeared after 30 days.

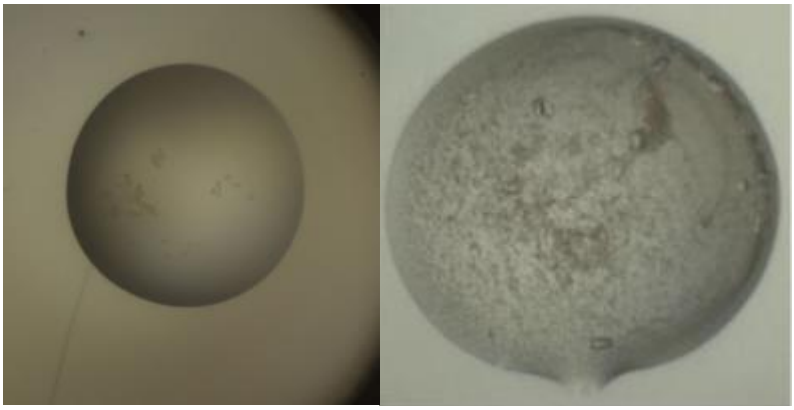


Figure 4-1. Crystallisation condition that yielded 2YD0 crystal (right) and exactly the same condition reproduced (left) with no sign of crystallisation at all.

Crystallisation trials were set up at 4^{o}C and 20^{o}C. Various batches of purified ERAP1-c200 with protein concentration between 10 and 20 mg/ml, preincubated with 2 mM bestatin, were set up using different coarse crystallization screens – JCSG, HIN, BCS, HCS – and the exact follow-up screen that was used to obtain the 2YD0 crystal. Despite the considerable effort, ERAP1-c200 crystals were not obtained either using the same condition as described before (Figure 4-1) or a new crystallisation condition. This was likely due to differences in protein quality between the obtained samples and those that were successfully co-crystallised previously, even though such differences have not been identified. In addition, the heterogeneity of the obtained samples due to the presence of several ERAP1 glycoforms complicated the task. The detailed analysis of ERAP1-c200 samples' quality is described in (chapter 2.2.3).

After numerous failed attempts to obtain the full-length ERAP1-c200 crystals, it was decided to try crystallising a shorter ERAP1 construct that was as close as possible to the amino acid sequence modelled in the published crystal structure 2YD0 (residues 46-940). The idea was to reduce the number of potentially labile residues in the protein, thus improving its crystallisability, therefore the construct ERAP1-c201 was chosen. Upon TEV-cleavage of its N-terminal tag the remaining sequence contained just a few residues outside the modelled sequence of 2YD0 (Table 4-1). ERAP1-c201 crystallisation trials were set both at 4°C and 20°C. Various batches of purified ERAP1-c201 with protein concentration between 6 and 22 mg/ml, preincubated with 2 mM bestatin, were crystallised using coarse crystallization screens – JCSG, HIN, BCS, HCS – and the exact follow-up screen that was used to obtain 2YD0 crystal. Despite all the attempts no crystals were observed.

	N-terminus	C-terminus
ERAP1-200	MVFLPLKWSLATMSFLLSSLLALLTVSTPSWCQSTEASPKRSDGT	MA**ENLYFQ\S**HHHHHHHHHHDYKDDDDK
ERAP1-201	MVSAIVLYVLLAAAAHSAFAAAMGHHHHHHSSGVDLGT**ENLYFQ\S**MASPKRSDGT	M

Table 4-1. The N- and C-termini of the constructs ERAP1-c200 and ERAP1-c201. The sequence of the residues 46-940 is the same for both constructs. TEV-protease cleavage site is marked in green. Upon TEV-cleavage ERAP1-c201 consists almost entirely from residues that were modelled on the aforementioned crystal structure 2YD0.

4.3 Rationale for using LTA4H as an ERAP1 substitute enzyme to obtain structural information about the identified hits

Since the ERAP1 crystallisation campaign was unsuccessful for a long time, it was decided to try using a different enzyme, which would be easier to crystallise but with a very similar active site. All the identified hits are small fragments, so if they bind to the catalytic site of ERAP1, differences in the binding pockets of the two enzymes further away from the Zn ion should not be very important. The co-crystals with the identified hits would provide at least some structural information about their potential interactions with ERAP1 in this case.

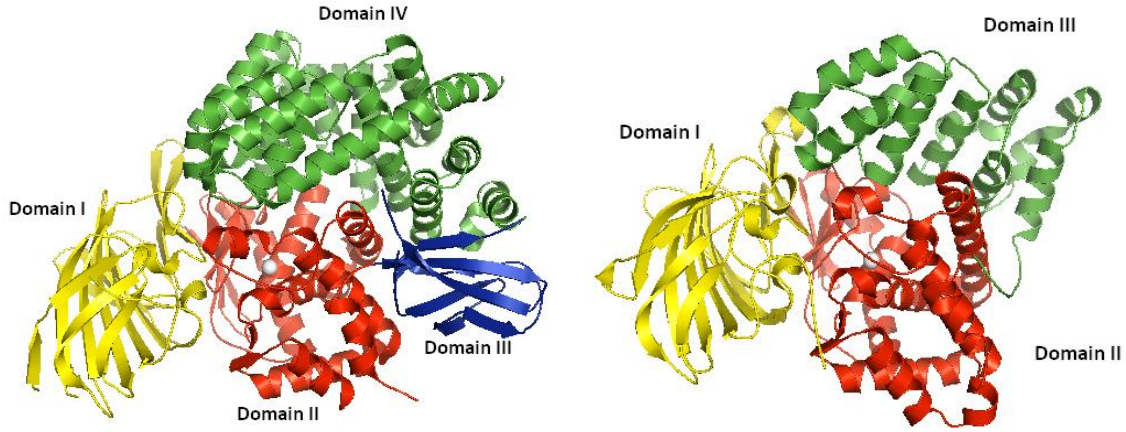


Figure 4-2. Closed state full-length ERAP1 (left, PDB: 2YD0) and LTA4H (right, PDB: 1HS6). Zn ion in the catalytic site is represented as a white sphere. Structurally similar domains are coloured the same. For a more detailed description of ERAP1 and LTA4H structures see Chapter 1.

LTA4H was chosen as a crystallisation substitute for several reasons. Firstly, it belongs to the same M1 aminopeptidase family with the same catalytic peptide cleavage mechanism as ERAP1. [28] Moreover, the key residues in its active site are not just the same as in ERAP1, they also have a very similar conformation to those in the ERAP1 closed state (Figure 4-4). Secondly, LTA4H had already been successfully used in FBDD co-crystallisation campaigns [59] with reproducible crystallisation conditions yielding high (<2Å) resolution crystals. Finally, this enzyme was expressed using the BL21 bacterial system, which allowed quicker and higher protein yield compared to the Sf9 insect cell system used for ERAP1 production. In addition, bacterially expressed proteins are not N-glycosylated, which makes it easier to achieve reproducibility of protein expression and to control homogeneity of the purified protein samples.

4.4 LTA4H crystallisation

Crystallisation trials were set up with both full-length and TEV-cleaved LTA4H in presence of bestatin in a set of coarse screens – JCSG7, HIN3 – and a follow-up screen – LTA4H-fu1 – based on the reported LTA4H crystallization conditions [46], [59], [35] containing 100 mM NaAc, 100 mM imidazole/HCl pH 6.2-7.4 and 14-36% PEG 8000 (wt/vol). Crystallisation experiments were performed both at 4^{o}C and 20^{o}C. All three cited papers reported a very similar crystallization condition with PEG 8000 as a precipitant, and sodium acetate, ytterbium chloride and imidazole as additives, which indicated that this condition ought to provide a robust crystallization method.

The follow-up conditions yielded a lot of needle-like crystals in many wells both with full-length and TEV-cleaved LTA4H after three days. The crystals from the latter had less nucleation and grew bigger (Figure 4-3), hence they were chosen for optimisation. The first round showed that the crystallisation was not affected by the pH, but higher PEG 8000 concentrations – 24%-36% – were preferred. Thus, a modified screen – LTA4H-fu2 – was created containing 100 mM NaAc, 100mM imidazole/HCl pH 6.8-7.2 and 26-36% PEG 8000 (wt/vol.).

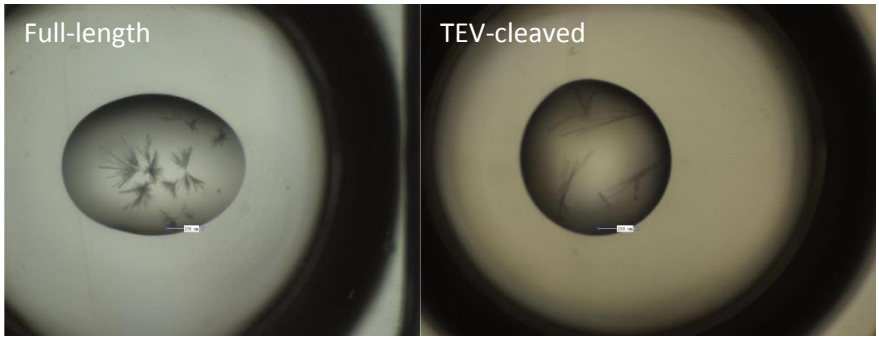


Figure 4-3. The first co-crystallisation round using full-length (left) and TEV-cleaved (right) LTA4H with 10 mM bestatin in the same LTA4H-fu1 condition – 28% PEG 8K, 0.1 M NaAc, 0.1 M imidazole pH 6.6.

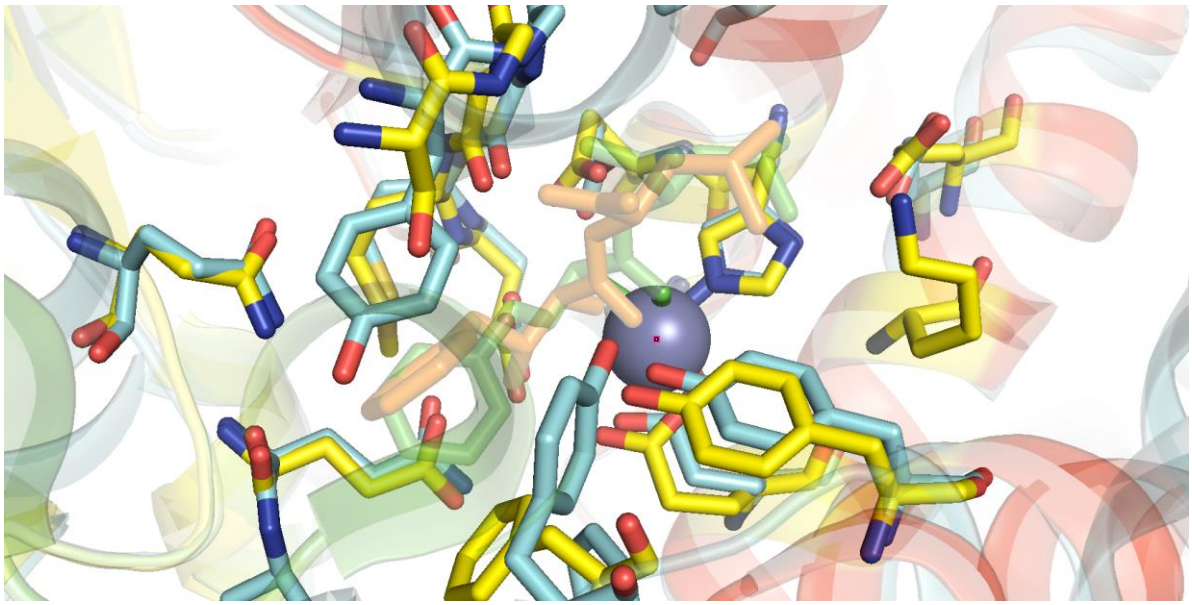


Figure 4-4. Active site residues within 4Å from bestatin in ERAP1 (PDB: 2YD0, yellow) and LTA4H (PDB: 1HS6, blue) crystal structures aligned using PyMOL software. Zn is shown as a grey sphere. Bestatin is represented as transparent sticks - orange from the ERAP1 structure and green from the LTA4H one. The image created using Pymol software.

The optimisation strategy concentrated on reducing nucleation to yield fewer but larger crystals in a drop. The obtained crystals were therefore used for micro-seeding in the subsequent crystallisation rounds exploring the effect of temperature and enzyme concentration on the growth of crystals using the modified screen (Figure 4-5). The best crystals were obtained from the crystallisation at 20^{o}C using 6 mg/ml enzyme sample, which corresponded to 40% less enzyme than in the initial crystallisation, and micro-seeding. As usual, the protein was preincubated with 5 mM bestatin prior to the crystallisation plate set up. The biggest crystal (580 µm) from the well with 100mM NaAc, 100 mM imidazole/HCl pH 7.2 and 36% PEG 8000 was mounted. It diffracted to 2.2Å at the Diamond Light Source (DLS), showing a good agreement with the published LTA4H-bestatin co-crystal 1HS6 (**Error! Reference source not found.**).

The obtained crystal was needle-like rather than plate-shaped as reported [35] and was of monoclinic P 1 2 1 point group, whereas 51 out of 52 published human LTA4H crystals belong to the same orthorhombic P 2_1 2_1 2_1 symmetry group.

Then, it was attempted to crystallise TEV-cleaved LTA4H using the optimised conditions in the presence of a set of the identified fragments hits instead of bestatin. There were either no crystals

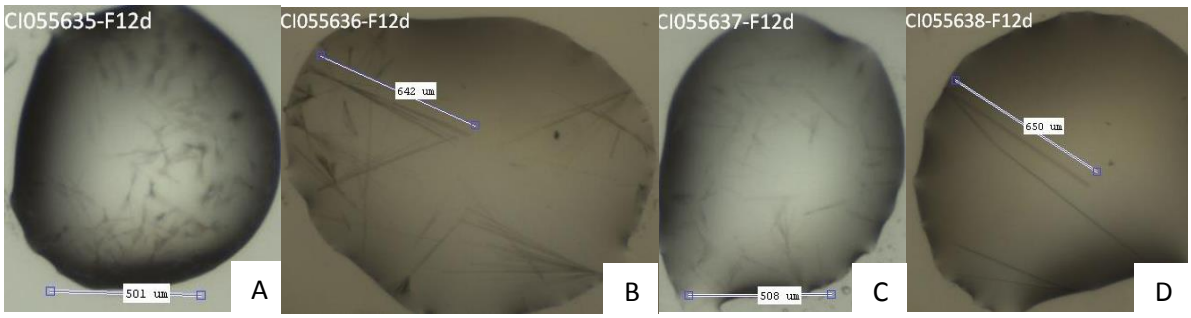


Figure 4-6. LTA4H crystals grown in 36% PEG 8K, 0.1M NaAc, 0.1M imidazole pH 7.0. The crystals obtained at 8 mg/ml (A and B) and 6 mg/ml (C and D) of the protein with micro-seeding. Crystallisation at 4 ^{o}C (A and C) and at 20^{o}C (B and D). The best crystals were obtained at 20^{o}C and lower enzyme concentration.

observed after a week or the obtained crystals were small and numerous, just as in the initial screen. The idea of optimising crystallisation conditions for every fragment was considered impractical. Consequently, it was decided to optimise apo crystals of LTA4H, which could be subsequently soaked with the desired compounds.

The apo crystals that were obtained using the previously optimised conditions with micro-seeding of the needle-like crystals were small and numerous. In the reported crystallisations LTA4H was stored after the purification in a much lower salt buffer compared to ours – 25 mM KCl versus 500 mM NaCl respectively. Hence crystallisation of the desalted enzyme samples was attempted. In addition, the effect of micro-seeding and addition of 5 mM $YbCl_3$ to the crystallisation conditions was explored. As a result plate-like apo crystals were obtained. The key factors for the successful crystallisation were: a) low salt concentration of the enzyme purification buffer b) 5 mM YbCl3 additive and c) seeding. The plate-like crystals were much easier to handle and diffracted in-house with resolution below 2.5Å on the in-house diffractometer (Figure 4-6).

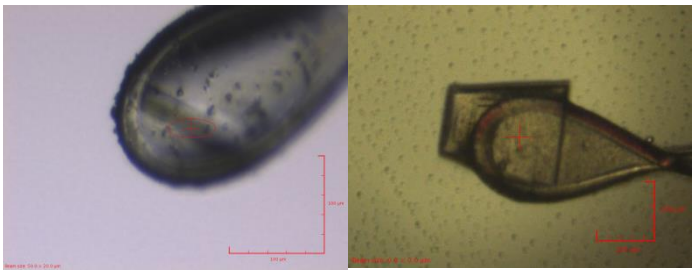


Figure 4-5. Needle-like (left) LTA4H crystal grown after protein sample preincubation with 50 mM bestatin and plate-like (right) LTA4H apo crystal after soaking with 50 mM bestatin in 30% DMSO.

Then one of the obtained plate-like crystals was soaked with 50 mM bestatin in 100% DMSO to test whether it survived soaking and if the ligand could diffuse into the active site of LTA4H. The soaked crystal diffracted to 2Å at the in-house diffractometer and the ligand indeed was observed in the active site. Thus, a LTA4H soaking system was successfully established.

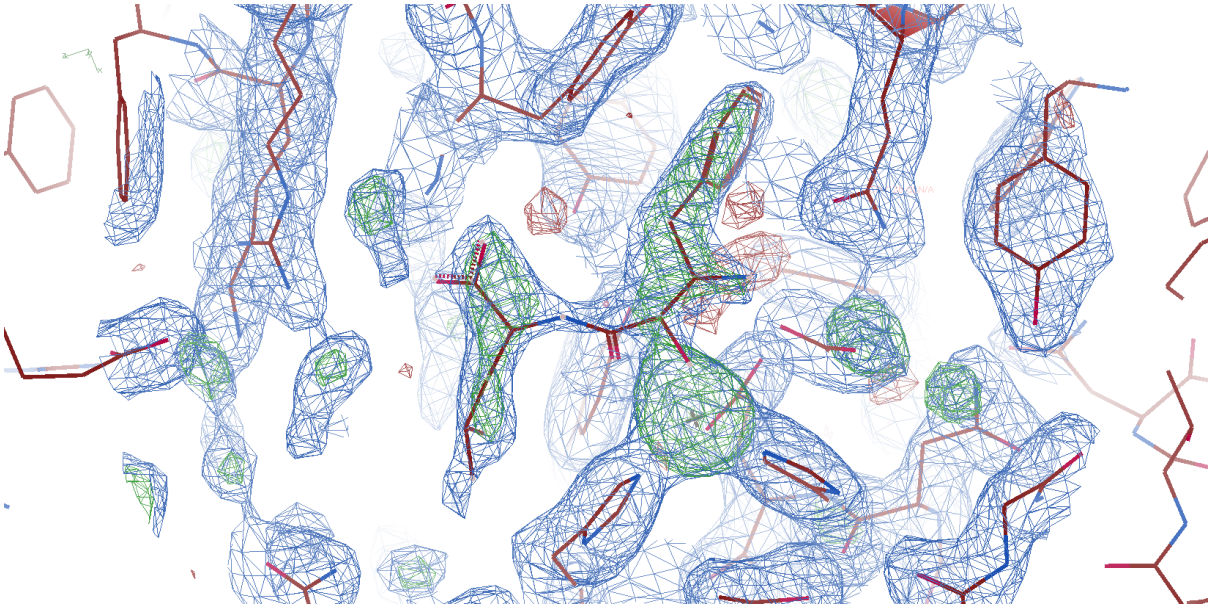


Figure 4-7.2F_o-F_c and F_o-F_c maps of LTA4H in complex with bestatin contoured to 1σ (blue) and +/-3σ (green/red), respectively. The bestatin molecule as well as the active site Zn ion (grey sphere) were not included in the refinement. Image created using Coot.

15 fragments were soaked into apo crystals (Figure 4-11). The soaked fragments were in 100% DMSO and their stock concentrations are reported in the Table 4-1. They were selected in order to cover the structural diversity of the identified hits with at least one selected molecule representing a particular hit structure scaffold. The set also contained several phenoxy-phenol type molecules that showed different degrees of LTA4H activation in the aminopeptidase activity assay. The latter were meant to provide insights about the nature of the observed activation effect.

Data for the soaked crystals were collected at DLS beamline I03 and the datasets were refined against the existing LTA4H model (PDB: 1HS6). The analysis of the electron density maps revealed that only five fragments – Ext_38, Ext_43, Ext_44, Ext_45 and Ext_47 – were successfully bound (Figure 4-11). No relevant electron density was observed for the rest of the soaked fragments. Data collection and refinement statistics of the soaked crystals are summarised in Appendix IV.

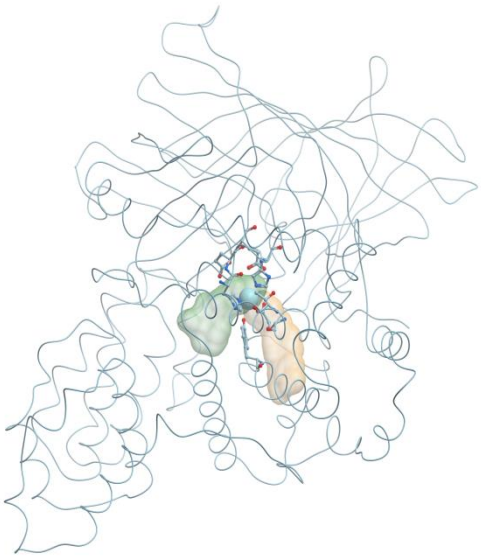


Figure 4-8. LTA4H peptide (green) and LTA_4 binding pockets (orange) based on the 4MKT LTA4H co-crystal structure with peptide and ARM-1 (PDB: 4MKT). The image created using ICM-pro.

The phenoxy-phenol type activating hits gave high resolution co-crystal structures with LTA4H (Figure 4-9). They were observed in the LTA_4 binding pocket of the enzyme (Figure 4-8), which does not exist in ERAP1.

