University of Southampton

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

In Vitro Characterisation of ERAP1 Mechanism and Antigen Processing

Mary E. Beton

May 2018

Abstract

The aminopeptidase ERAP1 is crucial for processing peptides for presentation by MHC class I, starting from sub-optimal N-terminally extended peptides. MHC I and ERAP1 both have genetic associations with autoimmune diseases and the reasons are poorly understood. Both proteins have peptide length and sequence preferences, and which protein ultimately determines the antigenic peptide repertoire is unknown. Presented here are studies through NMR and X-ray crystallography which investigate the importance of individual interactions between MHC I and the peptide, and how NMR can be used to probe the mechanism of ERAP1.

Crystal structures of four single chain constructs based on MHC I allele H-2K^{b} with A pocket mutations were solved. The mutations caused no difference to structure or peptide binding with respect to the unmutated form, and rearrangement of neighbouring side chains and water molecules in the A pocket were apparent in response to the removal of hydrogen bonding side chains. These structures enabled collaborators to perform assays and computational analysis which showed impaired binding of the peptide in the mutants permitted peptide trimming by ERAP1.

A peptide library was devised and characterised with proton chemical shifts assignments, and the peptides were expressed in *E. coli* with $^{13}C/^{15}N$ isotope labelling. HSQC spectra of an isotope-labelled peptide confirmed the proton assignments and offers potential for future experimentation with ERAP1 and MHC I.

From the peptide assignments, proton NMR reaction monitoring was used to comprehensively assay three ERAP1 variants to uncover differences in peptide specificity. This method offers scope to examine multiple steps in an aminopeptidase reaction which commonly used ERAP1 assays do not provide. A series of 1D proton NMR spectra obtained throughout the reaction were found to be an effective way of interrogating ERAP1 aminopeptidase activity. Wild type and 5SNP variant ERAP1 appeared to show different substrate length preferences. ERAP1-peptide binding was analysed with saturation transfer difference NMR and indicated the primary points of interaction were the peptide C-terminal and hydrophobic side chains. This matches findings on ERAP1 peptide preferences by other authors, but the interactions between peptide and ERAP1 have never been directly investigated before and have enabled tentative modelling of ERAP1-peptide binding.

Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration of Authorship

I, Mary Beton declare that this thesis entitled:

'In Vitro Characterisation of ERAP1 Mechanism and Antigen Processing'

and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Papakyriakou, Athanasios, Emma Reeves, Mary Beton, Halina Mikolajek, Leon Douglas, Grace Cooper, Tim Elliott, Jörn M. Werner, and Edward James. 2018. "The Partial Dissociation of MHC Class I– Bound Peptides Exposes Their N Terminus to Trimming by Endoplasmic Reticulum Aminopeptidase 1." *Journal of Biological Chemistry* 293 (20): 7538–48.

Signed: .....

Date: .....

Acknowledgements

First and foremost I must thank my supervisor, Dr Joern Werner, for his support and supervision throughout this project. I would not have been able to complete it without his encouragement and enthusiasm. In addition I would like to thank my co-supervisor Dr Tim Elliott for his knowledge and advice.

I would like to thank Dr Stuart Findlow for his early assistance with NMR, and Neville Wright for use of his extensive expertise, maintenance of the HPLC system and carrying out mass spectrometry for me. Thanks to Dr Patrick Duriez for his help and watchful guidance of the insect cell expression, and to Dr Emma Reeves for advice and editing. My thanks also to Dr Phil Williamson, for permitting us to use his spectrometer when our own machine broke.

Especial thanks go to Dr Halina Mikolajek, who so ably guided me through X-ray crystallography, and has supported me through any and every aspect of my PhD, even beyond the end of her time at the University of Southampton, and who constantly inspired with her artwork.

Thanks to the structural biology PhD students, who have made the lab experience more enjoyable, and to the cake club, whose solemn reverence towards baked things must surely be applauded. Thanks also to Tom and Mathilde, who, though separated by distance, were always near me.

I would like to extend enormous gratitude to my parents, who have sheltered me and supported me through difficulties, for without them it would have been impossible. And to my brother David for being alright too.

To the God beyond all understanding, whose work is laid out for us to marvel at and explore.

List of Abbreviations

AMC – 7-amino-4-methylcoumarin
AS – ankylosing spondylitis
at – acquisition time
CTL – cytotoxic T lymphocyte
dtSCT – disulphide-trap single chain trimer
ER – endoplasmic reticulum
ERAP1 – endoplasmic reticulum aminopeptidase 1
ERAP2 – endoplasmic reticulum aminopeptidase 2
HC – heavy chain
HLA – human leukocyte antigen
HSQC – heteronuclear single quantum coherence/correlation
IRAP – insulin regulated aminopeptidase
L-AMC – L-leucine-7-amido-4-methylcoumarin
MD – molecular dynamics
MHC I – class I major histocompatibility complex
NMR – nuclear magnetic resonance
NOESY – nuclear Overhauser effect spectroscopy
np – number of points
PLC – peptide loading complex
pMHC – peptide-MHC
RMSD – root-mean-square deviation
SCT – single chain trimer
SL8 – SIINFEKL peptide
SNP – single nucleotide polymorphism
STD NMR - saturation transfer difference nuclear magnetic resonance
sw – sweep/spectral width
TAP – transporter associated with antigen processing
TCR – T-cell receptor
TOCSY – total correlation spectroscopy
TOPS - TCR/pMHC Optimized Protein crystallization Screen
TSP – 3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt
wt – wild-type
β_2m – beta-2-microglobulin

1 Introduction

1.1 Antigen Processing

Peptides are routinely created within the cell as part of the process of proteolytic degradation. The majority of peptides created are 'self' peptides and are short fragments from unwanted proteins targeted for destruction. In cases of infection, however, the presence of unusual foreign proteins within the cell must be used to alert the immune system. Destruction of virally infected or abnormal/cancerous cells is the role of cytotoxic T lymphocytes (CTLs), which must be alerted to aberrant cell properties by presentation of foreign and unrecognised peptides on the outside surface of the cell (Young, Nathenson, and Sacchettini 1995).

Antigenic peptides within the cell are presented by the class I major histocompatibility complex (MHC class I, or MHC I) to the T cell receptor (TCR) of CD8+ CTLs (Young, Nathenson, and Sacchettini 1995). It is crucial that this process, classical antigen presentation, works well for CTL recognition of foreign, 'non-self', peptides, as well as those arising from over-expressed proteins in cancerous cells, and for 'self' peptides to be recognised as harmless lest they trigger autoimmune responses.

1.2 An Overview of MHC class I

1.2.1 Structure

MHC I consists of the light chain, beta-2-microglobulin (β_2m), and the heavy chain (HC), as well as the antigenic peptide without which the complex will not be stable (Young, Nathenson, and Sacchettini 1995).

Human heavy chains are human leukocyte antigen (HLA) molecules and consist of an α3 domain with the membrane-spanning region of the protein. Next to the α3 domain is β_2m, and above them are the α1 and α2 domains which form the two jaws around the peptide binding groove (Figure 1). Interestingly, there are over 8000 HC allotypes of the three HLA-A, HLA-B and HLA-C alleles, giving potential for very diverse peptide repertoires (Rossjohn et al. 2015). The highly polymorphic HCs are glycoproteins which exhibit huge flexibility in the sequences of peptides they can bind, but show length and sequence preferences. Optimal peptides bound to MHC I often have a low off-rate, and thus increase the chance of successful presentation to CTLs through stable cell surface expression. A significant feature of peptides that display low off rates is peptide length, as peptides which are too long or too short will not create a stable pMHC I complex. Ideal peptides are 8-10 amino acids long, depending on the heavy chain variant in question (Cresswell et al. 2005).

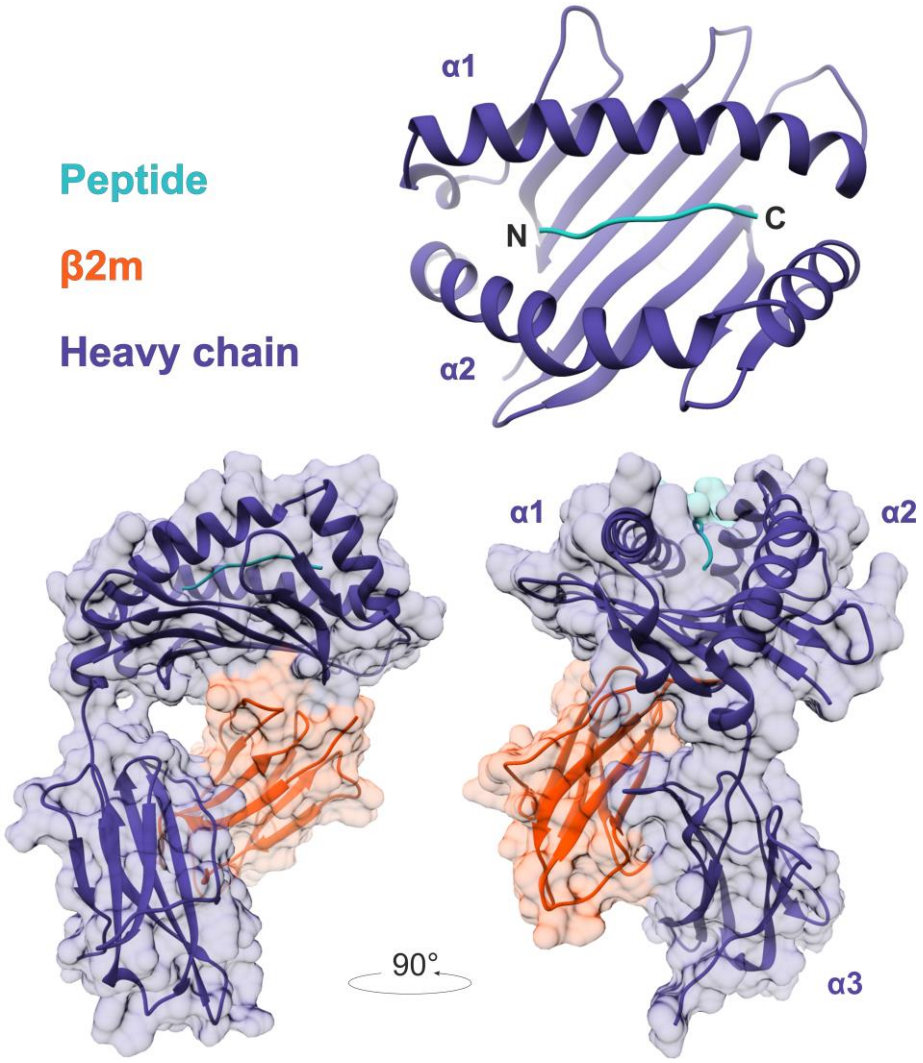


Figure 1 - MHC class I backbone structure ribbon cartoon, showing peptide binding domain secondary structure (top) and side views with transparent surface representation coloured by domain. Adapted from PDB ID: 3P9L (Denton et al. 2011).

1.2.2 Peptide binding

Peptides of optimal length (8-10 amino acids) bind within the binding groove formed between the HC α1 and α2 domains. Explanations for the specificity and versatility of peptide bindings can be found in the structure of the binding groove, with useful insights from those residues which are, or are not, conserved. There are six identified binding pockets named A to F in the binding groove of MHC I (Figure 2). The peptide N-terminal region is bound within the A pocket, and the C-terminal region binds within the F pocket. A feature of note is that the

structures of pockets A and F are well conserved throughout MHC I molecules, but B to E display much more variation (Matsumura et al. 1992). This observation is also seen when comparing murine and human MHC I alleles. Matsumura et al. (1992) described especial similarity between the α1 and α2 domains of H-2K^b and HLA-A2 over the rest of the protein, with these similar binding regions having an RMSD value of only 0.66 Å. Involved in the hydrogen bonding of peptide N- and C-termini are 8 invariant residues: Y7, Y59, Y159 and Y171 at the N-terminal, and Y84, T143, K146 and W147 at the C-terminal (Mitaksov and Fremont 2006).

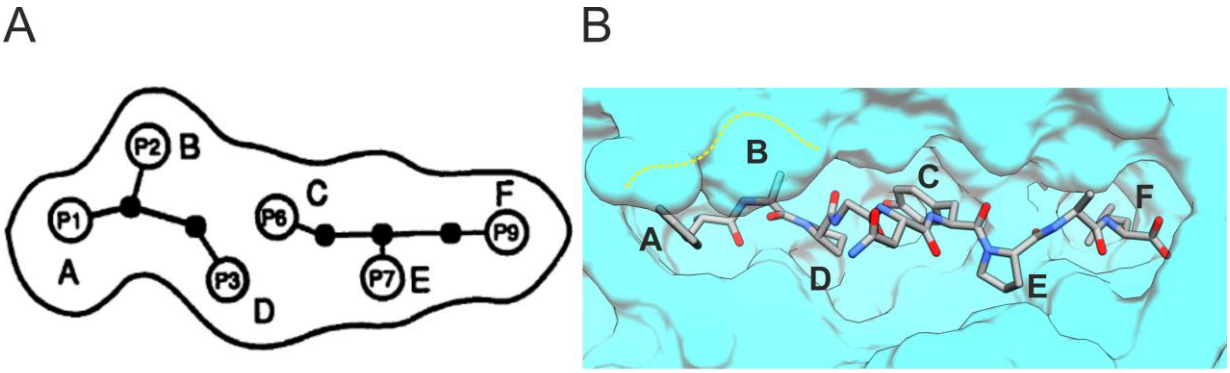


Figure 2 – H-2K^b peptide binding groove A) Diagram of pockets A to F within the H-2K^b binding groove with SEV-9 peptide backbone Cα positions (closed circles) and side chain positions (numbered open circles) shown. Taken from Matsumura et al. (1992). *B) SEV-9 peptide shown within transparent surface representation of H-2K^b binding groove with pocket positions labelled. Structure from* Matsumura et al. (1992)*, PDB ID: 2VAB.*

The structure of the murine MHC I HC H-2K^b has been extensively studied. Several structures exist of H-2K^b with peptides, including to the viral peptides RGYVYQGL (VSV8), and FAPGNYPAL (SEV9) (Fremont et al. 1992) and the ovalbumin-derived peptide SIINFEKL (SL8) (Denton et al. 2011; Fremont et al. 1995). Thus, commonality may be observed between these similar structures with different peptides bound. Within the H-2K^b binding groove, the N-terminal region of the peptide is described as forming hydrogen bonds to the conserved tyrosines at positions 7, 159 and 171, and also to the less-conserved E63, K66 and N70 positions (Matsumura et al. 1992). Interestingly the terminal amine

group is the hydrogen-bonding part of this pocket-peptide interaction, and the side chain is directed upwards away from the binding groove and towards the solvent. This allows the N-terminal residue to be non-specific. By contrast, the F pocket has the carboxy-terminal pointing towards the top of the pocket, and the side chain pointing downwards. This imposes steric preference, although still permits long residues of varying hydrophobicity and has preference flexibility (Matsumura et al. 1992).

The thermal stability produced by these A and F pocket interactions with the peptide termini was examined by Bouvier et al. (1994) by substituting methyl groups in place of the hydrogen bonding groups at each terminus of an antigenic peptide, GILGFVFTL, from influenza virus. By this, Bouvier et al., demonstrated that hydrogen bonding of either the N- or C-terminus alone with the MHC I binding sites was adequate for formation of a stable complex, although less stable than with both ends forming hydrogen bonds. This is unsurprising given that very short or extended peptides of a non-ideal length can still bind to the class I complex in some cases (Bouvier and Wiley 1994).

Much of the peptide preferences of various class I alleles is, therefore, a consequence of pockets B, C, D and E, and peptide-binding residues that are less conserved between alleles than those in the A and F pockets (Bouvier and Wiley 1994; Elliott and Williams 2005). These non-conserved pocket residues allow affinity for different peptide side chains (Bouvier and Wiley 1994). The natures of the six different pockets has permitted MHC I to possess peptide sequence affinities that lie between specific and indifferent, varying from one residue to the next. Therefore, they can form stable complexes with peptides of preferred and optimal sequences whilst still retaining the ability to bind a broad range of antigenic peptides.

1.3 Origin of peptide antigens

The initial stages of antigen processing occur in the cytosol when the peptides are first generated in their crudest form through proteolytic degradation. Most proteolytic cleavage in the cytosol is carried out by the 26S proteasome (Blum, Wearsch, and Cresswell 2013), a large tubular structure containing a 20S core capped with 19S complexes at each end which act as regulatory domains. These 19S subunits enact ATP-dependent unfolding of proteins and recognise ubiquitin-conjugated targets. Sandwiched between them is the 20S core of this protein, comprised of 4 rings (7 subunits each) which make a cylindrical structure. This has α-subunits at the ends and central β-subunits. Three of the β-subunits, β1, β2 and β5, have proteolytic activity. Stimulation with IFN-γ causes replacement of these three subunits with their counterparts, β1i, β2i and β5i (also known as PSMB9/LMP2, PSMB10/LMP10 and PSMB8/LMP7) leading to the formation of the immunoproteasome (Blum, Wearsch, and Cresswell 2013). The presence of the immunoproteasome, rather than the proteasome, may improve response to pathogens by CD8+ T cells by increasing epitope generation efficiency (Sijts and Kloetzel 2011).

Known MHC I-presented peptide antigens can be generated by either the proteasome or immnoproteasome, but intriguingly the immunoproteasome generates longer peptide products in comparison to those from the proteasome (Sijts and Kloetzel 2011; Cascio et al. 2001). Cascio et al. revealed that the final SIINFEKL peptide was extended by 1-7 amino acids at the N-terminus after degradation of ovalbumin by the immunoproteasome. By comparison, the production of N-terminally extended SIINFEKL was up to 4 times lower when ovalbumin was degraded by the proteasome (Cascio et al. 2001). The result of this proteolytic degradation, whether by the proteasome or immunoproteasome, is a range of N-terminally extended peptides which require further processing to stably bind to MHC I and be presented at the cell surface. Finessing of this peptide selection and peptide processing occurs within the endoplasmic reticulum (ER) by the peptide loading complex and related proteins.

1.3.1 Presentation of bound peptides

MHC I exits the ER and egresses to the cell surface via the golgi secretory pathway (Spiliotis et al. 2000). Once at the cell surface it is available for recognition by CD8+ T cells. The heterodimeric T cell receptor, comprised of an α and β chain, is responsible for recognition of the cell-surface pMHC I complexes. A number of TCR-pMHC I complexes have been solved and are uniform in their arrangement, with the TCR sitting diagonally across the binding cleft of MHC I (Rossjohn et al. 2015). The low-variability germline CDR1 and CDR2 loops make contacts to MHC I itself and the variable CDR3 loops make peptide contacts (Garboczi et al. 1996). The TCR Vα-chain is orientated over α2 helix of MHC I, and α2 below the TCR Vβ-chain (Rossjohn et al. 2015). This provides the mechanism for diverse antigen recognition within the same MHC allele. The affinity (~35 μM) is considered low, which is unsurprising as the interface between the two complexes is not a close fit (Rossjohn et al. 2015). In order to generate stable pMHC I complexes for cell surface presentation, a process of folding and selection takes place within the ER.

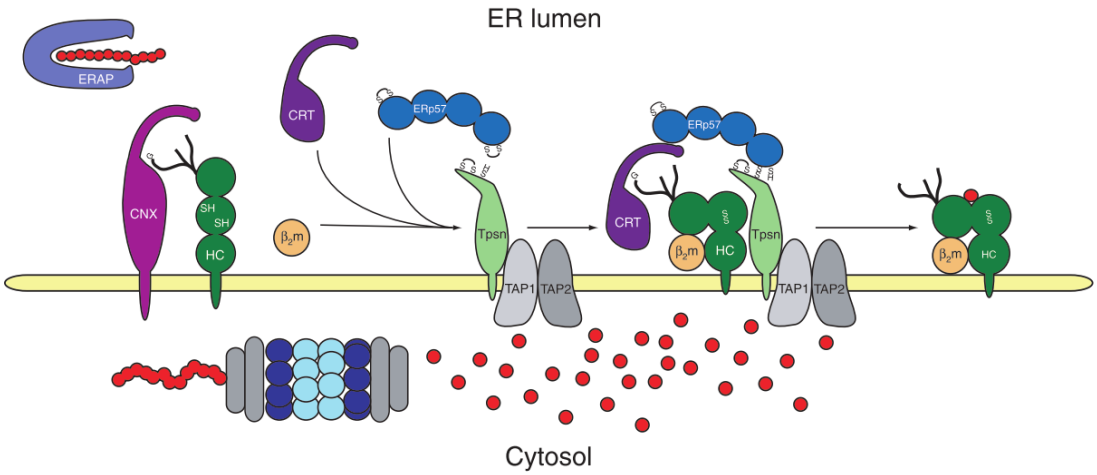


Figure 3 - Antigen processing and the MHC class I assembly pathway. Extended peptides/proteins (red) are degraded by proteasomes in cytosol to smaller peptide fragments which can enter the ER. Abbreviations: 'ERAP'=ERAP1/2, 'CNX'=calnexin, 'HC'=MHC class I heavy chain, 'CRT'=calreticulin & 'Tpsn'=Tapasin. Taken from Cresswell, Ackerman, Giodini, Peaper, & Wearsch, 2005.

1.3.2 Components of the Peptide Loading Complex

The peptide loading complex (PLC) is primarily responsible for assembly of peptide-MHC I complexes in the ER and shows a degree of conservation across higher vertebrates (Hinz et al. 2014). The processes of folding and loading MHC I are linked, and to form a stable interaction between β_2m and the heavy chain, peptide binding is required (Williams, Au Peh, and Elliott 2002). Components of the PLC, shown in Figure 3, are TAP, tapasin ('Tpsn' in Figure 3), ERp57 and calreticulin ('CRT' in Figure 3) as well as the MHC I molecule itself (HC and β_2m). The complex structure has been modelled from electron microscopy data by Blees et al., showing arrangement of two MHC-editing units (Figure 4C) which sit atop TAP in the ER membrane (Blees et al. 2017).

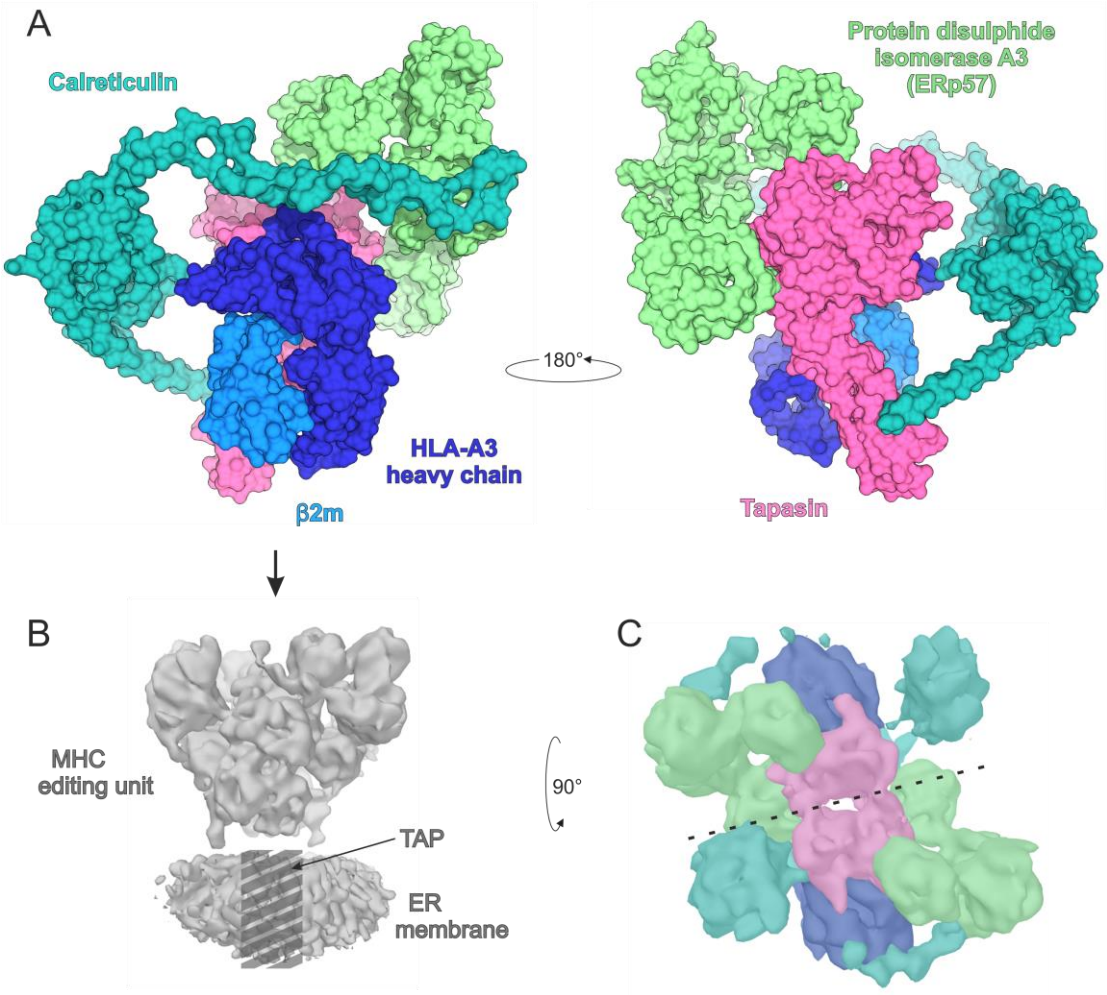


Figure 4 - PLC structure from EM data (PDB ID: 6ENY) by Blees et al. (2017). *A) MHC-editing unit B) EM data (EMDB: 3904) contoured at 0.022 C) Top view of PLC EM data showing rotational symmetry*

1.3.2.1 TAP

Transport of precursor peptides into the ER is executed through the transporter associated with antigen processing (TAP), a heterodimeric complex comprised of TAP1 and TAP2 that spans the ER membrane (Sadasivan et al. 1996). Because TAP is an ATP-Binding Cassette (ABC) transporter, both TAP1 and TAP2 monomers possess nucleotide-binding domains within the cytoplasm which carry ATP molecules (Grossmann et al. 2014). Crucially, TAP possesses a transmembrane domain through which peptides are translocated, and preferentially translocates peptides of 8 to 13 residues, although peptides outside this range can be transported with less efficiency (Parcej and Tampe 2010). TAP also seems partially responsible for some peptide selection, with murine TAP being observed to have preference for aromatic or hydrophobic C-termini (Neisig, Roelse, and Sijts 1995).

1.3.2.2 Tapasin

Tapasin is an important chaperone of stable pMHC I generation. Both TAP1 and TAP2 bind to tapasin, which stabilizes TAP and the PLC around it (Raghuraman, Lapinski, and Raghavan 2002; Leonhardt et al. 2005). These two tapasin molecules of the PLC have a crucial role in optimising peptide selection by MHC I (Grandea et al. 2000; Williams et al. 2002). However the dependency upon tapasin function varies widely between MHC I alleles, with some more dependent on it for peptide editing than others (Hermann, Trowsdale, and Boyle 2015).

1.3.2.3 ERp57

ERp57 is a thiol oxidoreductase chaperone, capable of forming disulphide bonds with tapasin, enabling them function together as a stable heterodimer (Dick et al. 2002). The Cys57 residue usually forms transient interactions with substrate proteins but in the case of tapasin the strength of the disulphide bond allows

longer lasting complex formation. Crystal structures show that tapasin interacts with the *a* (through the disulphide bridge) and *a'* domains of ERp57, blocking the ERp57 active sites and retaining the disulphide bond to tapasin residue Cys95 (Dong et al. 2009). The interaction between ERp57 and tapasin is crucial for generation of optimal and stable pMHC I complexes. Absence of the ERp57-tapasin dimer results in poorly folded and unstable pMHC I, as well as impaired recruitment of MHC I for peptide loading in cells lacking ERp57 expression (Dick et al. 2002); (Wearsch and Cresswell 2007).

1.3.2.4 Calreticulin and calnexin

Calnexin and calreticulin are ER resident molecular chaperones that aid MHC I folding and assembly. Calnexin, unlike calreticulin, is not thought to be part of the PLC, however these two proteins are functionally similar chaperones which collaborate with ERp57 to fold MHC I (Hermann, Trowsdale, and Boyle 2015). Calnexin, a membrane protein, performs the first heavy chain folding, after which the HC associates with β_2m and the rest of the PLC, including calreticulin (Hermann, Trowsdale, and Boyle 2015). Calreticulin is a soluble protein which does not bind to the free heavy chain but acts as a chaperone when MHC I is in the PLC and loading can commence (Williams, Au Peh, and Elliott 2002).

1.4 An Overview of endoplasmic reticulum aminopeptidases

In 2001, length differences were observed between peptides entering the ER through TAP and those peptides presented by MHC I at the cell surface, indicating a role for an unknown aminopeptidase within the ER (Serwold, Gaw, and Shastri 2001). The proteosomes and immunoproteosome produce a low proportion of optimal length peptides suitable for MHC I binding, with many precursors entering the ER having ideal C-terminal residues for MHC I binding, but N-terminally extending amino acids (Cascio et al. 2001). In 2002, the ER resident trimming enzyme was isolated from murine tissues and named ER

Aminopeptidase associated with Antigen Processing (ERAAP) as well as being identified in humans as Endoplasmic Reticulum AminoPeptidase 1 (ERAP1) (Serwold et al. 2002). ERAP1 had already been identified as adipocyte-derived leucine aminopeptidase (A-LAP) (Hattori et al. 1999), puromycin-insensitive leucine aminopeptidase (PILSAP) (Schomburg et al. 2000) and aminopeptidase regulator of TNFR1 shedding (ARTS1) (Cui et al. 2002) depending on its different reported cellular functions (Serwold et al. 2002; Saric et al. 2002).

ERAP1 belongs to the M1 aminopeptidase family, and specifically to the M1 oxytocinase subfamily, and contains both a conserved zinc binding and GAMEN motif within the active site. Other members of this subfamily of aminopeptidases include Insulin Regulated AminoPeptidase (IRAP) and ERAP2, and both have approximately 50% sequence homology to ERAP1 (Stratikos and Stern 2013). The three proteins all exhibit intracellular localisation, ER localisation in the cases of ERAP1 (for its role in antigen processing) and ERAP2, and in vesicles and the cell membrane for the membrane-protein IRAP (Tsujimoto and Hattori 2005). ERAP1/2 have a key role in endogenous antigen processing, whereas IRAP has a role in generating peptides for cross-presentation (Stratikos and Stern 2013). ERAP2 is not present in mice, however the in humans it appears to work in concert with ERAP1 to provide complementary substrate specificities (Lorente et al. 2013; Saveanu et al. 2005). Heterodimers of ERAP1/2, stable enough for purification, have been identified although sadly no crystal structure of these dimers exists and how they dimerise is currently unknown (Evnouchidou et al. 2014). The absence of ERAP2 in mice together with the strong links between ERAP1 and disease (see section 1.4.3) (D. M. Evans et al. 2011) have naturally resulted in deeper investigation of ERAP1.

1.4.1 ERAP1 structure

The structure of human ERAP1, solved first by Kochan et al. in 2011, was followed soon after by Nguyen et al. in the same year (see Figure 5). These crystal structures reveal both closed and open conformations of ERAP1, shown

in Figure 6, which crystallised in space groups P622 and P212121 respectively, as well as ERAP1 in a semi-open form (Figure 5) crystallised in space group P21. The structures show substantial similarity to other M1 aminopeptidase family members. Interestingly, a third structure paper was published in the same year, but presented solely domains III and IV, from residue 529 onwards, and examined the binding of a 'peptide C-terminal' to this region of the protein, enabled through crystal packing (Gandhi et al. 2011).

As with other members of the M1 aminopeptidase family, ERAP1 possesses four domains. Domain I (residues 0-254) is likely to dock the peptide N-terminal and is comprised of a saddle-shaped 8-strand beta sheet (beta 1, 2, 4, 5, 8 and 11, 13, 14) which abuts beta sheets 3, 6 and 7 at one end, and 9, 10, 12 and 15 at the other. This latter region interacts with domain IV by means of the loop between beta strands 9 and 10. Domain I provides a crucial steric block above the catalytic site which is necessary for strict exopeptidase activity (Kochan et al. 2011; Nguyen et al. 2011).

Domain II (residues 255-529) is the catalytic region of the protein. A subdomain at the N-terminal carries several crucial residues: Tyr438, a conserved residue throughout the M1 family, uses its hydroxyl group to stabilize intermediates. Y438F mutations are known to cause 99.5% reduction in enzymatic activity, though in ERAP1 it is positioned further from the other catalytic residues than M1 family members with solved structures (Nguyen et al. 2011). Beta strands 16-20 form a beta sheet which carries the GXMEN (often GAMEN, as in ERAP1) motif conserved through the M1 family. The Asn321 residue of the motif is suggested to stabilize loops in the region and Glu320 is crucial for enzymatic activity and mutants of this residue are thus inactive (Nguyen et al. 2011).

Within domain II a catalytic zinc residue is bound by a conserved H-E-X-X-H-X_{18}-E motif. The floor of the channel around the catalytic site is helix 6, which also has His353 and His357 of the zinc-binding motif. Coordination of the zinc ion is by His353, His357 and Glu376. The nearby Glu354 and Glu320 hydrogen bond to

a water molecule which is utilised for nucleophilic attack on the peptide (Nguyen et al. 2011; Kochan et al. 2011).

Domain III (residues 530-614) is a small region of seven beta strands which form two facing beta sheets. This joins the much larger domain IV which is made up of sixteen alpha helices. Domain IV (residues 615-940) is a regulatory domain which is suggested to interact with the peptide substrate C-terminal (Chang et al. 2005). The substrate C-terminal interacting interior face of this domain consists of the even-numbered helices of the eight antiparallel helix-turn-helix repeats structure (Nguyen et al. 2011). This region contains several common SNPs (see Figure 8), some of which have been identified as having a genetic linkage to Ankylosing Spondylitis (AS). Two examples of SNPs in this domain are R725Q and Q730E, both suggested to confer protective effect with regards to AS but which differ in activity, with Q730E causing little change to activity but altering amino acid specificity, while R725Q is suggested to either decrease activity (García-medel et al. 2012) or increase it in combination with other SNPs (Reeves et al. 2014).

The internal cavity of the closed structure of ERAP1, largely accommodated within the shape of domain IV, is unusually large for an aminopeptidase of this family, possibly because of the length of the peptide substrates. Kochan et al. reported the cavity size as 2920 $Å^3$, larger than that of ePepN or LTA4H (2200 $Å^3$ and 1130 $Å^3$ respectively).

The two published full-length ERAP1 crystal structures reported slightly different N-glycosylation sites. Kochan et al. reported electron density for glycosylation at Asn70, Asn154 and Asn414 (Kochan et al. 2011) whereas Nguyen et al. reported glycosylation at Asn760 rather than Asn414 (Figure 5).

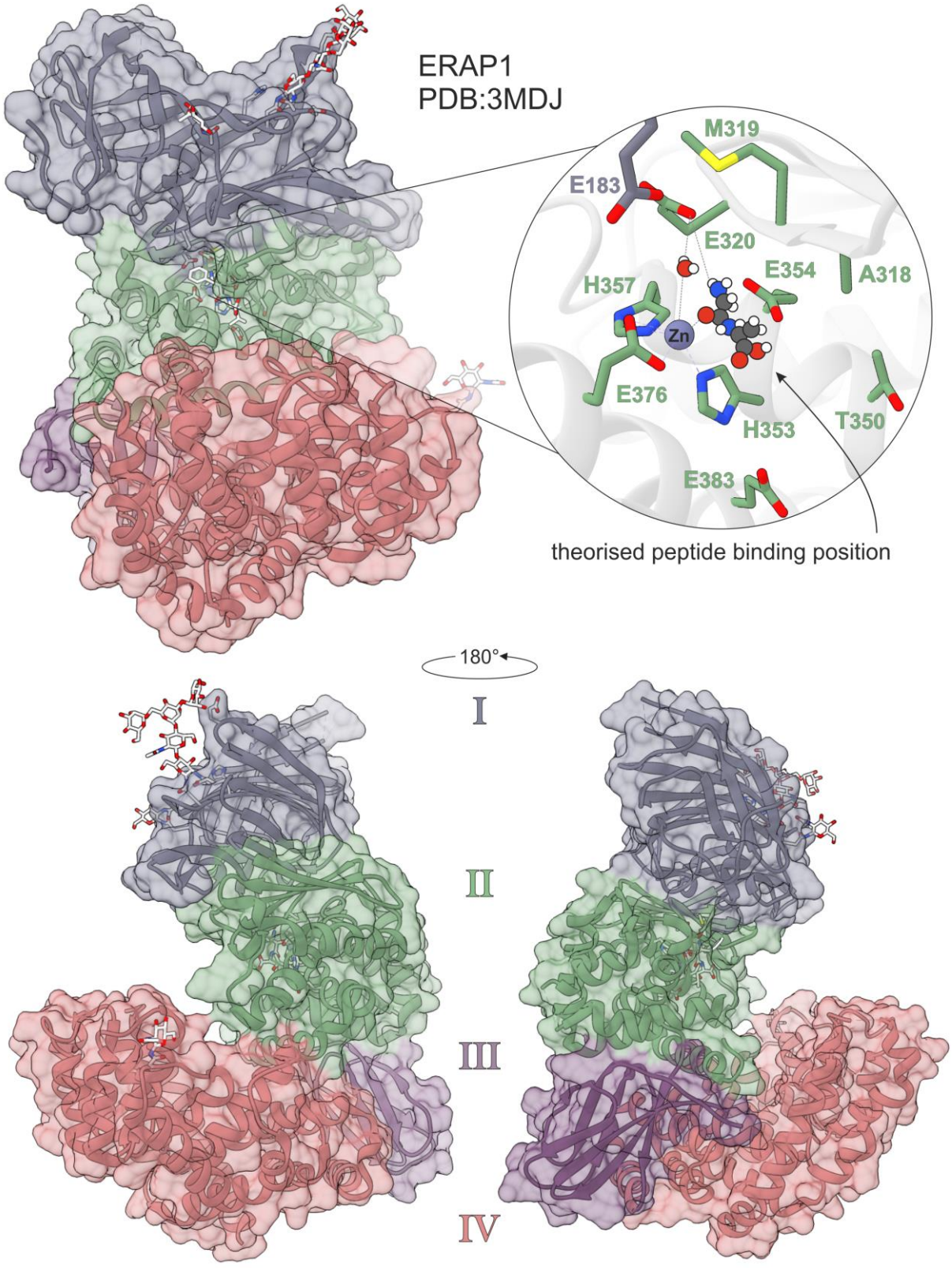


Figure 5 - The structure of human ERAP1 in an open conformation with transparent surface representation over ribbon cartoon. Coloured by domains (denoted by Roman numerals) with ball-and-stick representations of glycosylation sites. Expanded view shows active site residues and zinc atom. PDB ID: 3MDJ structure solved by Nguyen et al. (2011).