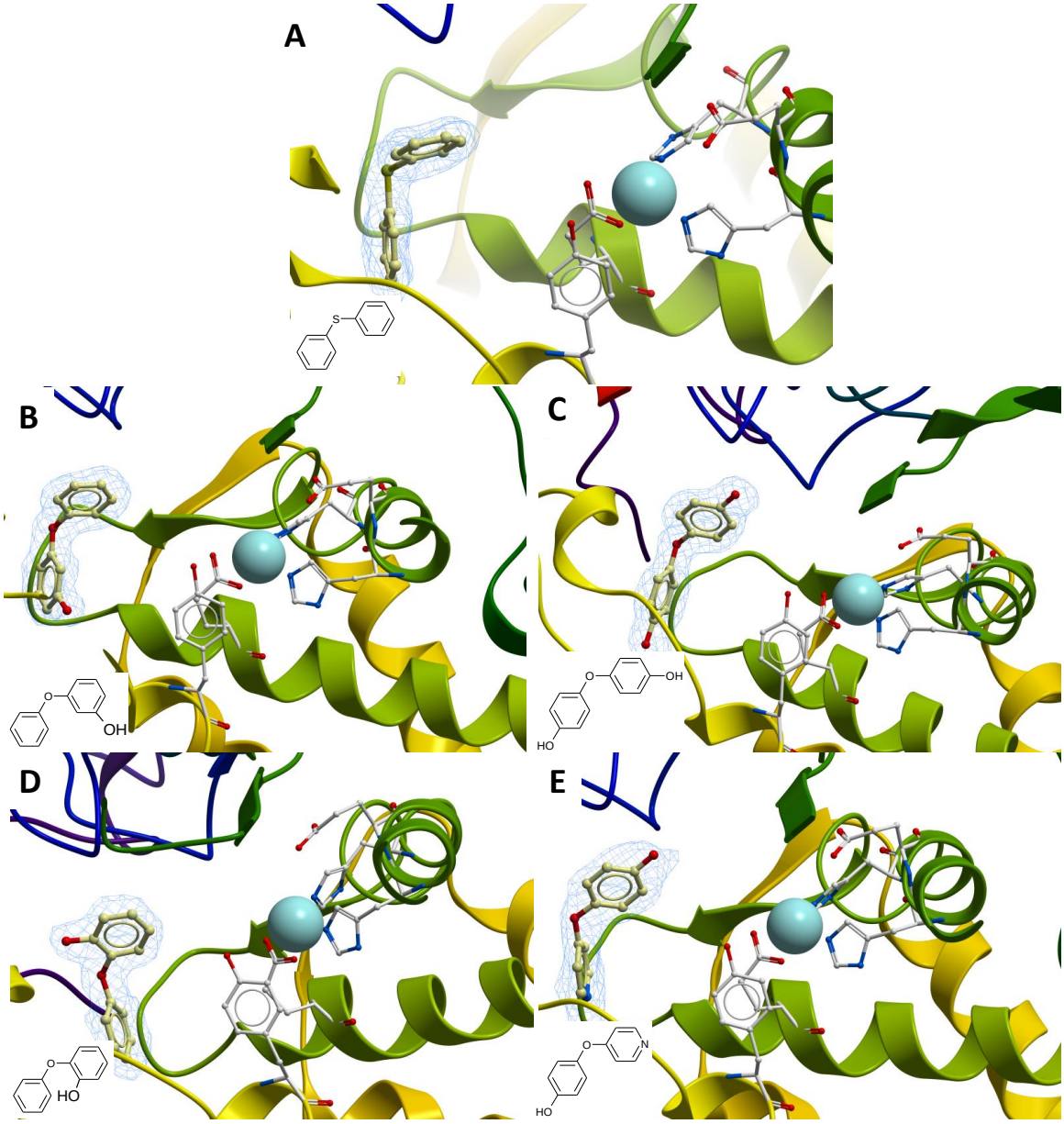


Figure 4-9. The refined crystal structures of Ext_38 (A), Ext_43 (B), Ext_44 (C), Ext_45 (D) and Ext_47 (E). All images are taken from the similar angle. $2F_o$-F_c electron density maps of the ligands are to 1σ. Zn ion represented by grey sphere, Zn-binding His^{295}, Glu^{296} and His^{299} as well as catalytically important Glu^{318} and Tyr^{383} are shown as sticks. The images were created using ICM-Pro software.

The binding modes of the co-crystallised ligands were in good agreement with that of the reported molecule ARM1, which represented the same scaffold (Figure 4-10). [46] Moreover, ARM1 also was reported to exhibit activating effect on LTA4H in the same colorimetric assay as was used in this work.

From the obtained co-crystal structures the following LTA4H aminopeptidase activation mechanism was hypothesised. The fragments bind to the LTA_4 binding tunnel near the enzyme active site, thus a creating a small hydrophobic cavity, which accommodates the N-terminal leucine of the substrate peptide. The floor of this cavity is formed by the phenyl group of the fragment. Indeed, fragments Ext_43 and Ext_47, which in the co-crystal structures have their polar hydroxyl groups pointing towards the active site thus disturbing such hydrophobic pocket, have little (x4) to none activating effect. This hypothesis is further supported by the fact that introduction of large hydrophobic groups, *e.g.* benzyl, to the N-terminus of a peptide substrate has been reported to significantly (up to 100 times) increase the rate of the aminopeptidase reaction. [88] This bulky group supposedly binds to the same hydrophobic pocket of the LTA_4 binding site, which the co-crystallised fragments occupy, thus improving the affinity of the substrate.

When it was decided to use LTA4H as a substitute for fragments co-crystallisations, the fact that the fragments which inhibited ERAP1 in the colorimetric assay showed no effect on LTA4H aminopeptidase activity in the same assay was not known. The latter implied that LTA4H was not a suitable enzyme to model the hits' interactions with ERAP1. Indeed, none of the ERAP1 inhibitor hits were observed in the obtained crystal structures.

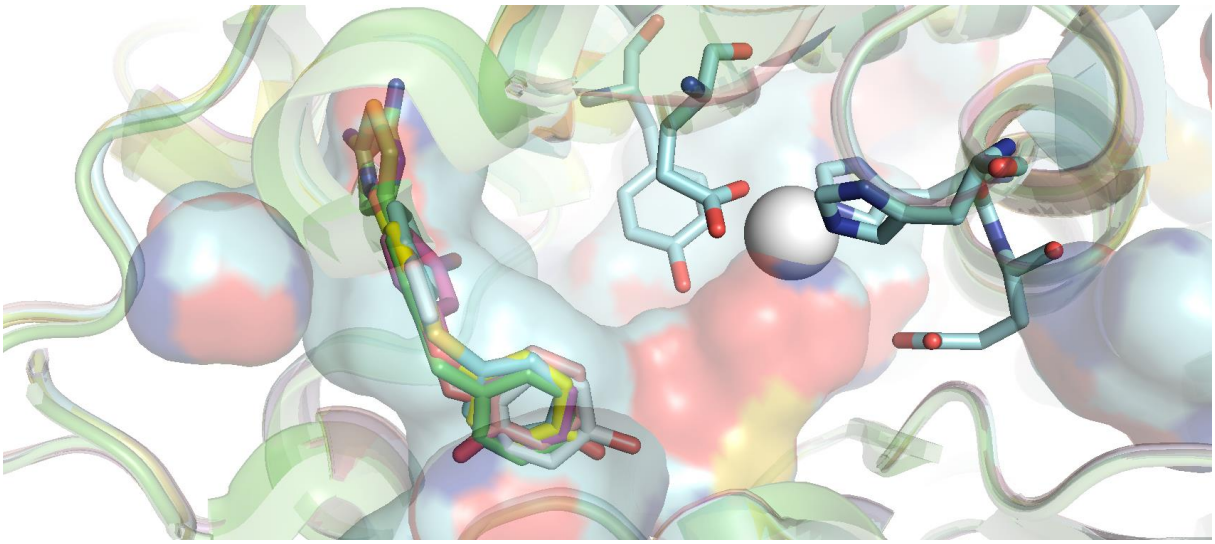


Figure 4-10. Aligned LTA4H co-crystal structures of Ext_38, Ext_43, Ext_44, Ext_45, Ext_47 and ARM-1 (PDB: 4L2L). All the ligands occupy the same pocket near the LTA4H catalytic site (Zn ion – white sphere, Zn-binding $His^{295, 299}$ and Glu^{296} as well as catalytic Glu^{318} and Tyr^{383} are shown as blue sticks). Image created using CCP4MG software.

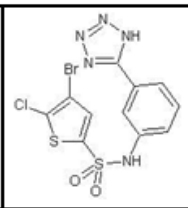
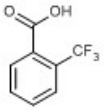
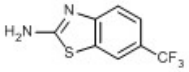
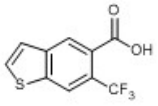
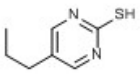
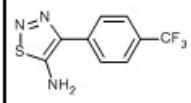
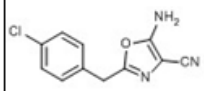
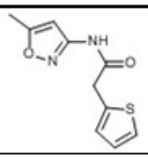
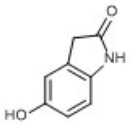
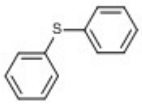
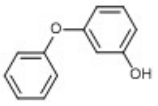
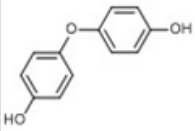
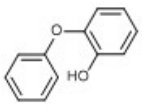
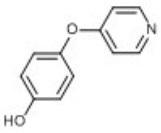
					
Fragment ID	**IRAP-I**	**GPA000553**	**AEC780**	**AKI824**	**Ext_6**
ERAP1, IC_{50} mM	**9 ± 6 µM**	**1.29 ± 2.37**	**4.1 ± 9.0**	**0.89 ± 1.61**	**Not active [MS]**
LTA4H, IC_{50} mM	**EC_{50} 0.15 ± 0.05 (x2)**	**Not active**	**1.08 ± 0.22**	**0.72 ± 0.85**	**EC_{50} Poor curve (x6)**
					
Fragment ID	**M_3**	**3D_12**	**3D_5**	**3D_4**	**M_20**
ERAP1, IC_{50} mM	**0.41 ± 0.25**	**ca 0.15**	**0.14 ± 0.72**	**0.08 ± 0.05**	**0.76 ± 0.65**
LTA4H, IC_{50} mM	**Not active**	**Not active**	**2.34 ± 5.11**	**Not active**	**Not active**
					
Fragment ID	**Ext_38**	**Ext_43**	**Ext_44**	**Ext_45**	**Ext_47**
ERAP1, IC_{50} mM	**Not active**	**ca. 1**	**Not active**	**Not active**	**Not active [MS]**
LTA4H, EC_{50} mM	**0.41 ± 1.58 (x50)**	**0.02 ± 0.01 (x45)**	**0.19 ± 0.17 (x4)**	**0.24 ± 0.09 (x30)**	**Not active**

Figure 4-11. A set of hits selected for soaking with LTA4H and their effect on aminopeptidase activity of ERAP1 and LTA4H in the colorimetric assay. (MS) implies that the fragment was not active in the ERAP1 Mass-Spectrometry activity assay which was used to confirm hits from the colorimetric one. Red frame highlights the fragments that were successfully co-crystallised with LTA4H.

4.5 Conclusion

All the attempts to reproduce full-length ERAP1-c200 crystallisation have failed. The attempts to crystallise a shorter ERAP-c201 construct, which was supposed to help crystallisation by reducing the length of flexible ends of the enzyme, have been also unsuccessful. This was attributed to the increased heterogeneity of ERAP1-c201 protein sample compared to that of ERAP1-c200 (see pg. 34).

A reproducible protocol for soaking fragments into apo LTA4H crystals has been successfully established. Unfortunately, no co-crystal structures have been obtained, when a set of ERAP1 hits has been soaked with the obtained LTA4H apo crystals. None of the ligands has been observed bound to LTA4H. In addition, most of the identified ERAP1 fragment hits did not exhibit any effect on LTA4H aminopeptidase activity in the same enzyme activity assay (see Table 3-24). Therefore, LTA4H has proved to be a poor substitute for modelling the fragment hit interactions with ERAP1, despite the structural similarity between two active sites.

4.6 Methods

Crystallisation of ERAP1

Crystallisation of ERAP1-c200 and ERAP1-c201 was performed with protein expressed and purified from Sf9 insect cells using 3 drop sitting vapour diffusion method at either 4^{o}C or 20^{o}C. Crystal drops were set up with the protein at a concentration of 10-20 mg/ml, which were pre-incubated with 2 mM bestatin for 30 minutes on ice. The protein compound mix was transferred onto a 96-well Swiss CI crystal plate (SWISSCI, Neuheim, Switzerland) with a mosquito robot (TTP Labtech) and mixed with precipitant solution in a 2:1, 1:1 and 1:2 ratio resulting in a final volume of 150 nl. JCSG, BCS, HIN and HCS sparse matrix screens, the follow-up screen as well as detergent and additive compound screens were obtained from Hampton Research. Crystallisation plates were imaged with a Minstrel HT system (Rigaku) and viewed in TexRank.

Crystallisation and structure determination of LTA4H

Crystallisation of LTA4H was performed with protein expressed and purified from BL21 *E. coli* cells using the 3 drop sitting vapour diffusion method at either 4^{o}C or 20^{o}C. Crystal drops were set up with the protein at a concentration of 6-14 mg/ml. For co-crystallisation with bestatin, protein samples were pre-incubated with 2 mM bestatin for 30 minutes on ice. The protein mix was transferred onto a 96-well Swiss CI crystal plate (SWISSCI, Neuheim, Switzerland) with mosquito robot (TTP Labtech) and mixed with precipitant solution in a 2:1, 1:1 and 1:2 ratio resulting final volume of 150 nl. JCSG, BCS, HIN and HCS sparse matrix screens, and the follow-up screen were obtained from Hampton Research. YbCl3 1.25 M stock solution was obtained from Alfa Aesar. Crystallisation plates were imaged with a Minstrel HT system (Rigaku) and viewed in TexRank. Crystals appeared after 1-5 days. Crystals were soaked in cryoprotectant containing 25% v/v ethylene glycol in well solution and then flash-cooled in liquid nitrogen at 100^{o}K. For fragment soaking, crystal drops were supplemented with 1 µl 100 mM fragment solution in DMSO and kept overnight. Crystals were screened for diffraction on the in-house rotating anode X-ray generator (Bruker). Diffracting crystals were selected for

screening and Data Collection at Diamond Light Source. The datasets were collected at DLS beamlines on samples at 100^{o}K.

All diffraction data were processed with XIA2 [89] as part of the auto-processing pipeline [90] at the Diamond Light Source. The structures were solved by molecular replacement with Phaser v2.5 [91] using PDB ID 1HS6 as a starting model. Ligand restraints were generated with ACEDRG [92]. The refinement of the datasets was done using the REFMAC [93] and iterative model building was performed with COOT. [94]

5. Inhibitor synthesis and analogue structure-activity relationship studies.

5.1 Introduction

5.1.1 Reported M1 class aminopeptidase inhibitors

M1 aminopeptidases have important functions in diverse systems in the human body, for example in the adaptive immune system. These enzymes have been associated with several diseases such as cancer or immune-mediated disorders, e.g. ankylosing spondylitis. Six out of eleven human M1 aminopeptidases – IRAP, ERAP2, AnPEP, EnPEP, LTA4H and ERAP1 – have been successfully crystallised, providing detailed information about their structure. Their catalytic mechanism has been elucidated. In addition, these enzymes have a well-defined active site, amenable for small molecule binding. [95] This makes them suitable targets for small molecule drug design. However, just a few reported selective inhibitors against mammalian M1 aminopeptidases have been disclosed to date.

Two groups have reported about development of selective AnPEP inhibitors [96], [97]. The first group reported K_i = 60 pM against AnPEP in the L-pNA colorimetric *in vitro* enzyme assay. The inhibitor did not show any effect on LTA4H in concentration up to 100 µM in the same assay. In addition, the inhibitor exhibited 10^6-fold selectivity against bovine cytosolic leucine aminopeptidase (LAP-C), an M17 class zinc-dependent aminopeptidase, and bacterial leucine aminopeptidase APaero. The second inhibitor had 850 nM reported potency in a cell-based assay and >1000 times selectivity against Matrix-Metalloproteinase 2, an MMP-class Zn-dependent aminopeptidase.

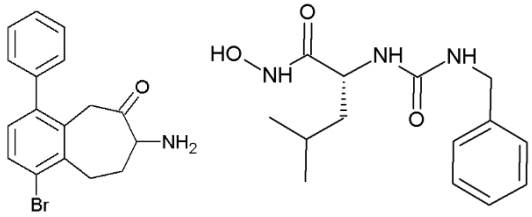


Figure 5-1. AnPEP inhibitors reported by C. Malereanu et al. [96] (left) and L. Su et al. [97] (right).

Two different IRAP inhibitors have been reported. [79], [98] The first inhibitor had a reported potency of 30 nM in the fluorogenic enzyme activity assay using L-leucine-4-methyl-7-coumarinylamide (L-MCA) as a substrate. No information was provided regarding its selectivity. The second inhibitor had 2.1 μM reported potency in the colorimentric enzyme activity assay using L-pNA. Again, no information about its selectivity was provided.

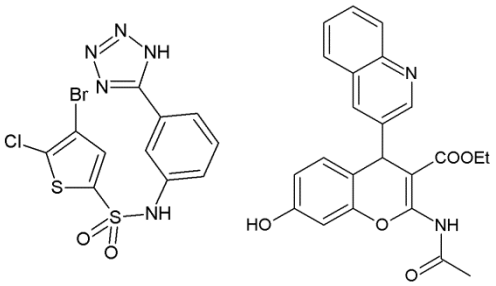


Figure 5-2. IRAP inhibitors reported by S. Borjade et al. [79] (left) and S. Mountford et al. (right). [98]

Two groups have reported LTA4H inhibitors with a related structure – DG-051 [59] and ARM-1. [46] DG-051 had the reported IC_{50} of 26 nM in the whole blood activity assay. No effect on the activity of AnPEP, ERAP2, APP1 (aminopeptidase P1), APA (aminopeptidase A) was observed. This molecule was shown to inhibit both LTA_4 hydrolytic and aminopeptidic activities of LTA4H. ARM-1 had a lower reported potency (IC_{50} = 500 nM *in vitro* enzyme activity assay), but was shown to specifically suppress hydrolase activity of LTA4H, leaving the aminopeptidase activity uninhibited.

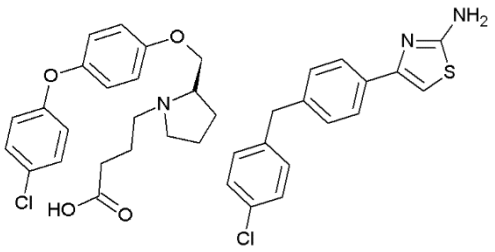


Figure 5-3. LTA4H inhibitors DG-051 [59] (left) and ARM1 [46] (right).

In this work DG-051 and IRAP-I were synthesised and screened against ERAP1. DG-051 was chosen, since its core was structurally similar to the M_18 and M_19 hits that had been identified during the screening campaign. The IRAP inhibitor was selected due to its novel and interesting structure, and its synthetic route that allowed quick development of a diverse set of analogues.

5.1.2 ICM Pro docking software overview

ICM-Pro software by Molsoft L.L.C. has been used for *in silico* docking of the suphonamide analogues. This software allows ligands docking and scoring of their binding energy to the target protein. It is good at predicting ligand binding conformation in the protein compared to other available programmes. [99] The calculated binding energy is less reliable though, which is the case for all software that implements the force field method for such computations. [100] ICM-Pro has been successfully used for *in silico* screening campaigns against diverse protein targets. [101], [102], [103]

5.1.3 Chapter overview

In this chapter, first the synthesis of the reported IRAP and LTA4H inhibitors is described. Then the results of *in silico* docking of sulphonamide inhibitor analogues using ICM-pro are reported. Three rounds of IRAP analogue synthesis based on the docking results are outlined. Finally, a set of IRAP inhibitor analogues were screened against ERAP1, thus attempting to design an inhibitor with < 1 µM potency.

5.2 Reported LTA4H and IRAP inhibitors synthesis

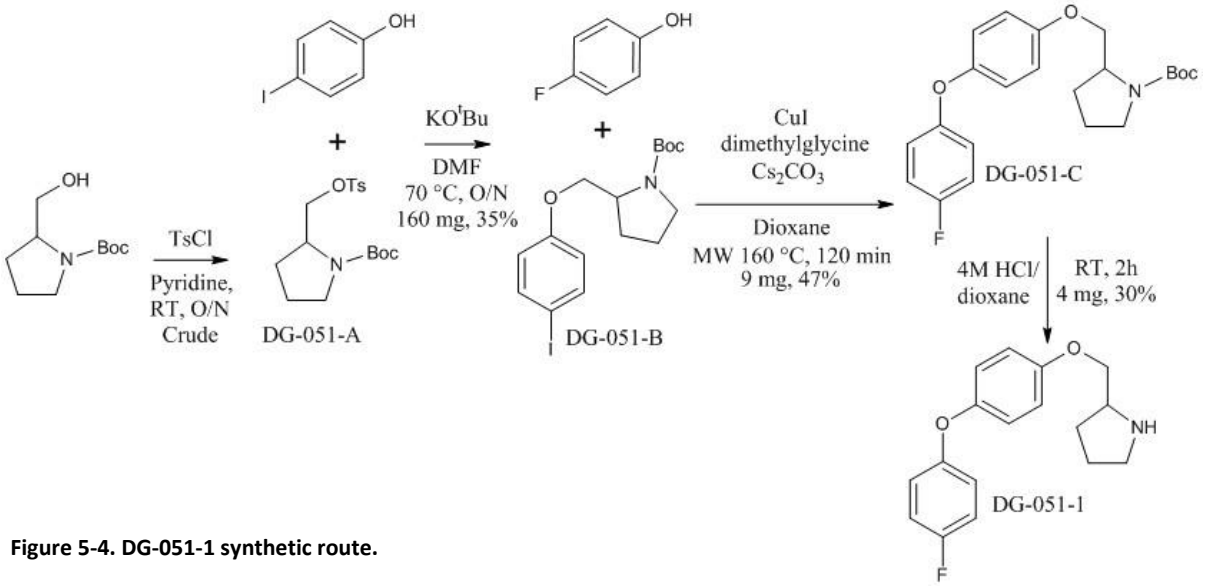


Figure 5-4. DG-051-1 synthetic route.

It was decided to resynthesise not the DG-051, but one of its reported analogues [59], which had showed 49 nM potency against LTA4H. This was due to the latter having fewer synthetic steps, thus allowing us a quicker access to this scaffold (Figure 5-4). In addition, chlorine was substituted for fluorine in the synthesised analogue structure DG-051-1. According to the paper, the chlorine was not involved in any binding interactions, but the improved molecule's pharmacokinetic parameters, which were not important at this stage in our study. The introduction of fluorine would allow this analogue to be screened by ^{19}F NMR, if the molecule showed inhibitory effect on ERAP1 activity (see Table 3-20).

The IRAP inhibitor (IRAP-I) was successfully synthesised following the reported synthetic route (Figure 5-10).

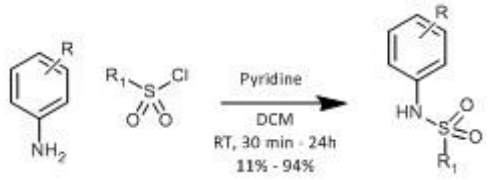


Figure 5-5. The general synthetic route for obtaining IRAP–I and sulphonamide analogues.

5.3 Sulphonamide analogues of the IRAP inhibitor

Towards the end of the project there was still no structural information about the binding mode of the identified fragments using X-ray crystallography (see Chapter 4). The only relatively large ERAP1 inhibitor that had been identified in this study was the previously reported IRAP-I (Table 3-20). It was reasonable to suggest that it was binding to the Zn ion in the ERAP1 active site through its tetrazole group, a carboxylic acid isostere, which in turn was a reported Zn-binding group. Thus, there was at least some putative idea about the binding mode of this hit.

5.3.1 Virtual screening of the sulphonamide analogues

In order to reduce the number of synthesised IRAP-I analogues a virtual screening campaign using ICM-Pro was performed. The IRAP-I structure was divided into three parts: a putative Zn-binding

"war-head", a linker and a sulphonamide substituent (Figure 5-6). The sulphonamides were supposed to be obtained by nucleophilic substitution of the relevant sulphonyl chloride including a linker amine. The "war-heads" were supposed to be joined with the linker using Suzuki coupling of the relevant boronic acid or ester with the linker's bromine atom. Thus, eight commercially available boronic acids and esters for "war-heads", two regioisomers of bromo-amino-phenyl linker and ten different sulphonyl chlorides were selected.

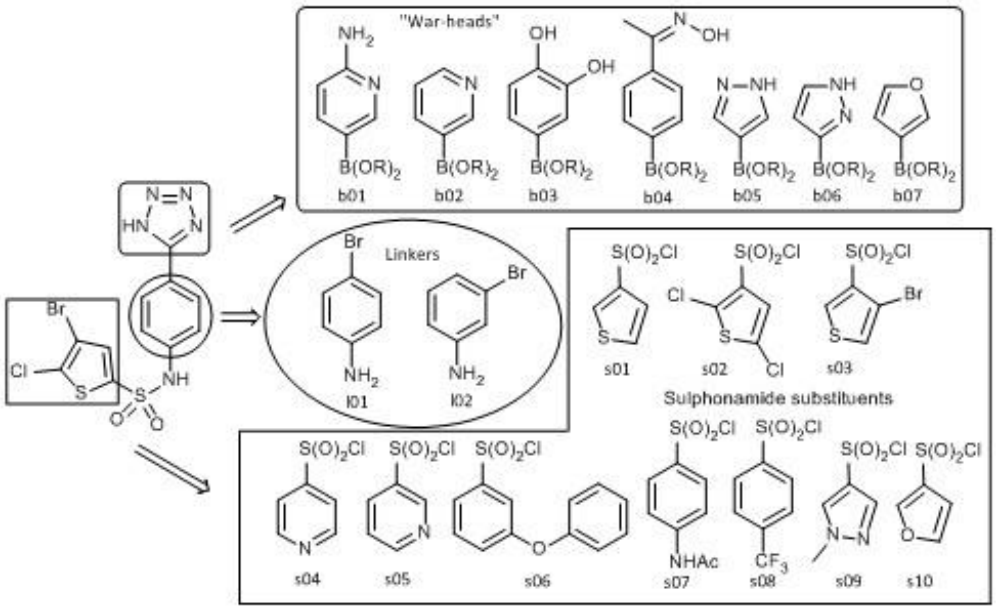


Figure 5-6. IRAP inhibitor structure was divided into three parts: the "war-head" (curved rectangle), the linker (ellipse) and the sulphonamide (rectangle). A set of commercially available fragments for each part has been developed.

All possible 160 combinations of the above fragments were docked *in silico* to the closed-state ERAP1 (PDB: 2YD0). Three parameters were considered, when fragments were selected for the synthesis of sulphonamide analogues. Firstly, it was the presence of interactions with Zn *via* the "war-head" or other group in any of the proposed conformations. Secondly, it was the presence of hydrogen bond interactions between a molecule and the enzyme residues in any of the proposed conformations. And finally, an arbitrary calculated binding energy cut-off of -25 units was introduced. If most molecules containing a certain fragment had no conformation with the calculated binding energy higher than the cut-off, such fragment was not selected for synthesis.

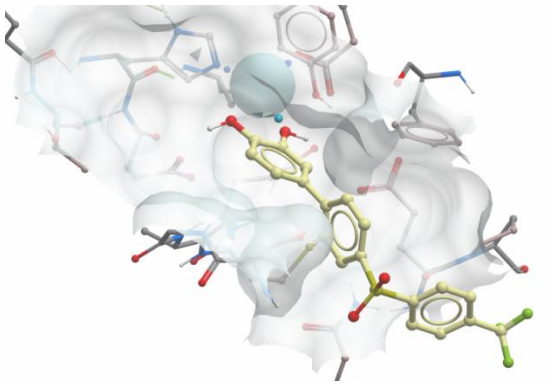


Figure 5-7. An example of the docking of the b03l01s08 molecule. In this conformation the catechol interacts with the Zn in the active site. Image created using ICM-Pro software.

As a result of the screening campaign fragments b04, s06 and s07 were removed from the panel. The vast majority of the analogues containing those fragments did not have interactions with Zn or ERAP1 residues and the calculated binding energy of most such molecules was below the cut-off value. In addition, furan based fragments b07 and s10 were removed. The proposed conformations were similar to the b05 and s01 containing analogues. However, the calculated binding energy of the proposed conformation was typically lower. Thus, the analogues containing these fragments were not believed to have superior potency or to increase the diversity of the accessible binding modes with the enzyme. Finally, the manufacturer reported issues with the synthesis of s04, so that molecule was not obtained.

5.3.2 Synthesis of the sulphonamide analogues

The analogues were synthesised using the same reaction between a relevant amine and sulphonyl chloride, which had been used for the IRAP-I synthesis (Figure 5-5). For the analogues, where a required amine was not available, an extended synthetic route using a Suzuki coupling reaction was applied (Figure 5-8). Initially, the $Pd(PPh_3)_4$ catalyst was used for the Suzuki coupling. However, little product was obtained with the unprotected amine-containing boronic acids, such as b01, b05 or b06, and it was hard to remove the reaction by-products. Therefore, the $PdCl_2$(dppf) was used as a catalyst instead, which gave much better yields of the desired product at a considerably shorter reaction time.

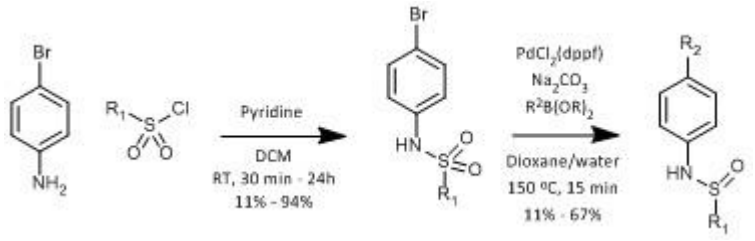


Figure 5-8. The amended synthetic route for sulphonamide analogues, when the required starting amine was not commercially available.

There were not many reported examples of Suzuki coupling with unprotected catechol boronic acids. Moreover, the reported yield of such reactions was low. [104] Hence, for the analogues with the catechol "war-head" b03, methyl ether protected catechol boronic acid was used instead. It was decided to use mild conditions for the alcohol deprotection [105] using *in situ* generated iodotrimethylsilane (Figure 5-9). [106]

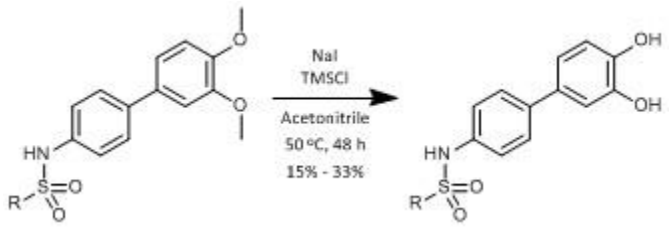


Figure 5-9. Alcohol deprotection using *in situ* generated TMSI yielding the analogues with catechol "war-head".

5.3.3 Screening of the sulphonamide analogues

The synthesised sulphonamide analogues inhibitory properties against ERAP1 were characterised using the colorimetric assay. The analogues were not characterised using the MS assay, since the stock of the purified ERAP1 enzyme depleted and there were issues with obtaining fresh protein samples (see Chapter 2.2.1). There were no interference or other issues with the IRAP-I characterisation using the colorimetric assay. Thus it was concluded that using only the colorimetric assay to characterise its analogues would still provide reliable results.

5 Inhibitor synthesis and analogue structure-activity relationship studies.

In the first round eleven sulphonamide analogues, containing variable sulphonamide groups were synthesised and characterised. In addition, triazole analogues and IRAP-I regioisomer were investigated (Figure 5-10). Both triazole analogues S09 and S10 were inactive, supporting the suggestion that the tetrazole group interacted with Zn. In difference to the tetrazole group, the triazole moiety could not be deprotonated. Thus, this group would not be able to form ionic interactions with the Zn cation in the ERAP1 active site. The regioisomer S11 exhibited reduced potency compared to the original compound. Hence, analogues with the 1,4-substituted aniline core were not further investigated. The lack of the potency of S8 suggested that a sulphonamide with a quite large hydrophobic group was important for the interactions with the protein. S5 showed comparable potency to the thiophene substituted sulphonamides.

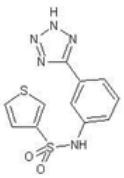
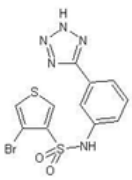
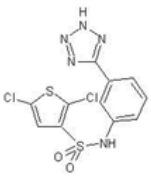
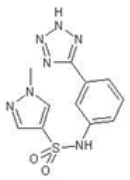
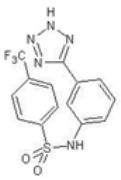
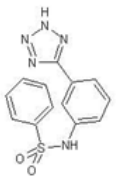
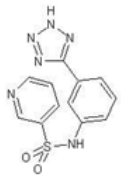
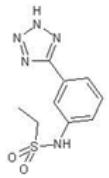
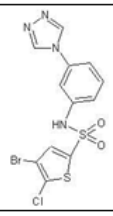
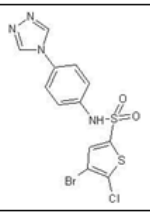
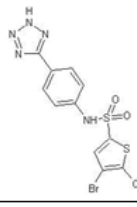
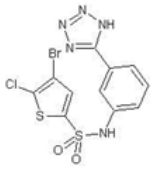
						
Fragment ID	**S01**	**S02**	**S03**	**S04**	**S05**	**S06**
Colorimetric IC_{50}, µM	**598 ± 970**	**48 ± 12**	**50 ± 52**	**ca. 1 mM**	**90 ± 20**	**396 ± 300**
						
Fragment ID	**S07**	**S08**	**S09**	**S10**	**S11**	**IRAP inhibitor**
Colorimetric IC_{50}, µM	**ca. 1 mM**	**Inactive**	**Inactive**	**Inactive**	**49 ± 51**	**9 ± 6**

Figure 5-10. Inhibitory properties of round 1 sulphonamide analogues in the ERAP1 colorimetric assay. S04 and S07 IC_{50} were not determined precisely, due to their solubility limit at approximately 1 mM

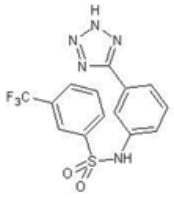
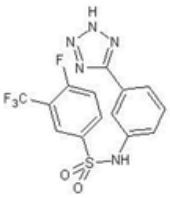
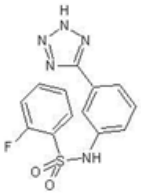
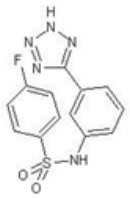
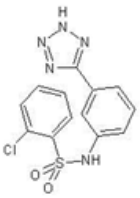
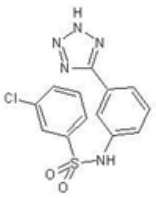
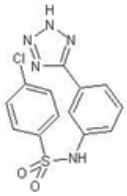
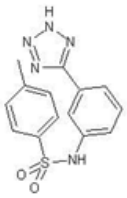
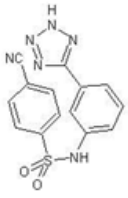
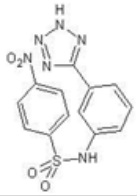
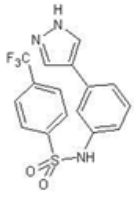
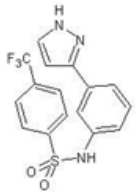
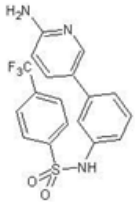
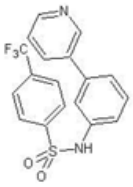
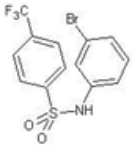
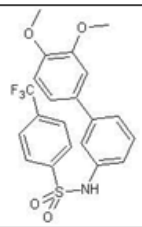
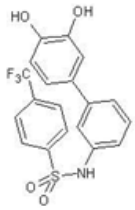
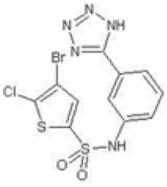
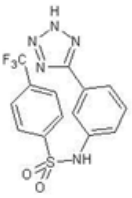
					
Fragment ID	**S12**	**S13**	**S14**	**S15**	**S16**
Colorimetric IC_{50}, µM	**192 ± 63**	**3 ± 5**	**318 ± 226**	**115 ± 41**	**157 ± 37µM**
					
Fragment ID	**S17**	**S18**	**S19**	**S20**	**S21**
Colorimetric IC_{50}, µM	**139 ± 142**	**880 ± 1820**	**358 ± 343**	**Inactive**	**68 ± 33**
					
Fragment ID	**S22**	**S23**	**S24**	**S25**	**S26**
Colorimetric IC_{50}, µM	**187 ± 91**	**40 ± 57**	**32 ± 34**	**92 ± 28**	**ca 100**
					
Fragment ID	**S27**	**S28**	**IRAP inhibitor**	**S05**	
Colorimetric IC_{50}, µM	**ca 40**	**10 ± 8**	**9 ± 6**	**90 ± 20**	

Figure 5-11. Inhibitory properties of round 2 sulphonamide analogues in the ERAP1 colorimetric assay. S26 and S27 IC_{50} were not determined precisely, due to their solubility limit at approximately 150 µM and 50 µM respectively

The diversity of available substituted phenyl sulphonyl chlorides – a starting material in the synthesis of sulphonamide analogues – was much larger than that of thiophenes. Hence, in the second round analogues with a set of substituted phenyl sulphonamides was synthesised and characterised. In addition, the effect of the tetrazole substitution with a set of different putative Zn-binding groups was investigated. In the second round seventeen IRAP inhibitor analogues were synthesised and characterised (Figure 5-11). The comparison of the potency of the chlorophenyl analogues (S16, S17 and S17) suggested that the substitution in *para* position gave the weakest compound. However, the

opposite trend was observed for the fluoro-substituted analogues (S14 and S15). The *para* analogue was more potent. The *para*-methylation of the phenyl ring (S19) did not lead to any significant change in potency (compared with S6). S13 turned out to be a very potent analogue, even slightly more potent than the original inhibitor. This was especially unexpected considering that S12 and S15 exhibited mediocre potency. Hence, the strong interaction of S13 with ERAP1 was not just due to the cumulative *meta*-trifluoromethyl and *para*-fluoro substituents' effects, but these two moieties conveyed some cooperative effect on binding. Comparing S22 and S23 pyrazole analogues of the original tetrazole, it was possible to suggest that the pyrazole was important for Zn-binding through ionic interactions. It could be that the delocalisation of the aromatic electrons in 3-substituted pyrazole did not extend to the adjacent phenyl ring as it did in the 4-substituted one. Thus, S23 higher potency compared to S22 could be due to a denser negative charge on the pyrazole upon deprotonation. Both 2-amino-pyridinyl (S24) and catechol (S28) analogues had improved potency compared to the original tetrazole analogue S5. 2-amino-pyridine (3D_22) and catechol (Zn_2) based hits had been previously identified by the screening campaign of the fragment libraries.

Based on the results from the two rounds of analogue screening it was concluded that the combination of the sulphonamide from S13 with the catechol from S28 was expected to yield the IRAP analogue with superior potency, hopefully with IC_{50} values in the sub µM range. Thus, the third round of analogue synthesis and characterisation was performed (Figure 5-12). Unfortunately, the substitution of the tetrazole with catechol did not lead to the expected increase in potency, as in the case of S28. S31 still had single digit micromolar potency, as the best analogues from the previous two rounds of synthesis. However, as a result of the analogue screening campaign a novel ERAP1 inhibitor chemical scaffold was developed.

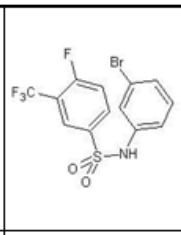
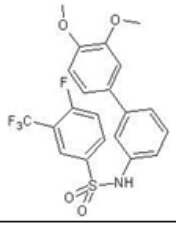
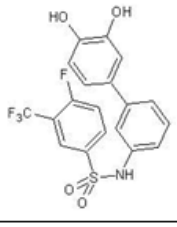
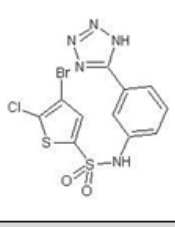
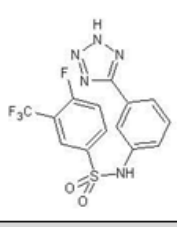
					
Fragment ID	**S29**	**S30**	**S31**	**IRAP inhibitor**	**S13**
Colorimetric IC_{50}, µM	**273 ± 950**	**9 ± 7**	**7 ± 3**	**9 ± 6**	**3 ± 5**

Figure 5-12. Inhibitory properties of round 3 sulphonamide analogues in the ERAP1 colorimetric assay.

5.4 Conclusion

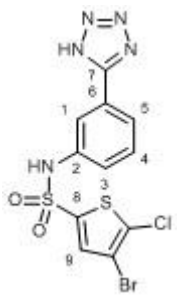
DG-051-1 LTA4H inhibitor and IRAP inhibitor have been successfully synthesised. In addition, 31 sulphonamide analogues were obtained for the SAR campaign. The screening of these analogues has provided insights about the structure-activity relationship of this scaffold. Four analogues with ≤ 10 µM potency have been identified – S13, S28, S30 and S31. It would be very interesting to find the explanation for the unexpected high potency of S13 once the structural information is available.

5.5 Experimental

All reactions were carried out under a nitrogen atmosphere. Water was deionised by an Elga DV 25 system. All other solvents and reagents were used as supplied (analytical or HPLC grade) without prior purification. Organic layers were dried over $MgSO_4$. Thin layer chromatography was performed on aluminium plates coated with 60 F_{264} silica gel. Plates were visualised using UV (254 nm). Flash column chromatography was performed on Biotage Isolera One flash column chromatography platform. Purification by HPLC was performed using a Waters SFO with 515 HPLC pump and Waters Binary Gradient 2545 Device (5%-95% solvent A (18% water, 80% acetonitrile, 2% 0.5 M ammonium acetate pH 6.0) in solvent B (93% water, 5% acetonitrile, 2% ammonium acetate pH 6.0)). Product was detected using a SQ Detector 2 and collected using a Waters Sample Manager 2767. Melting points were recorded on a Stuart SMP40 apparatus and are uncorrected. IR spectra were recorded on a Nicolet iS5 with an iD7 ATR module, neat. Selected characteristic peaks are reported in cm^{-1}.

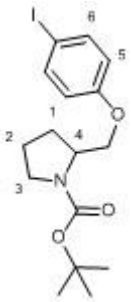
5 Inhibitor synthesis and analogue structure-activity relationship studies.

NMR spectra were recorded on a Bruker Avance spectrometer in deuterated d6-DMSO. The filed was locked by external referencing to the relevant deuteron resonance. Chemical shifts (δ) are reported in ppm and coupling constants (J) in Hz to the nearest 0.1 Hz. Spectra were recorded at room temperature. When peak multiplicities are reported, the following abbreviations are used: s – singlet, d – doublet, t – triplet, m – multiplet, b- broadened. m/z values are reported in Daltons. Compound purity was assessed by LCMS using UV absorbance at 254 nm or appropriate ESI TIC (SQ Detector 2). The detection method was chosen to obtain the strongest signal. Analytical LCMS was performed on the following system: Kinetex 5µ EVO C18 100A 100 x 2.0 mm column using a linear gradient of solvent A (93% water, 5% acetonitrile, 2% of 0.5 M ammonium acetate pH 6.0), eluting at a flow rate of 2 ml/minute: 5% B over 0.35 minute, 5% B to 95% B over 1 minute, 95% B over 0.75 minute, 95% to 5% B over 0.1 minute and 5% B over 0.8 minute. LRMS were recorded by LCMS on a Waters SQ Detector 2; data acquisition and processing was performed using Waters FractionLynx software. HRMS was performed on an Agilent 6530 Accurate Mass Q-TOF; data acquisition and processing were performed using Agilent MassHunter Workstation software.



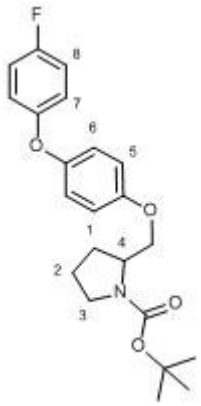
(N-(3-(1H-tetrazol-5-yl)phenyl)-4-bromo-5-chlorothiophene-2-sulfonamide), **IRAP-I**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.18 mmol, 30 mg) 4-bromo-5-chlorothiophene-2-sulfonyl chloride (1.2, 0.22 mmol, 65 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 0.677 mmol, 55 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in DMSO and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow waxy solid (34 mg, 44%). ^{1}H NMR (400 MHz) δ 7.91 (s, 1H, C(9)H), 7.82 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.70 (s, 1H, C(1)H), 7.55 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.35 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 156.8 (C7), 138.7 (C2), 138.1 (C8), 133.9 (C9), 132.7 (C6), 130.9 (C4), 127.7 (CCl), 123.7 (C3), 122.8 (C5), 119.3 (C1), 111.9 (CBr); m/z

(ESI^-) 418.0, 70%; 420.0, 100%; 422.0, 30% ($[M-H]^-$); HRMS (ESI^-) $C_{11}H_6BrClN_5O_2S_2^-$, ($[M-H]^-$) requires 417.8841, found 417.8863.