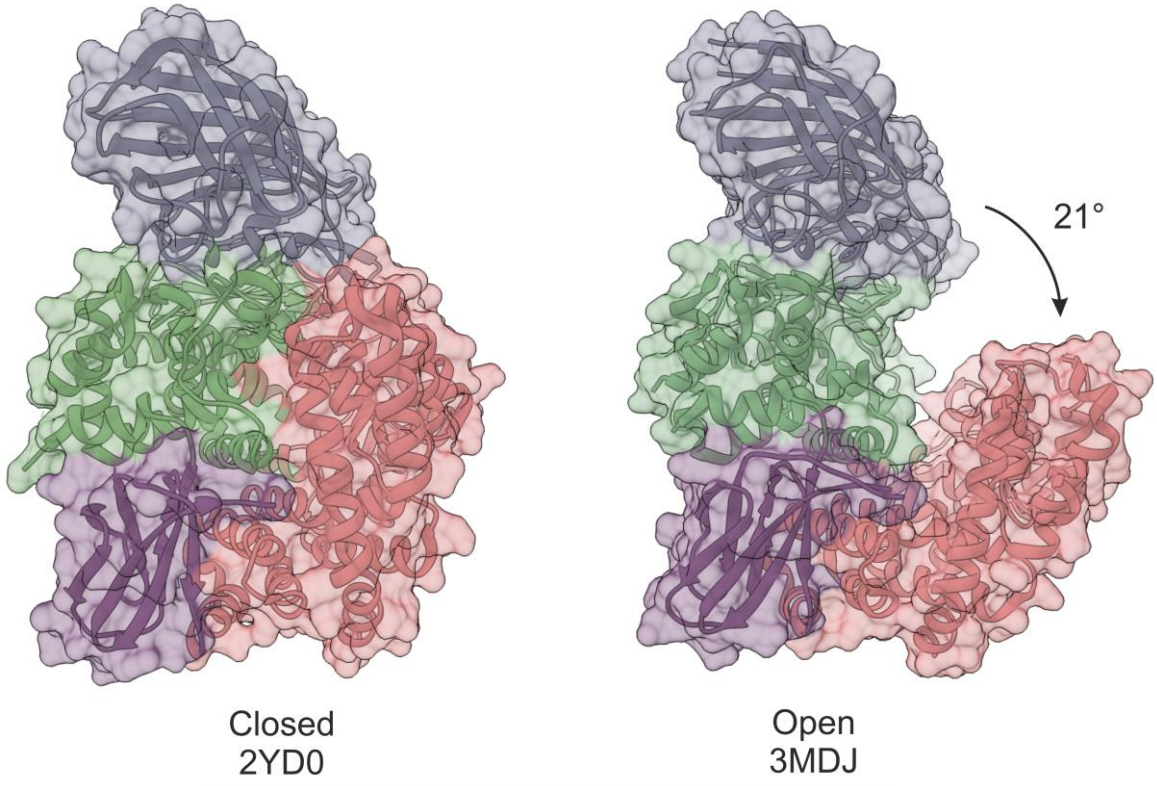


Figure 6 - Closed and open forms of ERAP1 with transparent surface representation over ribbon cartoon. Coloured by domain as in Fig. 5. Structures from Kochan et al. (2011) and Nguyen et al. (2011).

1.4.2 The M1 family and mechanism of action

To date, no substrate:ERAP1 pre-catalysis structures have thus far been published, which is likely a result of the substantial technical challenges of crystallising such a complex. However, ERAP1 has been crystallised with the inhibitor bestatin (more amenable than a cleavable peptide) in the closed structure 2YD0 but not the concomitantly solved open structure 3QNF (shown in Figure 6) (Kochan et al. 2011) and the published structure 3MDJ (Nguyen et al. 2011). These crystal structures suggest the location of peptide binding within the active site. A cartoon of the theorized peptide N-terminal pre-catalysis binding is shown in the magnified active site of Figure 5. Additionally, other members of the M1 family are well studied with multiple M1 aminopeptidase structures published with bound bestatin (Ito et al. 2006; Thunnissen, Nordlund, and Haeggstrom 2001), and other peptide analogues (Mcgowan et al. 2009). Also

published are LTA4H structures with substrate tripeptides in the active site (Tholander et al. 2008) and a crystal structure of Human Aminopeptidase N with the 6-mer angiotensin IV in the active site (Wong, Zhou, and Rini 2012). Collectively, these studies suggest extensive investigation into the structural properties of this family of metalloproteases. The mechanism by which these enzymes work has been convincingly detailed alongside the widespread structural analysis of the active site.

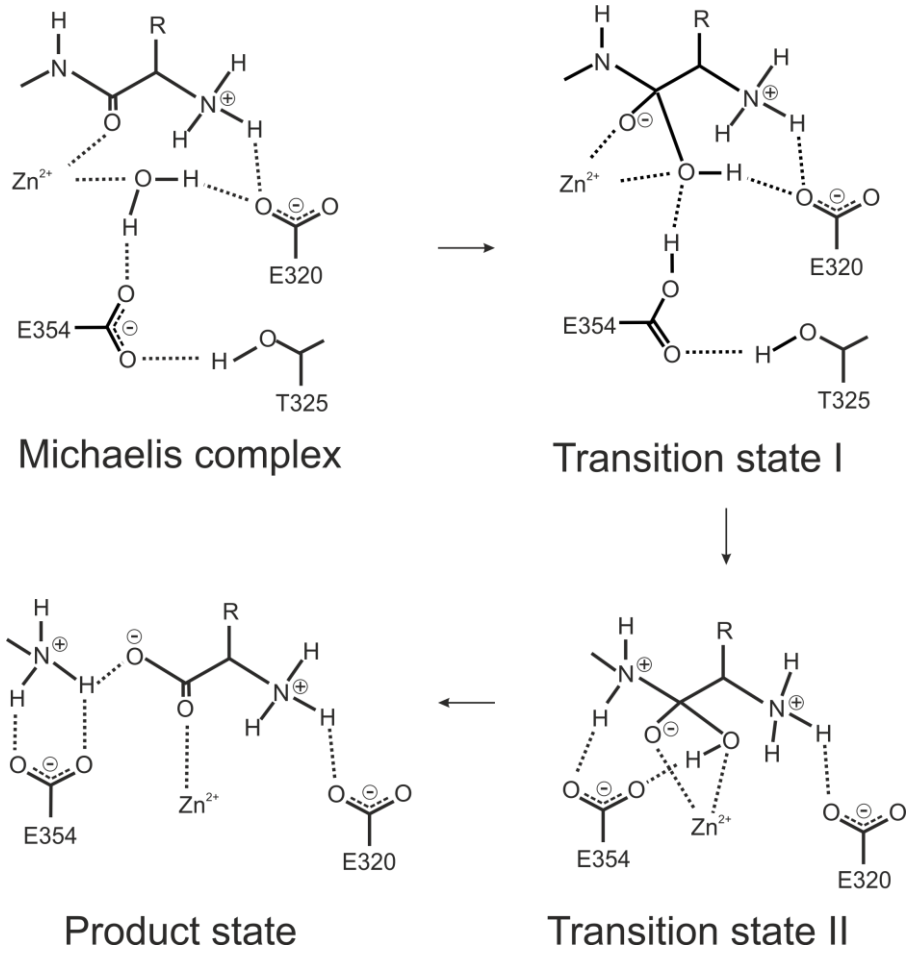


Figure 7 - Probable mechanism of ERAP1 catalytic activity, adapted from Jones et al. (2011).

Only the closed form of ERAP1 is believed to be active, and the protein opens to facilitate the binding or release of the substrate/products (Kochan et al. 2011). The reaction is thought to utilise a water molecule which hydrogen bonds to the side chain of E320 and E354 prior to the reaction beginning. A substrate peptide interacts with the zinc atom by its carbonyl group and E320 with its terminal amine group. The reaction (shown in Figure 7) begins with the zinc interacting with the carbonyl of the peptide N-terminal residue, which polarises. This enables a nucleophilic attack from the water molecule which is facilitated by E354 acting as a base and accepting a hydrogen from the water molecule. The unstable transition state then becomes a new N-terminus and free amino acid (Tholander et al. 2008; Jones et al. 2011). Through this process peptides are cleaved at the N-terminal residue, although whether ERAP1 releases peptides between trimming events is unclear.

1.4.3 ERAP1 single nucleotide polymorphisms and disease association

Many studies have identified the presence of single nucleotide polymorphisms (SNPs) within ERAP1. Genome-wide association studies (GWAS) have identified that ERAP1 SNPs carry the second-strongest genetic linkage to the inflammatory disease of the spine, ankylosing spondylitis (AS), after the MHC I allele HLA-B27. Evans et al. gave the AS association from combined studies between HLA-B27 and rs3018 (which is R528) as $P = 7.3 \times 10^{-6}$ (D. M. Evans et al. 2011). Notably, however, these disease associated SNPs in ERAP1 only increase the chance of AS onset in those individuals expressing the HLA-B27 allele (D. M. Evans et al. 2011). Overall, ERAP1 accounts for 26% of the genetic risk for AS, and the combination of both ERAP1 SNPs and HLA-B27 accounts for 70% (Seregin et al. 2013). A GWAS study of AS (D. M. Evans et al. 2011) indicated 102 SNP-SNP interactions linked to the disease, although this was later reduced to 8 significant interactions amongst several genes (Bessonov, Gusareva, and Van Steen 2015). Interestingly, 6 out of 8 pairs contain an ERAP1 SNP, with the interacting SNP being in either HLA-B or MICB.

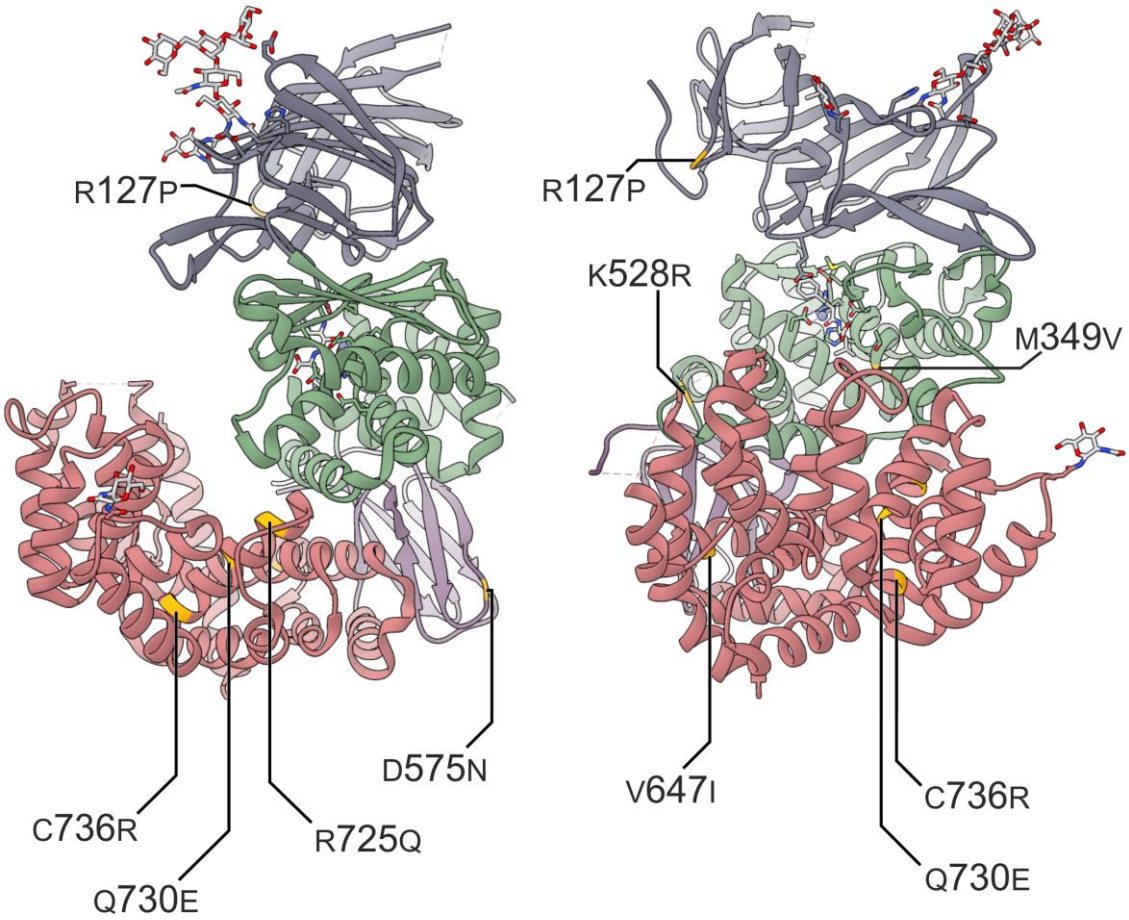


Figure 8 – Locations of ankylosing spondylitis-linked mutations in ERAP1 (shown in orange). (Harvey et al. 2009; Kochan et al. 2011; Nguyen et al. 2011)

AS ERAP1 association SNPs include R127P, M349V, K528R, D575N, V647I, R725Q, Q730E, and C736R, shown in Figure 8 (Kochan et al. 2011; Nguyen et al. 2011). These are not exclusively limited to a single area of the protein and their disease-causing mechanism is unclear. Position 349 is close enough to the active site to suggest it plays a role in binding and trimming the peptide substrates. Residues lining the internal peptide-binding face of domain IV, such as 725, 730 and 736, might bind a peptide C-terminus, or residues within hinge regions could mediate the ability of the enzyme to close around the substrate. R127, on a domain I loop, might be involved in interactions which close the protein, although this is not apparent in the 2YD0 structure. 5 of the mutation sites are on the exterior surface of ERAP1 in its closed form and this suggests something beyond the peptide processing ability of ERAP1 is linked to AS for these.

In addition to AS, the mutations shown in Figure 8 are also known to be factors in other autoimmune diseases. The most widely documented SNP position, 528, is associated with a number of diseases. K528 is a contributory factor to multiple sclerosis, as well as psoriasis when MHC allele HLA-Cw*0602 is also being expressed, whereas R528 is linked to an increased risk of hypertension (Stratikos et al. 2014). Another SNP, Q725, conveys an increased risk of Behçet's disease, along with the expression of the HLA-B51 allele (Stratikos et al. 2014).

Table 1 – Residue substitutions in ERAP1 functional variants presented by Reeves et al. (2014) are not a ready predictor for variant activity. Variants are classed by their peptide trimming rates (see colour key) and amino acids at each position given by single-letter code.

	Position/amino acid														
allotype	82	102	115	127	199	349	528	575	581	725	727	730	737	752	874
001	V	I	L	P	F	V	R	N	L	Q	L	E	V	R	M
002	**I**	**L**	**P**	**R**	**S**	**M**	**K**	**D**	**S**	**R**	**L**	**Q**	**A**	**R**	**V**
003	I	I	P	R	S	M	K	D	S	R	L	Q	A	G	V
004	I	I	P	R	S	M	K	D	S	R	A	Q	A	R	V
005	I	I	P	R	S	M	R	D	S	R	L	Q	A	R	V
006	I	I	P	R	S	M	K	D	S	Q	L	E	A	R	V
007	I	I	P	R	S	M	R	D	S	Q	L	Q	A	R	V
008	**I**	**I**	**P**	**R**	**S**	**M**	**R**	**D**	**S**	**R**	**L**	**E**	**A**	**R**	**V**
009	V	L	L	P	F	V	R	D	S	R	L	Q	A	R	V
010	I	L	P	R	S	V	K	N	S	Q	L	Q	A	R	V
011	**V**	**I**	**L**	**R**	**F**	**M**	**R**	**D**	**L**	**R**	**L**	**E**	**V**	**R**	**V**
012	I	I	P	R	S	V	K	D	S	R	L	Q	A	R	V
013	I	I	P	P	S	M	K	D	S	R	A	Q	A	R	V

Hypertrimmer: Hypotrimmer: intermediate trimmer:
efficient trimmer: unknown activity: ?

Further investigation into ERAP1 SNP variants revealed that imputed combinations of SNPs within an ERAP1 gene are known to act as AS-protective or causative haplotypes, however the reasons for this are not fully clear. K528, D575 and E730 are suggested to be an AS risk haplotype, and P127, I276 and R528 might be protective (Stratikos et al. 2014). Reeves et al. identified multiple ERAP1 SNP combinations from individuals and linked distinct ERAP1 allotypes to substantially different trimming activities: variants were divided by overall trimming behaviour, whether efficient, hyper-functional, hypo-functional,

intermediate or unknown (Table 1) (Reeves et al. 2014). This suggested the overall trimming activity of ERAP1 drives the disease mechanism, with hypofunctional variants failing to generate a pool of peptides short enough for optimal MHC I to bind strongly. Hyper-functional variants would, by contrast, over-trim the peptides, destroying the repertoire for MHC I binding. This would result in a sub-optimal range of peptides within the ER, resulting in aberrant HLA-B27 surface presentation.

1.5 The deciding factor for peptide trimming

A subject of discussion is whether ERAP1 or MHC I determines the final selection of antigenic peptides in the ER. Evidence of peptide specificity of both exists, and perhaps the most controversial idea within the debate is whether or not ERAP1 can trim a peptide bound within the MHC complex.

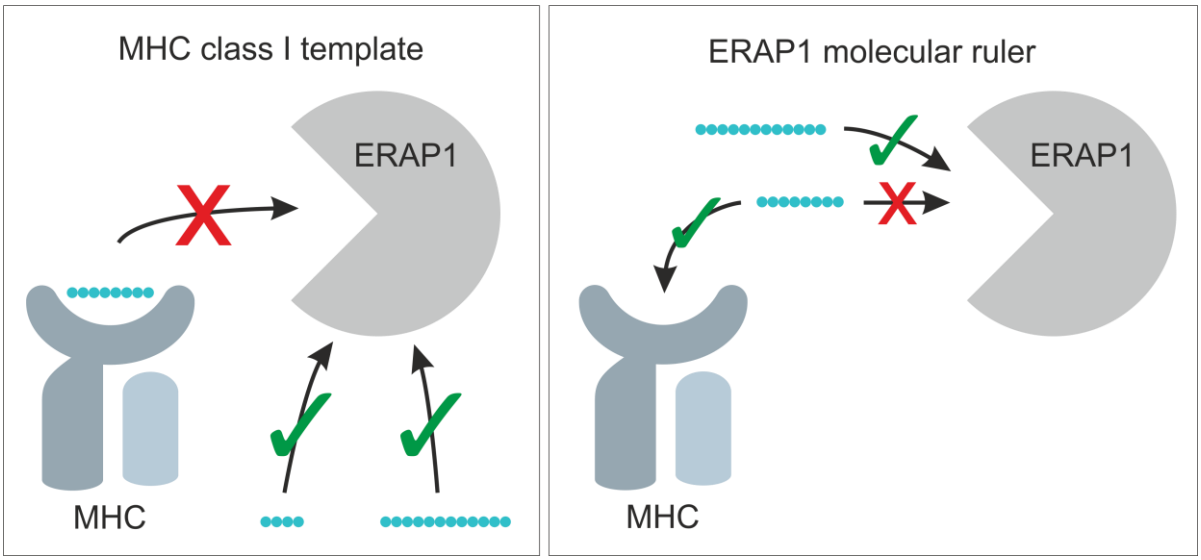


Figure 9 - Template and ruler theories of peptide length and sequence determination. Peptides in turquoise. Template theory: MHC I protects optimal peptides from degradation. Sub-optimal peptides are degraded. Molecular ruler theory: Peptide specificity of ERAP1 causes preferential degradation of peptides sub-optimal for presentation. Optimal presentation peptides are sub-optimal for ERAP1 and therefore released to MHC I.

1.5.1 ERAP1: the 'molecular ruler' hypothesis

An aspect of ERAP1 trimming mechanism which has piqued interest for many is the documented 'molecular ruler' behaviour. It has been observed that MHC I need not be present for the correct epitopes to be generated (York et al. 2002). The term 'molecular ruler' originates from ERAP1 having much lower trimming activity for <8 residue peptides (Chang et al. 2005; Nguyen et al. 2011). This makes ERAP1 well adapted to generating optimal-length 8 or 9 residue peptides for MHC I, however ERAP1 has been shown to destroy the peptide supply by trimming peptides beyond the optimal length (Kanaseki et al. 2006). Without ERAP1 present the peptides presented are substantially different, although this is unsurprising given the importance of its role.

Perhaps the most unusual feature of ERAP1 is the reported ability to select substrate peptides not merely by the N-terminal residue, but also by internal residues (Evnouchidou et al. 2008). This is unusual in an aminopeptidase, and suggests it is uniquely qualified to select the final range of antigens available to MHC I. The surface of the cavity within ERAP1, where the middle and C-terminal of the peptide are presumed to bind, carries an overall negative charge, which correlates with the preference of ERAP1 for positive residues (Evnouchidou et al. 2008; Nguyen et al. 2011). Several shallow hydrophobic pockets at varying distances from the catalytic site are suggested as C-terminal binding sites for peptides of different lengths (Nguyen et al. 2011).

1.5.1.1 Substrate peptide preferences of ERAP1

ERAP1 is known to have a substrate preference for N-terminal hydrophobic residues, with methionine being the most readily trimmed and proline the poorest trimmed (Reeves et al. 2013). Hydrophobic residues such as leucine and methionine are also rapidly cleaved from single amino acid-linked fluorogenic substrates (Hattori et al. 1999). ERAP1 specificity for peptides, compared to short fluorogenic substrates, is different and more complex as a result of the length and internal residue preferences (Evnouchidou et al. 2014).

No carboxyl group is actually required at the peptide C-terminus for ERAP1 activity (Chang et al. 2005). ERAP1 does however favour aromatic or hydrophobic C-termini and this is similar to MHC I preferences. In terms of peptide length, ERAP1 shows highest activity for 10-14 residue peptides, with activity dropping sharply from 9 residues and below. Activity is extremely low for peptides longer than 18 residues. 10-residue peptides containing prolines towards the N-terminal will be degraded at similar rates to 8/9-mers, which correlates with MHC I and TAP preferences, and the increased difficulty of trimming a proline residue for ERAP1 (Chang et al. 2005).

1.5.2 ERAP1: MHC I as an antigen 'template' for trimming

An alternative theory to the 'ERAP1 molecular ruler' hypothesis is that MHC I acts as a template, protecting high affinity peptides from degradation. It has also been postulated that ERAP1 trims peptides which are partially bound to MHC I. Evidence of this happening has been reported by Reeves, Edwards, Elliott, & James, (2013) and Papakyriakou et al. (2018) where detection of N-terminal cleaving on a C-terminally disulphide trapped 'peptide' region of an artificial single chain trimer form MHC I indicated ERAP1 could access peptides bound by MHC I. Another study, this time using HLA-B*0801 complexed with 14- and 15-mer peptides, showed trimming by ERAP1 and ERAP2 (Chen et al. 2016) using mass spectrometry to analyse fragments. In the MHC-free experiment peptides were trimmed to a range of short peptides down to 4-mers. By contrast the MHC-bound peptides were trimmed to either 8- or 9-mers, depending on the peptide used and no further peptide degradation was observed (Chen et al. 2016).

Beyond these observations, evidence of direct trimming of an MHC-bound peptide by ERAP1 is scant, although it is known that MHC I protects optimal peptides from destruction by ERAP1 (Kanaseki et al. 2006). Little explanation of how ERAP1 could approach MHC I is readily forthcoming: molecular modelling indicated that an extension of 6 N-terminal residues would be required for an MHC-bound peptide to approach the ERAP1 catalytic site (Nguyen et al. 2011).

These docking experiments show the closest approach would still leave a distance of 20 Å (Nguyen et al. 2011). It has been suggested that another thus-far unobserved super-open conformation would exist to facilitate this by permitting close approach to a pMHC I (Stratikos and Stern 2013).

1.5.3 The relationship between ERAP1 and MHC

Stratikos and Stern discuss the 'molecular ruler' and 'MHC template effect' in their review of ERAP1 structural work (Stratikos and Stern 2013). Several investigations into either ruler and template theories are mentioned. Although they refer to MHC/ERAP1 'synergy' they do not directly explore the idea of competition between the two. This might perhaps be close to one of several hypotheses presented by Falk et al. before the discovery of ERAP1:

'[The model] assumes that the polypeptide is degraded by specific endopeptidases independent of MHC. The resulting peptides are rapidly degraded by exopeptidases unless protected by binding to MHC.' (Falk, Rötzschke, and Rammensee 1990)

This did not take into account the location of antigen processing in the ER but nonetheless holds true. Prior to trafficking into the ER, suboptimal peptides are generated by the (immuno)proteasome (much of which is endopeptidase activity) but the exopeptidase ERAP1 is also known to possess peptide specificity. It could be considered possible for ERAP1 (and ERAP2 in a lesser capacity) to adopt both the specificity required to generate the epitope and the role of degrading unprotected peptides. Indeed ERAP1 has been shown to destroy unprotected peptides of optimal length as well as having some length preference (York et al. 2002). As has been suggested, what may occur is:

'-a balance between ERAP1 and MHC I for peptide binding based on affinities.' (Reeves et al. 2013)

Considered in a systems context and as competing entities, the ERAP1-ruler and MHC-template ideas co-exist well. Here they become, simultaneously, both trivial

and crucial: neither the ERAP1-ruler or MHC-template are solely responsible within the system for the peptides which are presented. But as the length of a peptide decreases, its off-rate for ruler-behaviour ERAP1 increases, making it more available for competitive MHC I. Conversely peptides poorly bound by MHC due to unfavourable length or sequence are more accessible to ERAP1. Unlike an *in vitro* situation, within antigen processing MHC I acts as a sink for optimal peptides, trafficking them to the cell surface and away from ERAP1.

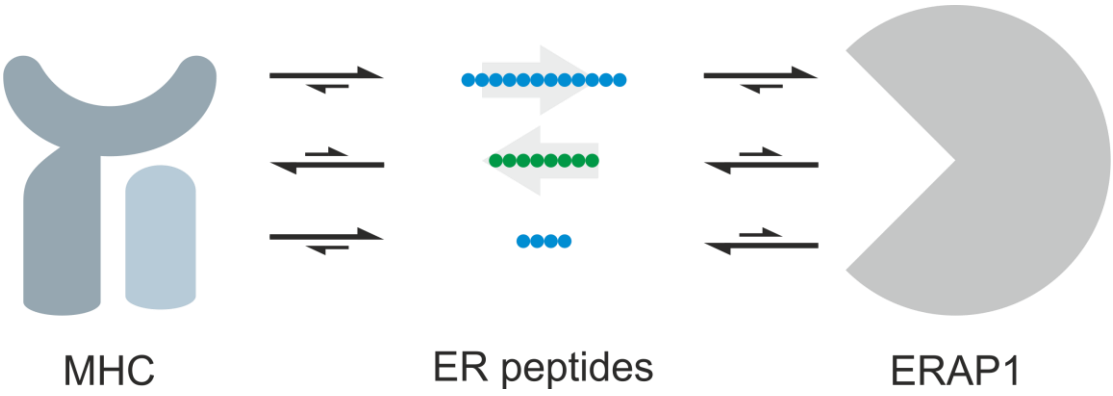


Figure 10 - Overview of concerted MHC I and ERAP1 peptide preferences. Optimal-length peptides for presentation (green) are bound by MHC I with low off-rates whereas sub-optimal peptides (too long/short) are ultimately degraded by ERAP1. Other factors such as peptide sequence preference and MHC I egress make the situation more complex in reality.

This provides an explanation for the reported observations of ERAP1 trimming MHC-bound peptides, suggesting these peptides ought not be thought of as irreversibly attached to MHC I, but as entities with on- and off-rates, which indicate the extent to which they are available to ERAP1. They might even perhaps undergo partial unbinding, where a free N-terminus but bound C-terminus of a sub-optimal peptide allows ERAP1 to approach and trim, since stable complexes with impaired N- or C-terminal binding have been reported in the past (Bouvier and Wiley 1994). Whether or not this occurs, ERAP1 will still act upon the pool of free peptides (Evnouchidou et al. 2008), and MHC still binds strongly and protectively to the optimal peptide (Kanaseki et al. 2006). The system of peptide optimisation could work around which one binds a peptide first and takes longer to release it. If this were so it would be easy to see that a

suboptimal ERAP1 and MHC could result in the pathogenesis of those diseases with strong ERAP1 and HLA genetic associations, such as AS or Behçet's disease.

1.6 Aims and objectives of this research

The aims and objectives for the work laid out in this thesis are to examine the peptide-MHC I, and peptide-ERAP1 relationships with a view to understanding how ERAP1 and MHC I shape the antigenic peptide repertoire. Firstly, this work aims to investigate the structure and peptide binding of an artificial pMHC construct by X-ray crystallography to look at peptide N-terminal binding interactions and whether the construct is suitable for assessing pMHC trimming by ERAP1. Secondly the characterisation of a peptide library by proton NMR assignments will be presented and the possibility of differentiating between related peptides and amino acids explored. Thirdly cell expression of peptides is shown to provide a means of isotope-labelling the peptides. As well as furthering characterisation of peptides from the proton assignments, this aims to verify proton assignments and provide scope for further experimentation. Fourthly, NMR is presented as an unusual means for monitoring peptide trimming at various stages of ERAP1 activity, to compare ERAP1 variants and to examine how peptides are bound by ERAP1.

2 Crystallisation of dtSCT mutants

2.1 Introduction

The concept of ERAP1 being able to direct trimming of antigenic peptides with an innate length and sequence specificity has been examined and championed by numerous studies which have provided decisive and convincing evidence (York et al. 2002; Chang et al. 2005; Evnouchidou et al. 2008; Gandhi et al. 2011). MHC class I acting as a template to specify the peptide characteristics may originally have been seen as a competing theory, but observations of ERAP1 trimming peptides apparently bound to MHC I continue to emerge (Kanaseki et al. 2006; Infantes et al. 2010; Reeves et al. 2013; Chen et al. 2016). How these observations could arise is uncertain. It is known that the key anchor residues for an antigenic peptide within the class I binding groove are the N- and C-terminal residues, although one of these alone may be enough for stability of the complex(Bouvier and Wiley 1994). The interactions between the MHC I A pocket and peptide N-terminal can be gleaned from examination of the wealth of crystal structures. The effects of individual conserved residues within the A pocket is not clear, and it is of interest to see whether removal of a single binding residue can create a less tightly bound peptide terminal.

It has often been stated that ERAP1 appears to have innate length and sequence preferences for substrate peptides and modelling indicates that the distance between the catalytic site of ERAP1 and the binding groove of MHC class I is too substantial for trimming of an MHC-bound peptide by ERAP1 (Stratikos and Stern 2013). Conversely, however, a publication by Chen et al examined the differential ERAP1/2 trimming of peptides bound to MHC versus free peptide (Chen et al. 2016). This might suggest ERAP1 can compete with MHC I for 'bound' peptides until the optimal final peptide is strongly bound to MHC. Strikingly, Reeves et al (Reeves 2013) observed ERAP1 trimming of a single chain form of MHC where the peptide was linked to MHC and held in the F pocket with a disulphide bridge. This finding suggests ERAP1 might be able to make a

close approach to MHC I, contradictory to models which indicate implausibility due to the size of ERAP1 (Nguyen et al. 2011).

The single-chain trimer (SCT) version of MHC I used by Reeves et al. (2013) as mentioned above is a concept originally used and described by (Lybarger et al. 2003). Their incorporation of an F pocket peptide/heavy chain disulphide bridge in addition to the flexible linker regions gives rise to a pMHC which is stable with a high likelihood of the peptide region being bound within the peptide groove as a consequence of the peptide C-terminal disulphide trap. They are known as disulphide trap single chain trimers, or dtSCT. The increased stability of peptide binding permits analysis of other structural alterations which might normally be too unfavourable to work with. Alterations to the A pocket of the binding groove, where the peptide N-terminal binds, become feasible.

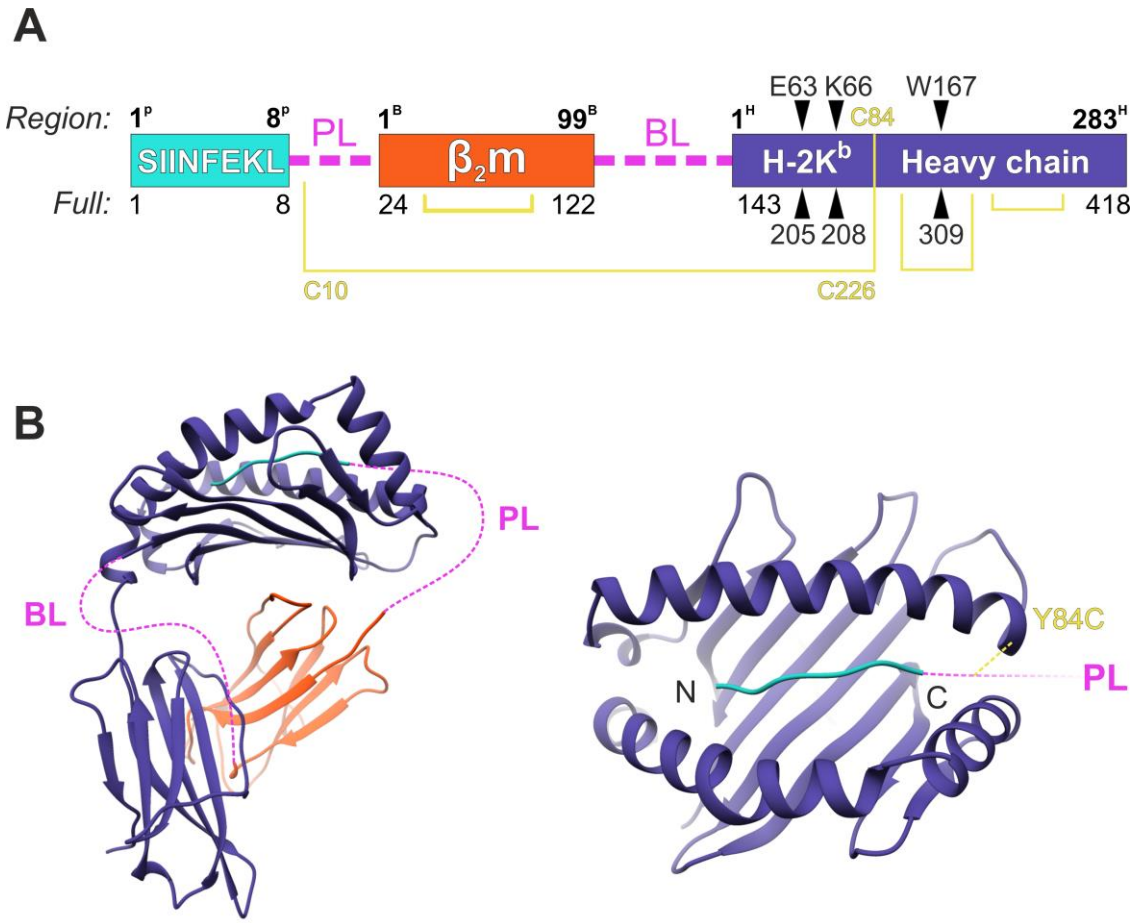


Figure 11 - Disulphide trap single chain trimer construct with peptide linker (PL) and β_2m linker (BL) (in pink), disulphide bonds (yellow), peptide (turquoise), β_2m (orange) and H-2K^b (indigo). A) Diagram of constructs showing numbering within region/subunit (top) or concurrent numbering within full protein sequence including linkers (bottom). B) Side and top views of disulphide trap single chain trimer design with approximate linker positions shown. With PDB 3P9L (solved by Denton et al. 2011).

Both SIINFEKL:H-2K^b:β_2m complexes and dtSCT of H-2K^b have been crystallised and published in the past (Denton et al. 2011; Mitaksov et al. 2007; Dam et al. 2003; Mareeva, Martinez-hackert, and Sykulev 2008; Deng et al. 2008). Mutations at other binding pockets have also been published, but there are no known publications of A pocket mutations intended to examine the contributions of individual residues to N-terminal peptide binding.

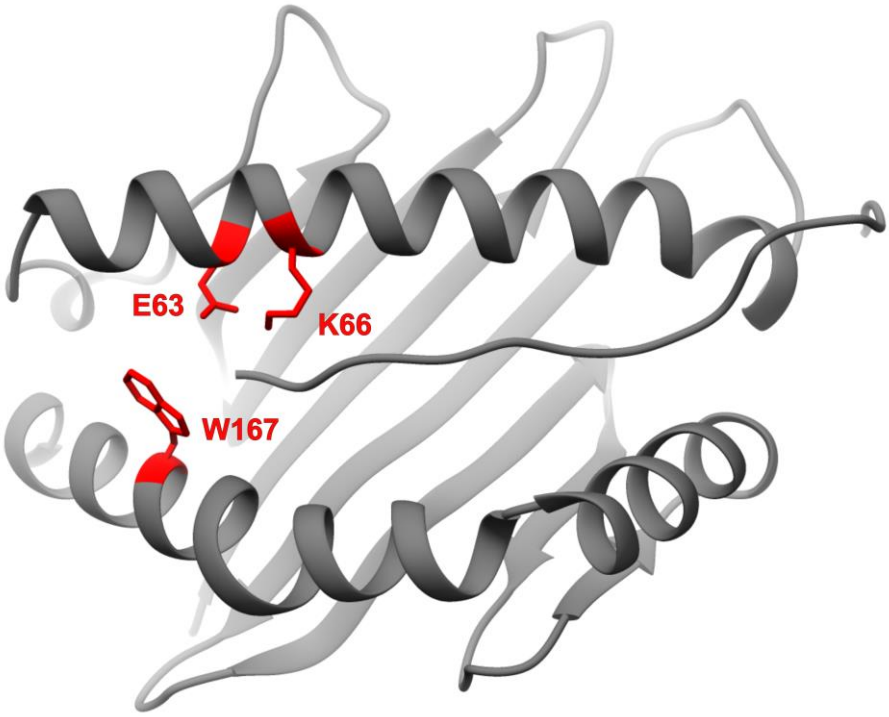


Figure 12 – Ribbon diagram of 'wild type' dtSCT binding groove showing sites of mutations (red) introduced to A pocket of peptide binding cleft.