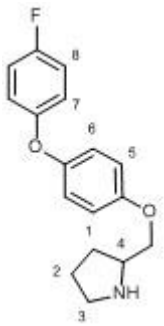


Tert-butyl 2-((4-iodophenoxy)methyl)pyrrolidine-1-carboxylate, **DG-051-B**. tert-butyl 2-(hydroxymethyl)pyrrolidine-1-carboxylate (1, 1.5 mmol, 300 mg) was dissolved in 0.25 ml pyridine, 4-methylbenzenesulfonyl chloride (1.1, 1.64 mmol, 315 mg) was dissolved in 0.25 ml pyridine, both were cooled on ice, then latter added dropwise to the former. The mixture was stirred for 90 minutes on ice, then at R.T. overnight. The solvent was evaporated and the crude residue dissolved in 2 ml EtOAc, washed with 1 ml 0.5M HCl, 1 ml sat. $NaHCO_3$ and 1 ml of brine. The organic phase was dried and concentrated to give the crude product (tert-butyl 2-((4-(4-fluorophenoxy)phenoxy)methyl)pyrrolidine-1-carboxylate), which was used for the next step without further purification. 4-iodophenol (1.1, 1.27 mmol, 280 mg) and KO^tBu (1.1, 1.27 mmol, 145 mg) were dissolved in 1 ml DMF and stirred at RT for 30 minutes under nitrogen atmosphere. Then the crude (tert-butyl 2-((4-(4-fluorophenoxy)phenoxy)methyl) pyrrolidine-1-carboxylate) was dissolved in 1 DMF and added to the reaction mixture. The mixture was heated to 70 ^oC and stirred overnight. Then the solvent was removed and the residue was partitioned between 10 ml cyclohexane and 5 ml water. The organic phase was washed with 0.2 M aq. NaOH (2 x 3 ml), water (5 ml) and 5% aq. LiCl (5ml), them dried and concentrated. The product was purified by flash-chromatography (silica gel, 10 g, 0-20% EtOAc in hexane). The product containing fractions were pulled together and concentrated to give the desired product as an off-brown waxy solid (160 mg, 34%). ^1H NMR (400 MHz) δ 7.58 (d, ^3J = 8.5 Hz, 2H, C(5)H), 6.82 (d, ^3J = 8.5 Hz, 2H, C(6)H), 4.01 (m, 2H, OCH_2), 3.88 (m, 1H, C(4)H), 3.22 (b, 2H, C(3)H_2), 1.90 (m, 4H, C(1)H_2, C(2)H_2), 1.40 (s, 9H, CH_3); ^{13}C NMR (101 MHz) δ 158.9 (C5*C*O, COO), 138.4 (C6), 117.8 (C5), 83.6 (CI), 79.1 (C4*C*O), 68.5 (*C*CH_3), 55.8 (C4), 46.8 (C3), 28.6 (CH_3), 26.8 (C1, C2); m/z (ESI^+) 404.1 ($[M+H]^+$, 100%).

5 Inhibitor synthesis and analogue structure-activity relationship studies.

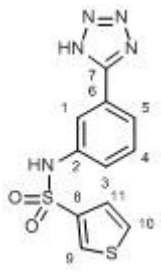


Tert-butyl 2-((4-(4-fluorophenoxy)phenoxy)methyl)pyrrolidine-1-carboxylate, **DG-051-C**. Copper(I) iodide (0.25, 0.031 mmol, 6 mg), dimethylglycine (0.5, 0.062 mmol, 7 mg), tert-butyl 2-((4-iodophenoxy)methyl)pyrrolidine-1-carboxylate (1, 0.124 mmol, 50 mg), 4-fluorophenol (3, 0.372 mmol, 42 mg) and Cs_2CO_3 (4, 0.5 mmol, 160 mg) were dissolved in 2 ml acetonitrile under N_2 and the reaction mixture was heated to 160^{o}C in a microwave for 30 minutes. Then the solvent was evaporated and the residue was partitioned between cyclohexane (10 ml) and water (5 ml), washed with 0.2 M NaOH (5 x 3 ml) and water (3 ml). The dried organic phase was concentrated and purified by chromatography (silica gel, 10 g, 0-15% EtOAc in hexane). Substance containing fractions were pulled together and concentrated, yielding the desired product as a dark brown waxy solid (20 mg, 42%). ^{1}H NMR (400 MHz) δ 7.18 (m, 2H, C(8)H), 6.96 (m, 6H, C(5)H, C(6)H, C(7)H), 4.03 (m, 2H, C(4)CH_2), 3.87 (b, 1H, C(4)H), 3.27 (b, 2H, C(3)H_2), 1.90 (m, 4H, C(1)H_2, C(2)H_2), 1.40 (s, 9H, CH_3); ^{13}C NMR (101 MHz) δ 158.2 (CF), 155.2 (COO, C5*C*O), 154.4 (C7*C*O), 150.5 (C6*C*O), 120.7 (C6), 119.6 (C5), 116.9 (C7), 116.7 (C8), 79.1 (C4*C*O), 68.8 (*C*CH_3), 56.1 (C4), 47.0 (C3), 28.6 (CH_3), 28.1 (C1), 23.7 (C2); m/z (ESI^{+}) 388.1 ([M+H]$^{+}$, 100%).

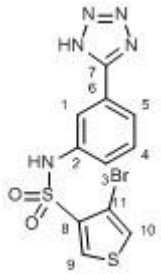


(2-((4-(4-fluorophenoxy)phenoxy)methyl)pyrrolidine), **DG-051-1**. Tert-butyl 2-((4-(4-fluorophenoxy)phenoxy)methyl)pyrrolidine-1-carboxylate (0.052 mmol, 20 mg) was dissolved in 1 ml 4N HCl in dioxane and stirred under nitrogen at R.T. for 2 h. Then the mixture was washed with 2 ml cyclohexane. The organic layer was washed with 2 ml 1M HCl. The aqueous layers were combined and neutralised with sat. Na_2CO_3, then extracted with cyclohexane (2 x 5 ml). The organic phase was dried and concentrated, yielding the desired product as a colourless waxy solid (4 mg, 27%). ^{1}H NMR (400 MHz) δ 7.18 (m, 2H, C(8)H), 6.96 (m, 6H, C(5)H, C(6)H, C(7)H), 3.78 (m, 2H, C(4)CH_2), 3.18 b, 1H, C(4)H), 2.82 (b, 2H, C(3)H_2), 1.65 (m, 4H, C(1)H_2, C(2)H_2); ^{13}C NMR (101 MHz) δ 158.4 (CF), 155.5 (C5*C*O), 154.5 (C7*C*O), 150.3 (C6*C*O),), 120.7 (C6), 119.6 (C5), 116.9 (C7), 116.2 (C8), 72.6 (C4*C*O), 57.0 (C4), 46.3 (C3), 28.6 (C1), 25.5 (C2);

m/z (ESI^+) 288.2 100% ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{17}H_{19}FNO_2{}^+$, ($[M+H]^+$) requires 288.1394, found 288.1396.

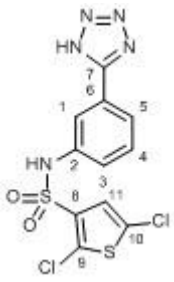


N-(3-(1H-tetrazol-5-yl)phenyl)thiophene-3-sulfonamide, **S1** . To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.456 mmol, 73.5 mg) and thiophene-3-sulfonyl chloride (1.2 ,0.548 mmol , 100 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 1.688 mmol, 0.14 ml) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in DMSO and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow waxy solid (40 mg, 28.5%). ^{1}H NMR (400 MHz) δ 8.20 (dd, ^4J = 2.8, 1.1 Hz, 1H, C(9)H), 7.85 (ddd, ^4J = 1.7, 1.1 Hz, 1H, C(1)H), 7.71 (m, 1H, C(10)H), 7.69 (m, 1H, C(5)H), 7.42 (dd, ^3J = 7.8 Hz, 1H, C(4)H), 7.28 (dd, , ^3J = 5.3 Hz, , ^4J = 1.1 Hz, 1H, C(11)H), 7.25 (dd, , ^3J = 7.8 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 156.6 (C7), 139.0 (C2), 138.5 (C8), 131.7 (C9), 130.0 (C4), 129.4 (C10), 127.4 (C6), 124.9 (C11), 122.3 (C5), 121.4 (C3), 118.0 (C1); m/z (ESI^-) 306.0 ($[M-H]^-$, 100%); HRMS (ESI^-) $C_{11}H_8N_5O_2S_2^-$, ($[M-H]^-$) requires 306.0125, found 306.0133.

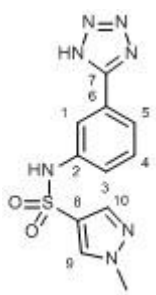


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-bromothiophene-3-sulfonamide), **S2**. To a solution of 3-(1H-tetrazol-5-yl)aniline ()1, 0.372 mmol, 60 mg) and 4-bromothiophene-3-sulfonyl chloride (1.2, 0.447 mmol, 117 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 1.377 mmol,0. 111 ml) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in DMSO and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a pink crystalline solid (120 mg, 83 %). Mp 196-197 ^{o}C; ν_{max} (FTIR) cm^{-1} 3253 (aromatic CH), 1563 (tetrazole), 1231, 1147 (O=S=O); ^{1}H NMR (400 MHz) δ 8.44 (d, ^4J = 3.5 Hz, 1H, C(9)H), 7.84 (m, 2H, C(1)H, C(10)H), 7.65 (dm, ^3J = 7.9 Hz, 1H, C(5)H), 7.37 (dd, ^3J = 7.9 Hz, 1H, C(4)H), 7.20 (dm, ^3J

= 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 157.7 (C7), 138.2 (C2), 137.4 (C8), 136.5 (C9), 130.0 (C4), 129.2 (C10), 122.4 (C5), 120.4 (C3), 117.7 (C1), 106.5 (C11); m/z (ESI$^-$) 383.9, 100%; 385.9, 90% ([M-H]$^-$); HRMS (ESI$^-$) $C_{11}H_7N_5O_2S_2Br^-$, ([M-H]$^-$) requires 383.9231, found 383.9248.

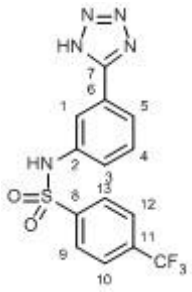


(N-(3-(1H-tetrazol-5-yl)phenyl)-2,5-dichlorothiophene-3-sulfonamide), **S3**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.248 mmol, 40 mg) and 2,5-dichlorothiophene-3-sulfonyl chloride (1.2, 0.298 mmol, 75 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.745 mmol, 60 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in DMSO and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow waxy solid (29 mg, 31%). ^{1}H NMR (400 MHz) δ 7.79 (m, 1H, C(1)H), 7.71 (dm, ^3J = 7.9 Hz, 1H, C(5)H), 7.31 (dd, ^3J = 7.9 Hz, 1H, C(4)H), 7.30 (s, 1H, C(11)H), 7.10 (dm, ^3J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 159.6 (C7), 137.0 (C2), 135.9 (C8), 132.7 (C10), 129.5 (C9), 129.4 (C4), 126.8 (C11), 126.4 (C6), 122.3 (C5), 119.4 (C30, 118.1 (C1); m/z (ESI$^-$) 373.9, 100%; 375.9, 70%; 377.9, 30% ([M-H]$^-$); HRMS (ESI$^-$) $C_{11}H_6N_5O_2S_2Cl_2^-$, ([M-H]$^-$) requires 373.9430, found 373.9368.

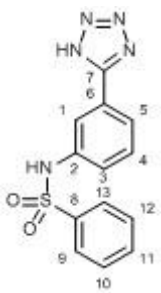


(N-(3-(1H-tetrazol-5-yl)phenyl)-1-methyl-1H-pyrazole-4-sulfonamide), **S4**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1.000, 0.372 mmol, 60 mg) and 1-methyl-1H-pyrazole-4-sulfonyl chloride (1.2, 0.447 mmol, 81 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 1.377 mmol, 0.111 ml) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in DMSO and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated,

yielding the desired product as a colourless waxy solid (107 mg, 94 %). ^{1}H NMR (400 MHz) δ 8.21 (s, 1H, C(9)H), 7.81 (s, 1H, C(1)H), 7.69 (s, 1H, C(10)H), 7.68 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.35 (dd, , ^{3}J = 7.9 Hz, 1H, C(4)H), 7.16 (dm, , ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 158.5 (C7), 138.4 (C2), 137.9 (C10), 132.6 (C9), 130.3 (C8), 129.5 (C4), 121.8 (C5), 121.1 (C6), 119.8 (C3), 117.6 (C1), 39.5 (CH3); m/z (ESI$^-$) 304.06, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{11}H_{10}N_7O_2S^-$, ([M-H]$^-$) requires 304.0622, found 304.0635.

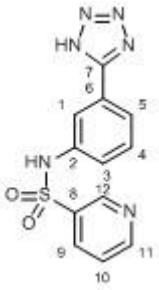


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-(trifluoromethyl)benzenesulfonamide), **S5**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-(trifluoromethyl)benzenesulfonyl chloride (1.1, 0.341 mmol, 83 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.931 mmol, 75 µ)*via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product an off-yellow crystalline solid (35 mg, 30%). Mp 199-200 ^{o}C; ν_{max} (FTIR) cm^{-1} 3245 (aromatic CH), 1565 (tetrazole), 1319, 1160 (O=S=O), 1060 (C-F); ^{1}H NMR (400 MHz) δ 8.00 (d, ^{3}J = 8.5 Hz, 2H, C(9)H, C(13)H), 7.96 (d, ^{3}J = 8.5 Hz,2H, C(10)H, C(12)H), 7.85 (m, 1H, C(1)H), 7.71 (dm, ^{3}J = 8.0 Hz,1H, C(5)H), 7.45 (dd, ^{3}J = 8.0 Hz, 1H, C(4)H), 7.23 (dm, ^{3}J = 8.0 Hz, 1H, C(3)H);); ^{13}C NMR (101 MHz) δ 156.2 (C7), 143.2 (C2), 138.0 (C8), 132.8 (C11), 130.3 (C4), 127.7 (C9, C13), 126.9 (C6), 126.7 (C10, C12), 122.8 (C5), 122.5 (CF_3), 122.0 (C3),118.4 (C1); m/z (ESI$^-$) 368.2, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{14}H_9F_3N_5O_2S^-$, ([M-H]$^-$) requires 368.0435, found 368.0455



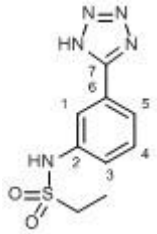
(N-(3-(1H-tetrazol-5-yl)phenyl)benzenesulfonamide), **S6**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.28 mmol, 40 mg) and benzenesulfonyl chloride (1.1, 0.273 mmol, 48 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.931 mmol, 75 µ) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated

and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a white crystalline solid (30 mg, 40%). Mp 177-178 ^{o}C; ν_{max} (FTIR) cm^{-1} 3248 (aromatic CH), 1477 (tetrazole), 1354, 1158 (O=S=O); ^{1}H NMR (400 MHz) δ 7.79 (s, 1H, C(1)H), 7.78 (dm, ^{3}J = 8.3 Hz, 2H, C(9)H, C(13)H), 7.62 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.58 (m, 1H, C(11)H), 7.54 (m, 2H, C(10)H, C(12)H), 7.33 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.15 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 157.8 (C7), 139.4 (C2), 138.2 (C8), 132.9 (C11), 129.7 (C4), 129.5 (C6), 129.3 (C10, C12), 126.6 (C9, C13), 122.0 (C5), 120.4 (C3), 117.9 (C1);); m/z (ESI^{-}) 300.0, 100% ([M-H]$^{-}$); HRMS (ESI^{-}) $C_{13}H_{10}N_5O_2S^{-}$, ([M-H]$^{-}$) requires 300.0561, found 300.0579.

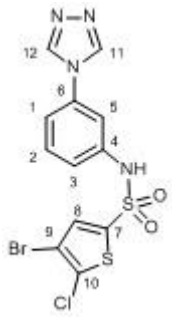


(N-(3-(1H-tetrazol-5-yl)phenyl)pyridine-3-sulfonamide), **S7**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.469 mmol, 76 mg) and pyridine-3-sulfonyl chloride (1.2, 0.563 mmol, 100 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7 eq., 1.736 mmol, 0.140 ml) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product an off-yellow waxy solid (60 mg, 42%). ^{1}H NMR (400 MHz) δ 8.89 (dd, , ^{4}J = 2.4, 1,6 Hz, 1H, C(12)H), 8.75 (dd, , ^{3}J = 4.9 Hz, , ^{4}J = 1.6 Hz, 1H, C(11)H), 8.11 (dm, , ^{3}J = 8.1 Hz, 1H, C(9)H), 7.75 (m, 1H, C(1)H), 7.67 (dm, , ^{3}J = 7.9 Hz, 1H, C(5)H), 7.59 (ddm, , ^{3}J = 8.1, 4.9 Hz, 1H, C(10)H), 7.27 (dd, , ^{3}J = 7.9 Hz, 1H, C(4)H), 7.06 (dm, , ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 159.8 (C7), 153.4 (C(11), 147.0 (C12), 137.2 (C2), 135.8 (C8), 134.7 (C9), 133.2 (C6), 129.3 (C4), 124.3 (C10), 122.2 (C5), 119.4 (C3), 118.1 (C1); m/z (ESI^{-}) 301.05, 100% ([M-H]$^{-}$); HRMS (ESI^{-}) $C_{12}H_9N_6O_2S^{-}$, ([M-H]$^{-}$) requires 301.0513, found 301.0524.

5 Inhibitor synthesis and analogue structure-activity relationship studies.

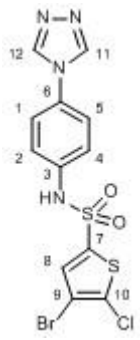


(N-(3-(1H-tetrazol-5-yl)phenyl)ethanesulfonamide), **S8**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.25 mmol, 40 mg) and ethanesulfonyl chloride (1.2, 0.3 mmol, 40 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 0.918 mmol, 75 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a colourless waxy solid (22 mg, 35%). ^{1}H NMR (400 MHz) δ 7.93 (m, 1H, C(1)H), 7.71 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.55 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.40 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H), 3.17 (q, ^{3}J = 7.9, 6.7 Hz, 2H, CH_2), 1.22 (t, ^{3}J = 6.7 Hz, 3H, CH_3); ^{13}C NMR (101 MHz) δ 155.5 (C7), 139.5 (C2), 130.5 (C4), 125.4 (C6), 122.0 (C5), 121.8 (C3), 117.3 (C1), 45.4 (CH_2), 8.0 (CH_3); m/z (ESI$^-$) 252.1, 100%([M-H]$^-$); HRMS (ESI$^-$) $C_9H_{10}N_5O_2S^-$, ([M-H]$^-$) requires 252.0561, found 252.0575.

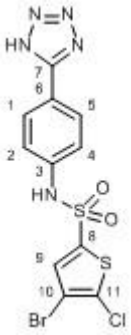


(N-(3-(4H-1,2,4-triazol-4-yl)phenyl)-4-bromo-5-chlorothiophene-2-sulfonamide), **S9**. To a solution of 3-(4H-1,2,4-triazol-4-yl)aniline (1, 0.25 mmol, 40 mg) and 4-bromo-5-chlorothiophene-2-sulfonyl chloride (1.1, 0.275 mmol, 81 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.75 mmol,61 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a colourless waxy solid (19 mg, 18%). ^{1}H NMR (400 MHz) δ 9.06 (s, 2H, C(11)H, C(12)H), 7.78 (s, 1H, C(8)H), 7.51 (dd, ^{3}J = 8.0 Hz, 1H, C(2)H), 7.46 (dm, ^{3}J = 8.0 Hz, 1H, C(1)H), 7.41 (m, 1H, C(5)H), 7.20 (dm, ^{3}J = 8.0 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 141.5 (C11, C12), 138.4 (C4), 138.1 (C7), 134.7 (C6), 133.9 (C8), 132.4 (C10), 131.0 (C2), 119.8 (C3), 118.0 (C1), 113.6 (C5), 111.6 (C9); m/z (ESI$^-$) 416.9, 75%; 418.9, 100%; 420.9, 30% ([M-H]$^-$); HRMS (ESI$^-$) $C_{12}H_7BrClN_4O_2S_2^-$, ([M-H]$^-$) requires 416.8888, found 416.8910.

5 Inhibitor synthesis and analogue structure-activity relationship studies.

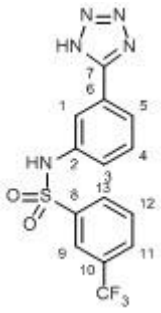


(N-(4-(4H-1,2,4-triazol-4-yl)phenyl)-4-bromo-5-chlorothiophene-2-sulfonamide), **S10**. To a solution of 4-(4H-1,2,4-triazol-4-yl)aniline (1, 0.312 mmol, 50 mg) and 4-bromo-5-chlorothiophene-2-sulfonyl chloride (1.2, 0.375 mmol, 111 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3.7, 1.155mmol, 93 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow waxy solid (20 mg, 15%). ^{1}H NMR (400 MHz) δ 9.04 (s, 2H, C(11)H, C(12)H), 7.71 (s, 1H, C(8)H), 7.65 (d, ^{3}J = 8.5 Hz, 2H, C(2)H, C(4)H), 7.30 (d, ^{3}J = 8.5 Hz, C(1)H, C(5)H); ^{13}C NMR (101 MHz) δ 141.4 (C11, C12), 138.3 (C3), 136.5 (C7), 133.6 (C8), 132.2 (C10), 130.7 (C6), 122.5 (C2, C4), 121.7 (C1, C5), 111.6 (C9); m/z (ESI$^-$) 416.9, 75%; 418.9, 100%; 420.9, 30% ([M-H]$^-$); HRMS (ESI$^-$) $C_{12}H_7BrClN_4O_2S_2^-$, ([M-H]$^-$) requires 416.8888, found 416.8906.

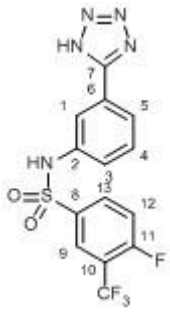


(N-(4-(1H-tetrazol-5-yl)phenyl)-4-bromo-5-chlorothiophene-2-sulfonamide), **S11**. To a solution of 4-(1H-tetrazol-5-yl)aniline (1, 0.217 mmol, 35 mg) and 4-bromo-5-chlorothiophene-2-sulfonyl chloride (1.1, 0.239 mmol, 71 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.651 mmol, 53 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow waxy solid (15 mg, 16%). ^{1}H NMR (400 MHz) δ 7.94 (d, ^{3}J = 8.3 Hz, 2H, C(1)H, C(5)H), 7.42 (s, 1H, C(9)H), 7.06 (d, ^{3}J = 8.3 Hz, 2H, C(2)H, C(4)H); ^{13}C NMR (101 MHz) δ 159.1 (C7), 130.9 (C9), 126.7 (C1, C5), 125.8 (C6), 121.1 (C2, C4), 112.7 (C11), 110.6 (C10);); m/z (ESI$^-$) 418.0, 70%; 420.0, 100%; 422.0, 30% ([M-H]$^-$); HRMS (ESI$^-$) $C_{11}H_6BrClN_5O_2S_2^-$, ([M-H]$^-$) requires 417.8841, found 417.8863

5 Inhibitor synthesis and analogue structure-activity relationship studies.

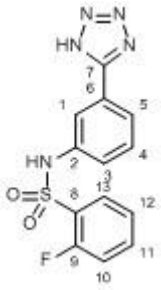


(N-(3-(1H-tetrazol-5-yl)phenyl)-3-(trifluoromethyl)benzenesulfonamide), **S12**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 3-(trifluoromethyl) benzenesulfonyl chloride (1.1, 0.34 mmol, 80 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow crystalline solid (51 mg, 45%). Mp 186-187 ^{o}C; ν_{max} (FTIR) cm^{-1} 3262 (aromatic CH), 1480 (tetrazole), 1322, 1102 (O=S=O) 1071 (C-F); ^{1}H NMR (400 MHz) δ 8.05 (s, 1H, C(9)H), 8.03 (m, 1H, C(11)H), 8.00 (d, ^{3}J = 8.0 Hz, 1H, C(13)H), 7.82 (s, 1H, C(1)H), 7.80 (dd, ^{3}J = 8.0 Hz, 1H, C(12)H), 7.72 (dm, ^{3}J = 7.7 Hz, 1H, C(5)H), 7.41 (dd, ^{3}J = 7.7 Hz, 1H, C(4)H), 7.19 (dm, ^{3}J = 7.7 Hz, 1H, C(4)H); ^{13}C NMR (101 MHz) δ 156.9 (C7), 140.4 (C2), 137.8 (C8), 131.0 (C12), 130.7 (C9), 130.1 (C4), 130.0 (C10), 128.1 (C13), 124.6 (C6), 123.2 (C11), 122.8 (C5), 122.3 (CF_3), 121.7 (C3), 118.5 (C1); m/z (ESI$^-$) 368.2 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{14}H_9F_3N_5O_2S^-$, ([M-H]$^-$) requires 368.0435, found 368.0451.

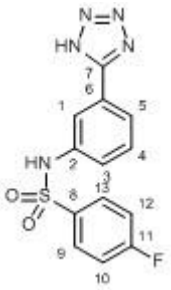


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-fluoro-3-(trifluoromethyl)benzenesulfonamide), **S13**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-fluoro-3-(trifluoromethyl)benzenesulfonyl chloride (1, 0.31 mmol, 81 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an orange crystalline solid (71 mg, 60%). Mp 118-119 ^{o}C; ν_{max} (FTIR) cm^{-1} 3074 (aromatic CH), 1420 (tetrazole), 1316, 1125 (O=S=O) 1055 (C-F); ^{1}H NMR (400 MHz) δ 8.10 (s, 1H, C(9)H), 8.09 (m, 1H, C(12)H), 7.81 (m, 1H, C(1)H), 7.74 (d, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.72 (m, 1H, C(13)H), 7.39 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.17 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101

MHz) δ 160.1 (C11), 157.5 (C7), 137.6 (C2), 136.3 (C8), 134.1 (C9), 130.0 (C4), 129.3 (C10), 126.2 (C13), 122.9 (C5), 121.5 (C3), 120.3 (CF_3), 119.0 (C12), 118.7 (C1); m/z (ESI^-) 386.1, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{14}H_8F_4N_5O_2S^-$, ($[M-H]^-$) requires 386.0340, found 386.0355.

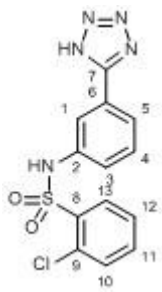


(N-(3-(1H-tetrazol-5-yl)phenyl)-2-fluorobenzenesulfonamide), **S14**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 2-fluorobenzenesulfonyl chloride (1.3, 0.411 mmol, 80 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-pink crystalline solid (31 mg, 31%). Mp 112-113 ^oC; ν_{max} (FTIR) cm^{-1} 3031 (aromatic CH), 1473 (tetrazole), 1328, 1152 (O=S=O) 1074 (C-F); ^1H NMR (400 MHz) δ 7.86 (dm, ^3J = 8.5 Hz, 1H, C(13)H), 7.82 (s, 1H, C(1)H), 7.67 (m, 2H, C(5)H, C(10)H), 7.40 (dd, ^3J = 8.5 Hz, 1H, C(12)H), 7.36 (m, 2H, C(4)H, C(11)H), 7,16 (dm, ^3J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 158.1 (C9), 157.9 (C7), 137.7 (C2), 136.0 (C8), 130.4 (C13), 129.7 (C4), 127.0 (C6), 125.0 (C12), 122.1 (C5), 120.1 (C3), 117.6 (C1), 117.4 (C11), 117.2 (C10); m/z (ESI^-) 318.0, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{13}H_9FN_5O_2S^-$, ($[M-H]^-$) requires 318.0466, found 318.0484.

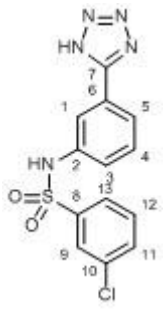


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-fluorobenzenesulfonamide), **S15**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-fluorobenzenesulfonyl chloride (1.1, 0.341, 67 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (3, 0.931 mmol, 75 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a colourless waxy solid (72 mg, 73%). ^1H NMR (400 MHz) δ 7.81 (m, 3H, C(1)H, C(9)H, C(13)H), 7.67

(dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.38 (m, 3H, C(4)H, C(10)H, C(11)H), 7.15 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 164.2 (C11), 163.1 (C7), 138.0 (C2), 135.8 (C8), 129.8 (C4), 129.7 (C9, C13), 123.9 (C6), 122.3 (C5), 120.1 (C3), 118.1 (C1), 116.5 (C10, C12); m/z (ESI$^-$) 318.1 100%([M-H]$^-$); HRMS (ESI$^-$) $C_{13}H_9FN_5O_2S^-$, ([M-H]$^-$) requires 318.0466, found 318.0489.

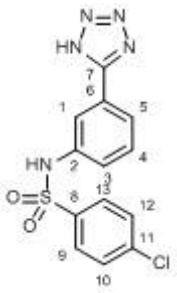


(N-(3-(1H-tetrazol-5-yl)phenyl)-2-chlorobenzenesulfonamide), **S16**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 2-chlorobenzenesulfonyl chloride (1.2, 0.38 mmol, 80 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-pink crystalline solid 33 mg, 32%). Mp 229-230 ^{o}C; ν_{max} (FTIR) cm^{-1} 3235 (aromatic CH), 1470 (tetrazole), 1368, 1167 (O=S=O);; ^{1}H NMR (400 MHz) δ 8.10 (d, ^{3}J = 7.8 Hz, 1H, C(13)H), 7.86 (s, 1H, C(1)H), 7.61 (m, 3H, C(5)H, C(10)H, C(12)H), 7.53 (ddm, ^{3}J = 7.5, 6.8 Hz, 1H, C(11)H), 7.45 (dd, ^{3}J = 8.0 Hz, 1H, C(4)H), 7.29 (dm, ^{3}J = 8.0 Hz,1H, C(3)H); ^{13}C NMR (101 MHz) δ 155.5 (C7), 138.0 (C2), 136.2 (C8), 134.9 (C10), 131.9 (C12), 121.6 (C13), 130.7 (C9), 130.4 (C4), 127.8 (C11), 125.5 (C6), 122.2 (C5), 121.3 (C3), 117.2 (C1); m/z (ESI$^-$) 334.0, 100%; 336.0, 35% ([M-H]$^-$); HRMS (ESI$^-$) $C_{13}H_9ClN_5O_2S^-$, ([M-H]$^-$) requires 334.0171, found 334.0185.



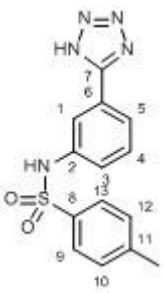
(N-(3-(1H-tetrazol-5-yl)phenyl)-3-chlorobenzenesulfonamide), **S17**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 3-chlorobenzenesulfonyl chloride (1.2, 0.38 mmol, 80 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an

off-orange crystalline solid (49 mg, 47%). Mp 185-186 °C; ν_{max} (FTIR) cm^{-1} 3063 (aromatic CH), 1408 (tetrazole), 1329, 1153 (O=S=O); ^{1}H NMR (400 MHz) δ 7.83 (s, 1H, C(1)H), 7.80 (s, 1H, C(9)H), 7.70 (m, 3H, C(5)H, C(11)H, C(13)H), 7.58 (dd, ^{3}J = 7.8 Hz, 1H, C(12)H), 7.42 (dd, ^{3}J = 8.0 Hz, 1H, C(4)H), 7.21 (dm, ^{3}J = 8.0 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 157.0 (C7), 141.1 (C2), 138.0 (C8), 133.9 (C10), 133.1 (C11), 131.4 (C12), 130.1 (C4), 128.0 (C6), 126.2 (C9), 125.4 (C13), 122.7 (C5), 121.5 (C3), 118.3 (C1); m/z (ESI^-) 334.0, 100%; 336.0, 35% ($[M-H]^-$); HRMS (ESI^-) $C_{13}H_9ClN_5O_2S^-$, ($[M-H]^-$) requires 334.0171, found 334.0185.

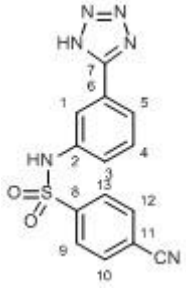


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-chlorobenzenesulfonamide), **S18**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-chlorobenzenesulfonyl chloride (1.2, 0.38 mmol, 80 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-pink crystalline solid (45 mg, 35%). Mp 139-140 °C; ν_{max} (FTIR) cm^{-1} 3261 (aromatic CH), 1416 (tetrazole), 1361, 1159 (O=S=O); ^{1}H NMR (400 MHz) δ 7.83 (s, 1H, C(1)H), 7.79 (d, ^{3}J = 8.6 Hz, 2H, C(9)H, C(13)H), 7.69 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.63 (d, ^{3}J = 8.6 Hz, 2H, C(10)H, C(11)H), 7.42 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.21 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 156.5 (C7), 138.2 (C11), 138.1 (C2), 137.9 (C8), 130.2 (C4), 129.5 (C10, C12), 128.6 (C9, C13), 127.3 (C6), 122.6 (C5), 121.7 (C3), 118.2 (C1); m/z (ESI^-) 334.0, 100%; 336.0, 35% ($[M-H]^-$); HRMS (ESI^-) $C_{13}H_9ClN_5O_2S^-$, ($[M-H]^-$) requires 334.0171, found 334.0185.

5 Inhibitor synthesis and analogue structure-activity relationship studies.

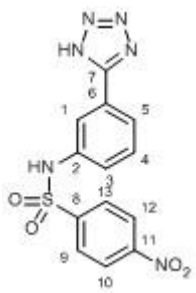


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-methylbenzenesulfonamide), **S19**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-methylbenzenesulfonyl chloride (1.1, 0.341 mmol, 65 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a white crystalline solid (48 mg, 49%). Mp 149-150 ^{o}C; ν_{max} (FTIR) cm^{-1} 3191 (aromatic CH), 1470 (tetrazole), 1330, 1157 (O=S=O); ^{1}H NMR (400 MHz) δ 7.86 (s, 1H, C(1)H), 7.69 (d, ^{3}J = 8.1 Hz, 2H, C(9)H, C(13)H), 7.66 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.44 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.34 (d, ^{3}J = 8.1 Hz, 2H, C(10)H, C(12)H), 7.26 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H), 2.30 (s, 3H, CH_3); ^{13}C NMR (101 MHz) δ 155.7 (C7), 143.5 (C8), 138.8 (C2), 136.4 (C11), 130.3 (C4), 129.8 (C10, C12), 126.7 (C9, C13), 125.7 (C6), 122.3 (C5), 121.9 (c3), 117.8 (C1), 20.9 (CH_3); m/z (ESI$^-$) 314.1, 100%([M-H]$^-$); HRMS (ESI$^-$) $C_{14}H_{12}N_5O_2S^-$, ([M-H]$^-$) requires 314.0717, found 314.0729.

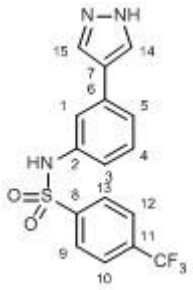


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-cyanobenzenesulfonamide), **S20**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 4-cyanobenzenesulfonyl chloride (1.1, 0.341 mmol, 69 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as a white crystalline solid (30 mg, 29%). Mp 144-145 ^{o}C; ν_{max} (FTIR) cm^{-1} 3211 (aromatic CH), 2235 (nitrile), 1483 (tetrazole), 1330, 1182 (O=S=O); ^{1}H NMR (400 MHz) δ 8.05 (d, ^{3}J = 8.7 Hz, 2H, C(9)H, C(13)H), 7.94 (d, ^{3}J = 8.7 Hz, 2H, C(10)H, C(12)H), 7.85 (s, 1H, C(1)H), 7.72 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.49 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.28 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 155.4 (C7), 143.2 (C8), 138.0 (C2), 133.6 (C10, C12), 130.6 (C4), 127.4 (C9, C13), 127.6 (C6), 123.0 (C5),

122.8 (C3), 118.5 (C1), 117.5 (CN), 115.6 (C11); m/z (ESI^-) 325.1, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{14}H_9N_6O_2S^-$, ($[M-H]^-$) requires 325.0513, found 325.0525.

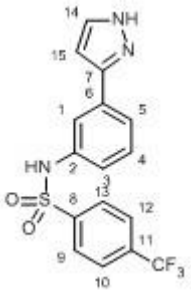


(N-(3-(1H-tetrazol-5-yl)phenyl)-4-nitrobenzenesulfonamide), **S21**. To a solution of 3-(1H-tetrazol-5-yl)aniline (1, 0.31 mmol, 50 mg) and 44-nitrobenzenesulfonyl chloride (1.1, .341 mmol, 76 mg) in anhydrous DCM (1 ml), under N_2 atmosphere, was pyridine (2.8, 0.865 mmol, 70 µl) *via* syringe. The reaction mixture was kept stirring at R.T. overnight. The next day reaction completion was confirmed by LC-MS. The solvent was evaporated and the residue dissolved in methanol and purified by prep-HPLC. Product containing fractions were combined and the solvent evaporated, yielding the desired product as an off-yellow crystalline solid (25 mg, 23%). Mp 219-220 ^oC; ν_{max} (FTIR) cm^{-1} 3105 (aromatic CH), 1524 (nitrate), 1416 (tetrazole), 1350, 1161 (O=S=O); ^1H NMR (400 MHz) δ 8.37 (d, ^3J = 8.5 Hz, 2H, C(9)H, C(13)H), 8.03 (d, ^3J = 8.5 Hz, 2H, C(10)H, C(11)H), 7.86 (s, 1H, C(1)H), 7.73 (d, ^3J = 7.9 Hz, 1H, C(5)H), 7.47 (dd, ^3J = 7.9 Hz, 1H, C(4)H), 7.27 (dm, ^3J = 7.9 Hz, 1H, C(3)H);); ^{13}C NMR (101 MHz) δ 155.8 (C7), 149.9 (C11), 144.6 (C8), 137.9 (C2), 130.5 (C4), 128.3 (C10, C12), 126.3 (C6), 124.8 (C9, C13), 123.1 (C5), 122.5 (C3), 118.6 (C1); m/z (ESI^-) 345.1, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{13}H_9N_6O_4S^-$, ($[M-H]^-$) requires 345.0411, found 345.0423.

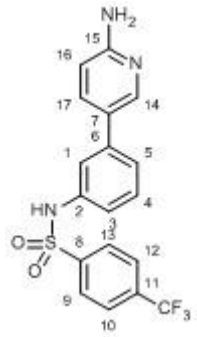


(N-(3-(1H-pyrazol-4-yl)phenyl)-4-(trifluoromethyl)benzenesulfonamide), **S22**. N-(3-bromophenyl)-4-(trifluoromethyl)benzenesulfonamide (1, 0.184 mmol, 70 mg), 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1H-pyrazole (1.5, 0.276 mmol, 54 mg) and $PdCl_2$(dppf) (0.1, 0.018 mmol, 14 mg) were dissolved in 2.5 ml dioxane, then supplemented with 300 µl 2M aq. Na_2CO_3. The reaction was run in MW for 15 min at 150°C. LC-MS showed the reaction was complete. The mixture was filtered through 45 um filter and concentrated, then redissolved in MeOH and purified by prep-HPLC. The product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow crystalline solid (8 mg, 11%). ^1H NMR (400 MHz) δ 8.06 (b, 1H, C(15)H), 8.00 (d, ^3J = 8.2 Hz,

2H, C(9)H, C(13)H), 7.95 (d, ^{3}J = 8.2 Hz, 2H, C(10)H, C(12)H), 7.76 (b, 1H, C(14)H), 7.30 (m, 2H, C(1)H, C(5)H), 7.21 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 6.88 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 143.4 (C2), 137.6 (C8), 136.0 (C14), 133.9 (C7), 132.5 (C11), 129.7 (C4), 127.7 (C10, C12), 126.6 (C9. C13), 126.0 (C15), 124.7 (CF_3), 121.7 (C5), 120.5 (14), 117.9 (C3), 117.2 (C1), m/z (ESI^-) 366.1, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{16}H_{11}F_3N_3O_2S^-$, ($[M-H]^-$) requires 366.0531, found 366.055.

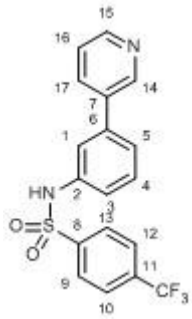


(N-(3-(1H-pyrazol-3-yl)phenyl)-4-(trifluoromethyl)benzenesulfonamide), **S23**. N-(3-bromophenyl)-4-(trifluoromethyl)benzenesulfonamide (1, 0.184 mmol, 70 mg), 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1H-pyrazole (1.5, 0.276 mmol, 54 mg) and $PdCl_2$(dppf) (0.1, 0.018 mmol, 14 mg) were dissolved in 2.5 ml dioxane, then supplemented with 300 µl 2M aq. Na_2CO_3. The reaction was run in MW for 15 min at 150°C. LC-MS showed the reaction was complete. The mixture was filtered through 45 um filter and concentrated, then redissolved in MeOH and purified by prep-HPLC. The product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow crystalline solid (20 mg, 30%). ^{1}H NMR (400 MHz) δ 7.99 (d, ^{3}J = 8.2 Hz, 2H, C(9)H, C(13)H), 7.94 (d, ^{3}J = 8.2 Hz, 2H, C(10)H, C(12)H), 7.72 (b, 1H, C(14)H), 7.59 (b, 1H, C(15)H), 7.47 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.29 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.03 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H), 6.58 (s, 1H, C(1)H) ; ^{13}C NMR (101 MHz) δ 143.5 (C2), 137.7 (C8), 132.5 (C11), 129.6 (C4), 127.7 (C9, C13), 126.6 (C10, C12), 123.3 (CF_3), 121.6 (C5), 119.4 (C3), 118.1 (C6), 117.1 (C15), 101.9 (C1); m/z (ESI^-) 366.0, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{16}H_{11}F_3N_3O_2S^-$, ($[M-H]^-$) requires 366.0531, found 366.0550.



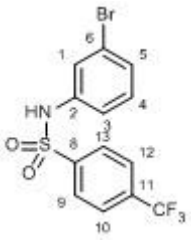
(N-(3-(6-aminopyridin-3-yl)phenyl)-4-(trifluoromethyl)benzenesulfonamide), **S24**. To a stirred solution of N-(3-bromophenyl)-4-(trifluoromethyl) benzenesulfonamide (1, 0.184 mmol, 70 mg) in 2 mL of toluene $Pd(Ph_3P)_4$ (0.07, 0.013 mmol, 15 mg), a solution of sodium carbonate (0.5, 0.092 mmol, 10 mg) in water (1 mL) and a solution of (6-aminopyridin-3-yl)boronic acid (2, 0.368 mmol, 51 mg) in EtOH (1 mL) under nitrogen atmosphere were added. The reaction mixture was stirred

vigorously under reflux overnight. LC-MS showed the reaction was complete. The mixture was cooled, supplemented with 5 ml water and the product was extracted with EtOAc (2 x 5 mL). The organic layers were combined and dried, then concentrated and purified by chromatography (10 g silica-gel, 10-40% EtOAc in hexanes). Product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow waxy solid (9 mg, 12%). ^{1}H NMR (400 MHz) δ 8.07 (d, ^{4}J = 2.2 Hz, 1H, C(14)H), 8.03 (d, , ^{3}J = 8.7 Hz, 2H, C(9)H, C(13)H), 7.50 (dd, ^{3}J = 8.7 Hz, ^{4}J = 2.2 Hz, 1H, C(17)H), 7,25 (m, 2H, C(4)H, C(5)H), 7.22 (s, 1H, C(1)H), 7.00 (dm, ^{3}J = 7.1 Hz, 1H, C(3)H), 6.50 (d, ^{3}J = 8.7 Hz, 1H, C(16)H), 6.11 (b, 2H, NH_2); ^{13}C NMR (101 MHz) δ 159.4 (C15), 145.7 (C14), 143.2 (C2), 139.2 (C8), 135.2 (C17), 132.5 (C11), 129.9 (C4), 127.7 (C9, C13), 126.5 (C10, C12), 123.1 (C6), 121.7 (C5), 118.3 (C3), 117.3 (C1), 107.9 (C16); m/z (ESI$^-$) 392.1, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{18}H_{13}F_3N_3O_2S^-$, ([M-H]$^-$) requires 392.0696, found 392.0701.

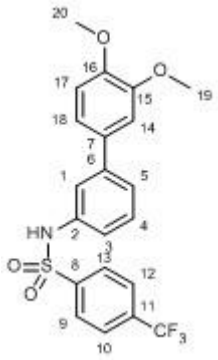


(N-(3-(pyridin-3-yl)phenyl)-4-(trifluoromethyl)benzenesulfonamide), **S25**. To a stirred solution of N-(3-bromophenyl)-4-(trifluoromethyl) benzenesulfonamide (1, 0.184 mmol, 70 mg) in 2 mL of toluene $Pd(Ph_3P)_4$ (0.07, 0.013 mmol, 15 mg), a solution of sodium carbonate (0.5, 0.092 mmol, 10 mg) in water (1 mL) and a solution of pyridin-3-ylboronic acid (2, 0.368 mmol, 45 mg) in EtOH (1 mL) under nitrogen atmosphere were added. The reaction mixture was stirred vigorously under reflux overnight. LC-MS showed the reaction was complete. The mixture was cooled, supplemented with 5 ml water and the product was extracted with EtOAc (2 x 5 mL). The organic layers were combined and dried, then concentrated and purified by chromatography (10 g silica-gel, 10-40% EtOAc in hexanes). Product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow crystalline solid (47 mg, 67%). Mp 188-189 ^{o}C; ν_{max} (FTIR) cm^{-1} 2716 (aromatic CH), 1322, 1163 (O=S=O); ^{1}H NMR (400 MHz) δ 8.73 (s, 1H, C(14)H), 8.57 (dm, ^{3}J = 4.9 Hz, 1H, C(15)H), 8.02 (d, ^{3}J = 8.2 Hz, 2H, C(9)H, C(13)H), 7.96 (d, ^{3}J = 8.2 Hz, 2H, C(10)H, C(12)H), 7.91 (dm, ^{3}J = 7.9 Hz, 1H, C(17)H), 7.48 (dd, ^{3}J = 7.9, 4.9 Hz, 1H, C(16)H), 7.41 (m 2H, C(3)H, C(4)H), 7.39 (s, 1H, C(1)H), 7.15 (dm, ^{3}J = 7.9 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 148.9 (C15), 147.5 (C14), 143.3

(C7), 138.2 (C2), 137.9 (C8), 134.1 (C6), 134.1 (C17), 132.6 (C11), 130.2 (C4), 127.7 (C9, C13), 126.6 (C10, C12), 123.9 (C16), 123.3 (C5), 123.1 (CF_3), 119.9 (C3), 118.7 (C1); m/z (ESI^-) 377.1, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{18}H_{13}F_3N_3O_2S^-$, ($[M-H]^-$) requires 377.0577, found 377.0591.