Three conserved residues in the A pocket of H-2K^b are E63, K66 and W167. E63 and K66 have direct hydrogen-bonding contact to the peptide N-terminal (Figure 31) and are thought to be crucial binding contacts for this reason. W167 forms no direct contact with the peptide N-terminal but is a large, steric and hydrophobic conserved residue above the A pocket which shows conservation (Matsumura et al. 1992; Achour et al. 2002). The structural effect of mutating these residues was unknown. Also unknown was the effect that removing these contacts with the peptide would have on the ability of ERAP1 to degrade the bound peptide. To assess the structural impact of mutating these residues, three mutants (E63A, K66A and W167A) as well as the 'wild-type' H-2K^b dtSCT were expressed, purified and crystallised for generation of crystal structures.

2.1.1 X-ray crystallography as a tool for structural biologists

2.1.1.1 History and use in biology

The basis for diffraction (of neutrons, electrons or X-rays) crystallography is fundamentally simple: that a beam passing through a multi-atom molecule will result in diffraction, that an ordered crystalline repeat of the same molecule can give a strong and recordable diffraction, and that the structure of the molecule can be calculated from the diffraction pattern.

This procedure is very simple for the crystal of an uncomplicated molecule. Salt molecules and anything with a small number of atoms will produce a handful of spots in a diffraction pattern. For biological molecules such as DNA and proteins, the challenges are markedly larger. The enormous number of atoms present in even a small protein will result in a huge number of diffraction spots which cannot be directly related to a protein structure without a large amount of computation. Additionally, biological molecules require careful handling. They operate under physiological conditions for the organism from which they were taken. Many are therefore extremely susceptible to changes in pH, temperature and salt concentration, and any move too far from optimal conditions can cause a protein to unfold, lose secondary structure, and aggregate. Further challenges are caused by flexible loop regions which do not readily crystallise, hydrophobic membrane proteins and co-crystallisation of multimeric binding partners. Despite these challenges the progress of biological crystallography has continued with an increasingly sophisticated range of techniques used to grow, diffract and solve these crystals (Rhodes 2010).

Although X-rays, neutrons and electrons are all used for crystallographic diffraction, X-ray crystallography is the most widely used technique. The usefulness of X-rays for this stems from their wavelengths being smaller than the atoms and bond lengths present in a molecule. Bond lengths present will be in the region of 1.5 Å which requires electromagnetic radiation of a wavelength

<1.5 Å, and X-rays fall between 0.1-100 Å, making high energy X-rays perfectly suited for crystallography (Rhodes 2010).

2.1.1.2 From protein to diffraction data

Crystallising a protein

There are many macromolecules which biological X-ray crystallography can help in solving the structure of, but this explanation will focus solely on soluble protein crystals, for which the techniques are well established and different to those of (for example) a membrane protein.

The optimal conditions for crystallization of a protein depends on the nature of the protein itself. How hydrophobic it is and whether it has disordered regions which do not form ordered crystals easily will change the required conditions and readiness of crystal formation. How likely it is that a protein will be denatured by a specific salt concentration or temperature depends on the stability of the protein. Screens are therefore constructed to find the optimum conditions adequate to nucleate ordered crystals without causing disorderly precipitation. Variable conditions include protein concentration, pH, temperature, and concentration of precipitant. These are set up as multi-well screens which is a rapid way of testing a range of conditions for the same protein.

A widely used technique for crystallisation is sitting drop plates. This method somewhat simplifies the optimisation process by using plates with differing protein concentrations in each micro-well. The purified protein solution is mixed with reservoir solution (comprised of buffer, precipitant and salt) in three different ratios, which are placed adjacent to the 40 μl reservoir and the plate is then sealed. The sitting drop method enables computer automation for well filling which gives the advantage of accurately and swiftly dispensing a large number of very small volumes but requires a cover is applied rapidly given that

the first wells to be filled will have had exposure to air by the time the final wells are filled. The much higher reservoir precipitant concentration drives diffusion of water vapour from the wells to the reservoir until an equilibrium is reached and crystals begin to nucleate. A metastable state of supersaturation occurs when despite being at a concentration high enough for precipitation, some proteins still remain in solution. The formation of a crystal allows the equilibrium to be reached again (Mcpherson 2004). The process allows proteins to come out of solution slowly and precipitate in an ordered crystal.

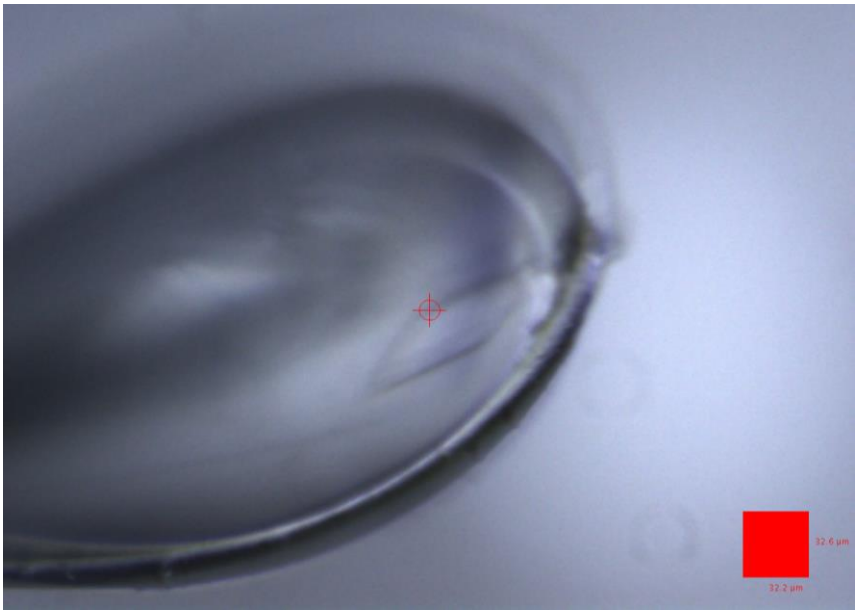


Figure 13 - Cryoprotected crystal within a loop (seen below crosshair - image author's own).

Cryo-protection of protein crystals

Cryo-protection is a crucial part of protein crystallography but occurs in a very small space of time, when a crystallographer transfers a mixture of glycerol and the mother liquor to a crystal-containing drop before fishing the crystal and rapidly transferring it to liquid nitrogen. The conditions of the crystallising drop change as soon as the well is opened, and a skilled crystallographer who acts quickly can significantly improve the chances of obtaining good data. Glycerol favours disordered solidification of water and thus is used to prevent the disruption caused to the crystal structure when water forms a lattice as it freezes

and expands within channels (Garman 1999). In addition to destruction of the crystal, the formation of ice can also lead to 'ice rings' on the collected diffraction image, posing difficulties. Freezing the crystals this way ensures the crystals are preserved by the low temperature (around 100 K) and can be stored for a longer period of time. The crystals are 'fished', cryo-stored and exposed to X-rays within the same loop, lowering the chance of damage (see Figure 13).

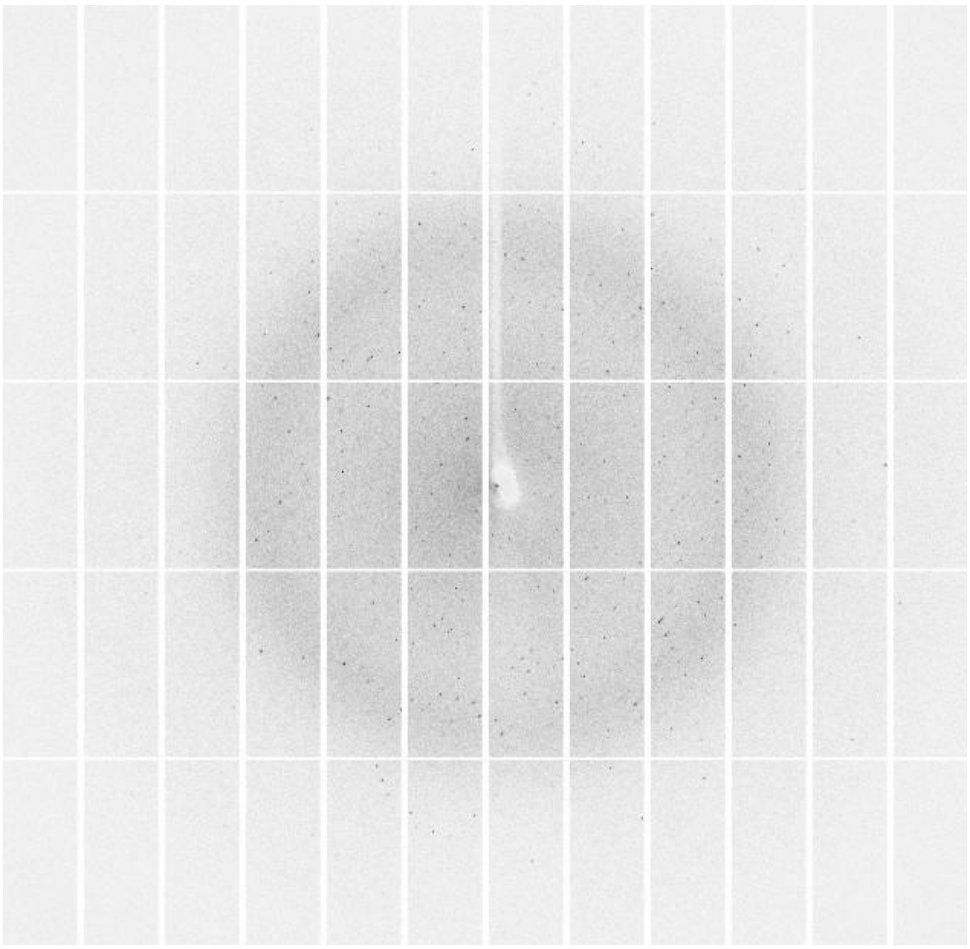


Figure 14 - Diffraction image from a crystal. Multiple similar images are recorded at different rotations of the crystal to obtain a full dataset. (Image author's own).

Data collection

It is important to be able to grow the best possible quality of crystal and cryo-protect effectively. The optimal protein would be sufficiently large and well-ordered to produce high quality diffraction data when subjected to X-ray. A large

number of planes in a crystal give rise to constructive interference, enhancing the diffraction data, which can be shown with Bragg's law:

$$n\lambda = 2\mathrm{d}\sin\theta$$

Where n is an integer, λ is the wavelength of the x-ray, d is the distance between planes in the lattice and θ is the angle of scattering for the diffracted wave (see Figure 15). Waves emerging from the lattice in phase will undergo constructive interference and produce a spot on the image, and this allows for calculation of the electron density when the phase problem (see below) has been overcome.

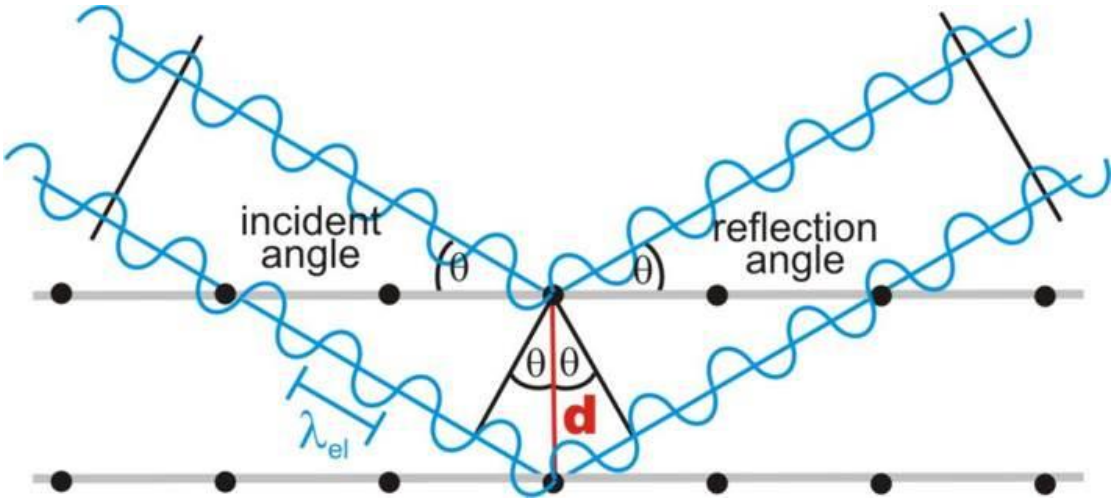


Figure 15 - Constructive interference of X-ray diffraction in a crystal lattice according to Bragg's Law. Image from www.microscopy.ethz.ch/bragg.htm (Krumeich 2015)

The X-rays diffract between the atoms and are scattered in a diffraction pattern which is recorded on a CCD detector. Because the protein is a three-dimensional molecule the diffracted X-rays are present in a reciprocal 3D space which is viewed by moving the crystal within the X-ray beam and capturing multiple images on the CCD detector. The nature of diffraction through a complex, multi-atom molecule means the 'real space' containing the protein is not mirrored identically by the reciprocal space, and consequently data processing is required to work back to the electron density which diffracted the X-rays originally (Rhodes 2010).

2.1.1.3 A structure model from diffraction data

The phase problem

The course of the X-ray beam through the crystal results in the reflections on the detector. In order to create an electron map from raw data, the data must be Fourier transformed, which describes the process of diffraction by breaking a complex (X-ray) wave into the constituent waves it is comprised of (Rhodes 2010). The waves analysed are three-dimensional with x-, y- and z-axis frequencies, and are described by the Fourier sum. In order to decompose these waves with Fourier transformation, the wave amplitude, frequency and phase are required. The difficulty for the crystallographer stems from the fact that, to interpret the reflections, the path that waves took through the structure must be known, but this cannot be known without using the reflections to solve the structure. This is referred to as the phase problem; of the three properties of the wave, only amplitude and frequency can be immediately determined from the reflection intensity and position.

The phases of the X-ray photons hitting the CCD are not detected or recorded by the CCD, but destructive interference of out-of-phase photons can cause loss of data and this must be understood for interpretation of the diffraction pattern.
It is possible to generate a Patterson map from the raw data, which is Fourier transformation of squared, unphased amplitudes of the reflections (Rhodes 2010). This does not produce an electron map, but a map of vectors between atoms. The vector map contains the same dimensions as the real unit cell but is otherwise insufficient alone to solve the structure; it can be combined with heavy atoms introduced to the crystal, because a crystal of single heavy atoms is simpler than a large, complex crystal of many atoms. If heavy atoms are not used, then, where possible, molecular replacement can solve the phase problem, relying on the ever-growing number of comparable protein models available to use. Molecular replacement uses a phasing model, which is a known protein structure related to the unsolved protein. Usually sequence homology gives an

indication of whether the two proteins will likely have structural similarity. When the phasing model is found, its reverse Fourier transform can be calculated and fitted against the data. The added complexity to this originates from the low chances of the phasing model having an identical orientation and unit cell (nonisomorphous phasing). The diffraction data reveals the unit cell of the crystal and aided by this the orientation of the protein is determined. A rotation function is used to compare orientations between the model and the target's Patterson map. The output of this is given by the log file from Molrep, where the comparison between data at different orientations is given in a scored table, and the best-matching orientation is used for the phases.

Refinement and model building

The output file from Molrep with the solved initial phases is used for model building and refinement of the experimental structure. The atomic positions from the phasing model have provided phases, structure factors (a function for the phase and amplitude of waves diffracted through the planes of the crystal) and an approximate structure for model building. Using this phasing model has potential for introducing bias, and it is important during model building and refinement to correct problems with the structure as they arise to prevent bias from incorrect phase determination. Two complementary quality values are calculated during each round of refinement. The residual factor (often referred to as R_{work}) is determined to show how much agreement there is between observed (initial data) and calculated (newly modelled) structure factor amplitudes. Lower sums denote greater agreement between the two. The free R-factor is used in conjunction with the R factor. This is determined with a test set of intensities which are never used during refinement rounds. An ideal final model will have low R_{work} and R_{free} values which show little divergence.

Refinement is a continuation from molecular replacement and improves upon the initial phases. A difference map mFo-DFc is determined, using the observations (F_o, or F_{obs}) and calculated structure factors (F_c, or F_{calc}). The term 'm' mean 'figure of merit' which accounts for the phase quality, and 'D' denotes

'Sigma-A weighting factor' which is weighting of the structure factors to better account for error (Read 1986; Adams et al. 2012). This results in a map with positive and negative contours which gives a clear indication of missing and incorrect parts of the model. Perhaps more useful is the 2mFo-DFc map also created during refinement. This is the map commonly depicted as blue mesh and used primarily for model building, as it gives an indication of electron density. Improvement of structure factors by building the model according to the 2mFo-DFc and mFo-DFc maps leads to improvement of maps, which is why repeat rounds of refinement and model building are carried out alternately (Rhodes 2010).

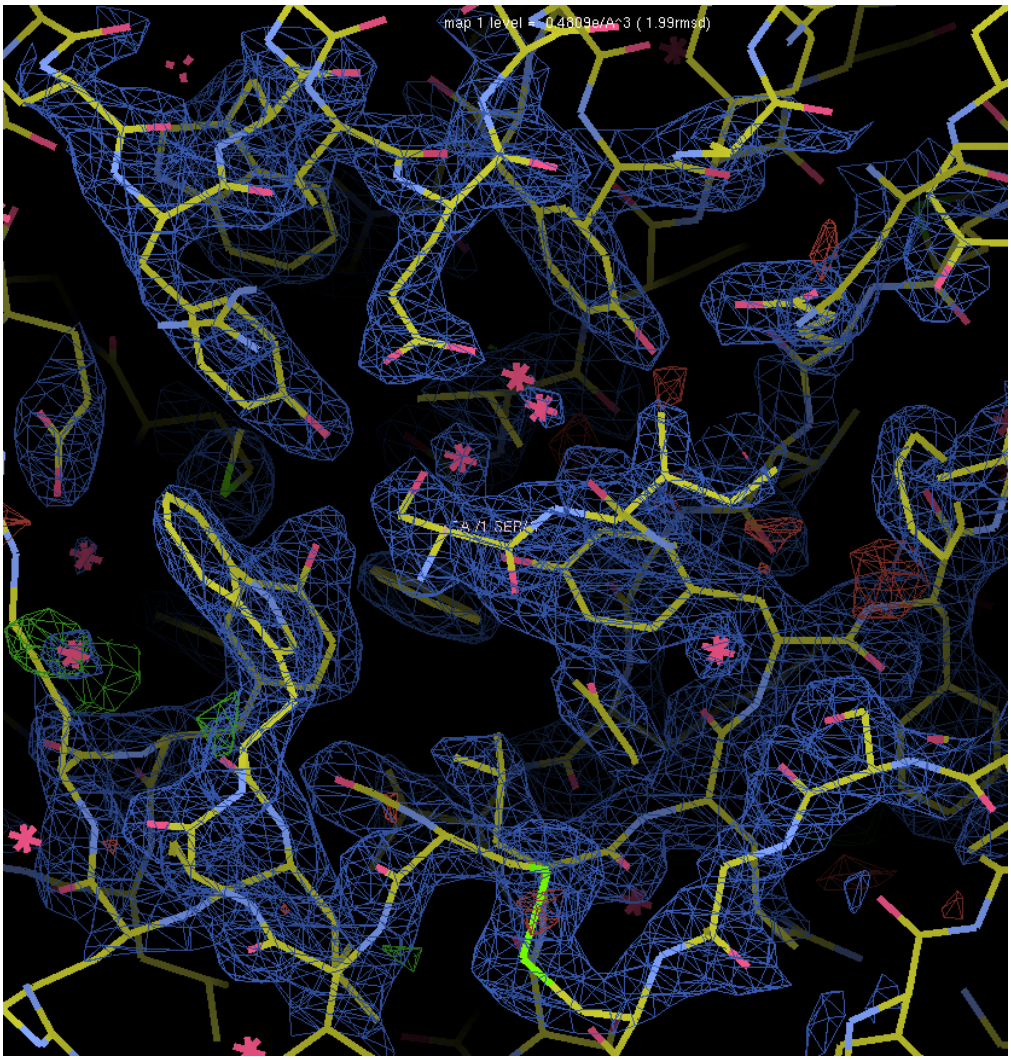


Figure 16 – 2mFo-DFc electron density map of a protein, contoured at 1.99 σ with built stick structure shown within. Water molecules shown as pink stars. (Image author's own).

The process of model building requires balance between fitting a structure within the electron density and optimising the geometry of the structure, keeping bond angles and lengths within acceptable parameters. This is most easily done with high resolution data which is unambiguous enough to justify abnormal sidechain conformations or less likely dihedral backbone angles, but which will obviously show, for the majority of a structure, geometries considered probable. A key marker for the quality of a structure is how it compares with all other published crystallographic models for R values and Ramachandran plot outliers, and this comparison forms part of the process of validating and depositing a structure in the worldwide Protein Data Bank (H. Berman, Henrick, and Nakamura 2003).

2.2 Methods

2.2.1 Expression and purification of dtSCTs

Expression, refolding and purification of W167A was carried out under the supervision of Dr Leon Douglas. Expression, refolding and purification of SL8-SCT, E63A and K66A were carried out by Dr Leon Douglas and Dr Halina Mikolajek.

2.2.1.1 Expression in *E. coli*

Glycerol stocks of *E. coli* BL21 RIPL cells transformed with plasmids containing the dtSCT insert were obtained from Dr Leon Douglas. LB media (lysogeny broth, Sigma Aldrich) was autoclaved. Selection antibiotics were added to this LB broth: ampicillin to 100 µg/ml, kanamycin to 50 µg/ml, chloramphenicol to 34 µg/ml, and streptomycin to 18.5 µg/ml. The glycerol cell stock was added

Streptomycin to 75 µg/ml was added to autoclaved LB broth. Well-isolated colonies were picked from agar plates and used to inoculate 5 ml LB/streptomycin media. The media was cultured for 4-6 hours at 37 °C in a shaking incubator. A highly turbid culture was selected and 100 µl added to 100ml LB media. 4 x 480ml LB/streptomycin media in 2L baffled flasks was prewarmed overnight to 37 °C.

The prewarmed media was inoculated with 20ml of overnight culture. The inoculated media was grown at 37 °C in in a shaking incubator at 200 rpm. When the culture reached an optical density absorbance value of 0.8 – 1 (at around 1 ½ hours) protein expression was induced by addition of IPTG (isopropyl β-D-1-thiogalactopyranoside) to 1mM. After a further 3 ¼ hours, the cells were extracted by centrifugation at 6000 rpm, 4°C for 10 minutes. The used media was discarded and the cell pellet stored at -20 °C before purification.

2.2.1.2 Inclusion body purification

The frozen cell pellet was kept on ice and 5 ml of BugBuster (obtained from Merck) was added per 1 g harvested cell pellet from expression along with 10 µl lysozyme (1000 kU/ml stock), 50 µl DNase I (2 mg/ml stock) and 50 µl $MgCl_2$ (1M stock). The pellet was resuspended by thorough homogenisation in a glass homogeniser, then incubated on rollers at 4 °C for 30 minutes. 10 mM DTT was added to the mixture before incubation on rollers, 4 °C, 10 minutes. The mixture was centrifuged at 15 000 rpm, 15 minutes, 4 °C and the supernatant removed.

The pellet was resuspended in 5 ml per g BugBuster with the same concentrations of lysozyme, DNase I and DTT given above and resuspended before incubation 30 minutes, 4 °C. The mixture was centrifuged at 15 000 rpm, 15 minutes, 4 °C and the supernatant removed.

The pellet was resuspended in 5 ml per g 1/10 diluted BugBuster and resuspended. The mixture was centrifuged at 18 000 rpm, 15 minutes, 4 °C and the supernatant removed. This step was repeated. The pellet was stored overnight at -20 °C before resolubilisation and refolding.

2.2.1.3 Protein refolding

Pelleted purified inclusion bodies were resolubilised with 15 ml resolubilisation buffer.

Resolubilisation buffer:

8M urea
500 mM MES (pH 6.5)
0.1 mM EDTA (from 500 mM stock)
1 mM DTT (from 1M stock)

They were then incubated at 4 °C for >1 hour. This was then centrifuged at 18 000 rpm for 20 minutes at 4 °C. Supernatant was removed and filtered with a 0.45 μm filter. 1L refold buffer was prepared like so:

Refold buffer:

100 mM Tris-HCl (pH 8)
400 mM L-arginine
2 mM EDTA
5 mM reduced glutathione
0.5 mM oxidised glutathione

Three additions of resolubilised protein were made to the gently stirred refold buffer at 4 °C. These additions were separated by 16 and then 6 ½ hours, to a final maximum concentration of 10 μM resolubilised inclusion body protein. The refold was stirred at 4 °C overnight.

2.2.1.4 Purification of folded protein

The refolded protein was centrifuged then filtered with a 0.45 μm filter to remove precipitate before concentration in a stirred Amicon filter. The his-tagged protein was purified by size exclusion chromatography in de-gassed phosphate buffered saline (from tablet form, Sigma Aldrich) on a HiLoad Superdex 75 pg 16/600 column (GE Healthcare). Nickel column purification followed this, with a 20 mM Tris (pH 8) + 300 mM NaCl buffer run over the HisTrap excel 1ml Ni sepharose column (GE Healthcare) before elution with the same buffer + 250 mM imidazole. Another gel filtration size exclusion chromatography was performed on an HiLoad Superdex 75 pg 16/600 column in tris-buffered saline (20 mM Tris (pH 8) + 25 mM NaCl). The protein was then purified by anion exchange purification with a 1 ml HiTrap Q HP column (GE Healthcare) in the same buffer (20 mM Tris (pH 8) + 25 mM NaCl) before elution with a high-salt buffer (20 mM Tris (pH 8) + 1M NaCl). A final size exclusion chromatography step was performed on the HiLoad Superdex column to buffer

exchange the sample into HEPES-containing crystal trial protein buffer for crystal trials:

Crystal trial HEPES buffer:

20 mM HEPES (pH 7.45)
200 mM KCl
0.5 % v/v glycerol
~0.1% w/v Sodium azide

This was made up to 1L, and filtered and de-gassed before use on the gel filtration column.

2.2.1.5 SDS-PAGE for protein expression and purification analysis

12% SDS polyacrylamide gels were cast in Bio-Rad 0.75 mm mini-PROTEAN plates. A gel buffer for the SDS stacking and running gels was made:

Gel buffer:

SDS (0.3%)
Tris (3M), pH 8.45

The running gel was mixed to the following recipe:

12% SDS running gel:

3.33 ml 30% bis acrylamide (37.5:1)
3.33 ml dH2O
3.33 ml gel buffer:
10 µl TEMED

This was mixed before 50 µl ammonium persulphate (100 mg/ml stock) was added. After rapid and thorough mixing, the mixture was distributed between 3 sets of casting plates. Isopropanol was added on top to a few millimetres and the

gels were left for ~30 minutes to set. The isopropanol was removed by blotting with filter paper.

A stacking gel was then made to create wells in the polyacrylamide gels. This was mixed to the following recipe:

Stacking gel:

660 µl 30% bis acrylamide (37.5:1)
2.9 ml dH2O
1.24 ml gel buffer:
10 µl TEMED

This was mixed before 25 µl ammonium persulphate (100 mg/ml stock) was added. The stacking gel mixture was then rapidly and thoroughly mixed again, before being distributed atop the cast running gels. Well combs were inserted (usually 10 wells) and the gels left to set for another 30 minutes.

To prepare samples for running on a gel the samples were mixed with sample running buffer.

4x Sample buffer:

200 mM Tris-HCl (pH 6.8)
8% w/v SDS
0.8% v/v bromophenol blue
40% v/v glycerol
400 mM β-mercaptoethanol
ddH_2O to 10 ml

To check efficacy of inclusion body expression, 500 µl samples of LB media pre- and post-induction were centrifuged at 13000 rpm in a benchtop Eppendorf centrifuge for five minutes. The supernatant was removed. 25 µl of 4x sample buffer was used to resuspend the pellet. The sample was heated at 95 °C in a heat

block for 10 minutes. 5 µl was inserted to a well on the SDS-PAGE. Protein marker (ThermoFisher Spectra Multicolor Broad Range Protein Ladder) was added to the first well. SDS gels were also used to check the protein at each stage of purification. These samples were mixed with the 4x sample buffer and loaded on the gel, with 15 µl loaded for weak samples and 5 µl for highly concentrated samples.

The gels were run in a Bio-Rad mini-PROTEAN tank according to the manufacturer's instructions. The anode buffer was 200 mM Tris (pH 8.9). The cathode buffer was made to the following recipe:

Cathode buffer:

100 mM Tris (pH 8.2)
100 mM Tricine
0.1% SDS

2.2.2 Crystallisation and data collection

The mutants were crystallised at room temperature with three well sitting-drop vapour diffusion. To crystallise E63A, K66A and W167A the 'TOPS' and 'TOPS2' screens (Bulek et al. 2012) were used (see Appendices). These screens had been successfully used by Dr Halina Mikolajek to crystallise SL8-SCT. The crystals were grown at pH 6-6.5.

Data was collected at the Diamond Light Source at the I03 beamline (with a Dectris Pilatus3 S 6M detector) for SL8-SCT, I24 beamline (with a Dectris Pilatus3 S 6M detector) for E63A and K66A, and I04 beamline (with a Dectris Pilatus 6M-F detector) for W167A.

2.2.3 Data handling and model building

All data handling for SL8-SCT was carried out by Dr Halina Mikolajek.

Data handling, scaling and reduction were performed with the Xia2 software suite (Winter 2010) with Aimless (Evans and Murshudov 2013).

2.2.3.1 Molecular replacement

Molrep (Vagin and Teplyakov 1997) was used for molecular replacement with published dtSCT structure PDB ID:2QRI (Mitaksov et al. 2007) as the phasing model.

2.2.3.2 Refinement and model building

Refinement and model building was carried out with the help of Dr Mikolajek, who executed most of the model building and refinement for SL8-SCT.

Structures were refined with Phenix (Adams et al. 2010) using default parameters. For the final rounds of refinement with model E63A bond length errors were corrected by optimised X-ray/stereochemistry weighting automated function. Model building was carried out in Coot (Paul Emsley and Cowtan 2004).

2.2.4 Solved model analysis

2.2.4.1 Structure validation

The models were validated throughout model building with integrated validation features in COOT (Ramachandran plots, geometry checks, electron density fit analysis and others) (P Emsley et al. 2010). Additionally the stringent wwPDB web validation tool (H. Berman, Henrick, and Nakamura 2003) was used at the end of model building and refinement to check the plausibility of the final model. Where this highlighted issues such as problematic bond lengths the models were returned to Coot and Phenix for further correctional model building and refinement until the wwPDB structural probability requirements were satisfied.

2.2.4.2 Hydrogen bonding analysis

'Optimal' and 'Potential' hydrogen bonding analysis was applied to the hydrogen-free structures using the WHAT IF web interface tools(Hooft, Sander, and Vriend 1996; Hekkelman et al. 2010).

2.3 Results

2.3.1 Expression and purification of dtSCT mutants

Expression of dtSCTs as inclusion bodies yielded a clear inducible protein which was evident on an expression gel (see Figure 17). Purification of the inclusion bodies showed little carry-over of dtSCT to the wash supernatant, and a strong band on the gel for the purified inclusion bodies (see Figure 18).

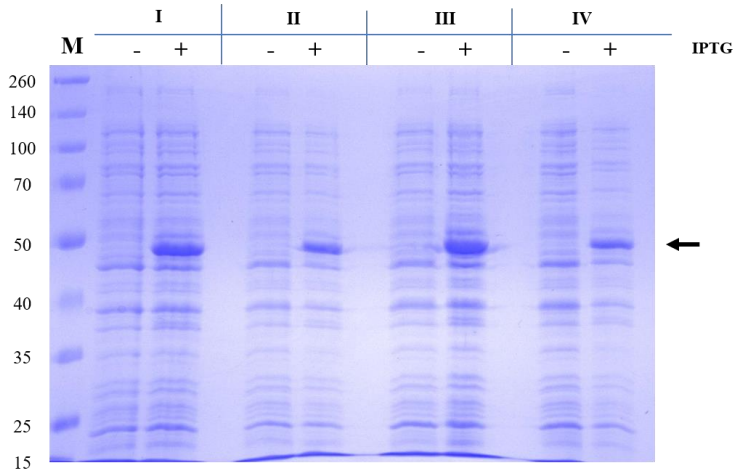


Figure 17 - 12% SDS-PAGE of W167A inclusion body overexpression. From four flasks, shown pre- and -post induction. The inclusion body expression band can be seen near the 50 kDa marker in post-induction samples (arrow). Protein marker weights (kDa) shown to left (marker: ThermoFisher Spectra Multicolor Broad Range Protein Ladder).

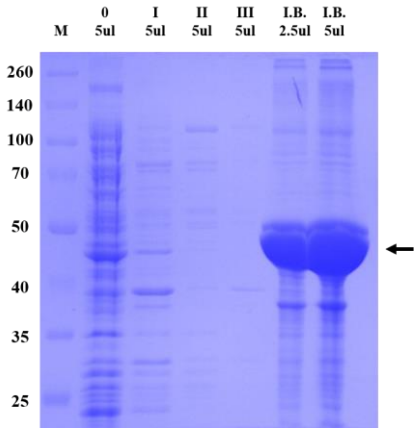


Figure 18 - SDS-PAGE of purified W167A inclusion bodies. Lanes are: 1, protein marker; 2, expressed unpurified sample; 3, purification supernatant I; 4, purification supernatant II; 5, purification supernatant III; 6, 2.5 µl of purified inclusion bodies; 7, 5 µl purified inclusion bodies. Protein marker weights (kDa) shown to left (marker: ThermoFisher Spectra Multicolor Broad Range Protein Ladder).

Initial purification of dtSCTs was carried out by size-exclusion chromatography to remove impurities. A clear peak was seen for the protein in the gel filtration trace, along with neighbouring impurity peaks.

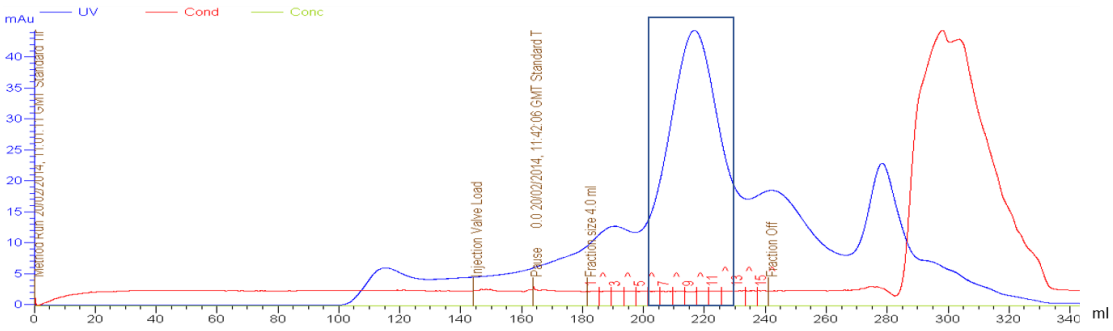


Figure 19 - Size exclusion chromatography of dtSCT W167A. Absorbance at 280 nm (blue) shows a large peak for the protein of interest. A black box indicates the pooled fractions which were used for His-tag purification.

The His-tag was used to selectively purify the tagged protein from the pooled size exclusion chromatography fractions. A trace for W167A nickel purification can be seen in Figure 20. Despite lower impurity levels due to the size exclusion step, some impurity bands are visible on the gel. The gel also shows some of the protein of interest within the flow-through, therefore some of the protein was lost from only partial sample binding to the column. Nonetheless the pooled fractions showed strong bands on the gel for W167A (Figure 20).

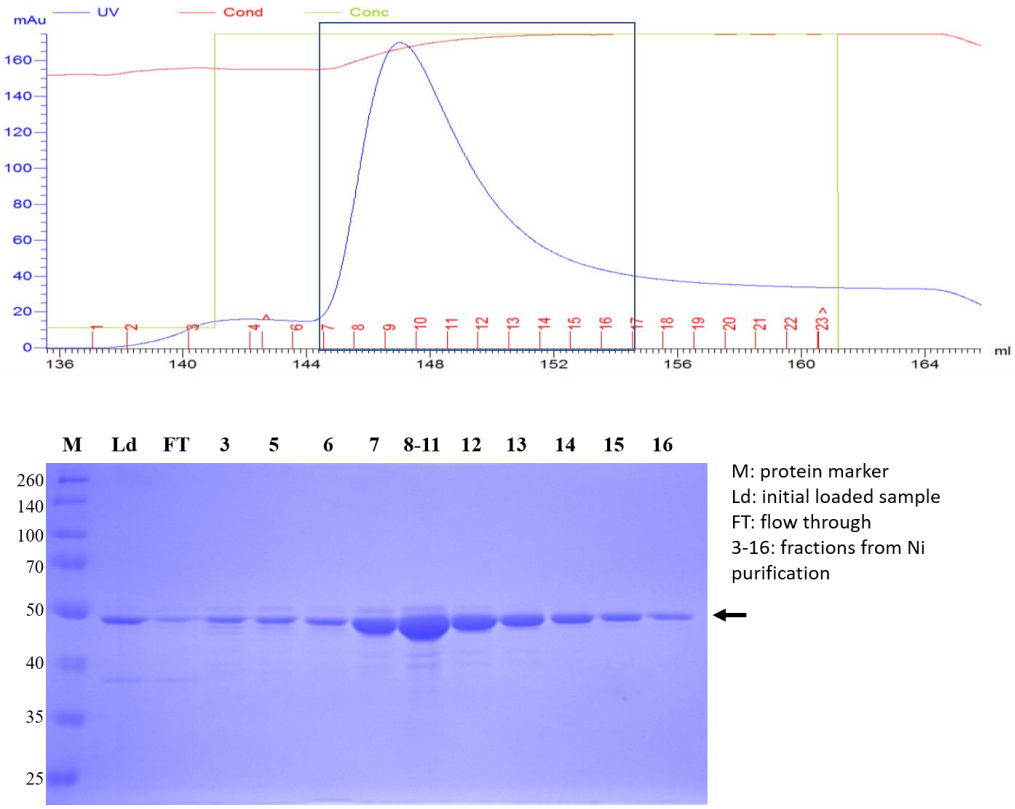


Figure 20 - Ni-column purification of W167A. Above, the purification trace, with a black box around the pooled fractions of the 280 nm peak. Below, an SDS-PAGE of the flow through and collected fractions, showing a high concentration of W167A (indicated by arrow) in fractions 8-11 but additional impurity peaks. Protein marker weights (kDa) shown to left (marker: ThermoFisher Spectra Multicolor Broad Range Protein Ladder).