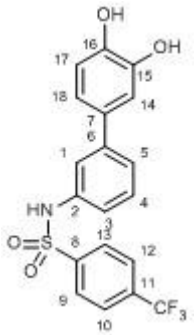


(N-(3-bromophenyl)-4-(trifluoromethyl)benzenesulfonamide), **S26**. To a solution of 4-(trifluoromethyl)benzenesulfonyl chloride (1.1, 2.05 mmol, 500 mg) and 3-bromoaniline (1, 1.86 mmol, 320 mg) in anhydrous DCM (4 ml), under N2 atmosphere, was pyridine (3, 5.57 mmol, 0.451 ml) *via* syringe. The reaction mixture was kept stirring at R.T for 1 hour. Then solvent evaporated and the crude product was purified by flash-chromatography (25 g silica-gel, 10-40% EtOAc in hexanes). Product containing fractions were pulled together and solvent evaporated, yielding the desired product as a white crystalline solid (360 mg, 51%). Mp 122-123 ^oC; ν_{max} (FTIR) cm^{-1} 3240 (aromatic CH), 1319, 1163 (O=S=O); ^1H NMR (400 MHz) δ 7.98 (s, 4H, C(9)H, C(10)H, C(11)H, C(12)H, C(13)H), 7.27 (s, 1H, C(1)H), 7,25 (dm, ^3J = 7.8 Hz, 1H, C(5)H), 7.22 (dd, ^3J = 7.8 Hz, 1H, C(4)H), 7.11 (dm, ^3J = 7.8 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 142.9 (C2), 138.8 (C8), 132.7 (C11), 131.4 (C4). 127.6 (C9, C13), 127.2 (C5), 126.7 (C10, C12), 123.2 (CF_3), 122.4 (C1), 121.2 (C6), 118.8 (C3); m/z (ESI^-) 378.0, 95%; 38.0, 100% ($[M-H]^-$); HRMS (ESI^-) $C_{13}H_8BrF_3NO_2S^-$, ($[M-H]^-$) requires 377.9417, found 377.9430.

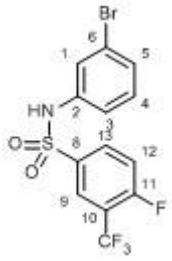


(N-(3',4'-dimethoxy-[1,1'-biphenyl]-3-yl)-4-(trifluoromethyl) benzenesulfonamide), **S27**. To a stirred solution of N-(3-bromophenyl)-4-(trifluoromethyl)benzenesulfonamide (1, 0.263 mmol, 100 mg) in 2 mL of toluene was added $Pd(Ph_3P)_4$ (0.07, 0.018 mmol, 21 mg), a solution of sodium carbonate (0.5, 0.132 mmol, 14 mg) in water (1 mL) and a solution of (3,4-dimethoxyphenyl)boronic acid (2, 0.526 mmol, 96 mg) in EtOH (1 mL) under nitrogen atmosphere. The reaction mixture was stirred vigorously under reflux overnight. LC-MS showed the reaction was complete. The mixture was cooled, supplemented with 5 ml water and the

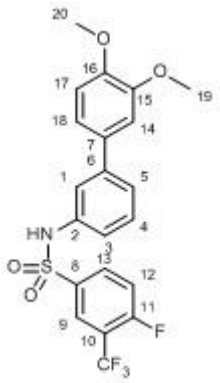
product was extracted with EtOAc (2 x 5 mL). The organic layers were combined and dried with MgSO4, then concentrated and purified by chromatography (10 g silica-gel, 10-40% EtOAc in hexanes). Product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow crystalline solid (103 mg, 90%). Mp 131-132 ^{o}C; ν_{max} (FTIR) cm^{-1} 3228 (aromatic CH), 1323, 1159 (O=S=O), 1131 (ether); ^{1}H NMR (400 MHz) δ 8..01 (d, ^{3}J = 8.7 Hz, 2H, C(9)H, C(13)H), 7.97 (d, ^{3}J = 8.7 Hz, 2H, C(10)H C(12)H), 7.31 (m, 2H, C(4)H, C(18)H), 7.29 (s, 1H, C(14)H), 7.04 (m, 3H, C(3)H, C(5)H, C(17)H), 7.02 (s, 1H, C(1)H), 3.80 (s, 3H, C(20)H_3), 3.78 (s, 3H, C(19)H_3); ^{13}C NMR (101 MHz) δ 149.0 (C16), 148.9 (C15), 143.3 (C6), 141.3 (C2), 137.6 (C8), 132.5 (C11), 132.2 (C7), 129.8 (C4), 127.7 (C9, C13), 126.7 (C10, C12), 123.2 (CF_3), 122.8 (C18), 118.8 (C5), 118.4 (C3), 112.2 (C17), 110.2 (C1), 55.5 (C20), 55.4 (C19); m/z (ESI$^-$) 436.1, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{21}H_{17}F_3NO_4S^-$, ([M-H]$^-$) requires 436.0836, found 436.0865.



(N-(3',4'-dihydroxy-[1,1'-biphenyl]-3-yl)-4-(trifluoromethyl)benzenesulfonamide), **S28**. To NaI (20, 3 mmol, 445 mg) suspension in 1ml AcN, TMS-Cl (20, 3 mmol, 0.38 ml) was added under nitrogen at 0°C. Solution of N-(3',4'-dimethoxy-[1,1'-biphenyl]-3-yl)-4-(trifluoromethyl) benzenesulfonamide (1, 0.1 mmol, 65 mg) in 1 ml AcN was added and the reaction was allowed to warm to the room temperature. After 48h the reaction mixture quenched with 10 ml sat. aq. $Na_2S_2O_5$, then the product was extracted from the aqueous phase using EtOAc (2 x 5 ml). The organic fractions were pulled together, dried concentrated. Crude product purified by prep-HPLC, yielding the product as a dark brown waxy solid (20 mg, 33%). ^{1}H NMR (400 MHz) δ 7.96 (m, 4H, C(9)H, C(10)H, C(12)H, C(13)H), 7.23 (m, 3H, C(4)H, C(14)H, C(18)H), 6.98 (d, ^{3}J = 8.0 Hz, 1H, C(5)H), 6.91 (s, 1H, C(1)H), 6.77 (m, 2H, C(3)H, C(17)H); ^{13}C NMR (101 MHz) δ 145.6 (C16), 145.5 (C15), 143.8 (C6), 141.5 (C2), 138.2 (C8), 132.4 (C11), 130.8 (C7), 129.6 (C4), 127.7 (C9, C13), 126.5 (C10, C12), 123.0 (CF_3), 121.9 (C18), 118.4 (C5), 188.0 (C14), 117.6 (3), 116.1 (C17), 113.8 (C1); m/z (ESI$^-$) 408.1, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{19}H_{13}F_3NO_4S^-$, ([M-H]$^-$) requires 408.0523, found 408.0538.

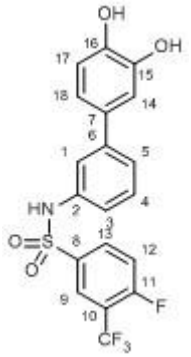


(N-(3-bromophenyl)-4-fluoro-3-(trifluoromethyl)benzenesulfonamide), **S29.** To a solution of 4-fluoro-3-(trifluoromethyl)benzenesulfonyl chloride (1.1, 1.5 mmol, 400 mg) and 3-bromoaniline (1, 1.38 mmol, 240 mg) in anhydrous DCM (3 ml), under nitrogen atmosphere, was pyridine (1.5, 2 mmol, 0.17 ml) via syringe. The reaction mixture was kept stirring at R.T. After 1 hour the solvent was evaporated and the crude product was purified by flash-chromatography (25 g silica-gel, 10-40% EtOAc in hexanes). The product containing fractions were pulled together and the solvent evaporated, yielding the desired product as a white crystalline solid (515 mg, 93%). Mp 100-101 ^{o}C; ν_{max} (FTIR) cm^{-1} 3253 (aromatic CH), 1323, 1127 (O=S=O); ^{1}H NMR (400 MHz) δ 8.11 (m, 1H, C(9)H), 8.08 (dm, ^{3}J = 6.2 Hz, 1H, C(13)H), 7.74 (dd, ^{3}J = 10.1, 6.2 Hz, 1H, C(12)H), 7.27 (m, 1H, C(5)H), 7.26 (m, 1H, C(1)H), 7.23 (dd, ^{3}J = 7.8 Hz, 1H, C(4)H), 7.11 (dm, ^{3}J = 7.8 Hz, 1H, C(3)H); ^{13}C NMR (101 MHz) δ 160.9 (C11), 138.6 (C2), 135.9 (C8), 134.2 (C9), 131.5 (C4), 127.6 (C5), 126.2 (C13), 122.9 (C1), 121.9 (C6), 121.2 (CF_3), 119.3 (C3), 118.2 (C12), 117.5 (C10); m/z (ESI$^-$) 395.9, 95%; 398.0, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{13}H_7BrF_4NO_2S^-$, ([M-H]$^-$) requires 395.9322, found 395.9338.



(N-(3',4'-dimethoxy-[1,1'-biphenyl]-3-yl)-4-fluoro-3-(trifluoromethyl) benzenesulfonamide), **S30**. N-(3-bromophenyl)-4-fluoro-3-(trifluoromethyl) benzenesulfonamide (1, 0.63 mmol, 250 mg), (3,4-dimethoxyphenyl)boronic acid (2, 1.26 mmol, 230 mg) and $PdCl_2$(dppf) (0.1, 0.063 mmol, 46 mg) were dissolved in 3 ml dioxane, then supplemented with 1 ml 2M aq. Na_2CO_3. The reaction was run in MW for 15 min at 150°C. Then the mixture was filtered through 45 um filter, concentrated and the residue dissolved in 5 ml DCM, subsequently filtered and dried. The crude mixture was purified by flash-chromatography (25 g silica-gel, 10-40% EtOAc in hexanes). The product containing fractions were pulled together and concentrated, yielding the desired product as an off-yellow crystalline solid (150 mg, 52%). Mp 89-90 ^{o}C; ν_{max} (FTIR) cm^{-1} 3238

(aromatic CH), 1316, 1137 (O=S=O); ^{1}H NMR (400 MHz) δ 8.10 (m, 2H, C(9)H, C(13)H), 7.73 (dd, , 1H, C(12)H), 7.36 (dm, ^{3}J = 7.9 Hz, 1H, C(5)H), 7.32 (dd, ^{3}J = 7.9 Hz, 1H, C(4)H), 7.28 (s, 1H, C(1)H), 7.00 (m, 4H, C(3)H, C(14)H, C(17)H, C(18)H), 3.81 (s, 3H, C(20)H), 3.78 (s, 3H, C(19)H), ^{13}C NMR (101 MHz) δ 161.2 (C11), 149.0 (C16), 148.8 (C15), 141.4 (C6), 137.4 (C2), 136.2 (C8), 134.3 (C9), 132.1 (C7), 128.8 (C4), 126.2 (C13), 123.0 (C5), 121.9 (CF_3), 119.3 (C3), 118.9 (C1), 188.8 (C12), 118.6 (C18), 117.6 (C10), 112.2 (C14), 110.2 (C17), 55.5 (C19, C20); m/z (ESI$^-$) 454.1, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{21}H_{16}F_4NO_4S^-$, ([M-H]$^-$) requires 454.0742, found 454.0755.



(N-(3',4'-dihydroxy-[1,1'-biphenyl]-3-yl)-4-fluoro-3-(trifluoromethyl) benzenesulfonamide), **S31**. To NaI (20, 4.4 mmol, 660 mg) suspension 1ml AcN, TMS-Cl (20, 4.4 mmol, 480 mg) was added under N2 at 0°C. Solution of the N-(3',4'-dimethoxy-[1,1'-biphenyl]-3-yl)-4-fluoro-3-(trifluoromethyl) benzenesulfonamide (1, 0.22 mmol, 100 mg) in 1 ml AcN was added and the reaction was allowed to warm to the room temperature. The reaction was left stirring at 50°C for 48 hours. The reaction mixture was poured into 10 ml sat. aq. $Na_2S_2O_5$ for quenching, then the product was extracted from the aqueous phase using EtOAc (2 x 5 ml). The organic fractions were pulled together, dried and concentrated. Crude product purified by prep-HPLC, yielding the product as a dark yellow waxy solid (14 mg, 15%). ^{1}H NMR (400 MHz) δ 8.08 (m, 2H, C(9)H, C(13)H), 7.71 (m, 1H, C(12)H), 7.26 (m, 1H, C(4)H), 7.21 (m, 2H, C(14)H, C(18)H), 6.97 (d, ^{3}J = 7.9 Hz, 1H, C(5)H), 6.91 (s, 1H, C(1)H), 6.76 (m, 2H, C(3)H, C(17)H); ^{13}C NMR (101 MHz) δ 161.1 C(11), 145.7 (C16), 145.5 (C15), 141.6 (C6), 138.1 (C2), 136.9 (C8), 134.1 (C13), 130.8 (7), 129.7 (C4), 126.1 (C9), 122.2 (C18), 121.7 (CF_3), 118.9 (C12), 118.8 (C5), 118.5 (C14), 117.5 (C3), 116.1 (C17), 113.8 (C1); m/z (ESI$^-$) 426.1, 100% ([M-H]$^-$); HRMS (ESI$^-$) $C_{19}H_{12}F_4NO_4S^-$, ([M-H]$^-$) requires 426.0429, found 426.0446.

6. Concluding remarks and future work

Screening of 1200+ compounds for ERAP1 inhibition from three different fragment libraries has yielded twelve structurally diverse hits (Table 6-1). The hits have been characterised by three orthogonal enzyme activity assays used in this work. Based on the identified hits a library of 52 follow-up fragments has been compiled and screened, providing initial structure-activity relationship information on the identified scaffolds. In addition, the reported IRAP inhibitor has been shown to inhibit ERAP1 with single digit micromolar potency in both colorimetric and MS-based enzyme activity assays (Fig. 6-6).

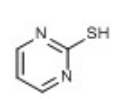
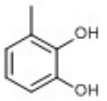
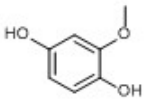
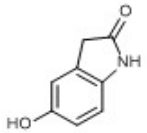
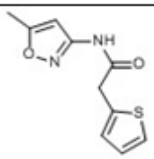
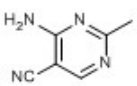
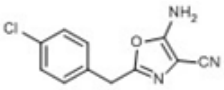
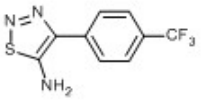
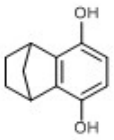
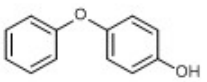
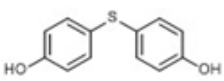
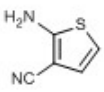
				
Fragment ID	**Zn_1**	**Zn_2**	**Zn_3**	**3D_4**
Colorimetric IC_{50}, mM	**2.23 ± 3.01**	**0.24 ± 0.21**	**0.001 ± 0.001**	**0.08 ± 0.05**
Fluorogenic IC_{50}, mM	**0.08 ± 0.07**	**Interfere**	**Interfere**	**0.08 ± 0.09**
MS, IC_{50}, mM	**0.79 ± 0.61**	**0.002 ± 0.003**	**0.90 ± 0.61**	**0.15 ± 0.21**
				
Fragment ID	**3D_5**	**3D_22**	**3D_12**	**M_3**
Colorimetric IC_{50}, mM	**0.14 ± 0.72**	**Not active**	**ca 0.15**	**0.41 ± 0.25**
Fluorogenic IC_{50}, mM	**0.05 ± 0.08**	**0.31 ± 0.22**	**0.25 ± 0.19**	**0.22 ± 0.16**
MS, IC_{50}, mM	**0.006 ± 0.013**	**0.43 ± 0.29**	**0.78 ± 3.71**	**0.92 ± 2.41**
				
Fragment ID	**M_7**	**M_18**	**M_19**	**M_20**
Colorimetric IC_{50}, mM	**0.07 ± 0.09**	**1.41 ± 2.15**	**1.64 ± 2.09**	**0.76 ± 0.65**
Fluorogenic IC_{50}, mM	**Interfere**	**0.31 ± 0.82**	**Interfere**	**0.07 ± 0.04**
MS, IC_{50}, mM	**0.1 ± 0.5 µM**	**0.85 ± 3.05**	**0.40 ± 0.53**	**0.31 ± 0.54**

Table 6-1. 12 hits identified from by the initial screening campaign were characterised using all three enzyme activity assays. In red are highlighted potential Pan Assay Interference Compounds (PAINS). [110]

Through collaboration with Novartis, a proprietary fluorinated fragments library has been screened by NMR, which has resulted in four more hits. The hits have been validated in the colorimetric and MS-based enzyme activity assays (Table 6-2). Some of these molecules share structural similarity with the hits identified initially. The fact that the fragments identified by different biophysical screening methods have common structural features is reassuring that they are genuine ERAP1 inhibitors and not assay artefacts. In the absence of the structural information, the NMR hits provide the first indication that some fragments do bind inside ERAP1.

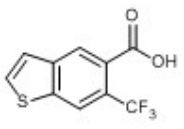
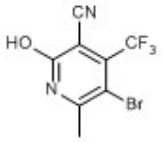
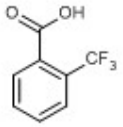
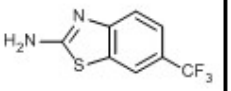
				
Fragment ID	**AKL824**	**CGA262442**	**GPA000553**	**AEC780**
Colorimetric IC_{50}, mM	**0.89 ± 1.61**	**0.66 ± 0.98**	**1.29 ± 2.37**	**4.1 ± 9.0**
MS IC_{50}, mM	**0.23 ± 0.18**	**0.33 ± 0.12**	**0.41 ± 0.14**	**0.60 ± 0.76**

Table 6-2. The results of the identified NMR hits characterisation using the colorimetric and MS enzyme activity assays.

One of the three selectivity panel enzymes – LTA4H – has been successfully cloned, expressed and purified. The cloning of the remaining two enzymes – ERAP2 and AnPEP – has been never achieved despite numerous attempts due to the unidentified reasons. Screening of the ERAP1 hits against LTA4H has shown that most of them do not affect its aminopeptidase activity, thus the identified scaffolds exhibit good selectivity against this enzyme (Table 6-3).

IRAPi	LTA4Hi	Zn_A7	3D_E10	Ext_2	Ext_6	Ext_7	Ext_8	Ext_9	Ext_10
0.15 (x2)	Inactive	0.10 (x10)	Inactive	Inactive	Inactive	0.02 (x10)	Inactive	Inactive	Inactive
AKL824	CGA262442	Ext_51	Ext_52	Zn_B4	Zn_D1	M_6D7	GPA00055	Ext_12	Ext_15
Inactive	0.97	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
Ext_16	Ext_17	Ext_18	Ext_19	3D_C4	Ext_22	3D_C7	3D_C9	Ext_27	Ext_31
Inactive	Inactive	Inactive	Inactive	2.4	Inactive	Inactive	Inactive	Inactive	Inactive
Ext_34	M_4B9	AEC780	M_12A2	Ext_40	M_10B4	M_10E3	Ext_41	Ext_43	Ext_46
Inactive	Inactive	1.1 (x10)	Inactive	Inactive	0.005 (x4)	0.009 (x6)	Inactive	0.04 (x45)	0.20 (x5)
Ext_47									
Inactive									

Table 6-3. The results of the selectivity screening of the identified ERAP1 hits against LTA4H using the same colorimetric assay (n = 3). The compounds that did not show any effect are in red. The compounds that showed inhibitory effect are in green, IC_{50} are given in mM. The compounds that showed activating effect (the speed of the reaction was higher than the control) are in yellow, EC_{50} are given in mM, in parenthesis is the activation effect at 2 mM compounds concentration.

Three different approaches have been tried in order to obtain structural information about how the identified hits interact with ERAP1. Initially, attempts were made to reproduce the previously obtained ERAP1 crystals or to find a new crystallisation protocol using the same enzyme construct.

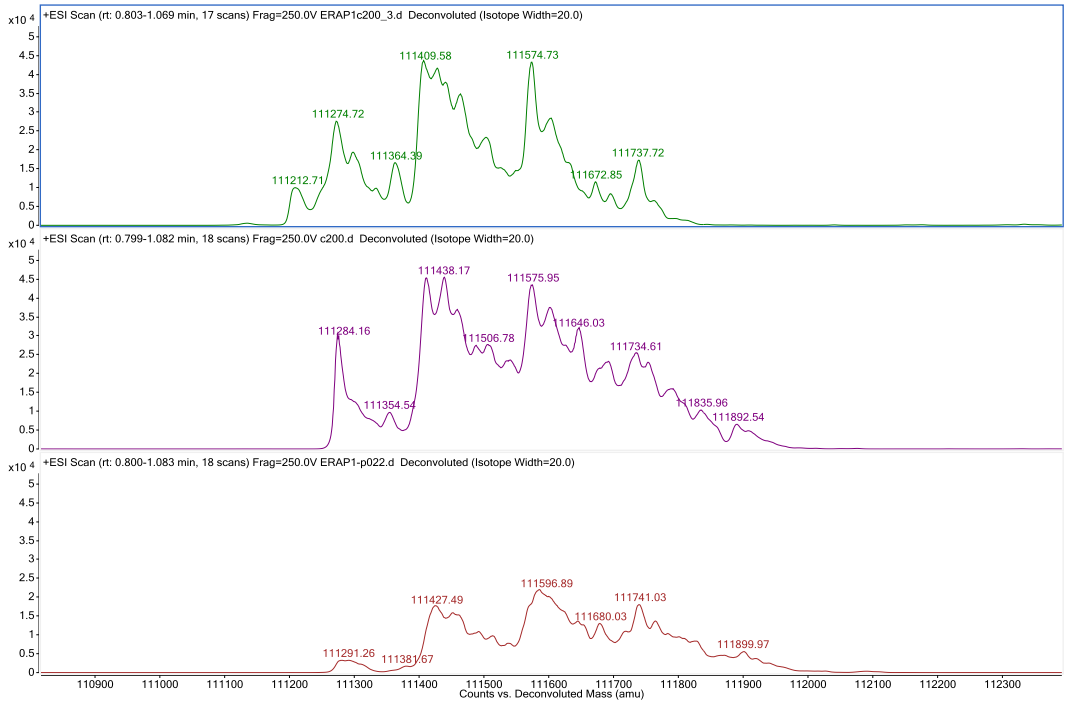


Figure 6-1. The deconvoluted intact MS spectra of three purified ERAP1-c200 samples from the different expression batches. The main peaks, each representing a particular glycoform, were the same for all samples. The distribution of the glycoforms in each protein sample varied from one expression batch to another

Then, crystallisation of the shorter ERAP1 construct was pursued. The glycosylation analysis of the protein samples from the both constructs has suggested that their heterogeneity due to presence of several glycoforms might be the reason why ERAP1 crystals have not been obtained (Fig. 6-1). Finally, LTA4H has been tried as an ERAP1 crystallisation substitute to elucidate the identified hits binding interaction with the enzyme. Hence, a LTA4H soaking system with the fragments was successfully developed. However, none of the ERAP1 inhibiting molecules have been observed on the obtained crystal structures (Table 6-5). Thus, both enzyme activity assay screening and crystallisation results have indicated that LTA4H is a poor ERAP1 substitute for this purpose.

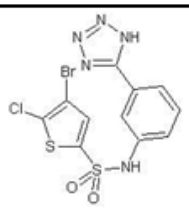
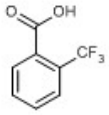
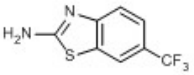
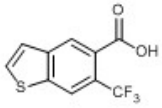
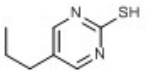
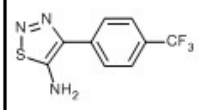
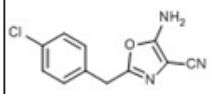
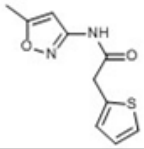
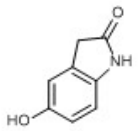
					
Fragment ID	**IRAP-I**	**GPA000553**	**AEC780**	**AKI824**	**Ext_6**
ERAP1, IC_{50} mM	**9 ± 6 µM**	**1.29 ± 2.37**	**4.1 ± 9.0**	**0.89 ± 1.61**	**Not active [MS]**
LTA4H, IC_{50} mM	**EC_{50} 0.15 ± 0.05 (x2)**	**Not active**	**1.08 ± 0.22**	**0.72 ± 0.85**	**EC_{50} Poor curve (x6)**
					
Fragment ID	**M_3**	**3D_12**	**3D_5**	**3D_4**	**M_20**
ERAP1, IC_{50} mM	**0.41 ± 0.25**	**ca 0.15**	**0.14 ± 0.72**	**0.08 ± 0.05**	**0.76 ± 0.65**
LTA4H, IC_{50} mM	**Not active**	**Not active**	**2.34 ± 5.11**	**Not active**	**Not active**

Table 6-4. A set of hits selected for soaking with LTA4H and their effect on aminopeptidase activity of ERAP1 and LTA4H in the colorimetric assay. (MS) implies that the fragment was not active in the ERAP1 Mass-Spectrometry activity assay which was used to confirm hits from the colorimetric one.

In the final stage of the project a structure-activity relationship studies campaign of the IRAP inhibitor was performed. Three rounds of synthesis have yielded 31 sulphonamide analogues that have been characterised in the colorimetric enzyme activity assay. The screening has provided some structure-activity relationship insights, for example unexpected strong potency of S13, which may be rationalised once structural information about the analogues is available. As the result of this campaign a single digit micromolar ERAP1 inhibitor with a novel scaffold S31 has been developed (Table 6-6).

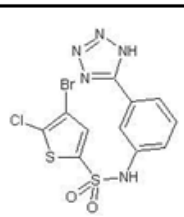
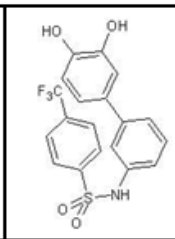
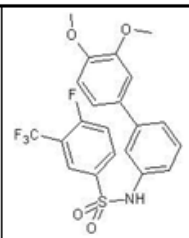
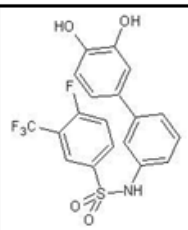
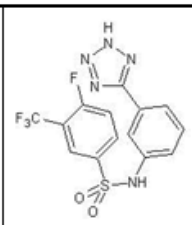
					
Fragment ID	**IRAP inhibitor**	**S28**	**S30**	**S31**	**S13**
ERAP1 Colorimetric IC_{50}, µM	**9 ± 6**	**10 ± 8**	**9 ± 7**	**7 ± 3**	**3 ± 5**

Table 6-5. The most potent sulphonamide analogues that have been identified during the SAR campaign.

The lack of structural information has been the main hurdle in this work. Thus, for the successful further development of this project a robust method for obtaining such information has to be devised. The first potential strategy would be to develop a different ERAP1 construct, which allows reproducible crystallisation. Alternatively, structural information can be obtained using a different biophysical method, for example NMR.

7. Bibliography

[1] T. Serwold et al., “ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum.,” *Nature,* vol. 419, pp. 480-483, 2002.

[2] T. Saric et al., “An IFN-gamma-induced Aminopeptidase in the ER, ERAP1, Trims Precursors to MHC Class I-Presented Peptides,” *Nat. Immunol.,* vol. 3, pp. 1169-1176, 2002.

[3] I. York et al., “The ER Aminopeptidase ERAP1 Enhances or Limits Antigen Presentation by Trimming Epitopes to 8-9 Residues,” *Nat. Immunol.,* vol. 3, pp. 1177-1184, 2002.

[4] A. Townsend et al., “The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides,” *Cell,* vol. 44, pp. 959-968, 1986.

[5] R. Kenneth et al., “Post-proteasomal antigen processing for major histocompatibility complex class I presentation,” *Nat. Immunol.,* vol. 5, pp. 670-677, 2004.

[6] P. Cascio et al., “26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide,” *EMBO,* vol. 20, pp. 2357-2366, 2001.

[7] F. Momburg et al., “Peptide Selection by MHC-encoded TAP Transporters,” *Curr. Opin. Immunol.,* vol. 6, pp. 32-37, 1994.

[8] S. Chang et al., “The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism,” *Proc. Natl. Acad. Sci. USA,* vol. 102, pp. 17107-17112, 2005.

[9] J. Yan et al., “In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules,” *J. Exp. Med.,* vol. 7, pp. 103-112, 2006.

[10] E. Firat et al., “The role of endoplasmic reticulum-associated aminopeptidase 1 in immunity to

infection and in cross-presentation.," *J. Immunol.,* vol. 178, pp. 2241-2248, 2007.

[11] G. Hammer et al., "Coping with loss of perfection in the MHC class I peptide repertoire," *Nat. Immunol.,* vol. 7, pp. 103-112, 2006.

[12] I. York et al., "Proteasome and Peptidase Function in MHC-class-I-mediated Antigen Presentation," *Immunol. Rev.,* vol. 172, pp. 49-66, 1999.

[13] N. Yamamoto et al., "Identification of 33 polymorphisms in the adipocyte-derived leucine aminopeptidase (ALAP) gene and possible association with hypertension," *Hum. Mutat.,* vol. 19, pp. 251-257, 2002.

[14] T. Akada et al., "Puromycin insensitive leucyl-specific aminopeptidase (PILSAP) is involved in the activation of endothelial integrins," *J. Cell. Physiol.,* vol. 193, pp. 253-262, 2002.

[15] Y. Watanabe et al., "Adipocyte-Derived Leucine Aminopeptidase Suppresses Angiogenesis in Human Endometrial Carcinoma via Renin-Angiotensin System," *Clin. Cancer Res.,* vol. 9, pp. 6497-6503, 2003.

[16] X. Cui et al., "Shedding of the Type II IL-1 Decoy Receptor Requires a Multifunctional Aminopeptidase, Aminopeptidase Regulator of TNF Receptor Type 1 Shedding," *J. Biol. Chem.,* vol. 171, pp. 6814-6819, 2003.

[17] I. Evnouchidou et al., "A New Role for Zn(II) Aminopeptidases: Antigenic Peptide Generation and Destruction," *Curr. Pharmacol. Design,* vol. 15, pp. 3656-3670, 2009.

[18] M. Brown, "Genetics and the pathogenesis of ankylosing spondylitis," *Curr. Opin. Rheumatol.,* vol. 52, pp. 318-323, 2009.

[19] W. Maksymowych et al., "Association of a specific ERAP1/ARTS1 haplotype with disease

susceptibility in ankylosing spondylitis," *Arthritis Rheum.,* vol. 60, pp. 1317-1323, 2009.

[20] J. Braun et al., "Ankylosing spondylitis.," *Lancet,* vol. 369, pp. 1379-1390, 2007.

[21] D. Evans et al., "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility.," *Nat. Gen.,* vol. 43, pp. 761-767, 2011.

[22] E. Fung et al., "Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus.," *Genes Immun.,* vol. 10, pp. 188-191, 2009.

[23] Editorial, "Towards a better understanding of type 1 diabetes," *Lancet,* vol. 387, p. 2264, 2016.

[24] A. Mehta et al., "Differences in genetic variation in antigen-processing machinery components and association with cervical carcinoma risk in two Indonesian populations," *Genes Chrom. Cancer,* vol. 46, pp. 577-586, 2007.

[25] X. Castellsague et al., "Worldwide Human Papillomavirus Etiology of Cervical Adenocarcinoma and Its Cofactors: Implications for Screening and Prevention," *J. Natl. Cancer Inst.,* vol. 92, pp. 914-921, 2006.

[26] N. Fowler et al., "Immunological and clinical significance of HLA class I antigen processing machinery component defects in malignant cells," *Gynecol. Oncol.,* vol. 92, pp. 914-921, 2008.

[27] D. Fruci et al., "Altered expression of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in transformed non-lymphoid human tissues.," *J. Cell. Physiol.,* vol. 216, pp. 742-749, 2008.

[28] G. Kochan et al., "Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming," *PNAS,* vol. 108, pp. 7745-7750,

2011.

[29] T. Nguyen et al., “Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1,” *Nat. Struct. Mol. Biol.,* vol. 18, pp. 604-613, 2011.

[30] M. Tsujimoto et al., “Biochemical and enzymatic properties of the M1 family of aminopeptidases involved in the regulation of blood pressure,” *Hear Fail. Rev.,* vol. 13, pp. 285-291, 2008.

[31] A. Hearn et al., “The Specificity of Trimming of MHC Class I-Presented Peptides in the Endoplasmic Reticulum,” *J. Immunol.,* vol. 183, pp. 5526-5536, 2009.

[32] Y. Goto et al., “Glutamine-181 Is Crucial in the Enzymatic Activity and Substrate Specificity of Human Endoplasmic-Reticulum Aminopeptidase-1,” *J. Biochem.,* vol. 416, pp. 109-116, 2008.

[33] A. Wetterholm et al., “Leukotriene A4 hydrolase: abrogation of the peptidase activity by mutation of glutamic acid-296.,” *Proc. Natl. Acad. Sci. USA,* vol. 89, pp. 9141-9145, 1992.

[34] M. Blomster et al., “Evidence for a catalytic role of tyrosine 383 in the peptidase reaction of leukotriene A4 hydrolase.,” *J. Eur. Biochem.,* vol. 231, pp. 528-534, 1995.

[35] M. Thunissen et al., “Crystal structure of human leukotriene A(4) hydrolase, a bifunctional enzyme in inflammation.,” *Nat. Struct. Biol.,* vol. 8, pp. 131-135, 2001.

[36] T. Kanaseki et al., “ERAAP synergizes with MHC class I molecules to make the final cut in the antigenic peptide precursors in the endoplasmic reticulum,” *Immunity,* vol. 25, pp. 795-806, 2006.

[37] S. Chang et al., “The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I

peptides by a “molecular ruler” mechanism,” *PNAS,* vol. 102, pp. 17107-17112, 2005.

[38] I. Evnouchidou et al., “The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1.,” *PLoS ONE,* vol. 3, 2008.

[39] A. Gandhi et al., “Structural insights into the molecular ruler mechanism of the endoplasmic reticulum aminopeptidase ERAP,” *Scientific Reports,* vol. 1, pp. 1-6, 2011.

[40] H. Umezawa et al., “Bestatin, an inhibitor of aminopeptidase B, produced by actinomycetes,” *J. Antibiot.,* vol. 29, pp. 97-99, 1976.

[41] E. Zervoudi et al., “Rationally designed inhibitor targeting antigen-trimming aminopeptidases enhances antigen presentation and cytotoxic T-cell responses,” *PNAS,* vol. 110, p. 19890–19895, 2013.

[42] L. Chen et al., “Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis,” *Ann. Rheum. Dis.,* vol. 75, pp. 916-923, 2016.

[43] V. Dive et al., “Phosphinic peptides as zinc metalloproteinase inhibitors,” *CMLS,* vol. 61, p. 2010–2019, 2004.

[44] E. Stamogiannos et al., “Screening Identifies Thimerosal as a Selective Inhibitor of Endoplasmic Reticulum Aminopeptidase 1,” *ACS Med. Che. Lett.,* vol. 7, pp. 681-685, 2016.

[45] K. Maynard et al., “Evolutionary analysis of the mammalian M1 aminopeptidases reveals conserved exon structure and gene death,” *Gene,* vol. 552, pp. 126-132, 2014.

[46] A. Stsiapanava et al., “Binding of Pro-Gly-Pro at the active site of leukotriene A4 hydrolase/aminopeptidase and development of an epoxide hydrolase selective inhibitor,”

PNAS, vol. 111, p. 4227–4232, 2014.

[47] J. Birtley et al., "The Crystal Structure of Human Endoplasmic Reticulum Aminopeptidase 2 Reveals the Atomic Basis for Distinct Roles in Antigen Processing," *Biochemistry,* vol. 51, p. 286−295, 2012.

[48] C. Schmitt, "Expression, purificacion et cristallisation de l'Aminopeptidase-H humaine (APN ou CD13); evaluation in vitro et in vivo d'inhibiteurs selectifs," *PhD thesis, University of Haute-Alsace, France,* 2012.

[49] M. Mueller et al., "Analysis of the molecular mechanism of substrate-mediated inactivation of leukotriene A4 hydrolase.," *J. Biol. Chem.,* vol. 273, pp. 11570-11575, 1998.

[50] X. Chen et al., "Role of leukotrienes revealed by targeted disruption of the 5-lipoxygenase gene.," *Nature,* vol. 372, pp. 179-182, 1994.

[51] R. Griffith et al., "Collagen-induced arthritis is reduced in 5-lipoxygenase–activating protein-deficient mice," *J. Exp. Med.,* vol. 185, pp. 1123-1129, 1997.

[52] P. Devchand et al., "The PPARalpha-leukotriene B4 pathway to inflammation control.," *Nature,* vol. 384, pp. 39-43, 1996.

[53] L. Saveanu et al., "Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum," *Nat. Immun.,* vol. 6, pp. 689-697, 2005.

[54] P. Mina-Osorio et al., "The moonlighting enzyme CD13: old and new functions to target.," *Trends Mol. Med.,* vol. 14, pp. 361-371, 2008.

[55] A. Wong et al., "The X-ray Crystal Structure of Human Aminopeptidase N Reveals a Novel Dimer and the Basis for Peptide Processing," *J. Biol. Chem.,* vol. 287, pp. 36804-36813, 2012.

[56] M. Dawson et al., “AT9283, a potent inhibitor of the Aurora kinases and Jak2, has therapeutic potential in myeloproliferative disorders.,” *Br. J. Haematol.,* vol. 150, pp. 46-57, 2010.

[57] M. Suires et al., “Biological characterization of AT7519, a small-molecule inhibitor of cyclin-dependent kinases, in human tumor cell lines,” *Mol. Cancer. Ther.,* vol. 8, pp. 324-332, 2009.

[58] C. Wada, “The Evolution of the Matrix Metalloproteinase Inhibitor Drug Discovery Program at Abbott Laboratories,” *Cur. Top. Med. Chem.,* vol. 4, pp. 1255-1267, 2004.

[59] V. Sandayaka et al., “Discovery of Novel Leukotriene A4 Hydrolase Inhibitors Based on Piperidine and Piperazine Scaffolds,” *Med. Chem. Lett.,* vol. 20, pp. 2851-2854, 2010.

[60] C. Bountra et al., “Open access chemical and clinical probes to support drug discovery.,” *Nat. Chem. Biol.,* vol. 5, pp. 436-440, 2009.

[61] M. Bunnage et al., “Target validation using chemical probes.,” *Nat. Chem. Biol.,* vol. 9, pp. 195-199, 2013.

[62] G. Seleznyov et al., “Escherichia coli physiology in Luria-Bertani broth.,” *J. Bacteriol.,* vol. 189, pp. 8746-8749, 2007.

[63] G. Rosano et al., “Recombinant protein expression in Escherichia coli: advances and challenges.,” *Front. Microbiol.,* vol. 5, pp. 1-17, 2014.

[64] S. Sahdev et al., “Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies,” *Mol. Cell. Biochem.,* vol. 307, pp. 249-264, 2008.

[65] F. Altmann et al., “Insect cells as hosts for the expression of recombinant glycoproteins,” *J. Glycoconj.,* vol. 16, pp. 109-123, 1999.

[66] X. Shi et al., “Protein N-Glycosylation in the Baculovirus-Insect Cell System,” *Curr. Drug. Targets,* vol. 8, pp. 1116-1125, 2007.

[67] E. Zervoudi et al., “Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides,” *J. Biochem.,* vol. 435, pp. 411-420, 2011.

[68] K. Sano et al., “Enhancement of protein expression in insect cells by a lobster tropomyosin cDNA leader sequence.,” *FEBS lett.,* vol. 532, pp. 143-146, 2002.

[69] Y. Watanabe et al., “Identification of an Alanine Aminopeptidase in Human Maternal Serum as a Membrane-Bound Aminopeptidase N,” *Biol. Chem. Hoppe-Seyler,* vol. 376, pp. 397-400, 1995.

[70] P. Hirel et al., “Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid.,” *Proc. Natl. Acad. Sci. USA,* vol. 86, pp. 8247-8251, 1989.

[71] V. Abet et al., “Biased and unbiased strategies to identify biologically active small molecules,” *Bioorg. Med. Chem.,* vol. 22, pp. 4474-4489, 2014.

[72] C. Schmitt et al., “Selective aminopeptidase-N (CD13) inhibitors with relevance to cancer chemotherapy,” *Bioorg. Med. Chem.,* vol. 21, pp. 2135-2144, 2013.

[73] I. Evnouchidou et al., “A continuous fluorigenic assay for the measurement of the activity of endoplasmic reticulum aminopeptidase 1: competition kinetics as a tool for enzyme specificity investigation.,” *Anal. Biochem. ,* vol. 395, pp. 33-40, 2009.

[74] A. Agraval et al., “Chelator Fragment Libraries for Targeting Metalloproteinases,” *Chem. Med. Chem.,* vol. 5, pp. 195-199, 2010.

[75] A. Morley et al., “Fragment-based hit identification: thinking in 3D.,” *Drug. Discov. Today,* vol. 18, pp. 1221-1227, 2013.

[76] M. Congreve et al., “A 'rule of three' for fragment-based lead discovery?,” *DrugDiscov. Today,* vol. 8, pp. 876-877, 2003.

[77] J. Dahlin et al., “Assay Guidance Manual,” *Eli Lilly & Company and the National Center for Advancing Trnslational Science,* 2015.

[78] J. Baell et al, “New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening,” *J. Med. Chem.,* vol. 53, pp. 2719-2740, 2010.

[79] R. Borhade et al., “Inhibition of Insulin-Regulated Aminopeptidase (IRAP) by Arylsulfonamides,” *ChemistryOpen,* vol. 3, pp. 256-263, 2014.

[80] L. Ornig et al., “The Bifunctional Enzyme Leukotriene-A, Hydrolase Is an Arginine Aminopeptidase of High Efficiency and Specificity,” *J. Biol. Chem.,* vol. 269, pp. 11269-11273, 1994.

[81] P. Hajduk et al., “A decade of fragment-based drug design: strategic advances and lessons learned.,” *Nat. Rev. Drug Discov.,* vol. 6, pp. 211-219, 2007.

[82] J. Schliebel et al., “High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits,” *Structure,* vol. 24, pp. 1398-1409, 2016.

[83] T. Blundell et al., “High-throughput X-ray crystallography for drug discovery.,” *Curr. Opin. Pharmacol.,* vol. 4, pp. 490-496, 2004.

[84] M. Winn et al., “Overview of the CCP4 suite and current developments.,” *Acta Cryst.,* vol. D67, pp. 235-242, 2011.

[85] P. Adams et al., “PHENIX: a comprehensive Python-based system for macromolecular structure solution.,” *Acta Cryst.,* vol. D66, pp. 213-221, 2010.

[86] T. Krojer et al., “The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination.,” *Acta Crystall.,* vol. D1, pp. 267-278, 2017.

[87] J. Oslob et al., “Discovery of a potent and selective aurora kinase inhibitor.,” *Bioorg. Med. Chem Lett.,* vol. 18, pp. 4880-4884, 2008.

[88] A. Byzia et al., “A remarkable activity of human leukotriene A4 hydrolase (LTA4H) toward unnatural amino acids,” *Amino Acids,* vol. 46, pp. 1313-1320, 2014.

[89] G. Winter et al., “xia: an expert system for macromolecular crystallography data reduction.,” *J. Appl. Cryst.,* vol. 43, pp. 186-190, 2009.

[90] G. Winter et al., “Automated data collection for macromolecular crystallography.,” *Methods,* vol. 55, pp. 81-93, 2011.

[91] A. McCoy et al., “Phaser crystallographic software.,” *J. Appl. Cryst.,* vol. 40, pp. 658-674, 2007.

[92] F. Long et al., “AceDRG: a stereochemical description generator for ligands.,” *Acta Cryst.,* vol. D73, pp. 112-122, 2017.

[93] A. Vagin et al., “Refinement of Macromolecular Structures by the Maximum-Likelihood method,” *Acta Cryst.,* vol. D53, pp. 240-255, 1997.

[94] P. Emsley et al., “Features and development of Coot.,” *Acta Cryst.,* vol. D66, pp. 486-501, 2010.

[95] N. Drinkwater et al., “M1 aminopeptidases as drug targets: broad applications or therapeutic niche?,” *FEBS,* vol. 284, pp. 1473-1477, 2017.

[96] C. Maieranu et al., “A novel amino-benzosuberone derivative is a picomolar inhibitor of

mammalian aminopeptidase N/CD13.," *Bioorg. Med. Chem.,* vol. 19, pp. 5716-5733, 2011.