Further purification by anion exchange chromatography yielded a shouldered peak with a good degree of purity seen in the SDS-PAGE of fractions (Figure 21).

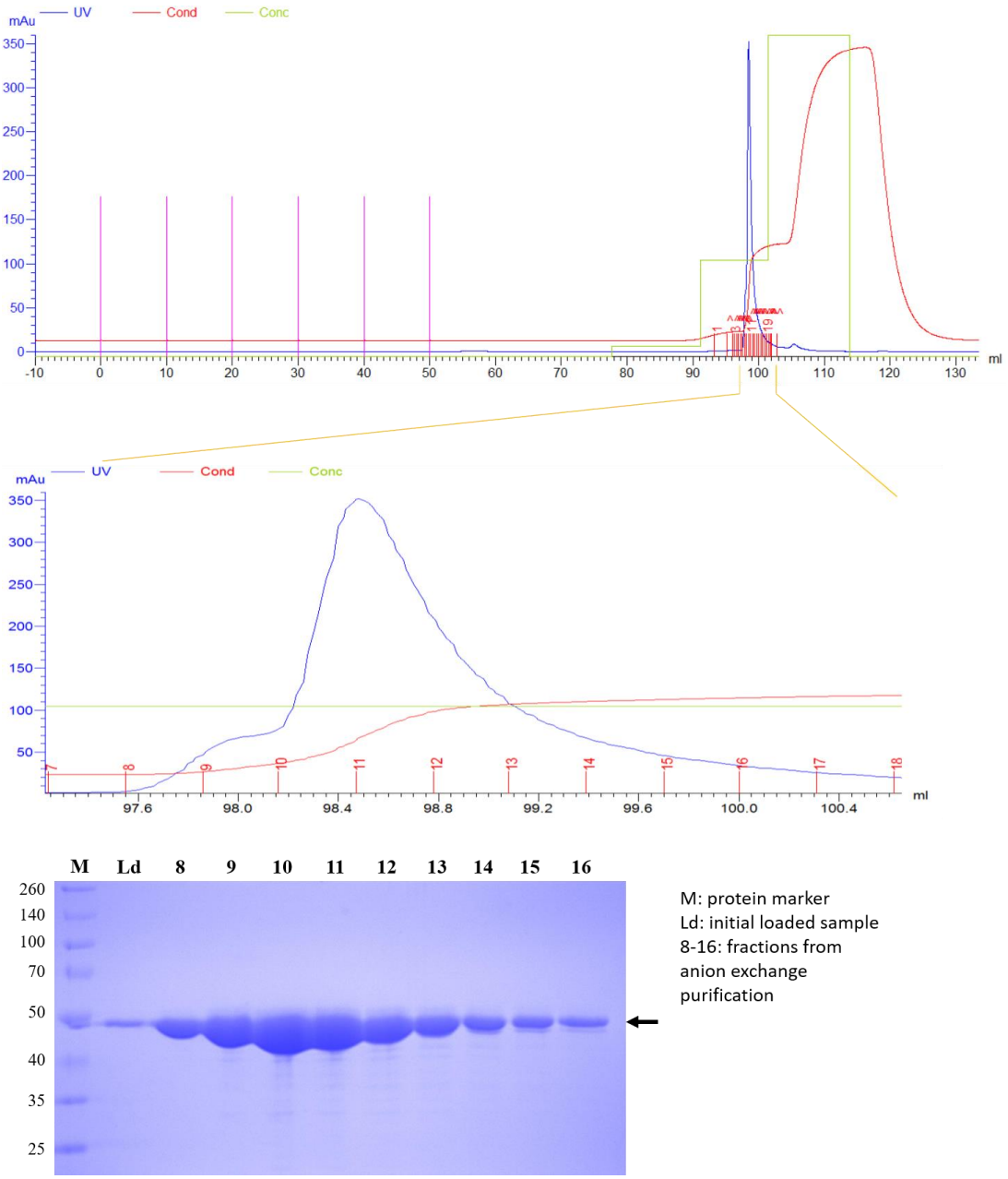


Figure 21 - Anion exchange purification of W167A with trace and expanded view (top) and SDS-PAGE (below). Few impurity peaks are seen at the 280 nm absorbance (blue) in the traces, and the SDS-PAGE similarly shows a large quantity of pure W167A (arrow) in collected fractions. Protein marker weights (kDa) shown to left of gel (marker: ThermoFisher Spectra Multicolor Broad Range Protein Ladder).

2.3.2 Crystallisation and X-ray data of dtSCT mutants

All four constructs were of sufficient stability to form crystals under nucleating conditions. Crystal trials conducted using the TOPS ('TCR/pMHC Optimized Protein crystallization Screen') screen (Bulek et al. 2012) at 21 °C yielded crystals within weeks. The successful crystallisation conditions for each structure can be found in Table 3. Crystals were approximately 200 x 100 x 100 µm in size and diffracted to 1.9-2.4 Å resolution (Table 3)

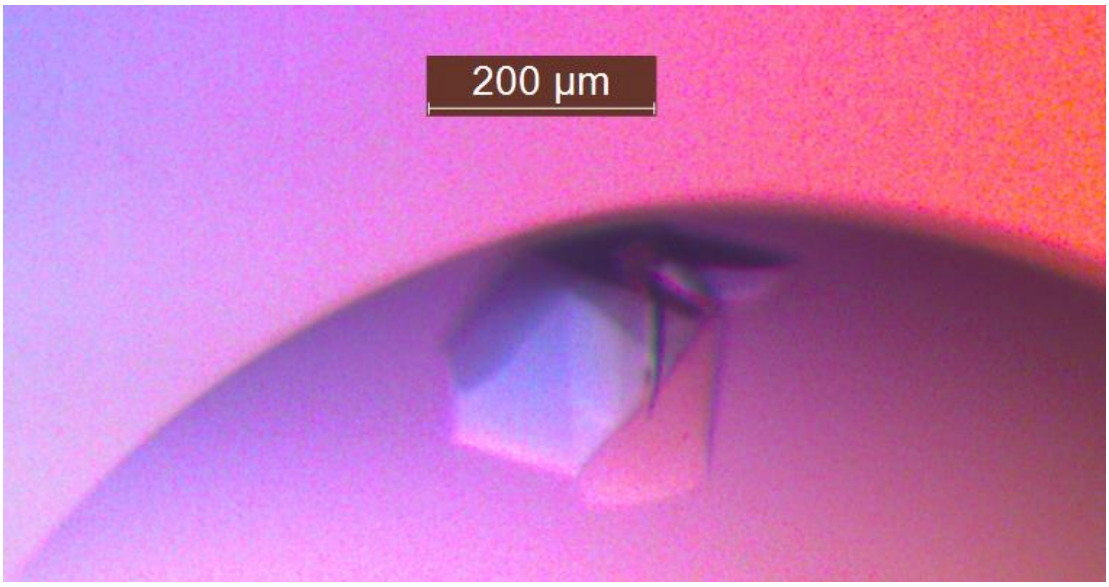


Figure 22 - Crystal of W167A grown with sitting drop vapour diffusion at pH 6.5

Indexing provided unit cell dimensions and a monoclinic $P2_1$ space group (Table 2). This matches the space groups of dtSCT structures published by Mitaksov et al. (2007).

Penalty	Lattice	a	b	c	α	β	γ	Bravais	Possible space groups
0	44	-1	0	0	-1	0	0	aP	P1
0	31	1	0	0	1	0	0	aP	P1
0	**34**	**0**	**0**	**0**	**0**	**1**	**0**	**mP**	**P 1 2 1, P 1 2_1 1**
8.8	39	0	0	0	1	0	0	mC	C 1 2 1
8.8	29	-2	0	0	-1	0	0	mC	C 1 2 1
8.8	38	-2	0	0	1	0	0	oC	C 2 2 2, C 2 2 2_1
14.3	33	0	0	0	0	1	0	mP	P 1 2 1, P 1 2_1 1
23.1	37	0	0	0	1	0	0	mC	C 1 2 1
23.1	28	-2	0	0	-1	0	0	mC	C 1 2 1
23.1	36	-2	0	0	1	0	0	oC	C 2 2 2, C 2 2 2_1

Table 2 - Penalty table for space group indexing.

2.3.3 Refinement and model building

Electron density was clear for unambiguous main and side chain placement in all structured regions of the protein. The flexible linker regions residues 120-144, which was not present in 2QRI, and 14-23 which was present in 2QRI did not have sufficient electron density. Also missing was the C-terminal tail of H-2K^b from residue 419 of the construct. Side chain conformations of the relevant peptide N-terminal and A pocket were readily apparent in the electron density maps. For W167A, of superior resolution, alternative conformations were visible for some side chains. This included N4^P of the peptide. Electron density at mutation sites clearly showed the mutations, with negative regions in the mFo-DFc maps for the large side chains in the structure model prior to the residue being switched to alanine. This was not seen at these residues in the 'wild-type' SL8-SCT structure.

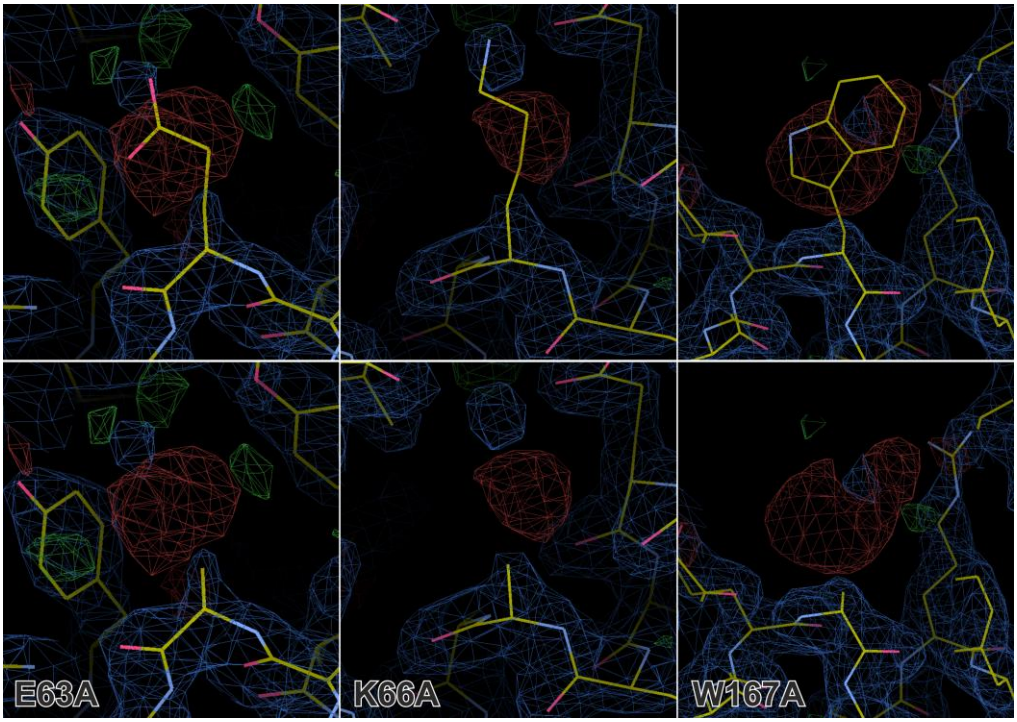


Figure 23 - Electron density of 2mFo-DFc and difference mFo-DFc maps with built structure before (top) and after (bottom) mutation to correct residues.

Structures were solved with a low number of Ramachandran outliers and sufficiently good geometry for them to be submitted to the Protein Data Bank (H. M. Berman et al. 2000). The number of water molecules is shown in Table 3.

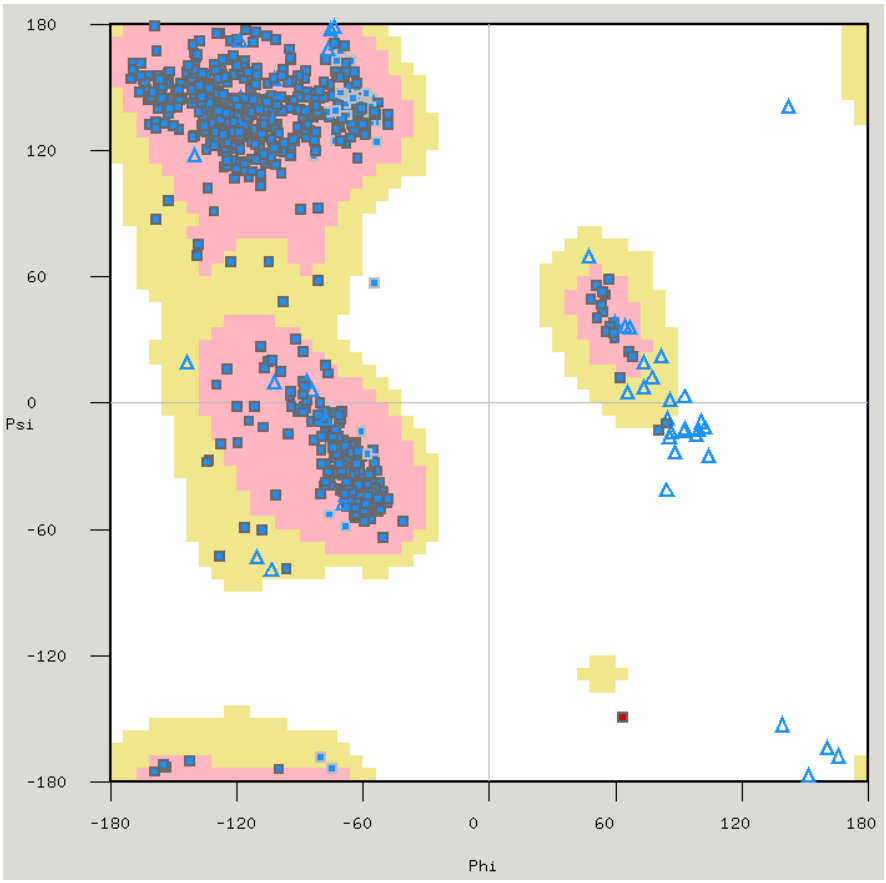


Figure 24 - Ramachandran map for final SL8-SCT structure showing single outlier (E196 in heavy chain region of chain A).

Table 3 - Data collection and refinement statistics for the four dtSCTs.

	SL8-SCT	E63A	K66A	W167A
Space group	P 21	P 21	P 21	P 21
Unit cell				
a, b, c (Å)	66.65, 90.08, 89.33,	67.13, 90.19, 89.45,	66.43, 89.32, 89.40,	66.43, 89.64, 89.45,
α, β, γ (°)	90.00, 111.23, 90.00	90.00, 111.22, 90.00	90.00, 111.24, 90.00	90.00, 111.24, 90.00
Upper resolution limit (Å)	2.27	2.40	2.05	1.9
Lower resolution limit (Å)	90.08	42.52	39.36	61.51
Multiplicity	2.8	3.8	3.8	7.5
Completeness (%)	98.6	97.9	98.2	99.8
R-merge (%)	7.7	10.3	7.4	7.0
Mean (I)/sd (I)	9.8 (2.0)	11.6 (2.1)	9.9 (2.4)	13.8 (2.8)
Reflections	44965	38114	59972	76876
R-factor	0.199	0.198	0.204	0.188
R-free	0.250	0.259	0.241	0.239
Crystallisation conditions	Am. sulphate , PEG 8K, Sod. cacodylate pH 6	PEG 4K (25% w/v), Glycerol (15% v/v), Sod. cacodylate (0.1M), pH 6.5	Am. sulphate (0.2M), PEG 8K (25% w/v), Sod. cacodylate (0.1M), pH 6	Am. sulphate (0.2M), PEG 8K (22.5% w/v), bis-tris (0.1M), pH 6.5
B-factor average, all atoms ($Å^2$)	38.0	38.0	35.0	38.0
N^o. water molecules	419	369	462	726

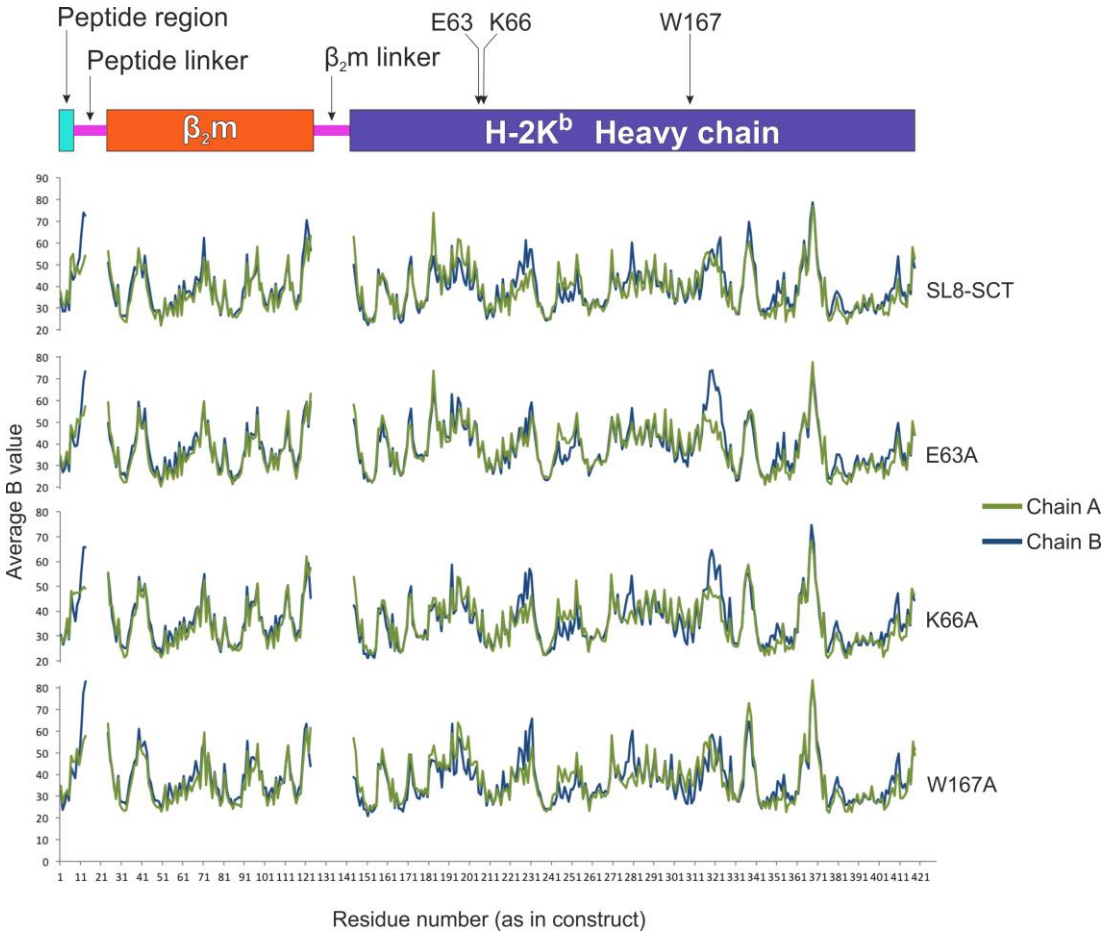


Figure 25 – Plot of overall (main and side chain) B-factor for each residue in the crystal structures with construct schematic above. Consistency between chains and mutants indicates little increase in mobility as a result of mutations.

2.3.4 Structures of dtSCT mutants

The structures are similar to those published by Mitaksov et al. (2007) such as 2QRT, which is also a dtSCT form of SIINFEKL and H-2K^b with the peptide region C-terminal constrained by a disulphide bond to prevent exit of the peptide. Unlike those structures the structures presented here contain A pocket mutations and are the first structures to our knowledge to do so. This is a unique approach to modulating a hypothetical interaction between ERAP1 and MHC. It also informs on how the N-terminal is held within the A pocket and the specific interactions which create this part of the binding site.

The dtSCTs are as previously described by Mitaksov et al. (2007). The complex has a 15 residue chain linking the C-terminus of the ovalbumin-derived SIINFEKL peptide to human β_2m and a 20 residue chain connecting the C-terminal of β_2m to the start of the heavy chain sequence at the base of the binding groove. In addition a Y84C mutation and a cysteine at position 2 in the peptide- β_2m linker (position 10 in the construct) provided a disulphide bond between the binding groove F pocket and the C-terminal of the peptide.

Unlike the dtSCT structures published by Mitaksov et al. (2007), the first linker region was not included in the structure model due to ambiguous electron density and difficulty placing the backbone chain within it. The areas flanking this linker have high B-factors (Figure 25) which is expected because it is a flexible region. The second linker region was also not resolved in the electron density, and not modelled as is the case for the Mitaksov et al. (2007) dtSCT structures. This flexibility indicated by this lack of clear placement in the electron density suggests the protein is not being constrained beyond normal parameters by these linkers.

2.3.5 Similarity of dtSCTs to H-2K^b complexes

A pocket mutant structures showed overall structural similarity to the wild-type complex of heavy chain, β_2m and peptide (see Figure 26). All four structures have a β_2m region with an identical structure to wild-type human β_2m comprised of six antiparallel beta strands in groups of three forming two beta sheet faces. The α3 region of the H-2K^b heavy chain similarly has seven antiparallel beta sheets and a small helical turn from residues 254-256 which are seen in our structures. The binding groove had similar structure to the wild-type with the base of seven antiparallel beta strands and the α1 and α2 helices on either side of the binding groove.

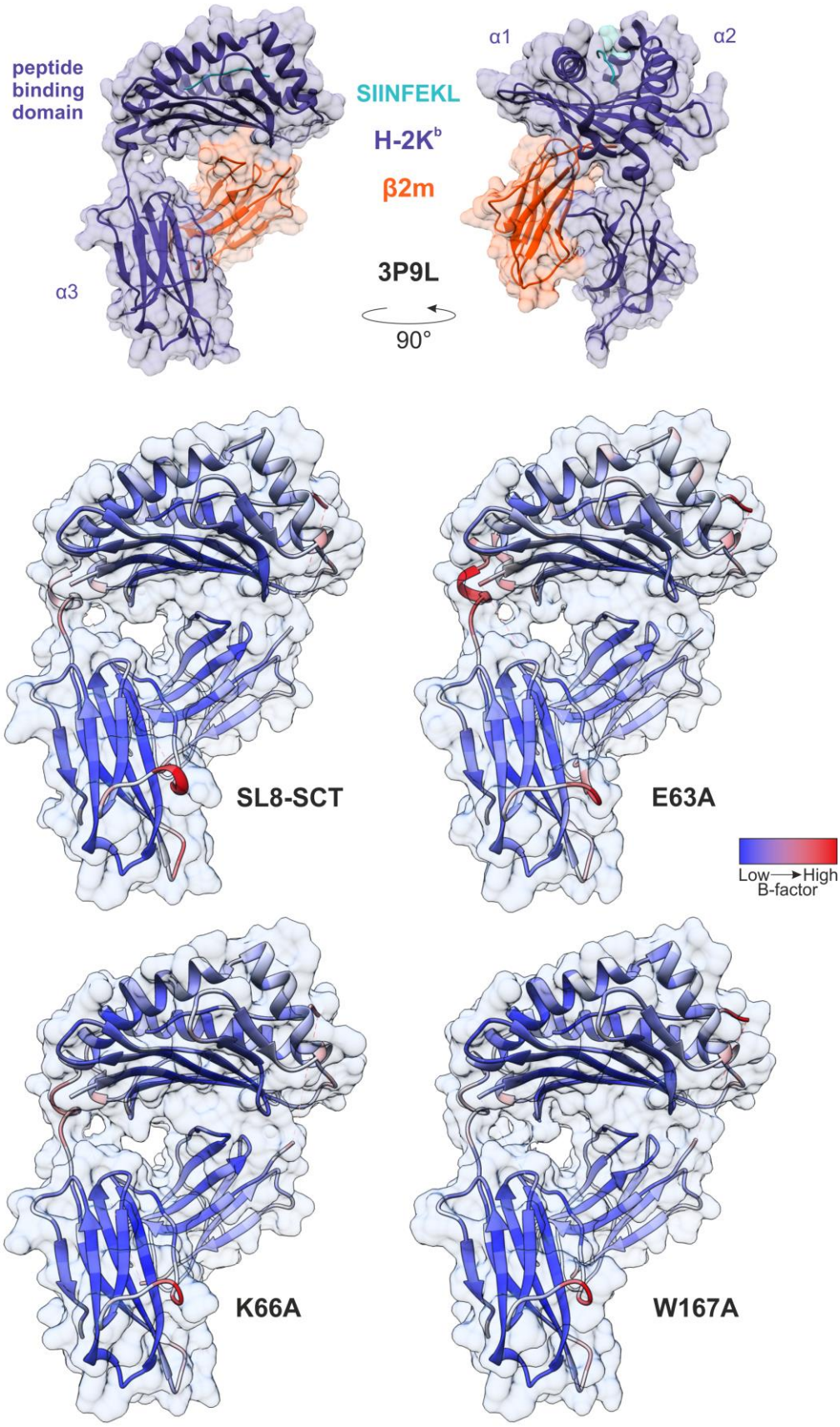


Figure 26 – Overall structures of dtSCT mutants as ribbon cartoon shown within transparent surface representation. Ribbons coloured by average B-factor. For comparison 3P9L complex (solved by Denton et al. 2011) is shown above with peptide, β_2m and H-2K^b indicated. Figure created in UCSF Chimera.

Within the binding groove of all four constructs the peptide backbone adopted the placement seen in published structures of H-2Kb complexed with SIINFEKL, such as 3P9L (see Figure 27). For the 'wild-type' SL8-SCT, binding of the peptide conformed to complexed trimeric peptide:H-2K^{b}:β_2m. For all constructs peptide I2, I3, F5 and L8 side chains were angled towards the beta-sheets at the base of the binding groove whereas N4, E6 and K7 pointed outwards. At the C-terminus, Y84C in the heavy chain replaced C-terminal interactions with the stable disulphide bond. There was little apparent difference in the more rigid parts of the protein with overall structures being similar (Figure 26). Cα RMSDs of the heavy chain region of the dtSCTs against an H-2K^{b} complex with SL8 (3P9L) were between 0.572 and 0.634 Å (see Table 4).

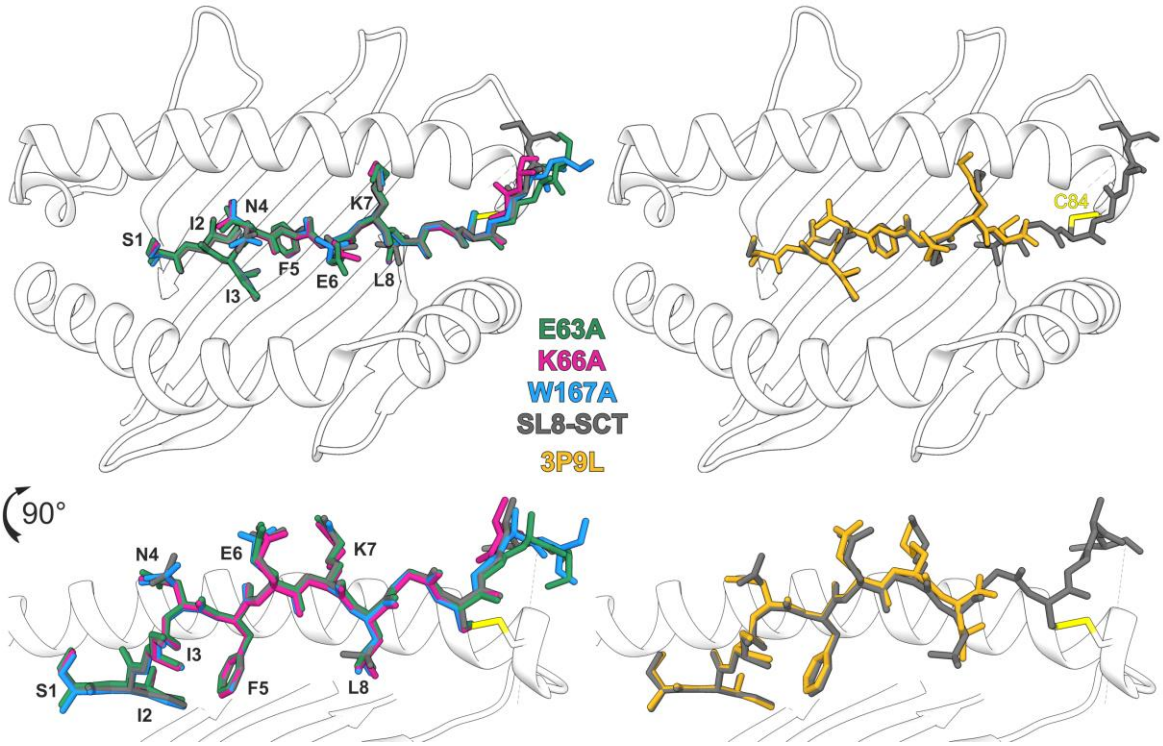


Figure 27 – Position of bound SIINFEKL peptide within the peptide binding groove for different structures. Top: aerial view. Bottom: side view from α2 helix (removed for clarity). Peptide backbones and side chains show little variation despite presence/absence of linkers and mutations. 3P9L solved by Denton et al. (2011). Figure created in UCSF Chimera.

Table 4 – Calculated alpha carbon RMSDs for dtSCTs and related structures. 2QRT (Mitaksov et al. 2007) *is a dtSCT (identical in sequence to SL8-SCT) and 3P9L* (Denton et al. 2011) *is a SL8/H-2K^b/β_2m complex*

Cα RMSDs between structures (Å)

	SL8-SCT	**E63A**	**K66A**	**W167A**
2QRT (whole construct, 388 residues)	0.425	0.484	0.336	0.401
2QRT (peptide region, 8 residues)	0.317	0.359	0.245	0.236
2QRT (heavy chain region, 276 residues)	0.439	0.493	0.343	0.414
3P9L (heavy chain region, 298 residues)	0.634	0.572	0.601	0.612
3P9L (peptide region, 8 residues)	0.175	0.297	0.205	0.310
SL8-SCT (whole construct, 388 residues)		0.464	0.335	0.373
SL8-SCT (heavy chain region, 276 residues)		0.436	0.329	0.346
SL8-SCT (peptide region, 8 residues)		0.246	0.224	0.214

2.4 Structure analysis and discussion

2.4.1 Effect of mutations on the F pocket

Introduction of a disulphide trap between the F pocket and peptide C-terminus is a strategy which has been used previously for single chain trimers as a means of preventing peptide escape and contributing to structural stability (Mitaksov et al. 2007). A Y84C mutation and the inclusion of a cysteine in the peptide-β_2m linker enables formation of a stabilizing cysteine bond (Figure 28). Investigations of this mutation have shown that it increases stability of the peptide-region binding and prevents binding groove entry by competitive peptides (Truscott et al. 2007). Addition of the peptide-β_2m linker and disulphide trap were reported not to have any major structural impacts. The first glycine of the peptide-β_2m linker ($G9^p$) was reported to replace a peptide C-terminal oxygen with its amide nitrogen in the same position and conserved water molecules in the F pocket were retained (Mitaksov et al. 2007). The authors also described an additional hydrogen bond between the peptide-β_2m linker N-terminal glycine and K146, which usually forms a hydrogen bond with Y84 in the wild-type structure and a new F pocket water molecule exhibiting hydrogen bonds with T80 and $C10^p$. They therefore asserted that in addition to electron density around the disulphide bond they observed other interactions which could be contributing to a stabilized F pocket binding (Mitaksov et al. 2007).

In order to assess contribution of the disulphide bond to stable peptide binding, Mitaksov et al. used N15 T cell hybridoma activation to determine the ability of the complex to resist competition from binding by VSV8 peptide (a peptide from vesicular stomatitis viral nuclear capsid protein, with peptide sequence RGYVYQGL). This indicated that dtSCT was 10^5 times more resistant to VSV8 binding than wild-type K^b, whereas the SCT without a disulphide trap was 1000 times more resistant to VSV8. Additionally, their thermal stability assays with CD spectroscopy gave the denaturation temperature of 49.55 °C for dtSCT and the lower temperature of 44.16 °C for a SCT without a disulphide trap.

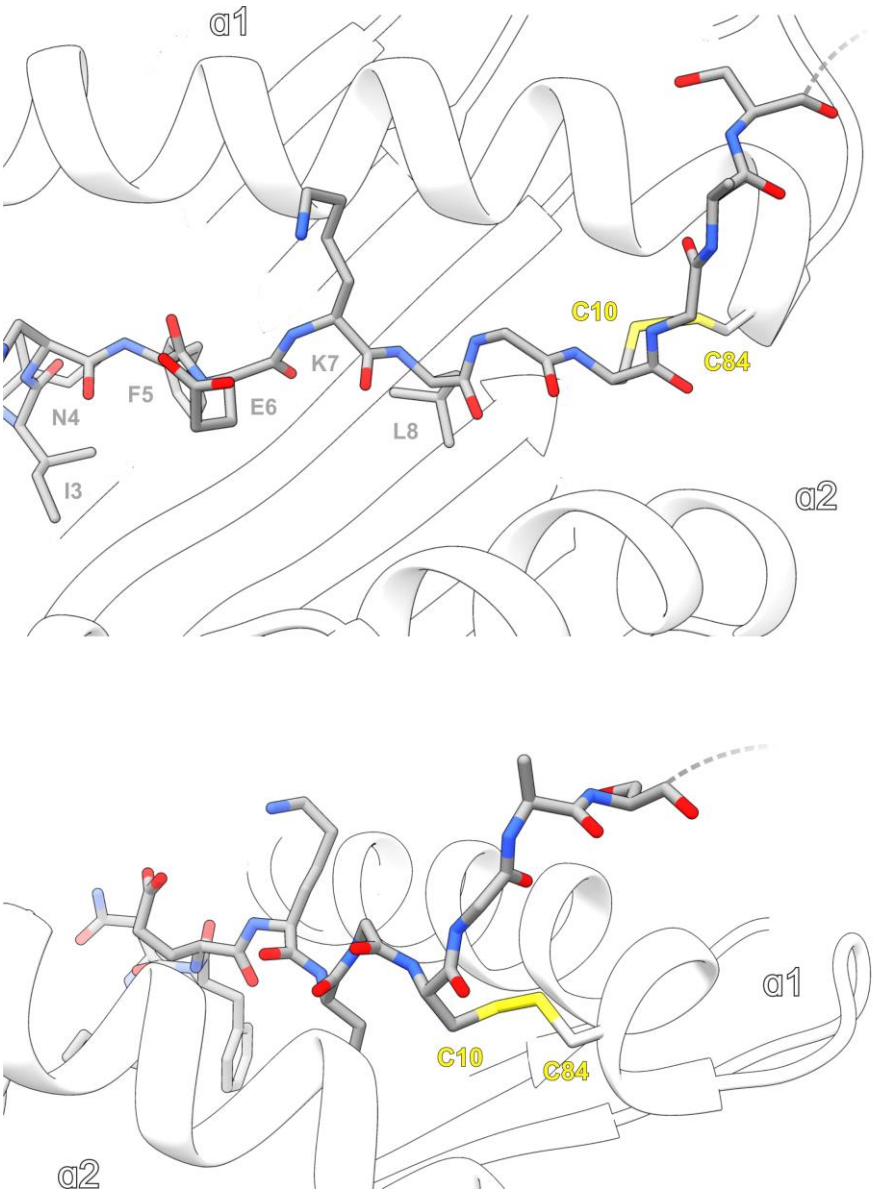


Figure 28 - 'C-terminal' placement of peptide region in binding groove F pocket, showing disulphide trap (yellow) with Y84C mutation. Figure created in UCSF Chimera.

The shape of the F pocket region appeared slightly affected by inclusion of the C-terminal cysteine clamp when compared to an unmutated complex PDB ID 3P9L (Figure 29). The RMSD difference between the SL8-SCT structure and H-2K^b of the 3P9L structure of complexed H-2K^b, SIINFEKL and murine β_2m was 0.634 (see Table 4).

There was less difference in the F pocket between the three point mutants reported here and the similar disulphide trap structure crystallised and reported by Mitaksov et al.

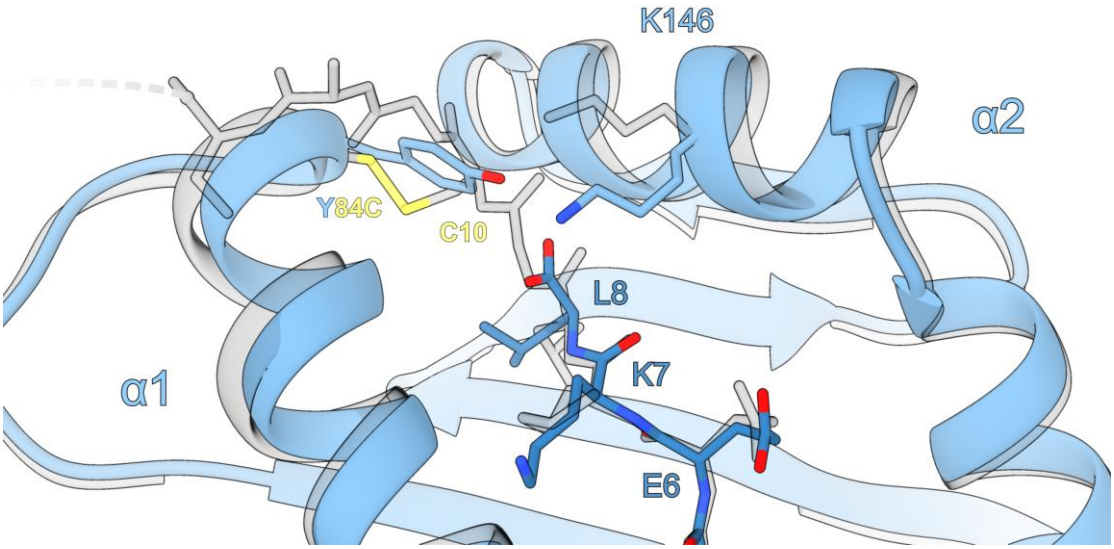


Figure 29 - F pocket of 3P9L (blue) with SL8-SCT as transparent overlay. Differences to side chains are shown, notably K146 and the mutated Y84C. 3P9L solved by Denton et al. (2011).

2.4.2 Binding of Peptide N-terminal

2.4.2.1 Hydrogen bonding

Hydrogen bonding is a phenomenon which has been known of for a substantial length of time but has slightly changed in definition over the years. The classical explanation for it is that disparities in electronegativity between, for example, a covalently bonded oxygen and hydrogen will lead to the electron-withdrawing oxygen holding a negative dipole, and a positive dipole for hydrogen. The O-H bond is therefore polar. This polarity makes it possible for the oxygen to attract a positive dipole hydrogen on a neighbouring molecule and the hydrogen to interact with a negative dipole atom. Although weaker in most cases than covalent bonds, they are usually much stronger (with dissociation energy of 3-5

kcal/mol for classical H-bonding) than Van der Waals interactions, leading them to be referred to as bonds themselves. Nonetheless they vary enormously in strength, length and bond angle. This geometric flexibility allows a single hydrogen to form one, two or three hydrogen bonds (multifurcation) which is often observed in protein structures. In a bifurcated hydrogen bond where one hydrogen bond is shorter than the other, this is referred to as major and the longer bond as minor (Steiner 2002).

The broad nature of hydrogen bonds means many definitions fail to encompass all possible forms. A definition given by Thomas Steiner in his review of the subject is:

"*An X–H···A interaction is called a "hydrogen bond", if 1. it constitutes a local bond, and 2. X–H acts as proton donor to A."*

This appropriately covers the range of possible hydrogen bonds without excluding by specifying geometry or atoms involved beyond hydrogen, but does exclude interactions such as Van der Waals by specifying a hydrogen donor/acceptor system. In 2011 an IUPAC task group assembled to establish a comprehensive definition:

"The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation." (Arunan et al. 2011)

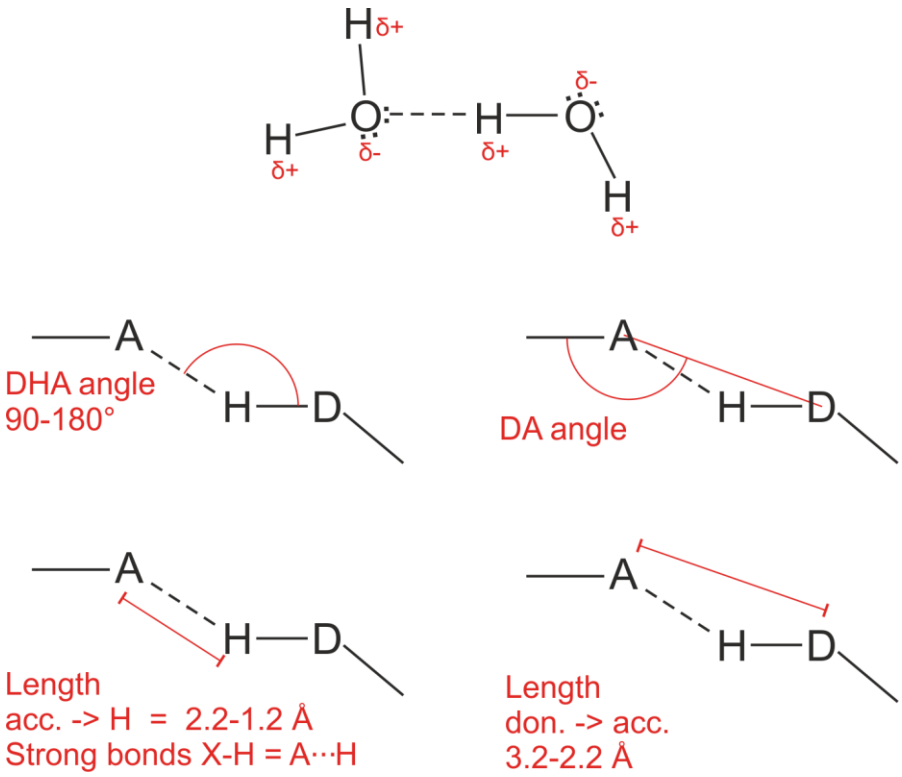


Figure 30 - Hydrogen bonding mechanism and geometry.

For many hydrogen bonds the A->H distance is shorter than the sum of Van der Waals orbitals but this is not always the case. Symmetric hydrogen bonds will have the hydrogen placed at equidistance between donor and acceptor, such that it might be impossible to say which is which. In this case the hydrogen involved can be said to exhibit a transition from hydrogen bond to covalent bond with two atoms (Steiner 2002). Stronger hydrogen bonds commonly exhibit a lengthening of the D-H bond length which is seen in infrared spectroscopy (Arunan et al. 2011).

The optimal DHA angle (see Figure 30) is 180° with the hydrogen bond being stronger the closer to 180° it is, and this is the hydrogen bonding angle most commonly observed in organic and inorganic crystal structures (Steiner 2002). Moreover the closer this angle is to linear, the shorter the H-A distance will be (Arunan et al. 2011). However, the optimal angle will be affected by the types of atoms involved and whether the hydrogen bond is bifurcated or not, as multifurcation will require accommodation of extra hydrogen bonds (Steiner 2002).

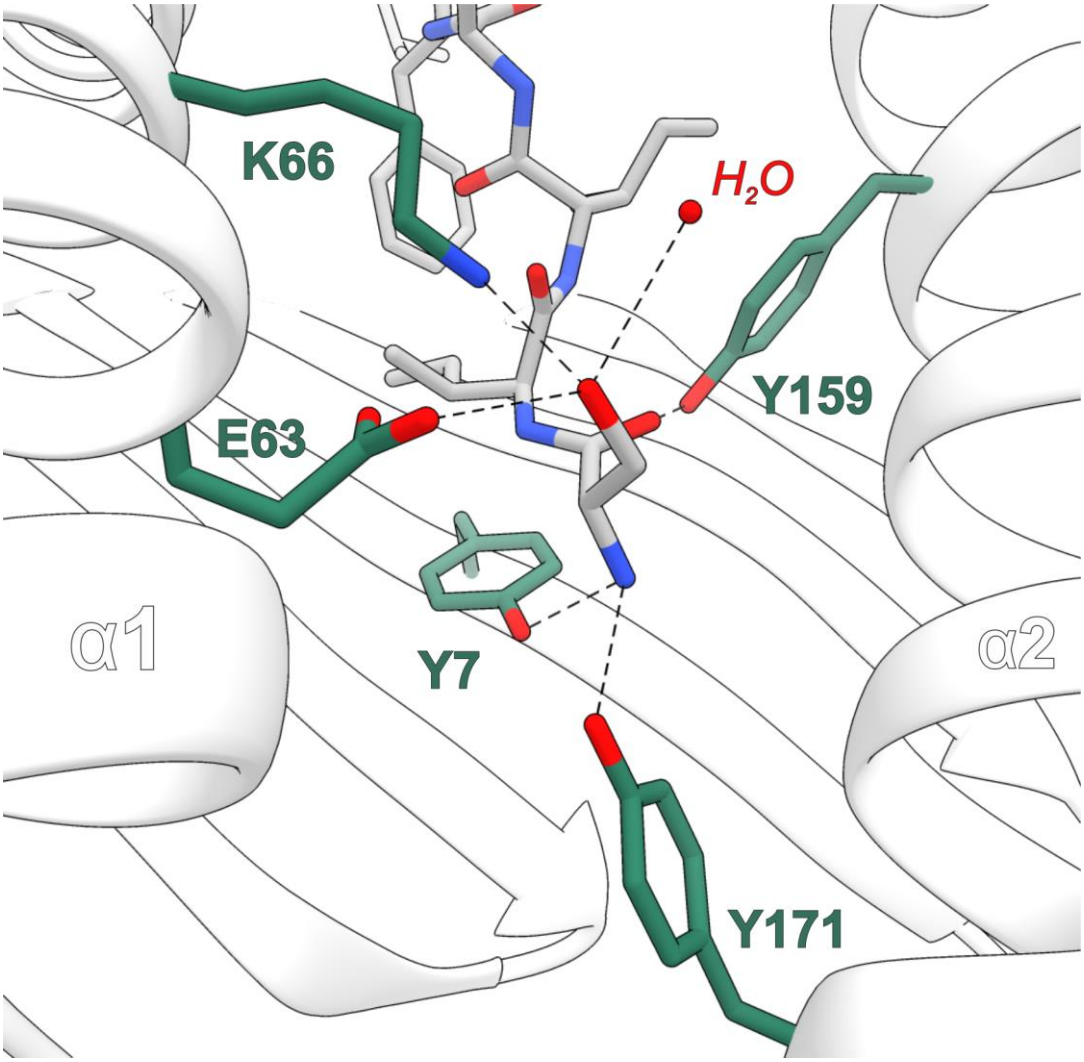


Figure 31 - SL8-SCT 'wild type' mutant hydrogen bonding between peptide N-terminal and A pocket residues. Figure created in UCSF Chimera.

2.4.2.2 N-terminal interactions with the A pocket

As there are already published examples of H-2K^b bound SIINFEKL peptide in the wild-type, complexed form, the hydrogen bonding of the N-terminal within the A pocket is known (Denton et al. 2011; Fremont et al. 1995). Fremont et al. (1995) described hydrogen bonds between the S1^P N-terminal amine and the hydroxyl groups of both Y7 on the heavy chain floor, and Y171. They also described a Y159 hydroxyl to serine carbonyl hydrogen bond, and a 'non-conserved' hydrogen bond between E63 Oε and S1^P hydroxyl. The 2.0Å structure (PDB ID: 3P9L) of H-

2K^b published by Denton et al. (2011) was depicted with similar hydrogen bonding although they did not explain how they had calculated this.

The interactions between the N-terminal and A pocket for the SL8-SCT trimeric form of H-2K^b are much the same as 3P9L due to similar positioning of peptide and side chains (Figure 26). Use of the WHAT IF online web service calculation of optimal hydrogen bonds (Hooft, Sander, and Vriend 1996; Hekkelman et al. 2010) for SL8-SCT yielded similar hydrogen bonds to those described above, with additional hydrogen bonds between the S1^P side chain hydroxyl and both the K66 Nζ and a neighbouring water molecule (Figure 31, Table 5). This water molecule and K66 were also predicted to hydrogen bond to the I2^P backbone NH and carbonyl groups respectively (Figure 32). which matches what Fremont et al. reported (Fremont et al. 1995).

2.4.2.3 Hydrogen bonding and side chain positioning in the mutant A pockets

N-terminal hydrogen bonding for the three mutant structures is predicted by WHAT IF hydrogen bond optimization to retain similarity to the SL8-SCT hydrogen bonding where bonding partners in the A pocket are unaltered. The exception to this is for K66A where no hydrogen bond was predicted for Y171 (Figure 32, Table 5). This hydrogen bond in the other structures is given a poorer quality score (high val.) than many of the other bonds, and may therefore have marginally exceeded the quality cutoff for K66A. Nonetheless, the Y171 hydrogen bond was predicted by the all potential hydrogen bonding WHAT IF calculation for K66A (see Appendices) so the absence is hard to count as significant.

The K66A mutation removes the possibility of the K66A Nζ hydrogen bonds to the S1^P hydroxyl and I2^P carbonyl. The gap created by this mutation not only prevents these interactions but also facilitates inclusion of a non-conserved water molecule (Figure 33). The water molecule appears to take over the hydrogen bond to the I2^P carbonyl from the removed lysine. In addition, another

water molecule appeared in the vicinity of E63, possibly as a result of the gap created.

E63A had greater apparent impairment of hydrogen bonding than K66A, because there was no evidence of additional water molecules adopting the created space. This may be in part due to poorer resolution for this structure than the other mutants and SL8-SCT. Perhaps due to this gap the K66 sidechain appeared in a different rotamer conformation in E63A (Figure 33) and was angled with Nζ towards $I2^P$ rather than between the first two N-terminal residues.

The W167A crystal structure presented an A pocket which most resembled SL8-SCT of the three mutants. The predicted optimal hydrogen bonding between A pocket and peptide N-terminal was similar between these two structures. An additional non-conserved water molecule in W167A was suggested to hydrogen bond with the $S1^P$ side chain hydroxyl. This would push the number of potential hydrogen bonding partners up to four for this moiety and therefore this hydrogen bond may be either an occasional bonding partner or enabled by multifurcation. Two other non-conserved waters were placed in the space vacated by the tryptophan mutation (Figure 34). As there is no evidence of the W167A mutation impacting on the hydrogen bonding network it must be assumed the role of this tryptophan is other than that of similarly conserved A pocket residues. It is likely to form a large steric block at the end of the binding groove which may help prevent peptide escape.

In addition to the key A pocket side chains involved in hydrogen bonding there were differences in the positions of local and peptide side chains, the causes of which are unclear. R62 is predicted to hydrogen bond to E63 and is angled towards the α2 helix in SL8-SCT (Figure 33). In E63A it appears in a similar conformation but hydrogen bonding to local waters instead. For both K66A and W167A however it points down the A pocket, away from the peptide. Although this is a change feasibly induced by the K66A mutation, it is less clear to see how this would be caused by W167A. The electron density for this side chain was

poorer than the electron density for other A pocket residues suggesting it may be a more flexible residue.

Another notable difference between structures was the $N4^P$ side chain. In SL8-SCT this was directed towards the α2. In E63A and K66A it was instead pointed towards α1 and the A pocket. In W167A electron density was observed for both conformations and it appeared to exist in an approximate 50:50 occupancy of the two states. What this signifies is unclear: W167A, being a higher resolution structure, has many dual conformation side chains which are not apparent in the electron density for the other structures. It is possible this side chain has some inherent flexibility. Position 4 in other octapeptides bound by H-$2K^b$ is known to sometimes provide a point of recognition for TCRs (Reiser et al. 2000). The published investigation into this by Reiser et al. states H-$2K^b$ bound dEV8 showed an interaction between K4 of the peptide and the 2C TCR whereas BM3.3 TCR interacted strictly with N6 and T7 with a substantial distance to position 4 which negated possibility of direct interaction. No TCR-SL8-H-$2K^b$ structure has been published, so the importance of $N4^P$ is unknown.

2.4.2.4 Crystal structure surface hydropathy

The hydropathy of the A pocket region (Kyte and Doolittle 1982) is not totally altered by the mutations, as you would expect and the overall surface hydropathy has not been changed (Figure 35). Despite the E63A mutation the hydrophilic area on the α1 helix remained. This is probably because E63A is fairly buried and could only contribute to a small hydrophobic area within the binding groove. Within the A pocket additional space was created by the mutation, and increased hydrophobicity as well. K66A by contrast faces outwards, and this mutation is shown to create a hydrophobic patch on the surface. W167A also results in a hydrophobic area, and the gap resultant from removing the large tryptophan side chain is clear in the surface rendering (Figure 35).

Given that the hydropathic effects of the K66A and W167A mutations are on the outwards-facing surfaces of the peptide binding region, they may have an effect on TCR binding. The gap created by W167A could feasibly alter peptide binding, TCR recognition, and stability of the protein, although it is hard to say for certain that this translates to local changes. Ensemble refinement indicates possible increase of side-chain flexibility for $S1^{P}$ and Y171 in the W167A mutant (Figure 36) but the crystal structure otherwise indicates no significant changes here (Figure 26).

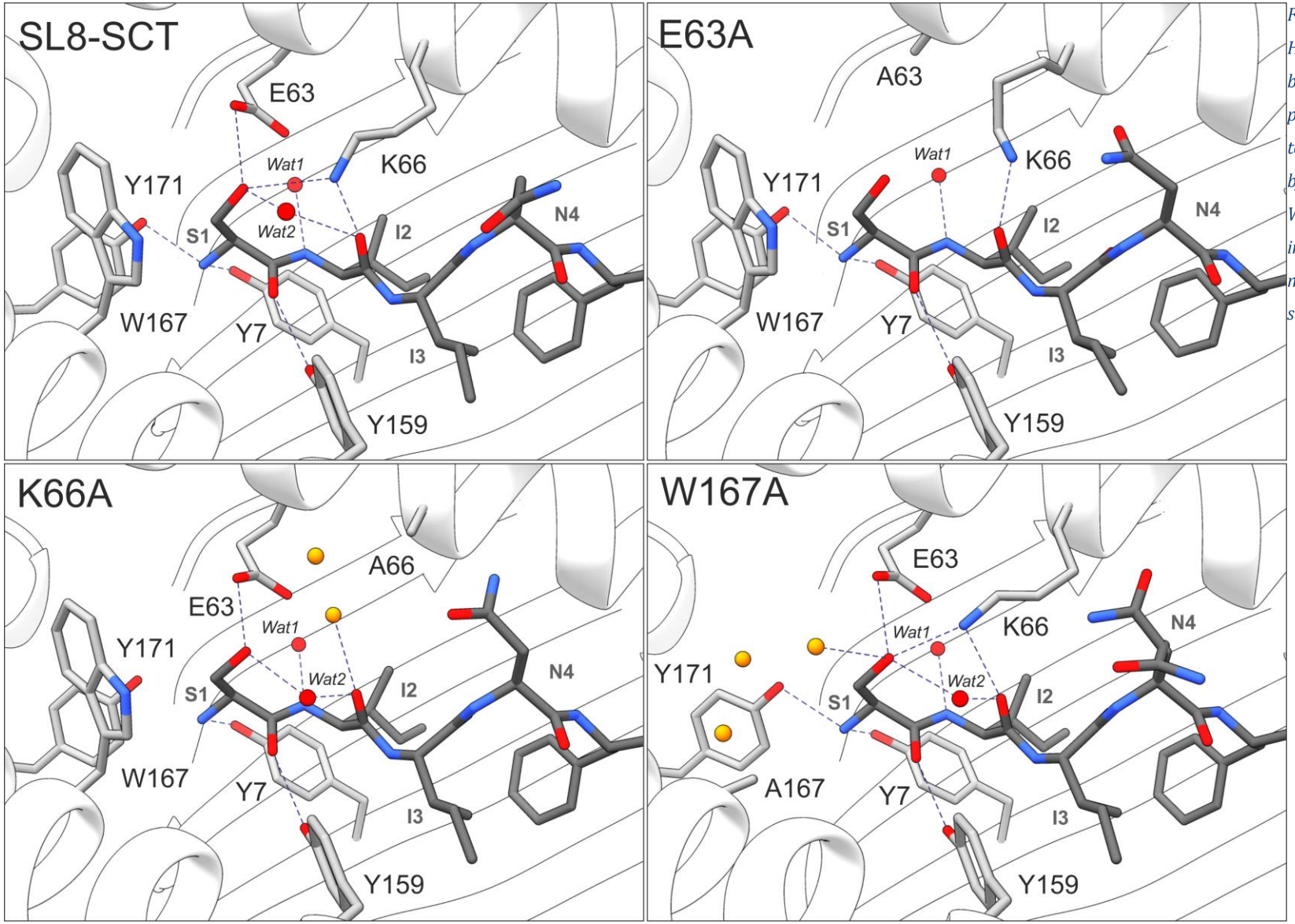


Figure 32 - Hydrogen bonding of peptide N-terminal (shown by dotted lines). Waters introduced by mutations are shown in orange.

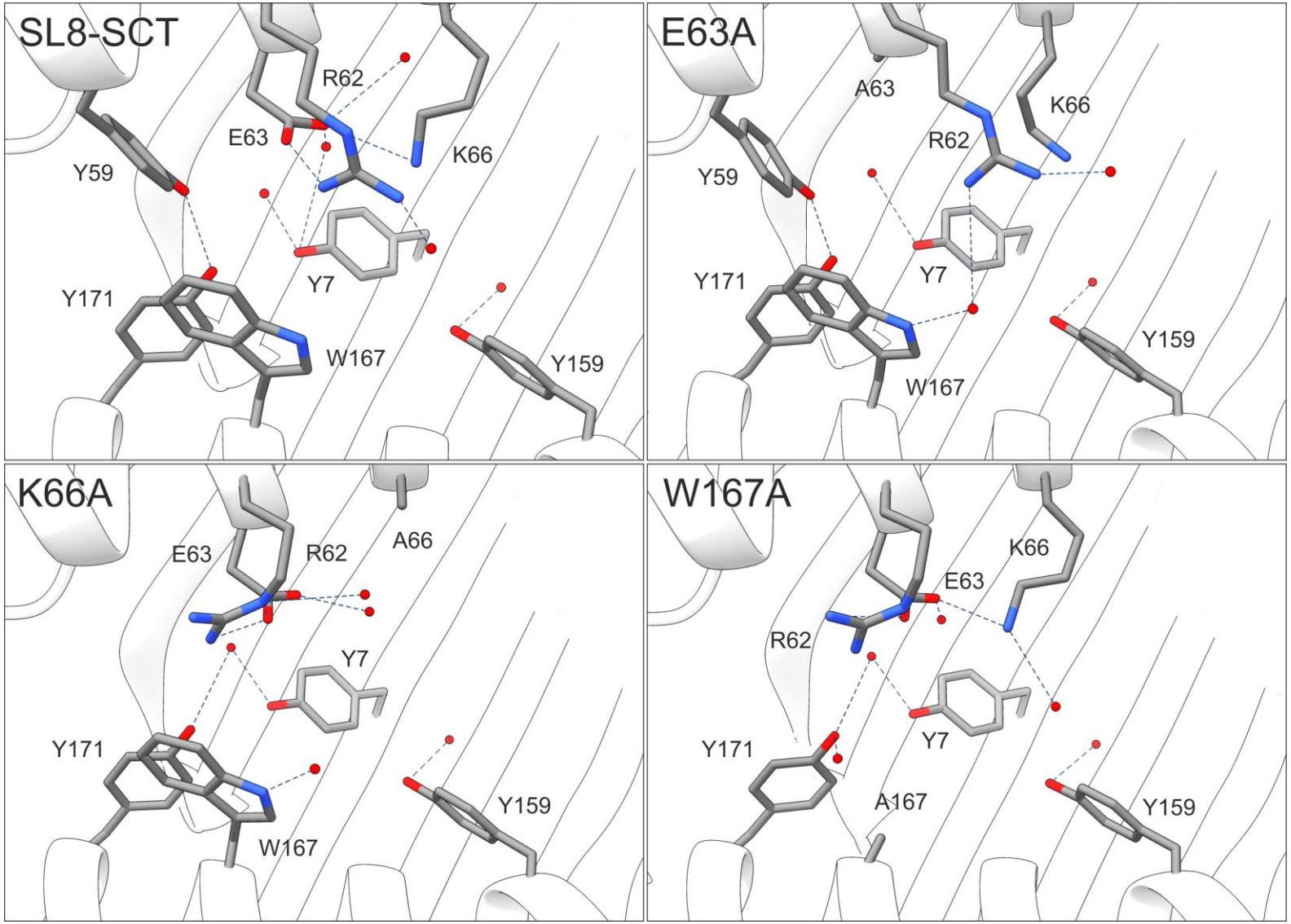


Figure 33 – Conformation of A pocket residue side chains and hydrogen bonding to neighbours.

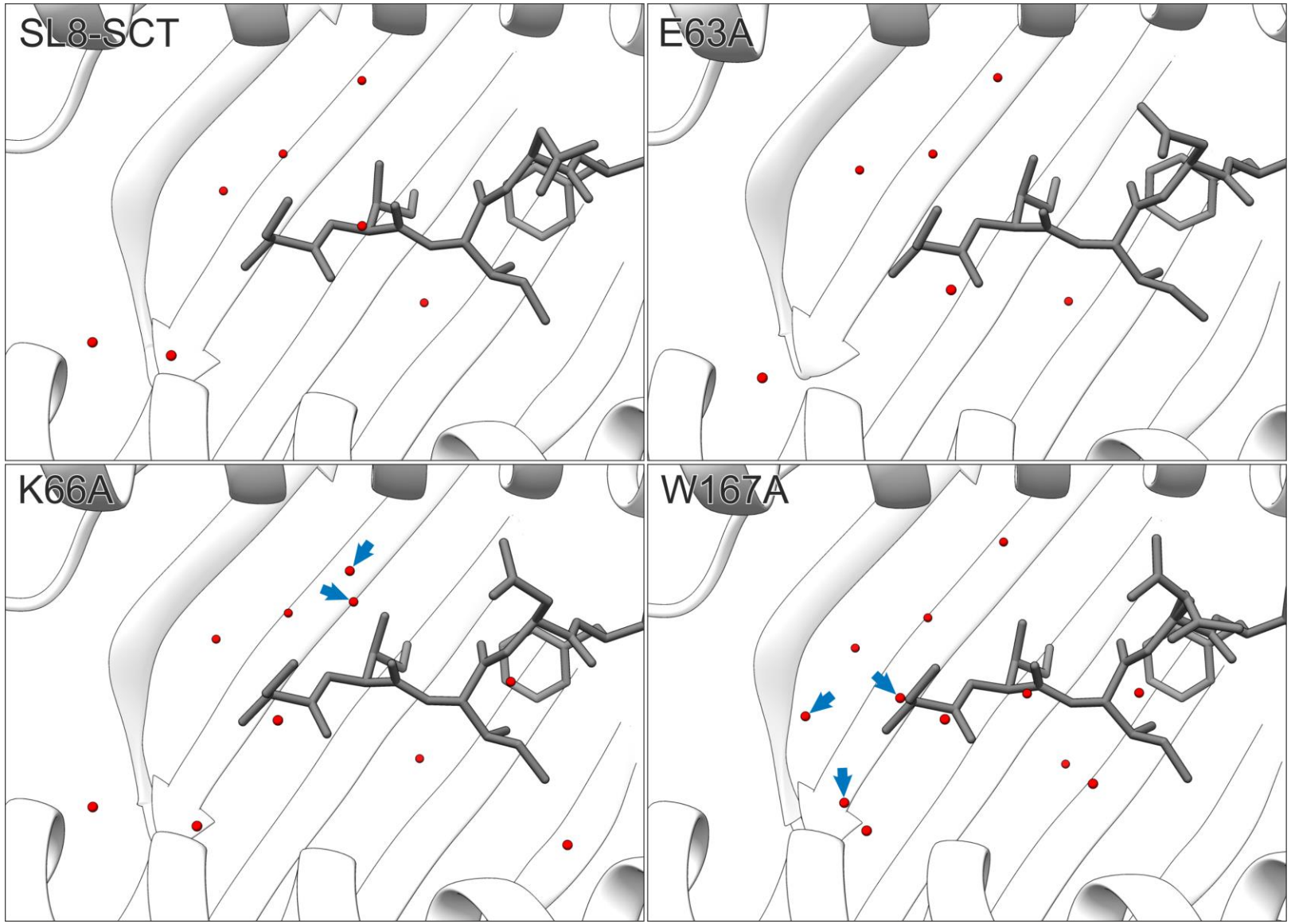


Figure 34 - Position of water molecules in binding groove of dtSCT mutants. Blue arrows indicate waters introduced where mutations have created spaces.

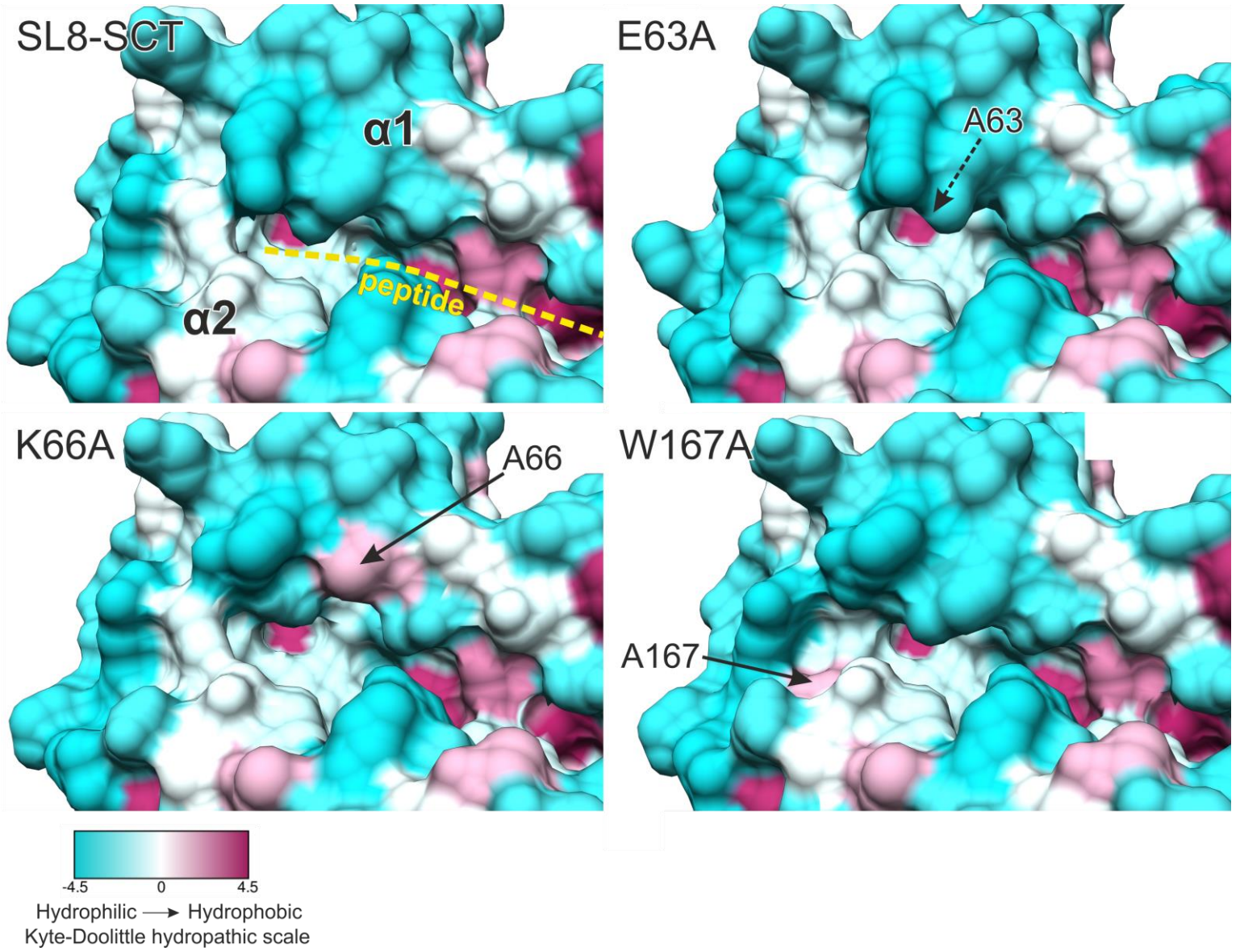


Figure 35 - Hydrophobicity surface rendering of binding grooves of dtSCT mutant structures showing changes to surface hydrophobicity introduced by mutations. Peptide position shown by yellow line.

Table 5 – Calculated optimal hydrogen bonding of peptide N-terminal residue in dtSCT mutant structures, calculated by WHAT IF (Hekkelman et al. 2010; Hooft, Sander, and Vriend 1996).

construct	Donor residue	Donor ID	Chain	donor	Acceptor Residue	Acceptor ID	Chain	Acceptor	Val	D->A distance	DHA angular error
SL8-SCT	1 SER	1p	A	N	283 TYR	171h	A	OH	0.75	2.83	20.02
	1 SER	1p	A	N	119 TYR	7h	A	OH	0.584	2.64	25.38
	178 LYS	66h	A	NZ	1 SER	1p	A	OG	0.345	3.05	64.34
	178 LYS	66h	A	NZ	1 SER	1p	A	OG	0.444	3.05	72.56
	271 TYR	159h	A	OH	1 SER	1p	A	O	0.65	2.73	16.26
	389 SER	1p	B	N	670 TYR	171h	B	OH	0.636	2.87	37.36
	389 SER	1p	B	N	506 TYR	7h	B	OH	0.84	2.93	15.57
	389 SER	1p	B	OG	562 GLU	63h	B	OE2	0.453	2.33	18.88
	565 LYS	66h	B	NZ	389 SER	1p	B	OG	0.389	3.23	39.26
	658 TYR	159h	B	OH	389 SER	1p	B	O	0.569	2.29	12.05
	776 HOH	366	C	O	389 SER	1p	B	OG	0.408	2.79	68.88
E63A	1 SER	1p	A	N	283 TYR	171h	A	OH	0.769	2.84	24.2
	1 SER	1p	A	N	119 TYR	7h	A	OH	0.563	2.72	24.73
	271 TYR	159h	A	OH	1 SER	1p	A	O	0.636	2.83	16.13
	389 SER	1p	B	N	507 TYR	7h	B	OH	0.808	3	14.29
	659 TYR	159h	B	OH	389 SER	1p	B	O	0.689	2.51	9.77
K66A	1 SER	1p	A	OG	175 GLU	63h	A	OE2	0.676	2.71	32.07
	271 TYR	159h	A	OH	1 SER	1p	A	O	0.712	2.63	6.57
	777 HOH	108	S	O	1 SER	1p	A	OG	0.749	3.12	23.32
	389 SER	1p	B	N	507 TYR	7h	B	OH	0.654	2.76	16.76
	659 TYR	159h	B	OH	389 SER	1p	B	O	0.734	2.55	2.27
W167A	1 SER	1p	A	N	283 TYR	171h	A	OH	0.802	2.82	18.55
	1 SER	1p	A	N	119 TYR	7h	A	OH	0.715	2.67	20.94
	1 SER	1p	A	OG	777 HOH	256	S	O	0.317	3.18	54.83
	1 SER	1p	A	OG	777 HOH	795	S	O	0.244	3.1	71.62
	178 LYS	66h	A	NZ	1 SER	1p	A	OG	0.411	2.59	63.85
	271 TYR	159h	A	OH	1 SER	1p	A	O	0.62	2.51	14.99
	389 SER	1p	B	N	671 TYR	171h	B	OH	0.766	2.87	21.85
	389 SER	1p	B	N	507 TYR	7h	B	OH	0.711	2.74	26.29
	389 SER	1p	B	OG	563 GLU	63h	B	OE2	0.515	2.59	42.71
	659 TYR	159h	B	OH	389 SER	1p	B	O	0.664	2.4	3.1
	777 HOH	327	S	O	389 SER	1p	B	OG	0.621	2.74	18.81
	566 LYS	66h	B	NZ	389 SER	1p	B	OG	0.505	3.31	39.24

2.4.2.5 MD refinement structure ensembles of SCT mutants

An additional ensemble refinement step from the deposited structure coordinates and electron density offered further analysis of flexibility within the structures which can be subjectively observed within the electron density. SL8-SCT and E63A were fitted with 30 ensemble models by the Phenix refinement, and K66A and W167A had 32 and 34 structures respectively. The ensembles showed improved R work and R free for all but W167A R work, indicating this refinement may have been unsuccessful for W167A (Table 6). This might partially be explained by the large number of alternative conformations already built into the W167A model which had lowered the R work and R free.

The ensemble structures show a predictable flexibility in solvent-exposed side chains and less mobility in the peptide binding region. There is little change to the backbone position except for modelled linker regions. Side chains of the peptide and A pocket are compared in Figure 36. These indicate substantial ranges of motion of $E6^P$ and $K7^P$ but much less for the peptide N-termini. It is also suggestive of possible flexibility for the $N4^P$ sidechain in all structures rather than only W167A, which would explain the alternative conformations modelled for K66A and E63A which were unalike the SL8-SCT $N4^P$.

Differences of A pocket side chains in these ensembles are difficult to ascertain. What look to be significant differences within one chain of a mutant are in most cases inconsistent when taking the other chain into account. Though this may be caused by differing crystal contacts it is probable that the ensemble refinement, being more a measure of uncertainty than an accurate depiction of structural motions, cannot give further information on mobility of this area. It does suggest, given the limited range of ensemble positions adopted by the A pocket side chains, that they have been appropriately modelled within the electron density and are unambiguously positioned in the crystal structures (that were solved prior to ensemble refinement). The changes to the hydrogen bonding network and peptide binding groove shape might then be enough to impair strength of

binding without causing enough flexibility to be apparent in the electron density and ensemble refinement.

Table 6 - Ensemble refinement quality statistics

		SL8-SCT	E63A	K66A	W167A
Ensemble size		30	30	32	34
R work					
	Previous	0.199	0.198	0.204	0.188
	New	0.196	0.186	0.190	0.191
R free					
	Previous	0.250	0.259	0.241	0.239
	New	0.248	0.250	0.233	0.235
Ramachandran outliers (%)		3.77	4.18	2.57	2.91
Ramachandran favoured (%)		88.29	86.86	90.87	90.59
Mean RMSD per structure					
	Bond (Å)	0.0076	0.0074	0.0074	0.0078
	Angle (°)	1.0520	1.0108	1.0434	1.1057
	Chirality (Å^3)	0.0535	0.0522	0.0546	0.0565
	Planarity (°)	0.0075	0.0074	0.0075	0.0082
	Dihedral (°)	18.919	18.305	18.152	19.775

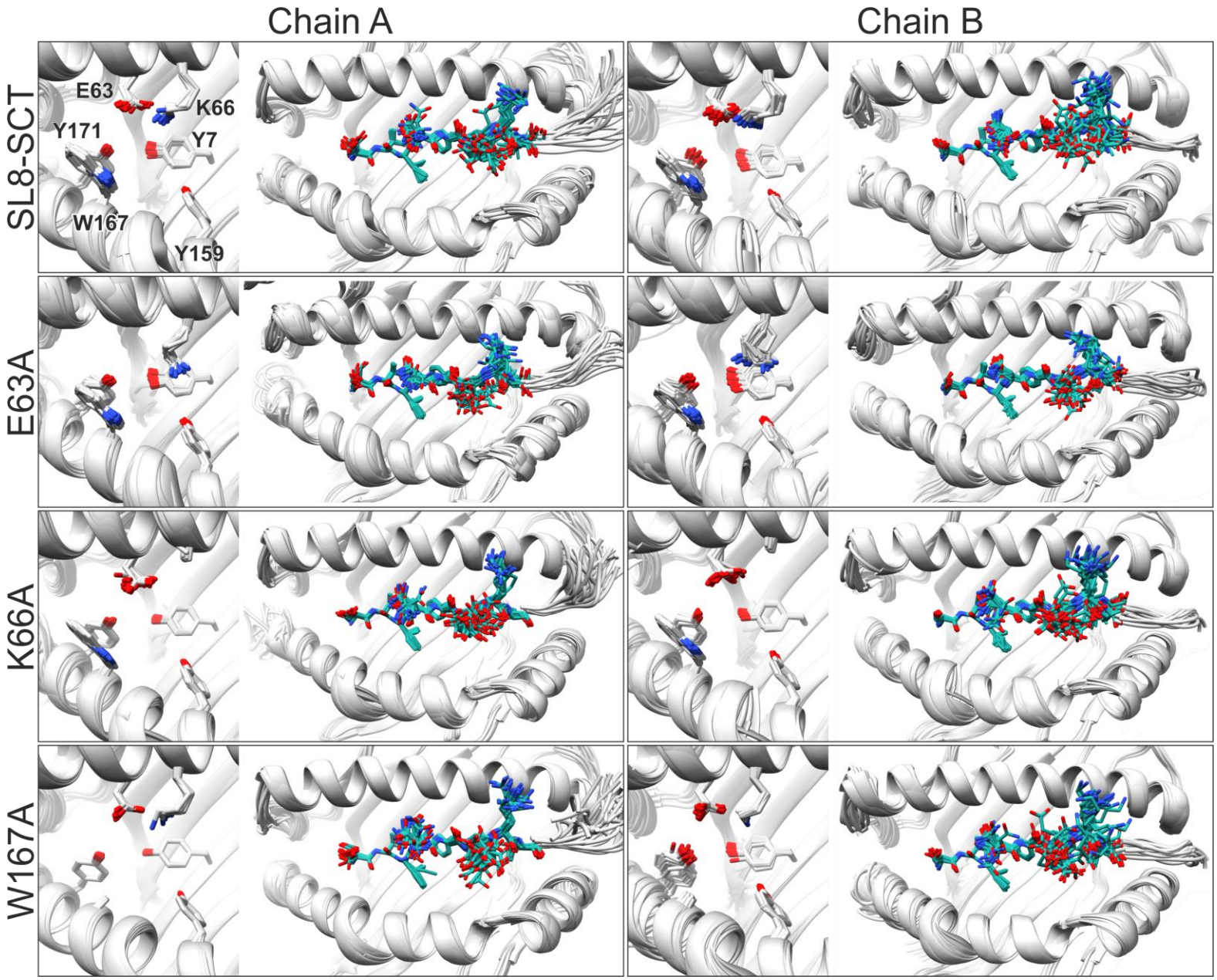


Figure 36 – Ensemble refinement of dtSCTs comparison between mutants and chains, showing close view of A pocket residues in addition to peptide within binding groove. No major side chain position differences are apparent as a result of A pocket mutations.