[97] L. Su et al., "Discovery of a synthetic Aminopeptidase N inhibitor LB-4b as a potential anticancer agent," *Bioorg. Med. Chem. Lett.,* vol. 23, pp. 2512-2517, 2013.

[98] S. Mountford et al., "Synthesis, Structure-Activity Relationships and Brain Uptake of a Novel Series of Benzopyran Inhibitors of Insulin-Regulated Aminopeptidase," *J. Med. Chem.,* vol. 57, pp. 1368-1377, 2014.

[99] H. Chen et al., "On evaluating molecular-docking methods for pose prediction and enrichment factors.," *J. Chem. Inf. Model,* vol. 46, pp. 401-415, 2006.

[100] K. Lipkowitz et al., Reviews in computational chemistry, New York: Wiley-VCH, 2001, p. pg. 26.

[101] V. Sorna et al., "High-Throughput Virtual Screening Identifies Novel N'-(1-Phenylethylidene)-benzohydrazides as Potent, Specific, and Reversible LSD1 Inhibitors," *J. Med. Chem.,* vol. 56, pp. 9496-9508, 2013.

[102] A. Wein et al., "Small Molecule Inhibitors of Bacillus anthracis Protective Antigen Proteolytic Activation and Oligomerization," *J. Med. Chem.,* vol. 55, pp. 7998-8006, 2012.

[103] A. Stelzer et al., "Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble.," *Nat. Chem. Biol.,* vol. 7, pp. 553-559, 2011.

[104] C. Whitehead et al., "Small molecule inhibitors of egfr and pi3k". Patent WO2016/100347 A2, 2016.

[105] N. Garg et al., "The Total Synthesis of (+)-Dragmacidin F," *J. Am. Chem. Soc.,* vol. 126, pp. 9552-9553, 2004.

[106] P. Perry et al., "The rapid reduction of α,α-diaryl alcohols to the corresponding alkanes using

iodotrimethylsilan," *Synth. Commun.,* vol. 26, pp. 101-111, 1996.

[107] F. Tholander et al., "Discovery of Leukotriene A4 Hydrolase Inhibitors Using Metabolomics Biased Fragment Crystallography," *Chem. Biol.,* vol. 15, pp. 920-929, 2008.

[108] S. Park et al., "Targeted Inhibitor Design: Lessons from Small Molecule Drug Design, Directed Evolution, and Vaccine Research," *Chem. Eng. Process. Tech.,* vol. 1, pp. 1-6, 2013.

[109] E. Mashalidis et al., "A three-stage biophysical screening cascade for fragment-based drug discovery," *Nat. Prot.,* vol. 8, pp. 2309-2324, 2013.

[110] J. Baell et al., "Chemistry: Chemical con artists foil drug discovery.," *Nature,* vol. 513, pp. 481-483, 2014.

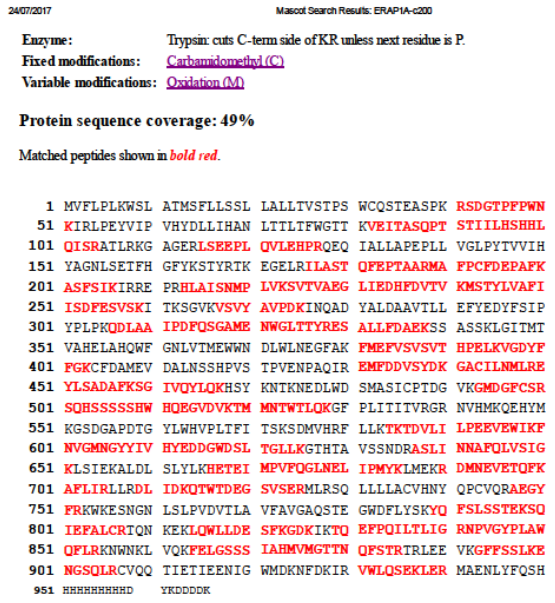
[111] D. Fusco et al., "Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study.," *Soft matter,* vol. 10, pp. 290-302, 2014.

1. Appendix I

MATLAB code for calculating reaction velocity v (abs. units / sec.) for each assay run

```
[Rows_Data, Columns_Data] = size(Kinetics);
%Kinetics - input matrix with experimental results

One_Measurement = zeros(Rows_Data,1);

Time = zeros(Rows_Data,1);

V_reaction = zeros(1,Columns_Data);

disp('vectors preallocated');

    for j=1:Rows_Data      %getting time measurements from the first column
        Time(j) = Kinetics(j, 1);
    end

 for i=2:Columns_Data

    for j=1:Rows_Data  %getting datapoints from one assay well
        One_Measurement(j) = Kinetics(j, i);
    end

    Linear_regression = polyfit(Time, One_Measurement, 1); %linear fitting

    if (Linear_regression(1)<0)
        disp('negative slope');
        i

    else
        V_reaction(i) = Linear_regression(1);

    end
end
```

2. Appendix II



24/07/2017 Mascot Search Results: ERAP1A-c200

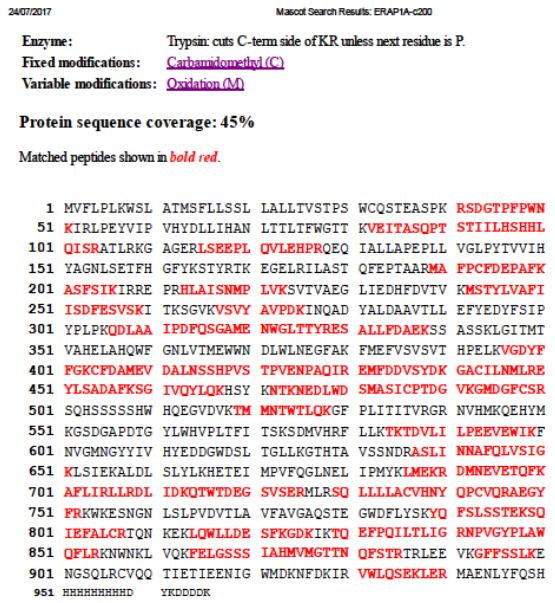
Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 49%

Matched peptides shown in ***bold red***.

```
   1 MVFLPLKWSL ATMSFLLSSL LALLTVSTPS WCQSTEASPK RSDGTPFPWN
  51 KIRLPEYVIP VHYDLLIHAN LTTLTFWGTT KVEITASQPT STILHSHHL
 101 QISRATLRKG AGERLSEEPL QVLEHPRQEQ IALLAPEPLL VGLPYTVVIH
 151 YAGNLSETFH GFYKSTYRTK EGELRILAST QFEPTAARMA FPCFDEPAFK
 201 ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV KMSTYLVAFI
 251 ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL EFYEDYFSIP
 301 YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS ASSKLGITMT
 351 VAHELAHQWF GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT HPELKVGDYF
 401 FGKCFDAMEV DALNSSHPVS TPVENPAQIR EMFDDVSYDK GACILNMLRE
 451 YLSADAFKSG IVQYLQKHSY KNTKNEDLWD SMASICPTDG VKGMDGFCSR
 501 SQHSSSSSHW HQEGVDVKTM MNTWTLQKGF PLITITVRGR NVHMKQEHYM
 551 KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI LPEEVEWIKF
 601 NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI NNAFQLVSIG
 651 KLSIEKALDL SLYLKHETEI MPVFQGLNEL IPMYKLMEKR DMNEVETQFK
 701 AFLIRLLRDL IDKQTWTDEG SVSERMLRSQ LLLLACVHNY QPCVQRAEGY
 751 FRKWKESNGN LSLPVDVTLA VFAVGAQSTE GWDFLYSKYQ FSLSSTEKSQ
 801 IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG RNPVGYPLAW
 851 QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE VKGFFSSLKE
 901 NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLER MAENLYFQSH
 951 HHHHHHHHHD  YKDDDDK
```

Sequence matching of the MS-MS results of the ERAP1-c200 tryptic digest.



24/07/2017 Mascot Search Results: ERAP1A-c200

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 45%

Matched peptides shown in ***bold red***.

```
   1 MVFLPLKWSL ATMSFLLSSL LALLTVSTPS WCQSTEASPK RSDGTPFPWN
  51 KIRLPEYVIP VHYDLLIHAN LTTLTFWGTT KVEITASQPT STILHSHHL
 101 QISRATLRKG AGERLSEEPL QVLEHPRQEQ IALLAPEPLL VGLPYTVVIH
 151 YAGNLSETFH GFYKSTYRTK EGELRILAST QFEPTAARMA FPCFDEPAFK
 201 ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV KMSTYLVAFI
 251 ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL EFYEDYFSIP
 301 YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS ASSKLGITMT
 351 VAHELAHQWF GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT HPELKVGDYF
 401 FGKCFDAMEV DALNSSHPVS TPVENPAQIR EMFDDVSYDK GACILNMLRE
 451 YLSADAFKSG IVQYLQKHSY KNTKNEDLWD SMASICPTDG VKGMDGFCSR
 501 SQHSSSSSHW HQEGVDVKTM MNTWTLQKGF PLITITVRGR NVHMKQEHYM
 551 KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI LPEEVEWIKF
 601 NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI NNAFQLVSIG
 651 KLSIEKALDL SLYLKHETEI MPVFQGLNEL IPMYKLMEKR DMNEVETQFK
 701 AFLIRLLRDL IDKQTWTDEG SVSERMLRSQ LLLLACVHNY QPCVQRAEGY
 751 FRKWKESNGN LSLPVDVTLA VFAVGAQSTE GWDFLYSKYQ FSLSSTEKSQ
 801 IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG RNPVGYPLAW
 851 QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE VKGFFSSLKE
 901 NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLER MAENLYFQSH
 951 HHHHHHHHHD  YKDDDDK
```

Sequence matching of the MS-MS results of the deglycosylated ERAP1-c200 tryptic digest.

24/07/2017 Mascot Search Results: ERAP1A-c201

ERAP1A-c201

Database: Constructs
Score: 262
Nominal mass (M_r): 108578
Calculated pI: 6.07

Sequence similarity is available as an NCBI BLAST search of ERAP1A-c201 against nr.

Search parameters

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 31%

Matched peptides shown in ***bold red***.

```
   1 MVSAIVLYVL LAAAAHSAFA AAMGHHHHHH SSGVDLGTEN LYFQSMASPK
  51 RSDGTPFPWN KIRLPEYVIP VHYDLLIHAN LTTLTFWGTT KVEITASQPT
 101 STIILHSHHL QISRATLRKG AGERLSEEPL QVLEHPRQEQ IALLAPEPLL
 151 VGLPYTVVIH YAGNLSETFH GFYKSTYRTK EGELRILAST QFEPTAARMA
 201 FPCFDEPAFK ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV
 251 KMSTYLVAFI ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL
 301 EFYEDYFSIP YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS
 351 ASSKLGITMT VAHELAHQWF GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT
 401 HPELKVGDYF FGKCFDAMEV DALNSSHPVS TPVENPAQIR EMFDDVSYDK
 451 GACILNMLRE YLSADAFKSG IVQYLQKHSY KNTKNEDLWD SMASICPTDG
 501 VKGMDGFCSR SQHSSSSSHW HQEGVDVKTM MNTWTLQKGF PLITITVRGR
 551 NVHMKQEHYM KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI
 601 LPEEVEWIKF NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI
 651 NNAFQLVSIG KLSIEKALDL SLYLKHETEI MPVFQGLNEL IPMYKLMEKR
 701 DMNEVETQFK AFLIRLLRDL IDKQTWTDEG SVSERMLRSQ LLLLACVHNY
 751 QPCVQRAEGY FRKWKESNGN LSLPVDVTLA VFAVGAQSTE GWDFLYSKYQ
 801 FSLSSTEKSQ IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG
 851 RNPVGYPLAW QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE
 901 VKGFFSSLKE NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLER
 951 M
```

Sequence matching of the MS-MS results of the ERAP1-c201 tryptic digest.

24/07/2017 Mascot Search Results: LTA4HA-c001

LTA4HA-c001

Database: Constructs
Score: 666
Nominal mass (M_r): 72420
Calculated pI: 5.89

Sequence similarity is available as an NCBI BLAST search of LTA4HA-c001 against nr.

Search parameters

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 48%

Matched peptides shown in ***bold red***.

```
   1 MHHHHHHSSG VDLGTENLYF QSMPEIVDTC SLASPASVCR TKHLHLRCSV
  51 DFTRRTLTGT AALTVQSQED NLRSLVLDTK DLTIEKVVIN GQEVKYALGE
 101 RQSYKGSPME ISLPIALSKN QEIVIEISFE TSPKSSALQW LTPEQTSGKE
 151 HPYLFSQCQA IHCRAILPCQ DTPSVKLTYT AEVSVPKELV ALMSAIRDGE
 201 TPDPEDPSRK IYKFIQKVPI PCYLIALVVG ALESRQIGPR TLVWSEKEQV
 251 EKSAYEFSET ESMLKIAEDL GGPYVWGQYD LLVLPPSFPY GGMENPCLTF
 301 VTPTLLAGDK SLSNVIAHEI SHSWTGNLVT NKTWDHFWLN EGHTVYLERH
 351 ICGRLFGEKF RHFNALGGWG ELQNSVKTFG ETHPFTKLVV DLTDIDPDVA
 401 YSSVPYEKGF ALLFYLEQLL GGPEIFLGFL KAYVEKFSYK SITTDDWKDF
 451 LYSYFKDKVD VLNQVDWNAW LYSPGLPPIK PNYDMTLTNA CIALSQRWIT
 501 AKEDDLNSFN ATDLKDLSSH QLNEFLAQTL QRAPLPLGHI KRMQEVYNFN
 551 AINNSEIRFR WLRLCIQSKW EDAIPLALKM ATEQGRMKFT RPLFKDLAAF
 601 DKSHDQAVRT YQEHKASMHP VTAMLVGKDL KVD
```

Sequence matching of the MS-MS results of the LTA4H-c001 tryptic digest.

3. Appendix III

Mascot Search Results

Protein View

Match to: **ERAP1A-c200** Score: **910**
ERAP1A-c200

Nominal mass (M_r): **111054**; Calculated pI value: **6.05**
NCBI BLAST search of ERAP1A-c200 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **49%**

Matched peptides shown in Bold Red

```
     1 MVFLPLKWSL ATMSFLLSSL LALLTVSTPS WCQSTEASPK RSDGTPFPWN
    51 KIRLPEYVIP VHYDLLIHAN LTTLTFWGTT KVEITASQPT STIILHSHHL
   101 QISRATLRKG AGERLSEEPL QVLEHPRQEQ IALLAPEPLL VGLPYTVVIH
   151 YAGNLSETFH GFYKSTYRTK EGELRILAST QFEPTAARMA FPCFDEPAFK
   201 ASFSIKIRRE PRHLAISNMP LVKSVTVAEG LIEDHFDVTV KMSTYLVAFI
   251 ISDFESVSKI TKSGVKVSVY AVPDKINQAD YALDAAVTLL EFYEDYFSIP
   301 YPLPKQDLAA IPDFQSGAME NWGLTTYRES ALLFDAEKSS ASSKLGITMT
   351 VAHELAHQWF GNLVTMEWWN DLWLNEGFAK FMEFVSVSVT HPELKVGDYF
   401 FGKCFDAMEV DALNSSHPVS TPVENPAQIR EMFDDVSYDK GACILNMLRE
   451 YLSADAFKSG IVQYLQKHSY KNTKNEDLWD SMASICPTDG VKGMDGFCSR
   501 SQHSSSSSHW HQEGVDVKTM MNTWTLQKGF PLITITVRGR NVHMKQEHYM
   551 KGSDGAPDTG YLWHVPLTFI TSKSDMVHRF LLKTKTDVLI LPEEVEWIKF
   601 NVGMNGYYIV HYEDDGWDSL TGLLKGTHTA VSSNDRASLI NNAFQLVSIG
   651 KLSIEKALDL SLYLKHETEI MPVFQGLNEL IPMYKLMEKR DMNEVETQFK
   701 AFLIRLLRDL IDKQTWTDEG SVSERMLRSQ LLLLACVHNY QPCVQRAEGY
   751 FRKWKESNGN LSLPVDVTLA VFAVGAQSTE GWDFLYSKYQ FSLSSTEKSQ
   801 IEFALCRTQN KEKLQWLLDE SFKGDKIKTQ EFPQILTLIG RNPVGYPLAW
   851 QFLRKNWNKL VQKFELGSSS IAHMVMGTTN QFSTRTRLEE VKGFFSSLKE
   901 NGSQLRCVQQ TIETIEENIG WMDKNFDKIR VWLQSEKLER MAENLYFQSH
   951 HHHHHHHHHD YKDDDDK
```

Mascot Search Results

Protein View

Match to: **ERAP1A-c010** Score: **836**
ERAP1A-c010

Nominal mass (M_r): **106437**; Calculated pI value: **6.05**
NCBI BLAST search of ERAP1A-c010 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **47%**

Matched peptides shown in Bold Red

```
     1 MGHHHHHHSS GVDLGTENLY FQSMASPKRS DGTPFPWNKI RLPEYVIPVH
    51 YDLLIHANLT TLTFWGTTKV EITASQPTST IILHSHHLQI SRATLRKGAG
   101 ERLSEEPLQV LEHPRQEQIA LLAPEPLLVG LPYTVVIHYA GNLSETFHGF
   151 YKSTYRTKEG ELRILASTQF EPTAARMAFP CFDEPAFKAS FSIKIRREPR
   201 HLAISNMPLV KSVTVAEGLI EDHFDVTVKM STYLVAFIIS DFESVSKITK
   251 SGVKVSVYAV PDKINQADYA LDAAVTLLEF YEDYFSIPYP LPKQDLAAIP
   301 DFQSGAMENW GLTTYRESAL LFDAEKSSAS SKLGITMTVA HELAHQWFGN
   351 LVTMEWWNDL WLNEGFAKFM EFVSVSVTHP ELKVGDYFFG KCFDAMEVDA
   401 LNSSHPVSTP VENPAQIREM FDDVSYDKGA CILNMLREYL SADAFKSGIV
   451 QYLQKHSYKN TKNEDLWDSM ASICPTDGVK GMDGFCSRSQ HSSSSSHWHQ
   501 EGVDVKTMMN TWTLQKGFPL ITITVRGRNV HMKQEHYMKG SDGAPDTGYL
   551 WHVPLTFITS KSDMVHRFLL KTKTDVLILP EEVEWIKFNV GMNGYYIVHY
   601 EDDGWDSLTG LLKGTHTAVS SNDRASLINN AFQLVSIGKL SIEKALDLSL
   651 YLKHETEIMP VFQGLNELIP MYKLMEKRDM NEVETQFKAF LIRLLRDLID
   701 KQTWTDEGSV SERMLRSQLL LLACVHNYQP CVQRAEGYFR KWKESNGNLS
   751 LPVDVTLAVF AVGAQSTEGW DFLYSKYQFS LSSTEKSQIE FALCRTQNKE
   801 KLQWLLDESF KGDKIKTQEF PQILTLIGRN PVGYPLAWQF LRKNWNKLVQ
   851 KFELGSSSIA HMVMGTTNQF STRTRLEEVK GFFSSLKENG SQLRCVQQTI
   901 ETIEENIGWM DKNFDKIRVW LQSEKLERM
```

Intact (top) and deglycosylated (bottom) ERAP1-c200 tryptic digest product MS-MS results. N-linked glycosylation was detected on N^{414}. CFDAMEVDALNSSHPVSTPVENPAQIR peptide present in the deglycosylated sample but not in the intact sample.

4. Appendix IV

Data collection and refinement statistics.

	X042	**X043**	**X044**
Ligand	Ext_38	Ext_43	Ext_44
Data collection			
beamline	DLS I03	DLS I03	DLS I03
Wavelength (Å)	0.97625	0.97625	0.97625
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a,b,c (Å)	77.93 86.90 98.72	77.74 86.83 98.64	77.89 86.88 98.71
α,β,γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	29.01– 1.51 (1.55-1.51)	28.96– 1.68 (1.72-1.68)	29.00– 1.68 (1.72-1.68)
Completeness (%)	99.6 (98.0)	99.8 (98.0)	99.8 (98.0)
Multiplicity	7.3 (7.4)	7.3 (6.9)	7.3 (7.0)
R_{sym}	0.066 (0.596)	0.107 (0.716)	0.107 (0.716)
Mean I/σ(I)	16.3 (2.9)	12.5 (2.3)	16.2 (2.6)
Refinement			
R_{cryst}	0.215	0.210	0.219
R_{free}	0.246	0.255	0.254
R.m.s.d.			
Bond length (Å)	0.011	0.015	0.013
Bond angle (°)	1.468	1.669	1.600
No. of atoms			
Protein	4846	4846	4846
ion	1	1	1
Water	648	550	406
ligand	14	15	16
$<B_{factor}>$ (Å^2)			
Protein	18.7	20.9	19.4
ion	12.62		13.5
Water	30.1	30.0	30.5
ligand	19.3	19.9	20.4

Numbers in parentheses refer to the highest resolution shell.
R_{sym} is the unweighted R-value on I between symmetry mates.
$R_{cryst} = \Sigma hkl \mid |F_{obs}(hkl)| - k|F_{calc}(hkl)| \mid / \Sigma hkl |F_{obs}(hkl)|$ for the working set of reflections; R_{free} is the R-value for 5% of the reflections excluded from refinement.

	X045	**X046**
Ligand	Ext_45	Ext_47
Data collection		
beamline	DLS I03	DLS I03
Wavelength (Å)	0.97625	0.97625
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a,b,c (Å)	98.46 77.77 86.82	77.85 86.86 98.54
α,β,γ (°)	90.00 90.00 90.00	90.00 90.00 90.00
Resolution (Å)	98.46-1.75 (1.84-1.75)	28.99-1.54 (1.58-1.54)
Completeness (%)	99.5 (98.9)	99.9 (98.5)
Multiplicity	7.2 (7.2)	7.3 (7.2)
R_{sym}	0.169 (0.964)	0.076 (0.682)
Mean I/σ(I)	8.7 (2.2)	15.7 (2.5)
Refinement		
R_{cryst}	0.235	0.209
R_{free}	0.277	0.239
R.m.s.d.		
Bond length (Å)	0.021	0.011
Bond angle (°)	1.957	1.461
No. of atoms		
Protein	4846	4857
ion	1	1
Water	406	592
ligand	15	15
$<B_{factor}>$ (Å^2)		
Protein	21.4	18.7
ion	14.72	12.01
Water	29.4	30.0
ligand	20.8	25.9

Numbers in parentheses refer to the highest resolution shell.
R_{sym} is the unweighted R-value on I between symmetry mates.
$R_{cryst} = \Sigma hkl \mid |F_{obs}(hkl)| - k|F_{calc}(hkl)| \mid / \Sigma hkl |F_{obs}(hkl)|$ for the working set of reflections; R_{free} is the R-value for 5% of the reflections excluded from refinement.