2.4.3 Functional effects of mutations

The crystal structures described here (those determined prior to ensemble refinement) have been deposited to the protein databank (Berman, Henrick, and Nakamura 2003) under PDB codes 5OQH, 5OQF, 5OQG and 5OQI. They were published alongside MD simulations of peptide escape and testing of ERAP1 activity on the bound peptide (Papakyriakou et al. 2018).

The ability of ERAP1 to access and trim SIINFEKL when bound to an A pocket mutant was tested by Dr Emma Reeves and Dr Edd James using B3Z assays and the 25-D1.16 antibody, which recognises H-2K^b/SIINFEKL in dtSCT form, but not H-2K^b/IINFEKL, indicating overtrimming by ERAP1 has occurred. This was used in HeLa ERAP1 knockouts, with 25-D1.16 highlighting similar expression levels of dtSCTs in all cells. When wild-type ERAP1 was expressed in the HeLa cells alongside the A pocket mutant dtSCTs, reduction in 25-D1.16 binding showed a drop of around 35-40%. This was not the case for SL8-SCT. An anti-mouse β_2m antibody indicated cell surface levels of dtSCTs had not changed, indicating the peptides being presented were probably shortened by ERAP1. E63A was the best of the three mutants for recognition by B3Z T cell hybridomas (all three showed poorer stimulation than SL8-SCT) and was tested with this as well. B3Z (as with 25-D1.16) responds to SIINFEKL but not IINFEKL, and in HeLa ERAP1 knockout cells, showed a 35% decrease in stimulation for E63A when ERAP1 was introduced, and no decrease for SL8-SCT. HeLa cells without an ERAP1 knockout showed reduction for both SL8-SCT (-15% stimulation) and E63A (-35% stimulation) upon transfection of ERAP1. The mutations in the A pocket seem to increase access for ERAP1 trimming of the peptide N-terminal which the unimpaired A pocket of SL8-SCT permits less readily (Papakyriakou et al. 2018).

The reason for the peptide N-terminal being accessed by ERAP1 more easily in the mutants despite the crystal structures suggesting no major structural changes was investigated by Dr Athanasios Papakyriakou, who used the crystal structure coordinates for MD simulation. RMSD values comparing the peptide binding domains Cα positions between simulation and crystal structure were

around 1.2 Å for SL8-SCT. This compares with RMSD values of 1.7-1.8 Å for the mutants, suggesting lower stability and greater fluctuation. RMSDs for the peptide were low for E63A and K66A (below 1 Å) but W167A had a much higher peptide RMSD of around 2 Å. These values do not match those of mutant/SL8-SCT peptide RMSDs described here, where E63A had the highest RMSD (0.246 Å) and W167A had the lowest (0.214 Å) because the crystal structures show little flexibility for this region as indicated by the low B-factors and ensemble refinement. MD simulations of the N-terminal showed mutants had lower free-energy barriers to N-terminal dissociation than SL8-SCT, of around 3-8 Kcal/mol. The free-energy barrier was lowest for E63A, then for W167A, with K66A the highest. Untethering the C-terminal made no meaningful difference to this (Papakyriakou et al. 2018).

The crystal structures of these dtSCT mutants show no major structural change, but rather small differences to the shape of the A pocket, and the hydrogen bonding within it. Though these changes could easily have no significant biological effect, B3Z assays and use of the 25-D1.16 antibody indicated that they permit over-trimming of the peptide by ERAP1, despite SIINFEKL being optimal, and over-trimming is not observed in SL8-SCT unless both endogenous and transfected ERAP1 are present. This matches a previous observation that ERAP1 can trim wild-type dtSCTs, although in that case the ERAP1 variant was a hyper-trimmer (Reeves et al. 2013). An attempt to replicate this in vitro using NMR reaction monitoring of E63A and wt ERAP1 was unsuccessful, possibly because conditions were unsuitable or a hyper-trimmer was required rather than wild-type ERAP1 to see a reaction. The outcomes of this experiment can be found in the appendices. The MD simulations and free energy dissociation calculations show that, despite the good resolution of the A pockets in the electron density data for these mutants, N-terminal binding is less strong for the mutants than for dtSCT, with higher RMSDs and readier dissociation as a result. This could also be suspected for sub-optimal peptides with similar weakened binding, perhaps at the N-terminal, allowing access by ERAP1.

3 Proton Resonance Assignments of a Peptide Series

3.1 Introduction

In order to fully characterise a series of N-terminally extended peptides for the purpose of using proton NMR to assess trimming by ERAP1, protons for these peptides were assigned. The peptides were a small library (see Figure 37) based on the ovalbumin-derived naturally occurring SIINFEKL peptide and N-terminal extensions of it. Originally included in the library were LSIINFEKL and LLSIINFEKL, with the aim of providing three types of N-terminal extension to test ERAP1 activity against: hydrophobic (leucine), charged (glutamic acid) and resistant to ERAP1 activity (proline). The single and double leucine N-terminal peptides were later substituted for lysine (see Figure 37) because this better fits published findings about the preferences of ERAP1 (Reeves et al. 2013). The resulting series of peptide extensions therefore are intended to look at an easily trimmed extension (glutamic acid) a moderately trimmed extension (lysine) and an untrimmed extension (proline) and the assignments of LSIINFEKL and LLSIINFEKL in 90% H_2O are not presented. Assigned chemical shifts for these peptides have not been published previously and were required for reliable identification of peptide species in a reaction. Both substrates and products of an aminopeptidase were assigned, from the initial 10- and 9-mer peptides and subsequently N-terminal shortened peptides, and also single amino acids.

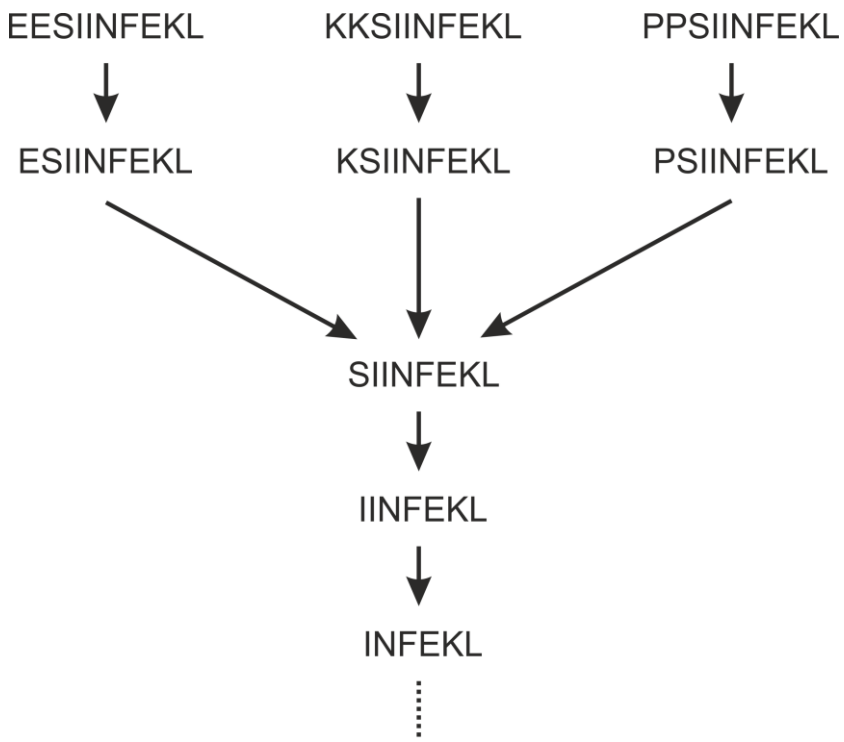


Figure 37 – Ovalbumin (SIINFEKL) N-terminally extended peptide library with single and double-residue N-terminal extensions based on known ERAP1 N-terminal preferences (Reeves et al. 2013) to exhibit varying N-terminal cleavage rates. Expression and assignments in 90% H_2O were also carried out for LSIINFEKL and LLSIINFEKL, but are not presented here.

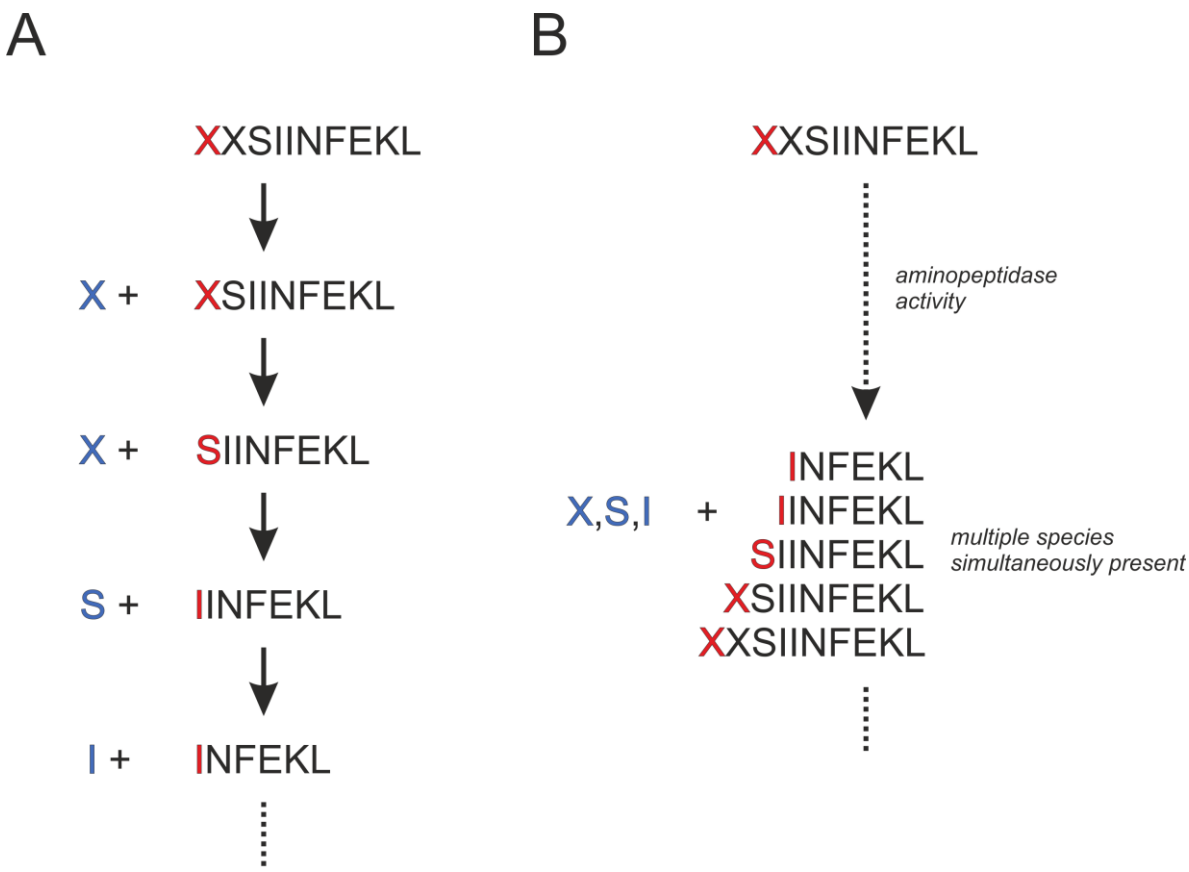


Figure 38 - Substrates and products of an aminopeptidase reaction. A) N-terminal residues (red) are cleaved to make subsequent product peptide and a free amino acid (blue). B) In reality the reaction would be a complex mixture with short peptides produced before removal of all long peptides. The species present, and their quantities, must be identified.

3.1.1 Nuclear Magnetic Resonance Spectroscopy as an analytical tool

Nuclear magnetic resonance (NMR) is a technique for detecting transitions of nuclear spins which give rise to magnetic moments. A strong B-field is used to excite a population of spins before detection of the magnetic moments from the spins returning to equilibrium in a free induction decay. The coil which generates the B-field can also be used for detection of the free induction decay. The simplest NMR experiment consists of a preparation period, a radio frequency pulse at the transition frequency of the spins, and detection of the subsequent detection of signal emitted by the sample (see Figure 39) (Wüthrich, 1986 p.45). This pulse sequence is then repeated for a number of transients to

build up signal-to-noise for the resulting spectrum. This will provide the most basic 1D spectrum of a given type of nucleus, and has been expanded with varying additions to the pulse sequence to provide a diverse range of experiments with different uses and capabilities.

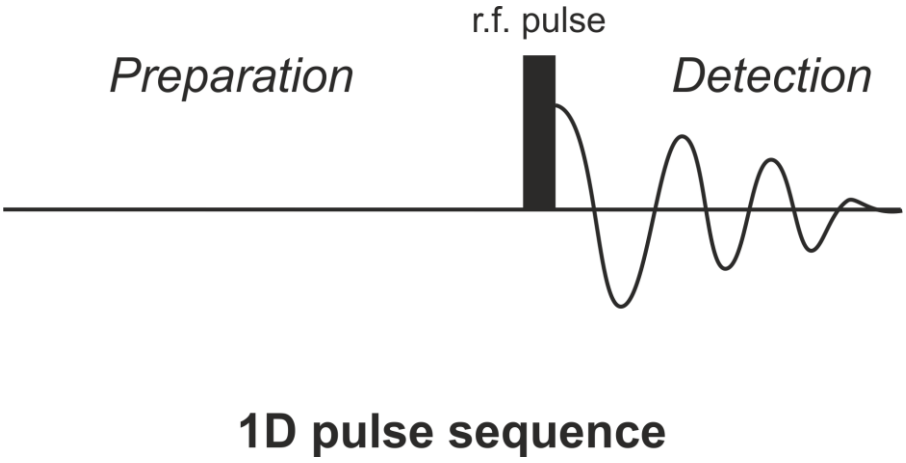


Figure 39 - Pulse sequence for simplest possible single pulse 1D experiment

NMR signals arise from specific nuclei within a molecule (e.g. ^{1}H nuclei) and are proportional to the number of nuclei creating the signal. This makes NMR a quantitative technique. The chemical shift values of these nuclei are highly specific to the environment the nucleus occupies, meaning NMR can be used to identify the molecules present (distinguishing them within a mixture) and indicating any changes to the chemistry or environment of the molecule.

3.1.2 Assignment with 2D homonuclear proton NMR

Homonuclear proton 2D spectra offer additional chemical and structural information on the tested molecule and are an extension of basic 1D techniques which offer enormous flexibility. Utilising T2 spin-spin relaxation means cross-relaxation between nuclei gives rise to cross-peaks, at the position of the two interacting protons on each axis. This can either be through-bond (e.g. COSY,

TOCSY) or through-space (e.g. NOESY, ROESY). The mixing period is used to establish cross-relaxation.

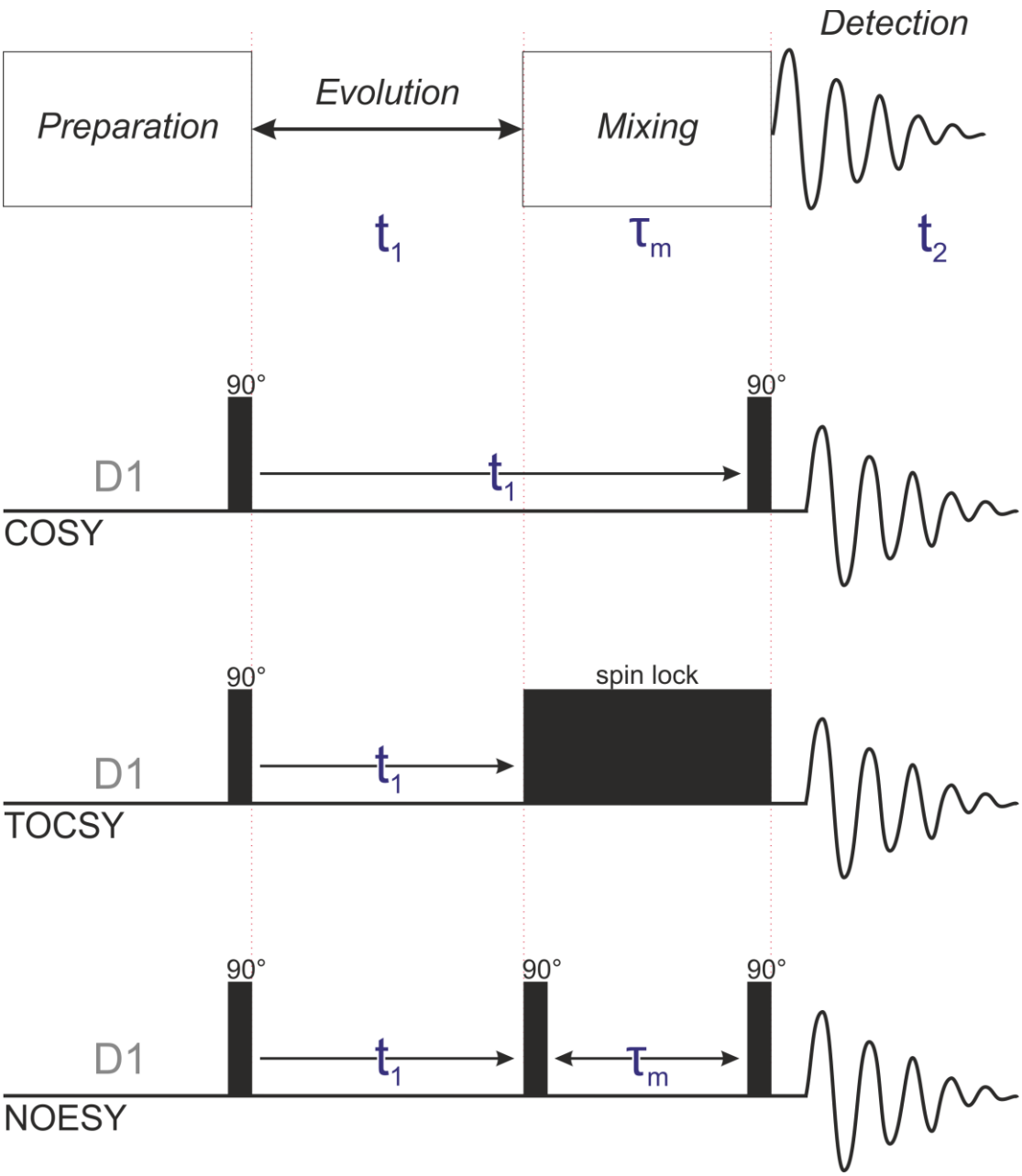


Figure 40 - Anatomy of a pulse sequence for simple 2D homonuclear experiments. The variable t1 evolution period (giving rise to 2D spectrum) is followed by τm mixing time and t2 free induction decay recording. A relaxation period (D1) is included at the start of the experiment to allow spins to return to equilibrium before a pulse is applied again.

For the COSY (Correlation Spectroscopy) experiment (designed but not published by Jean Jeener) two 90° pulses are performed, which allows

magnetisation transfer by J-coupling between two neighbours a maximum of 2-3 bonds apart (Cavanagh et al. 1996). Related is the TOCSY (Total Correlation Spectroscopy) experiment has a spin lock series of pulses instead of the second 90° pulse. A sufficiently long mixing period in a TOCSY experiment means magnetisation can transfer between up to 5-6 bonds, giving more cross-peaks than a COSY experiment. Helpfully (for protein/peptide assignment) magnetisation transfer is limited to connected protons within a residue (spin system) as heteronuclei will prevent transfer of magnetisation any further (Cavanagh et al. 1996).

NOESY experiments (Nuclear Overhauser Effect Spectroscopy) require different mixing period lengths for different sized molecules, as the cross-relaxation which occurs is dependent on the size of the molecule, but it is possible for magnetisation transfer to occur between protons up to 5 Å away from each other (Meier and Ernst 1979; Cavanagh et al. 1996). For very small peptides and single amino acids, simple 1D and 2D TOCSY spectra are adequate for proton assignments. Combining both through-bond and through-space experiments permits accurate proton assignments for many peptides and small proteins by identifying neighbouring and residue-specific nuclei through cross-peaks.

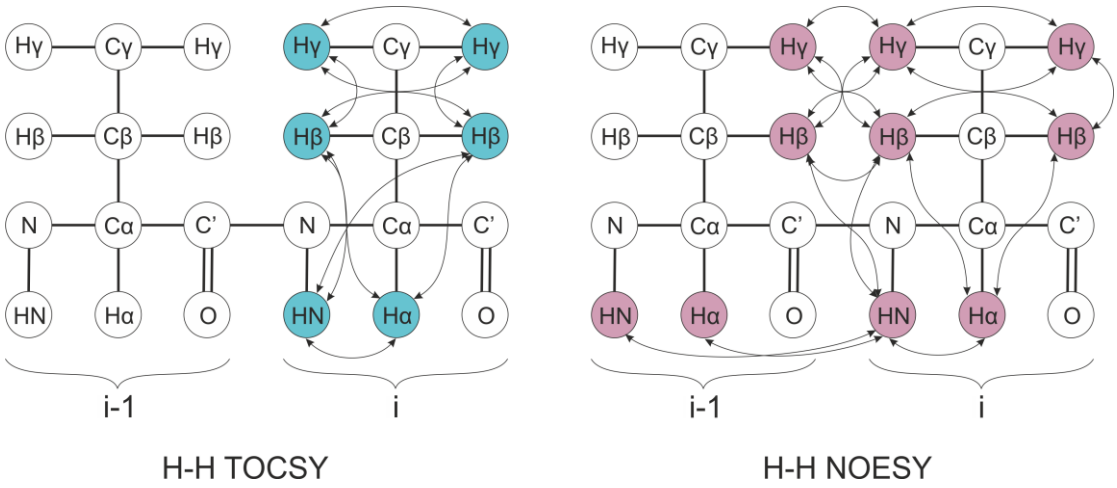


Figure 41 - Homonuclear TOCSY and NOESY experimental couplings. Proton TOCSY experiments make use of spin-spin coupling between protons within a residue to form cross peaks between protons (some transfers have been removed for simplicity). Proton NOESY experiments use the NOE effect to transfer magnetisation to protons in close proximity. Adapted from www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/h-h-noesy/ (Higman 2012)

3.1.3 Characterising a peptide library with reproducible chemical shift assignments

Because chemical shifts are a product of chemical environment, and each proton is affected by its surroundings. Beyond the chemistry of the molecule being examined, the surrounding solvent will influence chemical shift values, as can the temperature. Fortunately, the NMR spectrometer is a controlled environment. It is simple enough to reproduce the solvent environment for each spectrum and set a specific temperature. Doing so results in the highly reproducible chemical shifts which make NMR a powerful analytical tool. A characterised peptide library, therefore, is a large amount of information with great scope for its use. In addition to structural information about the peptide, the reproducibility of chemical shifts in identical samples enable quantitation. Described here is the characterisation of a peptide library by proton assignment for the purpose of informing other work.

3.2 Methods

3.2.1 Proton NMR of unlabelled peptides

3.2.1.1 Sample preparation

Synthesised peptides (purchased from ChinaPeptides and PPR Ltd.) were used for all experiments shown. Peptides were stored lyophilised at -20 °C, or for short periods in deuterated buffer at -20 °C, but rarely stored in water due to rapid degradation.

Peptides in 10% D_2O, 90% H_2O:

Peptides were dissolved at 1 mg/ml in ddH_2O and centrifuged after shaking to remove undissolved peptide. 10% deuterium oxide was added for the NMR sample, with 3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP) used as a reference. Samples were made to 180 µl in 3 mm NMR tubes.

Peptides in 100% D_2O:

A deuterated peptide buffer was prepared like so: 1M dibasic potassium phosphate K_2HPO_4 (to approx. 7.6 mM) was mixed with 1M monobasic potassium phosphate KH_2PO_4 (to approx. 12.4 mM) to pH 6.6 and a total concentration of 20 mM. The solution was dehydrated in a vacuum centrifuge and resuspended in the same volume of deuterium oxide, before being dried and reuspended again in the same manner resulting in a buffer at pD 7. Peptides were dissolved in this deuterated buffer at 1 mg/ml and a small pH probe used to check the pD. Peptides with a lower pD were buffer exchanged into the deuterated peptide buffer using 100-500 Da cutoff cellulose membranes. For NMR, 200 µM samples were prepared with 20 µM TSP and 0.04 % (w/v) sodium azide. Samples were made to 180 µl in 3 mm NMR tubes.

3.2.1.2 ^{1}H NMR experiments

NMR spectra were obtained on 600 MHz Varian Inova solution state cryoprobe spectrometer.

1D 1H spectra:

Presaturation pulse water suppression was used for suppression of residual water signal. Most spectra described here were taken with a 5 second relaxation delay and 0.32 second acquisition time (it will be noted where otherwise).

Inversion recovery experiments, from Varian BioPack software, with presaturation pulse water suppression were carried out with a 20 second relaxation delay. T1 was arrayed from 0.0625 to 32 seconds.

2D 1H spectra:

1H TOCSY spectra were MLEV17 pulse sequence experiments from the Varian BioPack software with an 80 ms mixing time. Sometimes watergate water suppression. These were taken at 10 °C (10% deuterated samples) with presaturation and PURGE solvent suppression or 25 °C (100% deuterated samples). The acquisition time was 0.15 seconds.

1H NOESY spectra with a 300 or 500 ms mixing time were taken at 10 °C using a Varian BioPack experimental pulse sequence. Gradient echo water suppression was used. The acquisition time was 0.107 seconds.

3.2.1.3 ^{1}H spectra Processing

Data conversion, phasing and processing of 1D and 2D spectra for assignment in CCPN Analysis was performed in nmrPipe and nmrDraw (Delaglio et al. 1995). For use outside CCPN Analysis, the MestReNova (MNova) software was used to process and display spectra.

Sequential assignment in CCPN Analysis (Vranken et al. 2005) was performed using overlaid 2D ^{1}H NOESY and TOCSY spectra. For each peptide a TOCSY spectrum was taken at 10°C to show intra-residue connections. Identifiable cross peaks on the TOCSY spectra arise through through-bond spin-spin coupling of protons. Carbonyl groups on the peptide backbone prevent coupling between residues and cross-peaks are therefore only seen between a set of protons within a residue. The NOESY spectra conversely provide cross-peaks for protons which are spatially close which arise due to cross-relaxation after a mixing time. Assignment by sequential walking is done through comparison of residue-confined TOCSY peaks and the NOESY cross-peaks between residues.

For PSIINFEKL a TOCSY spectrum was taken at 10°C to show intra-residue connections. This was compared with a 25°C TOCSY for PSIINFEKL and judged as having better resolution. Thereafter 2D spectra for proton assignment in H_2O were obtained at 10°C in order to perform accurate initial assignment. For assignments in deuterated buffer, a temperature of 25 °C was used so that peptide-ERAP1 experiments could take place using the same chemical shifts at a more optimal temperature for an enzyme assay.

Chemical shifts for the same residue within different peptides will often differ. Similarly, the same amino acid at different positions in the same peptide will often give rise to different chemical shifts because of their different environments within the peptide. This can be seen with isoleucines I3 and I4 of PSIINFEKL which give different NH shifts in H_2O (8.46 and 8.34 ppm respectively) and different chemical shifts for all other proton positions. The differences between isoleucine shifts within the same peptide and within

different peptides can be seen in Figure 45. Noticeable changes to chemical shift values between these peptides are extremely useful for NMR reaction monitoring of peptide species present in an aminopeptidase reaction. The final product which would be seen if isoleucine peaks were tracked during an ERAP1 NMR assay would be the free amino acid form of isoleucine, which again has different chemical shift values to the peptide residue form.

3.3 Results

The peptides used are short and unstructured and therefore give chemical shifts close to random coil values in solution. Random coil shifts are well studied and assignment can be partially carried out using established values as predictions (Wishart, Bigam, Holm, et al. 1995). Assignments were initially made with the peptide dissolved in water with 10% D_2O and then later in a fully deuterated 20 mM potassium phosphate buffer at pD 7 (pH 6.6) which was deemed suitable for maintaining ERAP1 in a folded and active state. Use of 100% D_2O was required to see Hα proton shifts near the highly distorting water peak in order to assign fully. Despite efforts to suppress the water signal, deuterated solvent ultimately gave far superior results.

Stepwise assignment of TOCSY/NOESY proton spectra was adequate for unambiguous assignments of most protons. BMRB database (Ulrich et al. 2008) chemical shift references were also used to check assignments and were adequate on their own for 1D ^1H amino acid spectra assignment.

Peptide TOCSY spectra showed clear peaks and low background noise. Suppression of the water signal allowed recovery to a small degree of peaks in the Hα region for samples in 90% H_2O. A fully deuterated buffer for the second set of assignments reduced the water peak further and allowed recovery of all Hα resonances in peptide. For greater accuracy of chemical shifts than could be obtained with solvent referencing, TSP was used as a standard reference. As well as assignments of peptides, 1D proton spectra of their composite L- form amino acids were recorded. Chemical shift values for these can be found in the Appendices.

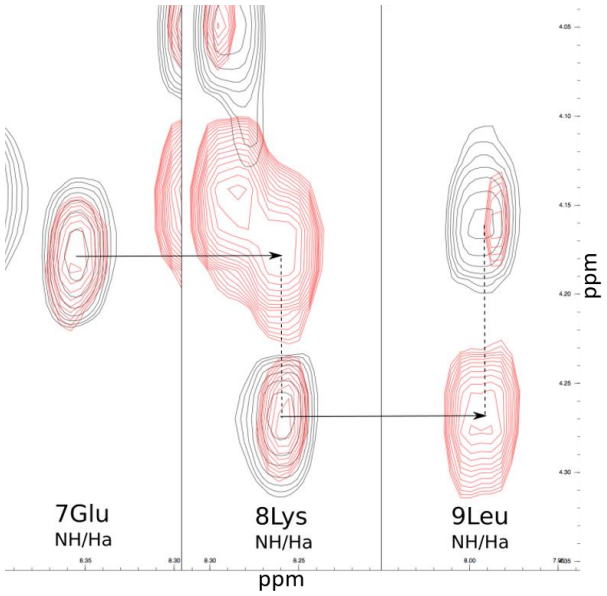


Figure 42 – Sequential proton assignment of PSIINFEKL resonances, shown by arrows on NOESY (red) and TOCSY (black).

	Assignments in 10% D_2O, 90% H_2O			Assignments in 100% D_2O, 20 mM potassium phosphate, pD 7 (non-labile protons only)		
Peptide/amino acid	*Number of assigned protons*	*Number of unassigned protons*	*Tentatively assigned peaks*	*Number of assigned protons*	*Number of unassigned protons*	*Tentatively assigned peaks*
SIINFEKL	61	5	2	56	0	0
ESIINFEKL	65	4	5	61	0	0
KSIINFEKL	77	2	0	65	0	0
PSIINFEKL	73	3	0	63	0	0
EESIINFEKL	78	2	0	66	0	0
KKSIINFEKL	88	2	0	74	0	0
PPSIINFEKL	82	2	0	70	0	0
IINFEKL	-	-	-	33	0	18
INFEKL	-	-	-	47	0	0

Table 7 - Proton NMR peak assignments for SIINFEKL-based peptide series.

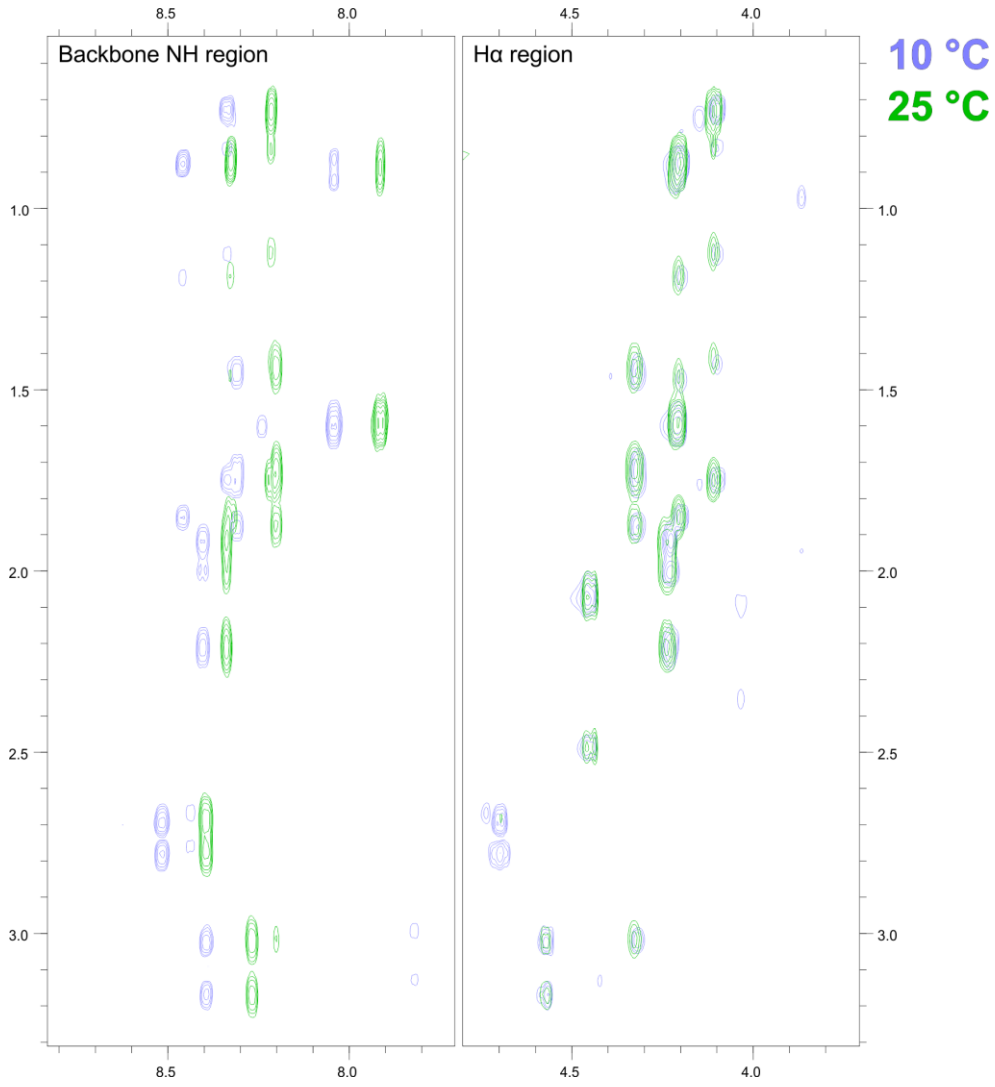


Figure 43 - Effects of temperature change on ^{1}H TOCSY spectrum peaks of PSIINFEKL in 10% deuterated water, pH7.

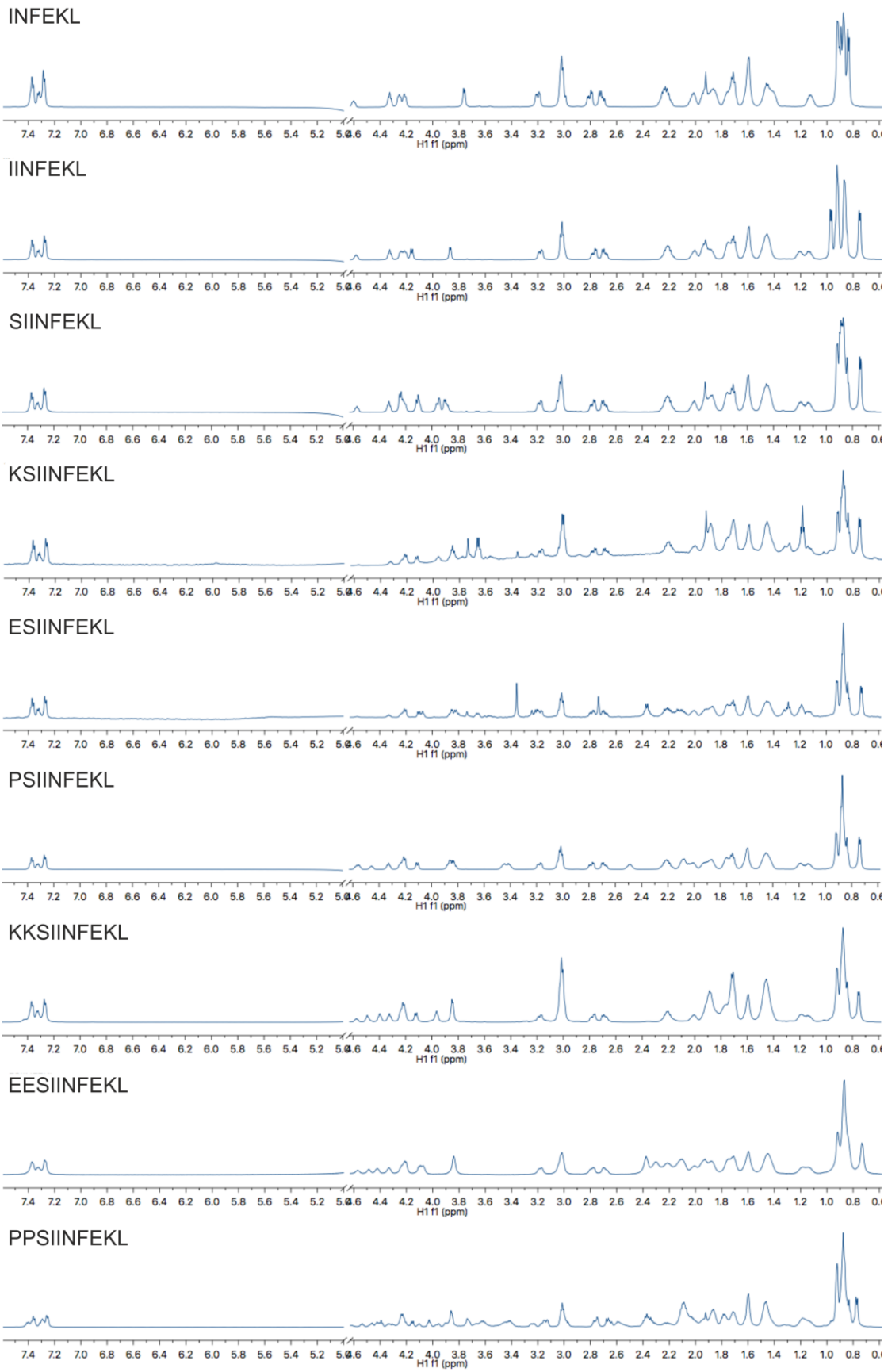


Figure 44 - 1D 1H NMR spectra of peptides (INFEKL to PPSIINFEKL) in 20 mM deuterated potassium phosphate buffer.

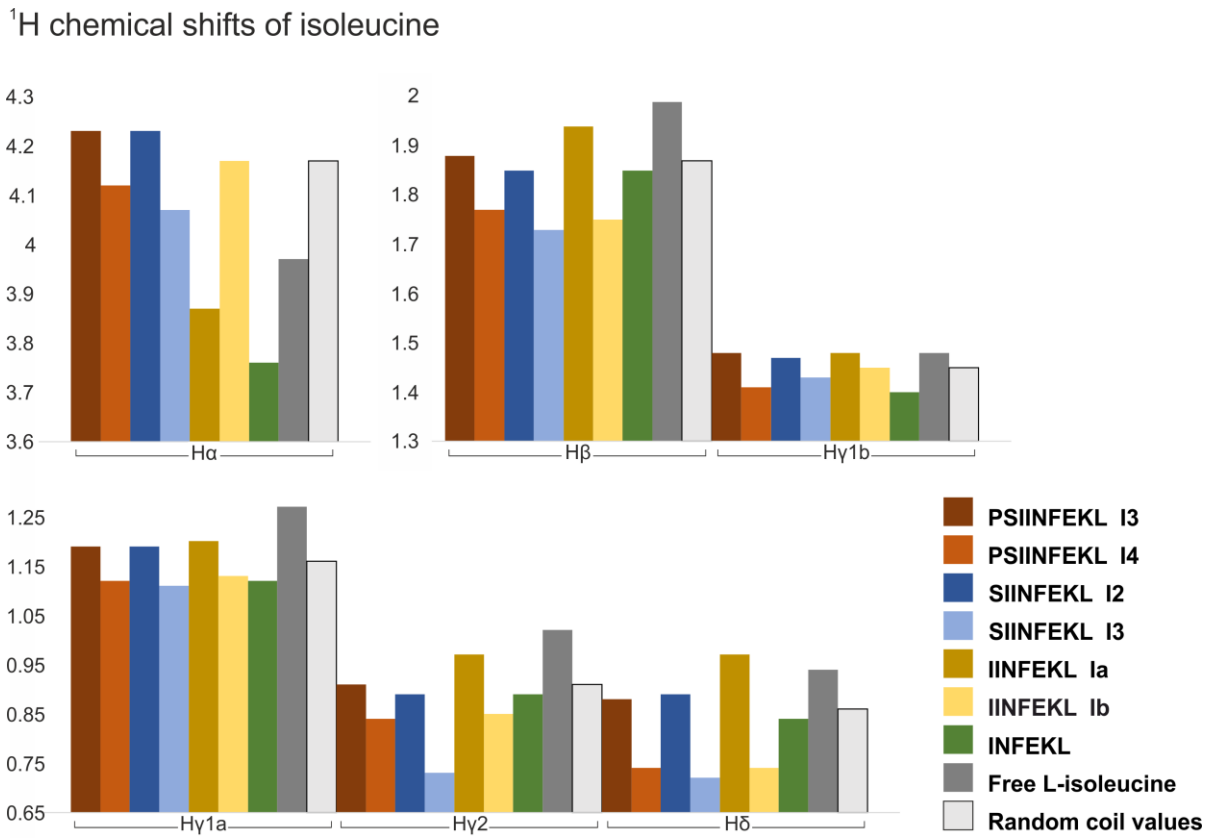


Figure 45 – Changing isoleucine chemical shift values for free isoleucine and isoleucine as a peptide residue

3.3.1 Peptide assignments in water

Proton assignments for all peptides were first carried out in 90% H_2O, 10% D_2O. The low solvent deuteration provided peaks for the labile backbone NH protons. This 'fingerprint' region is where assignments were predominantly carried out. Most protons were assigned for the peptides in 90% H_2O. An exception for all peptides was the N-terminal backbone NH protons, which were not expected to give a chemical shift due to rapid exchange with the solvent. No peak was assigned for the Serine -OH γ proton in any peptide. This was again presumed due to rapid exchange. Other backbone NH protons were assigned confidently in most cases. The second position residue backbone NH was unassigned for SIINFEKL (I2) and ESIINFEKL (S2). For SIINFEKL no TOCSY cross-peaks arose for I2 in the NH fingerprint region but all other I2 peaks were assigned. Putative S1 Hβ peaks were noted but gave no TOCSY or NOESY cross peaks to confirm

them. Similarly for ESIINFEKL no S2 NH cross-peaks were observed. A set of cross-peaks were observed at 4ppm which had the predicted shifts for glutamic acid but no NOESY cross-peaks were found to confirm this.

PPSIINFEKL chemical shifts were established for all protons of I4, I5, N6, F7, K9 and L10. The side chain hydroxyl γ proton of S1 could not be clearly identified, as was the case with all other peptides. Two sets of TOCSY NH cross-peaks for P1 appeared. These were linked by NOESY cross-peaks to chemical shifts for P2 and NOESY connections arose between these 'P2' peaks and S3 peaks. The position of E8 and P2 were initially ambiguous, because other than the chemical shift of Hα, chemical shifts were very similar. NH cross-peaks for E8 appeared which were impossible for P2 and so P2 peaks were initially considered to be multiple conformations of E8 due to the chemical shift overlap. However, E8 was distinguished from P2 by lack of key TOCSY cross-peaks (there was no peak well placed for the δ protons of P2) and the chemical shifts were adequately different in deuterated buffer for confident assignment. A NOE between P2 H∂ and P1 Hα indicated a trans conformation for P2 (Wüthrich 1986). No NOEs between P2 Hα and P1 Hα, and between P2 Hα and P1 NH showed, therefore no cis conformation was confidently assigned. The two assigned NH positions for P1 (see Figure 54) may arise from the cis and trans conformations.

3.3.2 Peptide assignments in deuterated buffer

Buffer deuteration prevented assignment of labile protons which were in exchange with the solvent and therefore not seen on spectra. Because the backbone NH peaks were absent, the fingerprint region was not used for assignment, unlike the assignments made in 90% H_2O. The α-hydrogen region between ~4ppm and the water peak was instead used. For SIINFEKL, ESIINFEKL, KSIINFEKL, EESIINFEKL, KKSIINFEKL and INFEKL, all non-labile protons were confidently assigned. PSIINFEKL provided two sets of chemical shifts each for what appeared to be E7 and K8 and what could be a small secondary 5AsnHα/Hb* cross peak can be seen in the TOCSY spectrum (Figure 51). NOESY spectra for PPSIINFEKL were poor, but the similarity between the chemical shifts

in water and in deuterated buffer permitted confident assignment without them. The only significant difference was improved separation of E8 and P2 peaks over the spectra in 90% H_2O, which aided the assignment.

IINFEKL and INFEKL were assigned from a TOCSY spectrum alone, as stepwise sequential assignment was deemed unnecessary. This enabled rapid and confident assignments. For IINFEKL, I1 and I2 peaks were identified as belonging to isoleucine, but it was uncertain which was the first position isoleucine and which was the second. To gain this information, a NOESY spectrum would be required to identify neighbouring residues Alternatively a spectrum in 90% H_2O would identify which of the two isoleucine residues did not have visible backbone NH peaks; if only one did, it would surely be the N-terminal residue involved in solvent exchange. However, the two short peptides can be differentiated from other peptides and amino acids by chemical shift values alone and are assigned as far as is deemed necessary.

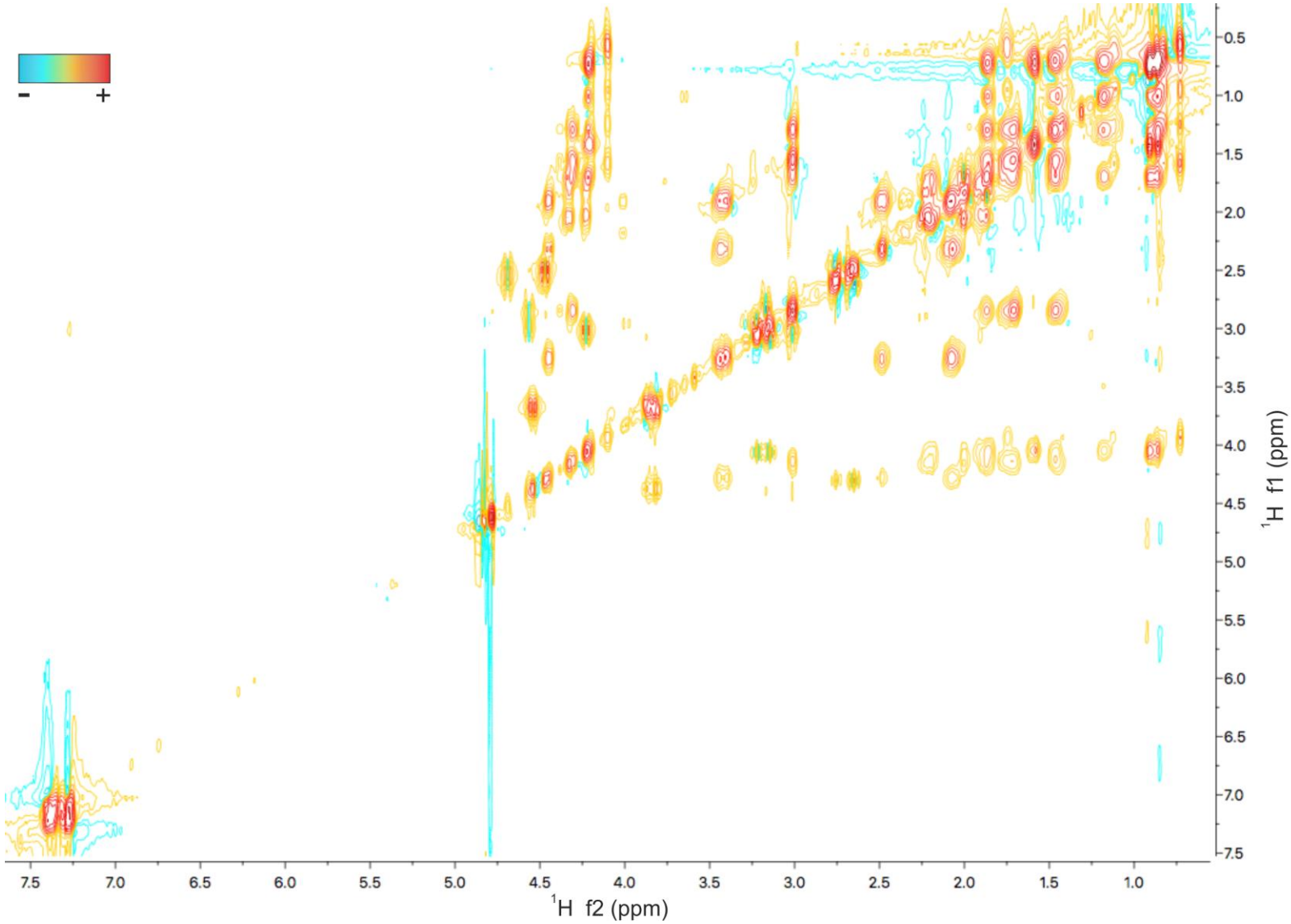
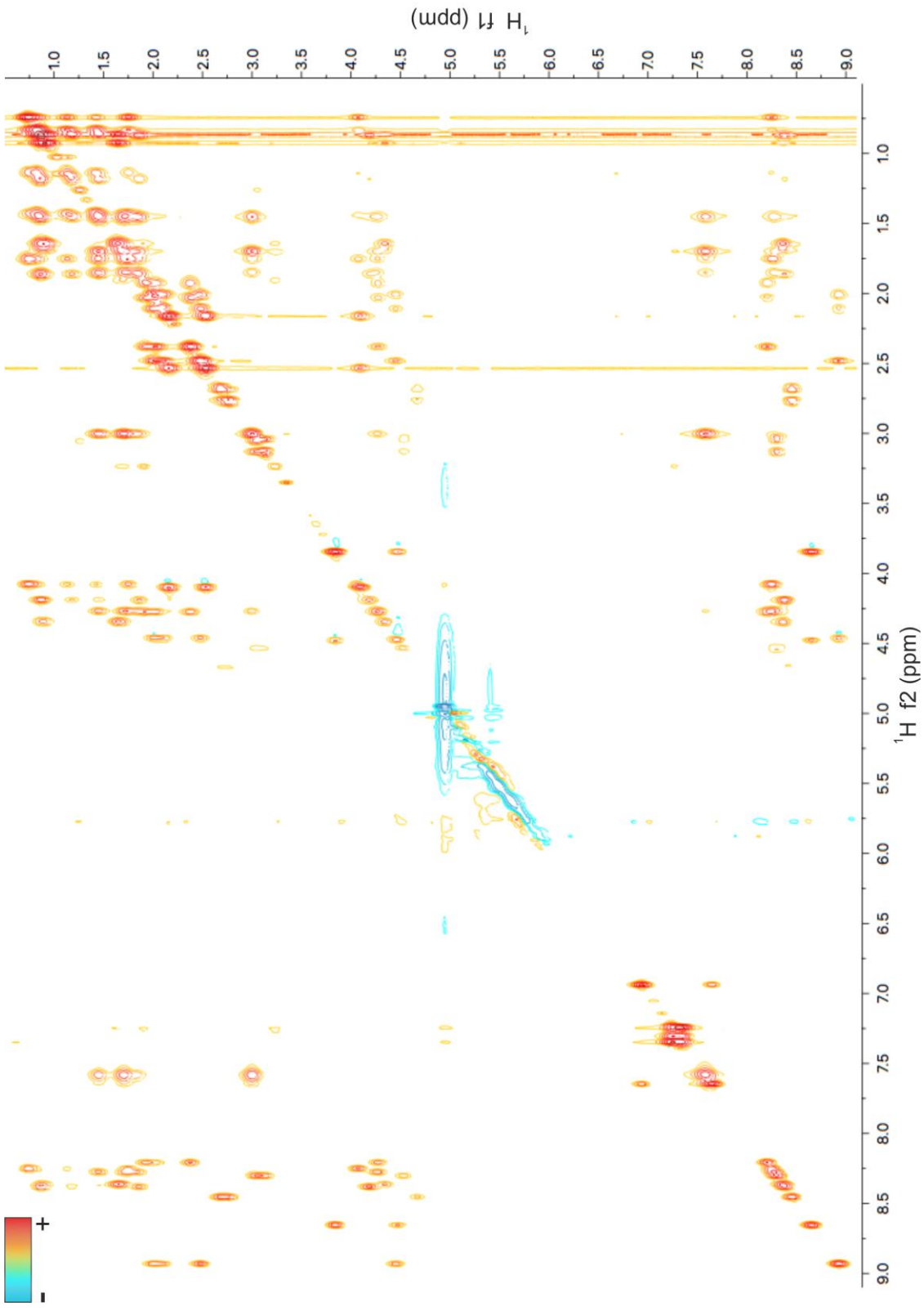
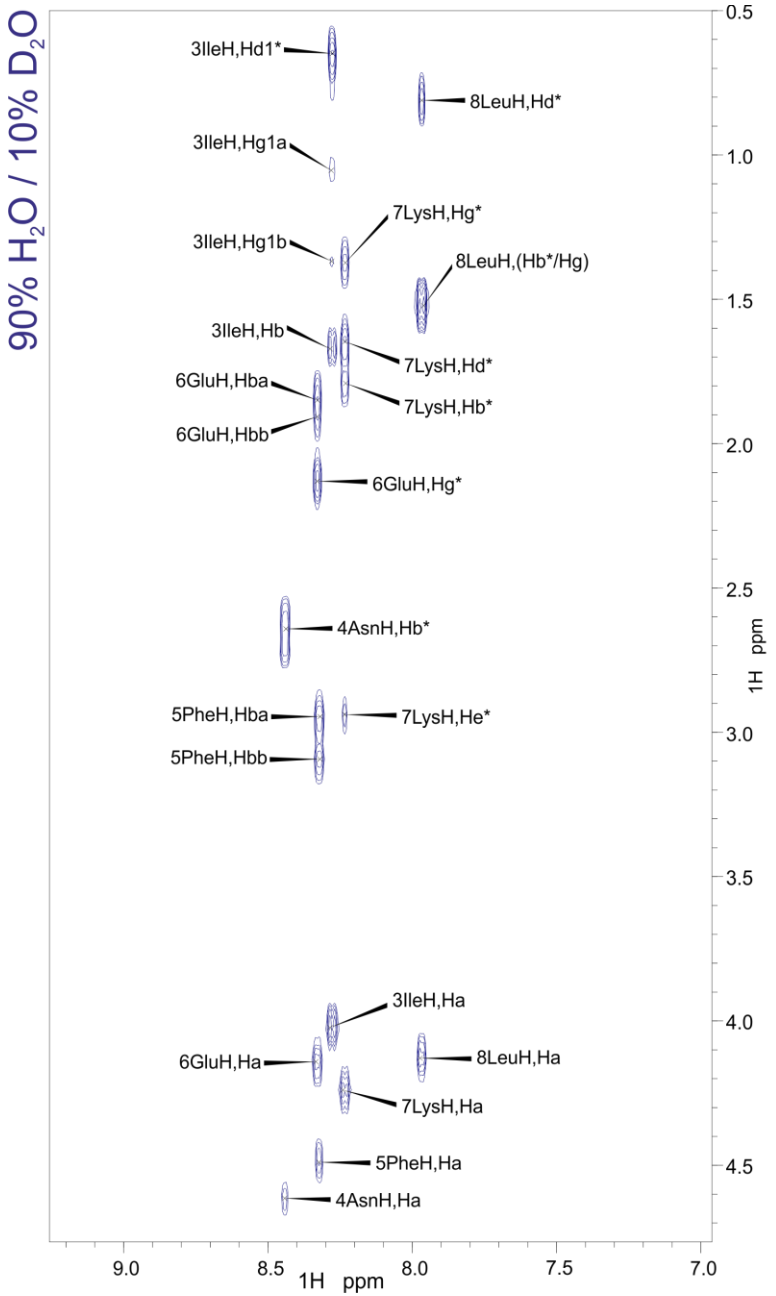


Figure 46 - 1H TOCSY spectrum of PSIINFEKL (≤ 2.5 mg/ml) in deuterated 20 mM potassium phosphate buffer (pD 7).

Figure 47 - 1H TOCSY spectrum of EESIINFEKL in 90% H2O, 10% D2O





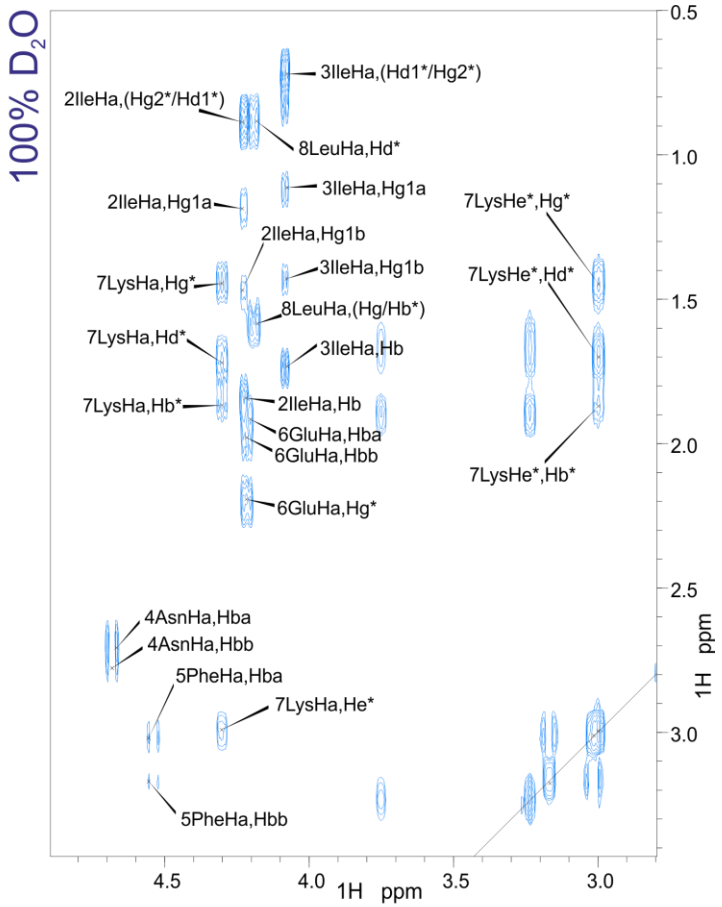


Figure 48 - 2D ^{1}H TOCSY spectra of ~1 mg/ml SIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

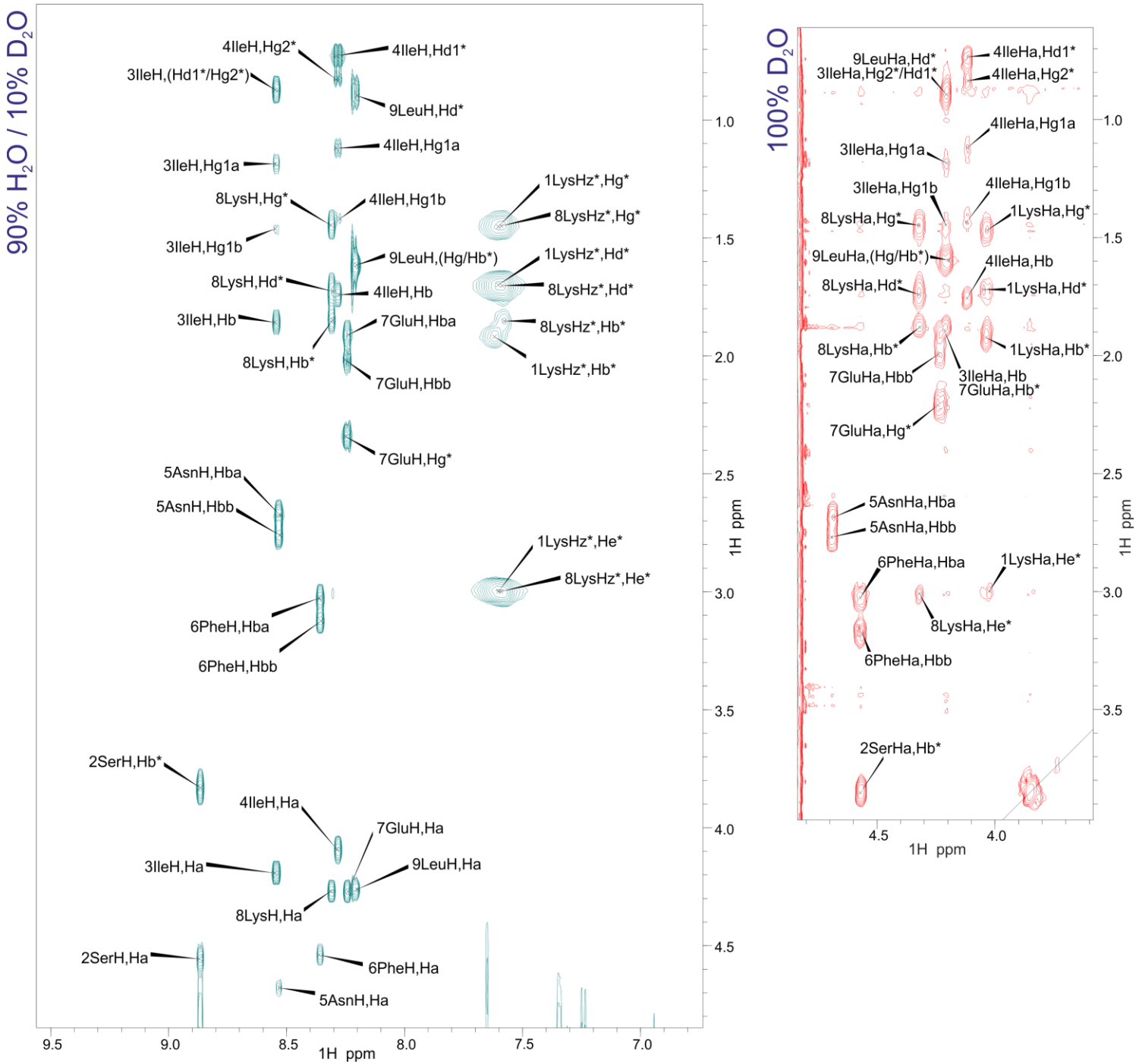


Figure 49 - 2D ^{1}H TOCSY spectra of ~1 mg/ml KSIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

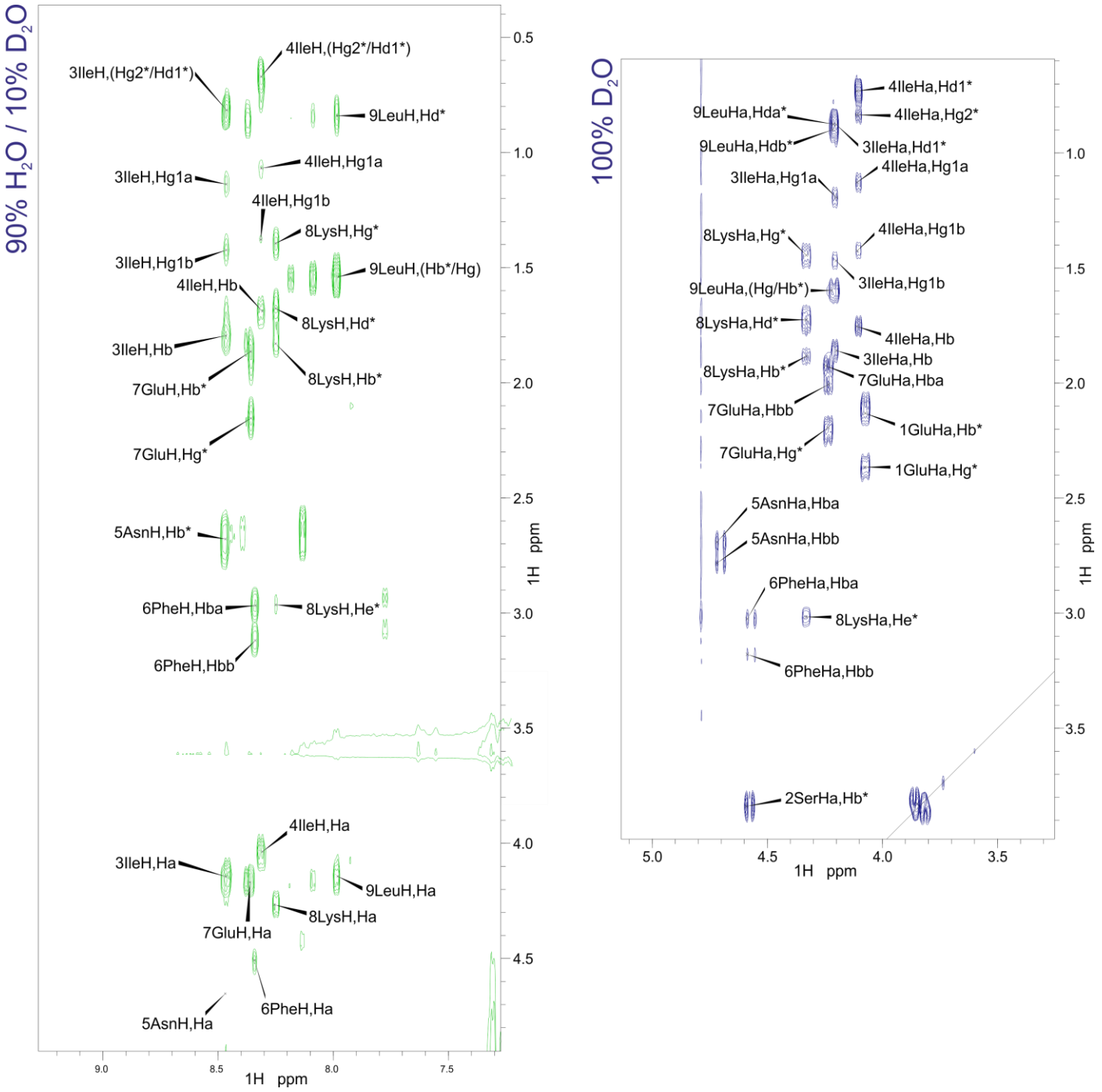


Figure 50 - 2D ^{1}H TOCSY spectra of ~1 mg/ml ESIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

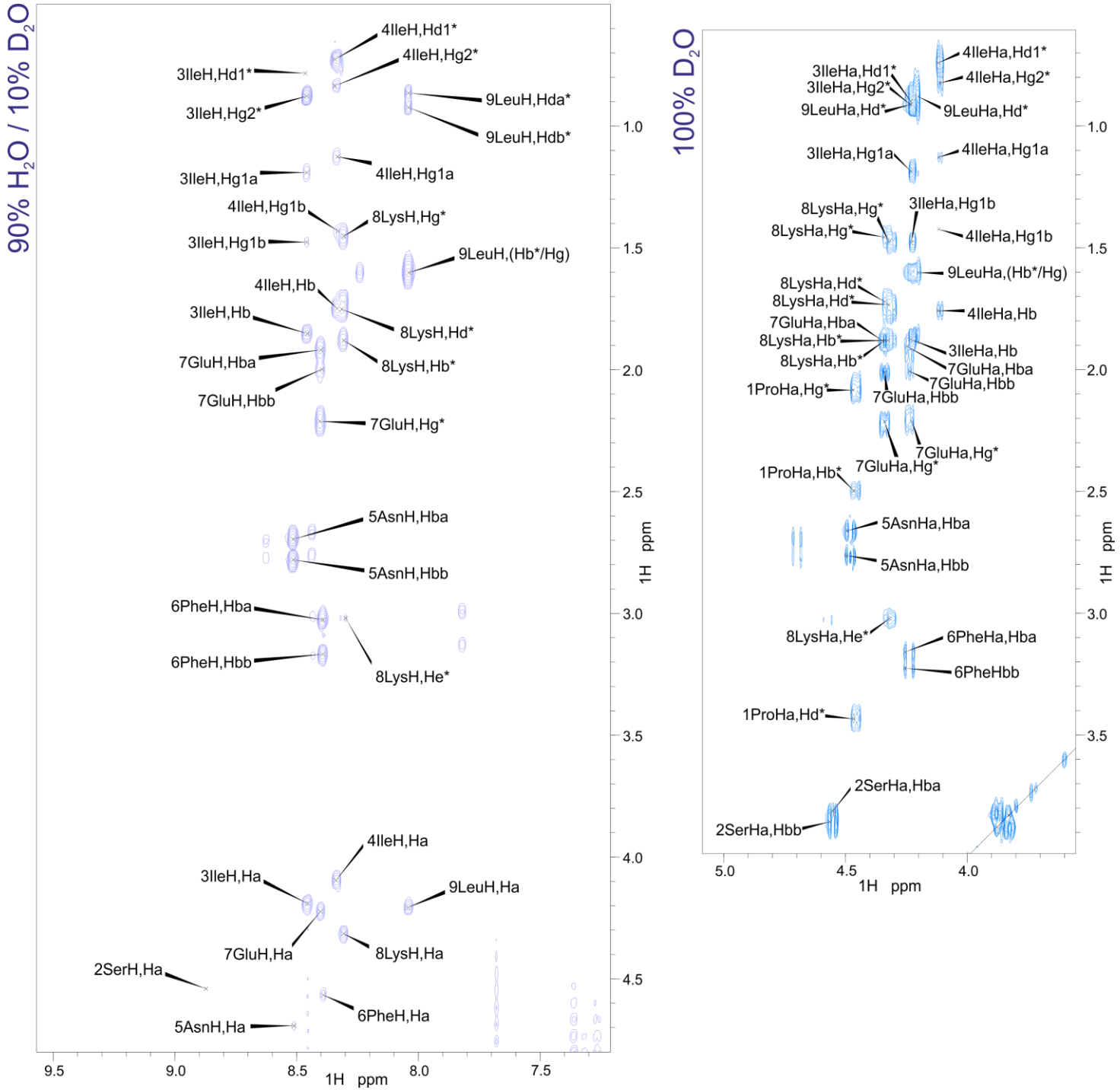
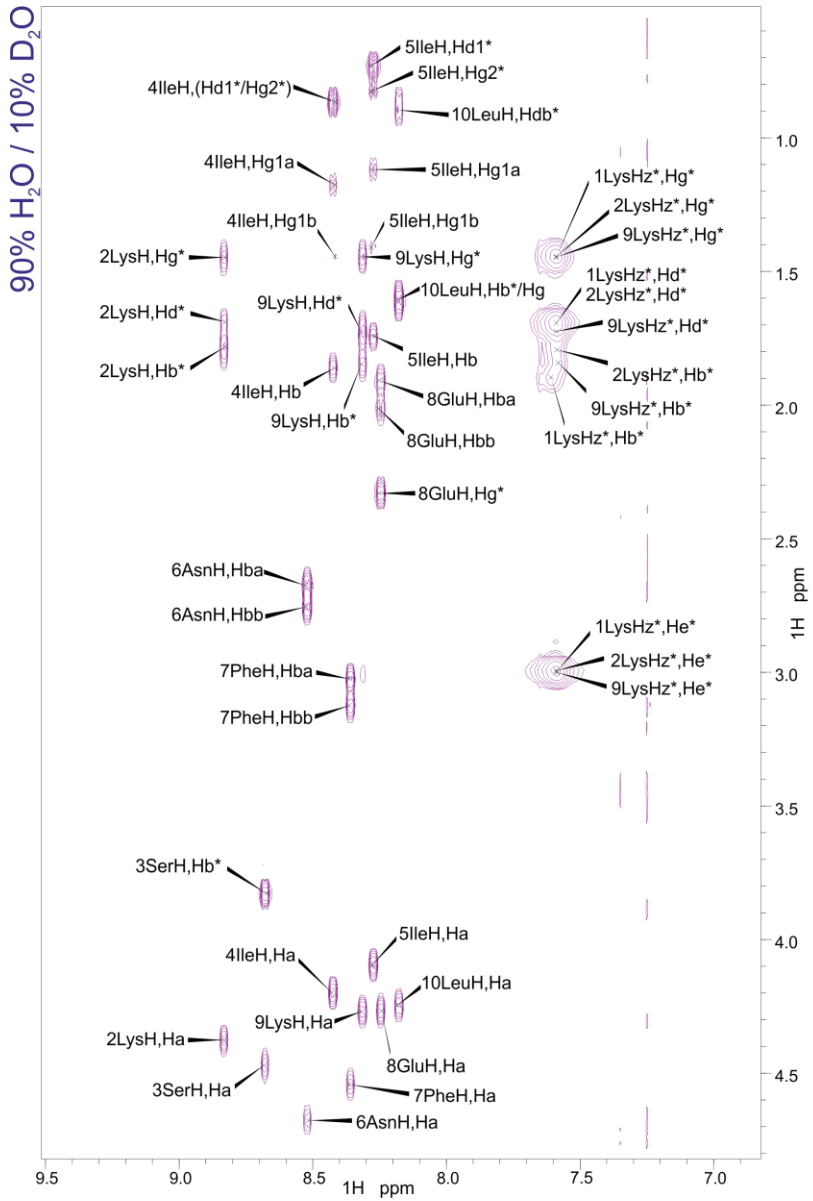


Figure 51 - 2D ^{1}H TOCSY spectra of ~1 mg/ml PSIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).



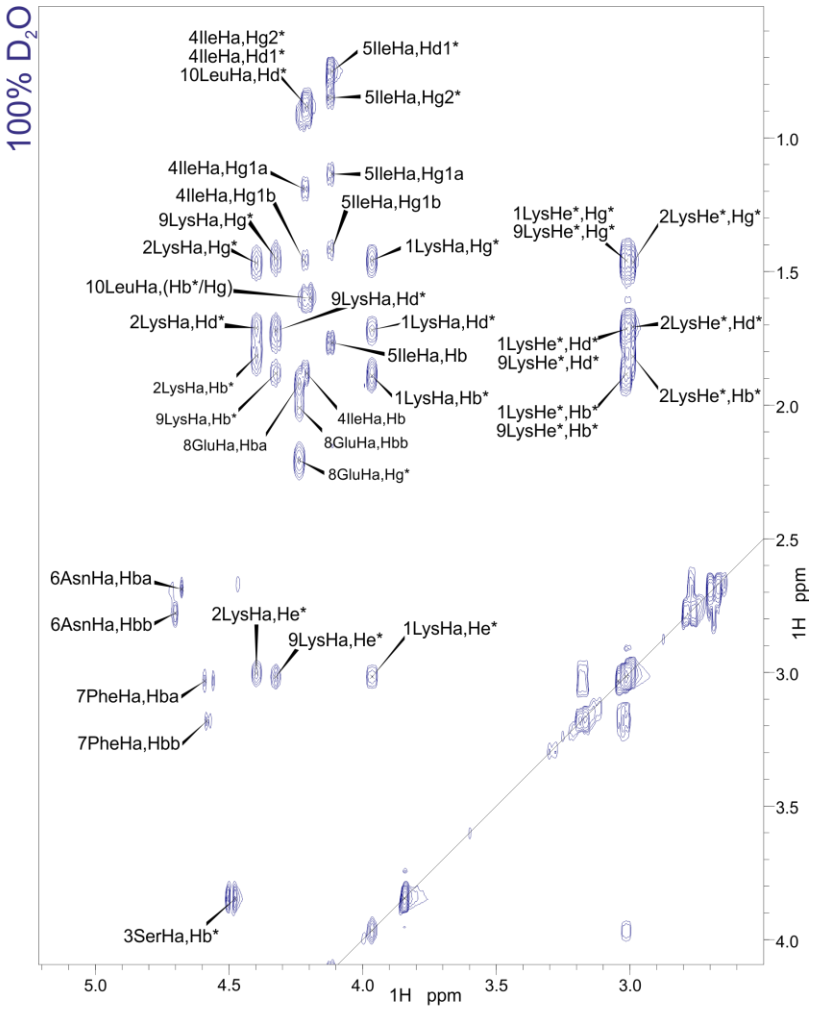
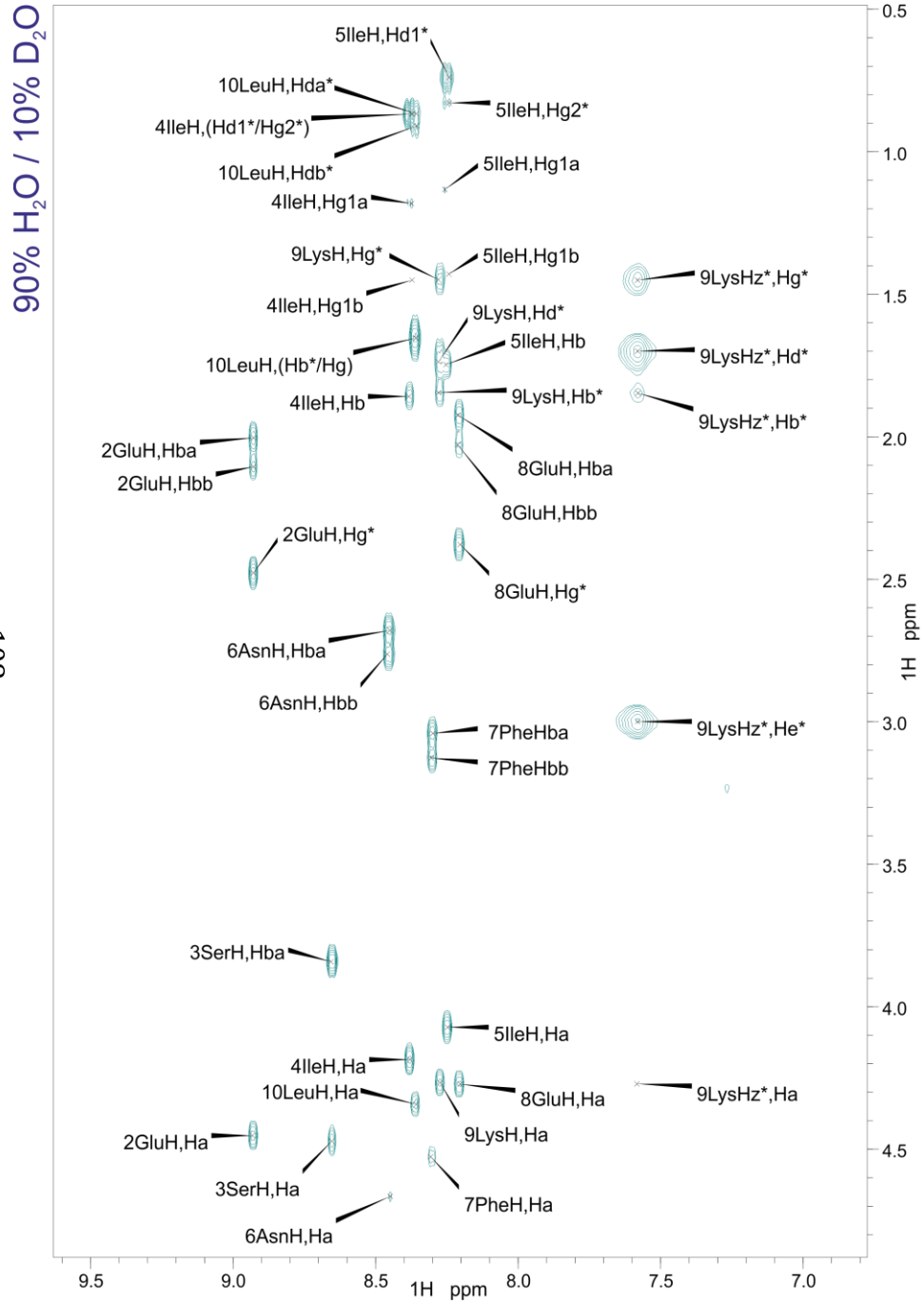


Figure 52 - 2D ^{1}H TOCSY spectra of ~1 mg/ml KKSIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).



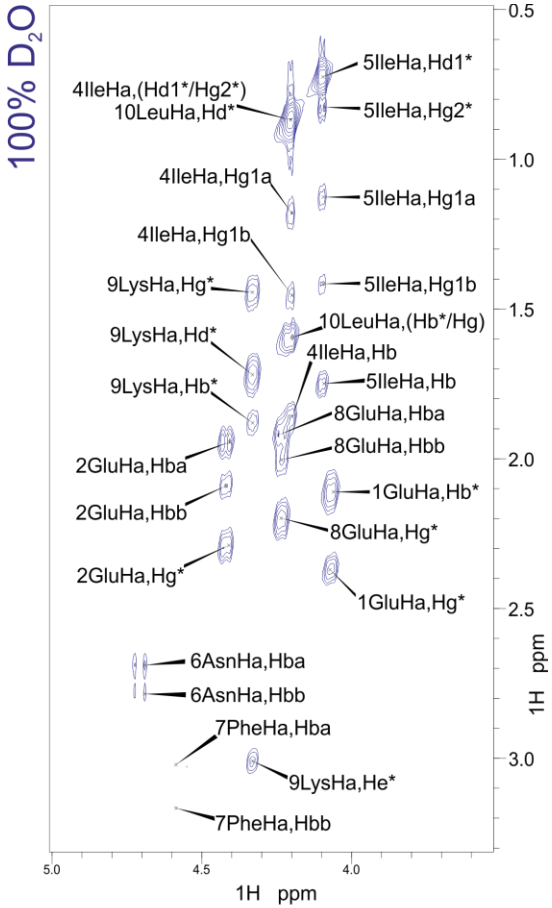


Figure 53 - 2D ^{1}H TOCSY spectra of ~1 mg/ml EESIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

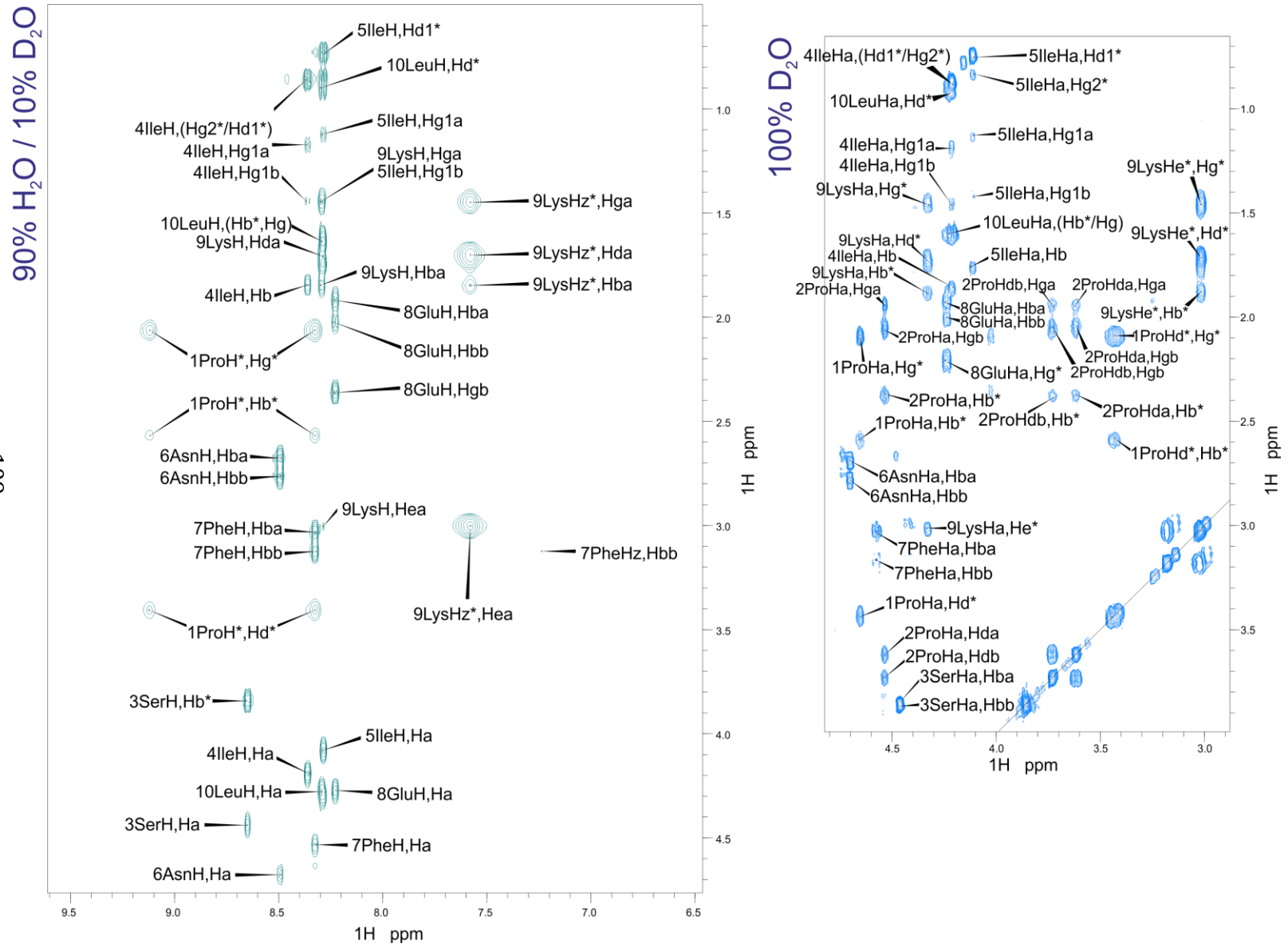


Figure 54 - 2D ^{1}H TOCSY spectra of ~1 mg/ml PPSIINFEKL with proton assignments shown. LEFT: in 90% H_2O, 10% D_2O at 10 °C (backbone NH fingerprint region). RIGHT: in 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

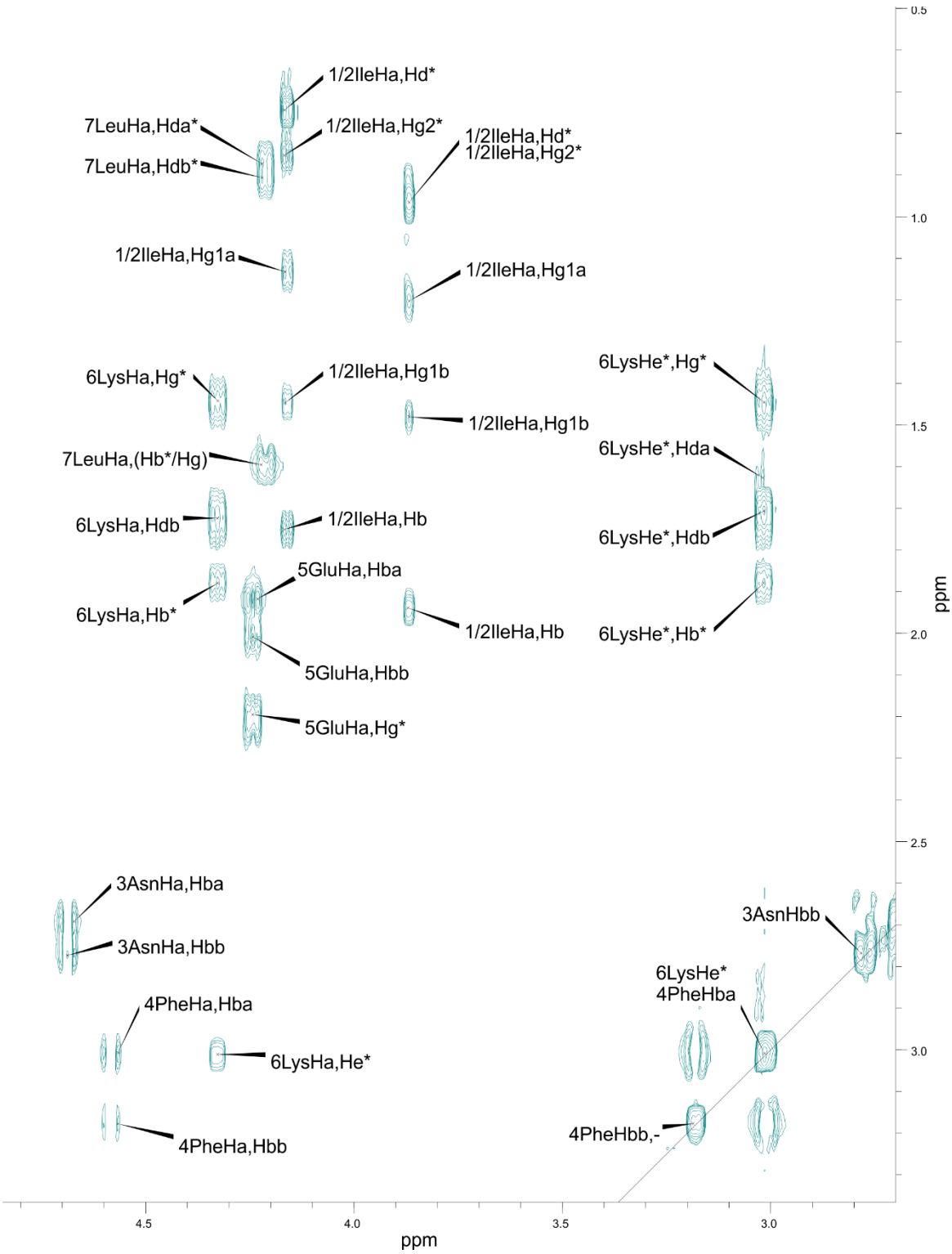


Figure 55 - 2D ^{1}H TOCSY spectra of ~1 mg/ml IINFEKL with proton assignments shown. 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

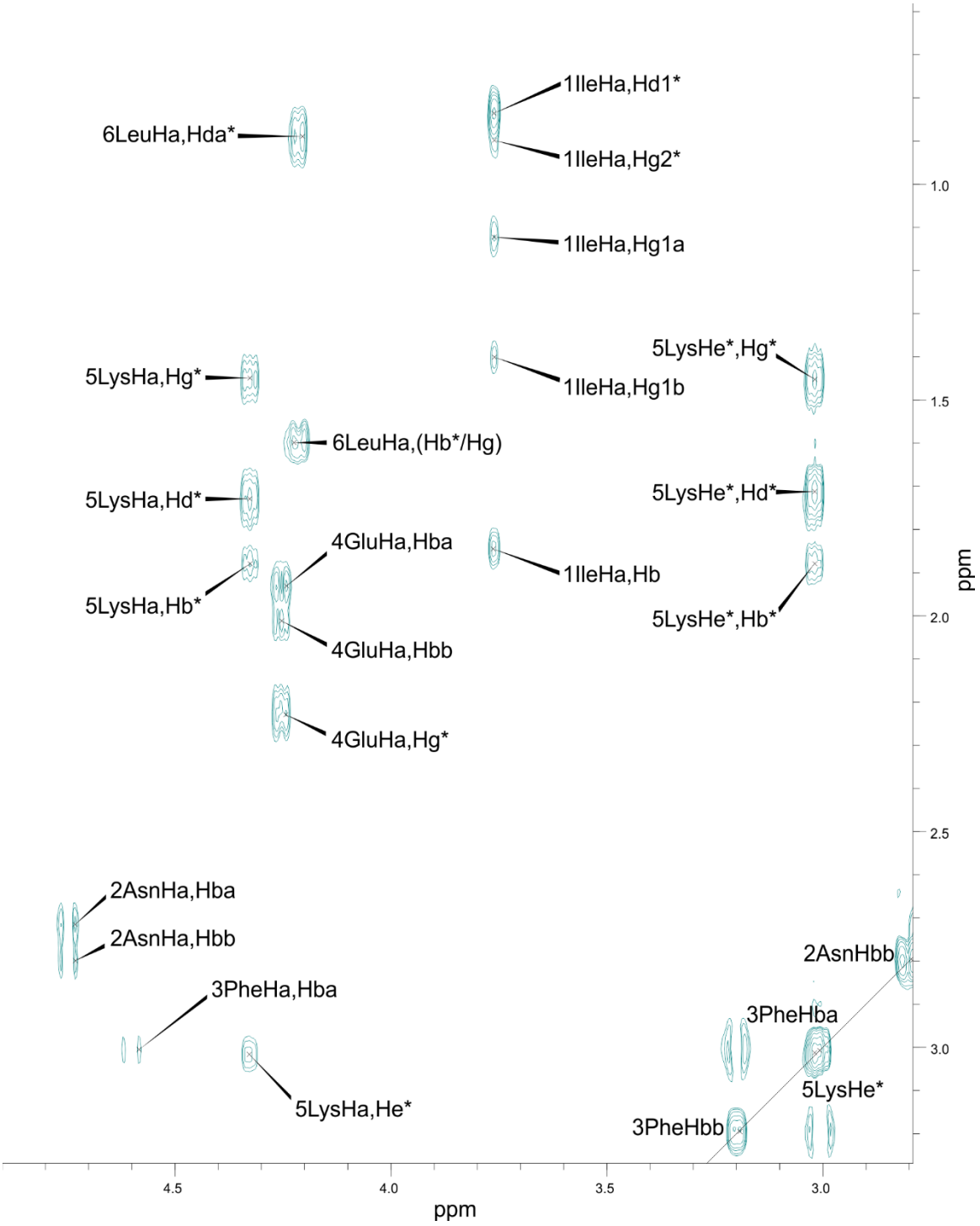


Figure 56 - 2D ^{1}H TOCSY spectra of ~1 mg/ml INFEKL with proton assignments shown. 100% deuterated 20 mM potassium phosphate, pD 7, at 25 °C (Hα region).

3.4 Discussion

PSIINFEKL gives additional Hα cross peaks for E7 and K8 in deuterated buffer. This may result from some unwanted modification to the peptide which was not excluded during purification, leading to the experienced environments for local protons in the peptide being altered. Similarly in 90% H_2O, PSIINFEKL has additional peaks near the 5AsnH,Hβa/Hβb and 6PheH,Hβa/Hβb cross peaks which might be minor alternative chemical shifts. Additional chemical shifts are also seen in the PPSIINFEKL 90% H_2O spectra for the backbone NH of 1P which may result from cis/trans isomerisation of the second position proline.

Most chemical shifts were mostly reasonably well dispersed though some overlaps did occur. In addition peptides adopted similar chemical shift values between them which made the process of assignment faster for each successive peptide. This can be readily seen when glancing at the spectra, for example in the isoleucine cross peaks for peptides in deuterated buffer which are clearly similar (for example in Figure 51, Figure 52 & Figure 53). The peptide chemical shifts were similar to random coil values as expected. Chemical shifts adjusted by secondary structure effects can move by as much as 2 ppm (Williamson and Asakura 1997). This was clearly not the case for these short unstructured peptides. Nonetheless the local environment has an effect on chemical shift values. The amino acids preceding or following a given residue will have an effect on the chemical shift, especially proline (Wishart, Bigam, Holm, et al. 1995).

For all the peptides, certain protons (such as isoleucine Hγ1a and Hγ1b, or asparagine Hβ1 and Hβ2) which yielded different chemical shifts but could not be definitively identified, were designated according to CCPN Analysis and based on common assignment databases. To gain accurate stereochemical assignments of these protons J-couplings would be required, but gaining this information was considered unnecessary. This was also initially the case for the aromatic ring protons of phenylalanine. In the NOESY spectrum of EESIINFEKL in deuterated buffer, an NOE arose between the phenylalanine H∂ and Hβ protons but not between Hβ and the other ring protons. This was considered evidence that the

assumption of the phenylalanine ring proton assignments in other peptides might be correct.

4 Expression, purification and NMR assignment of a $^{13}C/^{15}N$ isotope labelled peptide

4.1 Introduction

4.1.1 Bacterial cell expression of peptides

Commercially produced peptides are often produced *in vitro* with solid state synthesis. This is the usual way of making them and it yields high quality peptide in substantial quantities. For a peptide labelled with ^{13}C, ^{15}N or ^{2}H, the costs of synthesis of a bespoke peptide are prohibitive, especially if every protected amino acid reagent, supplied in excess, is isotope labelled. Cellular expression can yield labelled peptide of sufficient quantity for NMR data to be obtained, and through reverse phase HPLC (as with solid state synthesis) a sufficiently pure peptide can be produced.

Various sorts of labelling can be used to differentiate peptide from protein in a peptidase assay, and this includes isotope labelling of peptides with $^{13}C/^{15}N$ isotopes. Low natural abundance of ^{13}C and ^{15}N (at 1.07% and 0.37% respectively) means a protein or peptide to which they have not been introduced will not be apparent on a carbon or nitrogen HSQC spectrum, unlike the labelled peptide of interest (Rosman and Taylor 1998). Additionally, chemical shifts in both ^{1}H and $^{13}C/^{15}N$ dimensions of an HSQC provide more information about the peptide than proton NMR alone, and thus offer additional means of differentiating between similar peptides. Isotope-labelled peptides are therefore a useful tool.

The peptides cloned and expressed were SIINFEKL and 6 N-terminally extended derivative peptides as given in chapter 3, but including LSIINFEKL and LLSIINFEKL instead of KSIINFEKL and KKSIINFEKL. The oligonucleotide sequences for the cloning of all 9 peptides can be found in the appendices, though cloning was not completed for the lysine-extended peptides. Presented

here is the expression of isotope-labelled peptides and an example peptide characterised through heteronuclear NMR.

4.1.2 Heteronuclear correlation spectroscopy

Heteronuclear Single Quantum Coherence (HSQC) experiments make use of effective transfer of magnetisation between different nuclei. This is possible where the nuclei are directly bonded, and for this reason can be used to rapidly identify all connected nuclei of the relevant isotopes within the sample. For analysis of peptides and proteins, their chemical makeup being organic, a carbon and nitrogen labelled sample (providing isotopes at above natural abundance, and thus making the experiment rapid and sensitive) can reveal the $^{1}H/^{13}C$ or $^{1}H/^{15}N$ connections present.

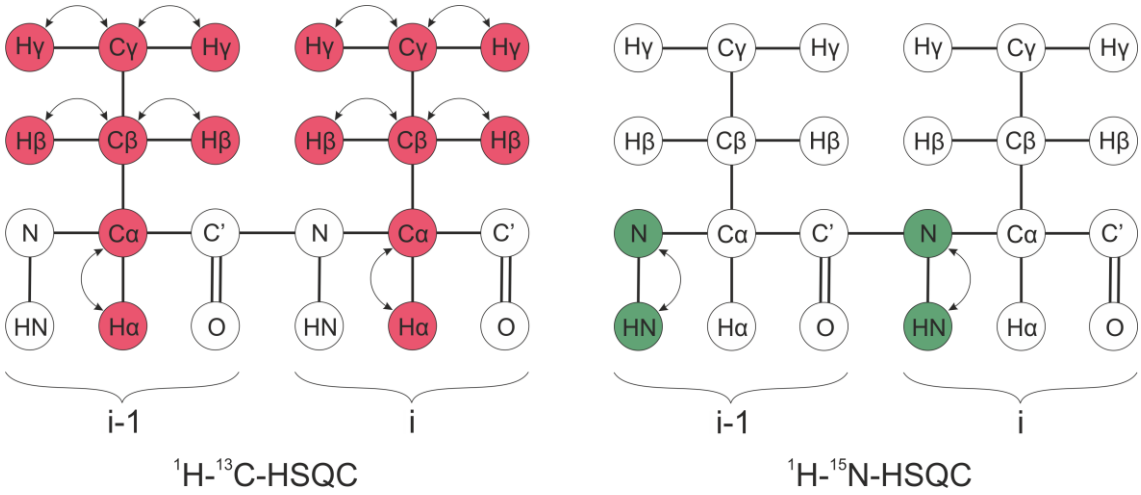


Figure 57 – ^{1}H-^{13}C-HSQC and ^{1}H-^{15}N-HSQC experiments. Magnetisation is transferred between the hydrogen and the attached $^{13}C/^{15}N$ atom by J-coupling. Adapted from www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/13c-hmqc/ and www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/1h-15n-hsqc/ (Higman 2012).

The process works much like homonuclear spectroscopy (see Chapter 3), except different nuclei can be independently probed by pulses at very different frequencies (Cavanagh et al. 1996). Excitation of protons can therefore be transferred to connected heteronuclei. Sensitivity differs between nuclei and

many experiments are designed with this consideration in mind: protons give signal-to-noise around 300 x greater than ^{15}N for example (Levitt 2008). An INEPT (insensitive nuclei enhanced by polarisation transfer) sequence bypasses this problem by magnetisation transfer from proton to heteronuclei during the evolution period then back to protons for detection, allowing the heteronuclei to be probed but maximising signal (Cavanagh et al. 1996).

Constant time experiments can be used for improved resolution and sensitivity. They have the additional bonus of aiding assignment: the number of connected carbon nuclei makes cross peaks either positive or negative due to coupling, so nuclei with 0 or 2 connected carbons will be opposite to those with 1 or 3.

4.2 Methods

4.2.1 A brief explanation of the expression system

Peptides were expressed as a fusion construct, with the ketosteroid isomerase protein (KSI) at the N-terminal with a linking methionine. This system enables bacterial peptide expression without degradation of the peptide by the expression organism, and the KSI protein can then be cleaved by a cyanogen bromide reaction after the linking methionine residue. The plasmid used (pET31b) is a low copy number plasmid, but can still provide sufficient antibiotic resistance for the ampicillin concentrations given below. For cloning, either a single enzyme (AlwNI) or double (XhoI or others; see Appendices for plasmid map) restriction digest can be used.

4.2.2 Cloning

4.2.2.1 Vector digestion:

The uncut pET31b plasmid was expressed in DH5α *E. coli*. Extraction of the plasmid DNA was performed with a QIAprep miniprep kit (from QIAGEN) according to the manufacturer's instructions. A digest reaction was mixed as follows:

Circularised plasmid double enzyme digest reaction mix:

500 ng vector
2 µl Tango buffer (2 ul)
1 µl AlwNI (Cail, supplied by ThermoFisher)
1 µl XhoI (supplied by ThermoFisher)
ddH_2O to 20 µl

The reaction was incubated at 37 °C for 1 hour. 2ul of the digested plasmid DNA was run on a 12% agarose gel to check for a digested band. A single band

indicated a linear plasmid, whereas two or three bands suggested incomplete digestion of the supercoiled bacterial plasmid.

4.2.2.2 Annealing of oligos:

The complementary single-stranded phosphorylated oligos were designed with AlwNI and XhoI restriction site compatible overhangs. A stop codon was included at the end of the peptide coding region and the obligatory cleavable methionine at the start. An annealing buffer for the oligos was prepared as follows:

10x annealing buffer:

0.485 g Tris pH 8
0.0952 g $MgCl_2$
0.292g NaCl
Made up to 100 ml with ddH_20

The oligos were mixed like so:

Oligo annealing reaction mix:

200 pmol oligo A
200 pmol oligo B
10 ul 10X annealing buffer
to 100 µl with ddH_20

The reaction was incubated in a thermocycler at 99 °C for 10 minutes, then cooled gradually to 30 °C, over 15 minutes, for optimal annealing without introduction of hairpins.

4.2.2.3 Ligation:

The inserts were combined with the linearised vector in the following mixture:

Plasmid/insert annealing reaction mix:

50 ng vector

100 / 200 ng insert (for a ratio of either 1:2 or 1:4)

The mixture was heated at 37 °C for 5 minutes in a thermocycler before being cooled on ice. At this point the ligase reaction was set up as follows:

Linear vector/insert ligation mixture:

1 µl ligase buffer

0.5 µl ligase

ddH_20 to 10 µl

The reaction was incubated in thermocycler at 16°C degrees overnight. After this, the ligase was denatured by heating at 65 °C for 10 minutes. The ligations were stored at -20 °C prior to further use.

4.2.2.4 DNA replication and sequencing:

The ligated plasmids were transformed into DH5α *E. coli* cells as follows: 2 µl of DNA was mixed with 20 µl of competent DH5α cells. The mixture was incubated on ice for 30 minutes. The cells were then heat shocked by incubation in water bath at 42 °C for 45 seconds. The mixture was returned to rest on ice for up to 2 minutes before 300 µl of LB (lysogeny broth – see recipe below) media was added, followed by incubation at 37 °C for 1 hour. 200 µl of transformed cells were then plated onto LB/ampicillin agar (with ampicillin at 100 µg/ml) and incubated at 37 °C overnight.

In order to check that the peptide-coding inserts had been correctly ligated into the vector, the DNA was extracted and sequenced like so: Firstly, well-isolated bacterial colonies were chosen from the agar plated cells and used to inoculate 10 ml of LB/ampicillin. This was cultured overnight (~16 hours) at 37 °C in a

shaking incubator. The plasmids were extracted from cells of the overnight culture by miniprep with a QIAprep kit (QIAGEN) according to manufacturer's instructions. The extracted DNA was dispatched for Sanger sequencing (Eurofins Sanger sequencing service) for verification of successful ligation.

4.2.3 Expression

4.2.3.1 Expression of unlabelled KSI fusion peptides

1L volumes of LB (lysogeny broth) media were prepared beforehand to the following recipe:

LB broth:

10 g tryptone
5 g yeast extract
10 g NaCl
Adjusted pH to 7, then made up to 1L and autoclaved

The peptides were expressed in inclusion bodies in *E. coli*. 2 µl of DNA was mixed with 20 µl of competent BL21 RIPL *E. coli* cells. The mixture was incubated on ice for 30 minutes. The cells were then heat shocked by incubation in water bath at 42 °C for 45 seconds. The mixture was returned to rest on ice for up to 2 minutes, then 300 µl of sterile LB media was added, followed by incubation at 37 °C for 1 hour. 200 µl of transformed cells were then plated onto sterile LB/ampicillin agar (with ampicillin at 100 µg/ml) and incubated at 37 °C overnight.

A well-isolated colony was picked from the agar plates and used to inoculate 100 ml LB/ampicillin media (ampicillin to 100 µg /ml). The media was cultured overnight (~16 hours) at 37 °C in a shaking incubator. Simultaneously, 900 ml volumes of sterile LB/ampicillin (100 µg /ml) media were prewarmed overnight to 37 °C in baffled flasks.

The prewarmed media was inoculated with 100 ml of overnight culture to make 1L volumes. The inoculated media was grown at 37 °C in in a shaking incubator at 200 rpm. When the culture reached an optical density absorbance value of 0.8 – 1 (usually between 1.5-2.5 hours) protein expression was induced by addition of IPTG (isopropyl β-D-1-thiogalactopyranoside) to 1mM. After a further 4 hours, the cells were extracted by centrifugation at 5000 rpm, 4°C for 20 minutes. The used media was discarded and the cell pellet stored at -20 °C before purification.

4.2.3.2 Expression of labelled KSI fusion peptides

The peptides were expressed in inclusion bodies in *E. coli*. 2 µl of DNA was mixed with 20 µl of competent BL21 RIPL *E. coli* cells. The mixture was incubated on ice for 30 minutes. The cells were then heat shocked by incubation in water bath at 42 °C for 45 seconds. The mixture was returned to rest on ice for up to 2 minutes, then 300 µl of sterile LB media was added, followed by incubation at 37 °C for 1 hour. 200 µl of transformed cells were then plated onto sterile LB/ampicillin agar (with ampicillin at 100 µg/ml) and incubated at 37 °C overnight.

A well-isolated colony was picked from the agar plates and used to inoculate 100 ml LB/ampicillin media (ampicillin to 100 µg /ml). The media was cultured overnight (~16 hours) at 37 °C in a shaking incubator.

M9 growth media and 10x M9 salts were prepared like so:

10x M9 salts:

60 g Na_2HPO_4

30 g KH_2PO_4

5 g NaCl

This was mixed and made up to 1 litre.

M9 minimal growth media (initial formulation):

1 ml 1M $MgSO_4$

0.3 ml 1 M $CaCl_2$

100 ml 10x M9 salts

pH 7.4, ddH_2O to 1L

The initial growth media was divided between baffled flasks, with 500 ml in each flask. The flasks were sealed and autoclaved.

A 100x trace elements solution was prepared to the following recipe:

100x trace elements solution:

5 g EDTA (pH 7.5)

0.83 g $FeCl_3.6H_2O$

84 mg $ZnCl_2$

13 mg $CuCl_2.2H_2O$

10 mg $CoCl_2.6H_2O$

10 mg H_3BO_3

1.6 mg $MnCl_2.6H_2O$

Made up to 1 litre, 0.2 μm filtered to sterilise.

The sterile M9 media was prepared for use in cell culture expression:

M9 growth media (final formulation):

M9 growth media (initial formulation) 500 ml

500 mg NH_4Cl (^{15}N labelled, supplied by Goss Scientific) 0.2 μm filter sterilised

1 g $^{13}C_6$ glucose (supplied by Goss Scientific) 0.2 μm filter sterilised

500 μl biotin stock (of 1 mg/ml) pH 7

500 μl ThiamineHCl stock (of 0.1 mg/ml)

5 ml of 100x trace elements

M9 media was prewarmed to 37 °C and ampicillin stock added to 100 μg/ml.

Overnight cultures were centrifuged at 4000 rpm, 10 minutes, 4 °C to pellet cells. The pellet was resuspended in 20 ml of prewarmed media, and 10 ml added per 490 ml of M9 media. Flasks were incubated at 37 °C in shaking incubators and induced with IPTG to 1 mM when a relative optical density absorbance of 1 was reached. The cultures were left to incubate at 37 °C in a shaking incubator overnight (~16 hours).

After the overnight incubation, an increase in the optical density of cultures was checked to verify that cells had grown overnight. Cells were pelleted by centrifugation at 5000 rpm, 20 mins, 4 °C. The pellet was weighed and frozen.

SDS-PAGE

12% SDS polyacrylamide gels were cast in Bio-Rad 0.75 mm mini-PROTEAN plates and used for analysis of expressed and purified protein.

A gel buffer was produced to the following recipe:

Gel buffer:

0.3 % SDS
3M Tris, pH 8.45

The running gel was made to the following recipe:

Running gel:

3.33 ml 30% bis acrylamide (37.5:1)
3.33 ml dH2O
3.33 ml gel buffer
10 μl TEMED

This was mixed before 50 µl ammonium persulphate (100 mg/ml stock) was added. After rapid and thorough mixing, the mixture was distributed between 3 sets of casting plates. Isopropanol was added on top to a few millimetres and the gels were left for ~30 minutes to set. The isopropanol was removed by blotting with filter paper.

A stacking gel was then made to create wells in the polyacrylamide gels. This was mixed to the following recipe:

Stacking gel:

660 µl 30% bis acrylamide (37.5:1)
2.9 ml dH2O
1.24 ml gel buffer (see above)
10 µl TEMED

This was mixed before 25 µl ammonium persulphate (100 mg/ml stock) was added. The stacking gel mixture was then rapidly and thoroughly mixed again, before being distributed atop the cast running gels. Well combs (usually of 10 wells) were inserted and the gels left to set for another 30 minutes.

To prepare samples for running on a gel the samples were mixed with sample buffer. A 4x Sample buffer was made like so:

4x sample buffer:

200 mM Tris-HCl (pH 6.8)
8% w/v SDS
0.8% v/v bromophenol blue
40% v/v glycerol
400 mM β-mercaptoethanol
ddH_2O to 10 ml

Pre- and post-induction samples from the LB *E. coli* expression of KSI-peptide inclusion bodies were run on a gel to check expression. 500 µl of each were centrifuged at 13000 rpm in a benchtop Eppendorf centrifuge for five minutes. The supernatant was removed. 25 µl of 4x sample buffer was used to resuspend the pellet. The sample was heated at 95 °C in a heat block for 10 minutes. 5 µl of protein ladder (Bio-Rad 'dual-color precision plus protein') was added to the first well. 5 µl of the protein expression sample was inserted to a well on the SDS-PAGE. Samples of purified inclusion bodies were similarly checked by mixing with sample buffer, heating at 95 °C in a heat block for 10 minutes and running 5 µl on a gel.

The gels were run in a Bio-Rad mini-PROTEAN tank according to the manufacturer's instructions. The running time was approximately 40 minutes. The anode buffer was 200 mM Tris (pH 8.9). The cathode buffer was made to the following recipe:

Cathode buffer:

100 mM Tris (pH 8.2)
100 mM Tricine
0.1% SDS

4.2.4 Peptide purification

4.2.4.1 Inclusion body purification

In order to purify the KSI-fusion-peptide inclusion bodies from the cells they were expressed in, a Triton buffer was made:

Triton buffer:

50 mM Tris-HCl (pH 8)
10 mM EDTA
0.5% v/v Triton X-100 (Sigma)

20 ml of Triton Buffer was added to harvested cell pellet from expression along with lysozyme (to 0.1 mg/ml), DNase I (to 0.02 mg/ml) and $MgCl_2$ (to 10 mM). The pellet was homogenised and left on a rocking shaker at room temperature for 20 minutes. The mixture was centrifuged at 12 000 x g, 15 minutes, 4 °C and the supernatant removed. This was repeated with the same quantities of Triton Buffer, lysozyme, DNase and $MgCl_2$ before homogenisation and 10 minute incubation at room temperature before an identical centrifugation step and removal of supernatant.

Three cycles of Triton Buffer addition on ice, homogenisation and centrifugation as above were carried out followed each time by removal of supernatant. This was repeated two times further with 5 ml dH_2O instead of Triton Buffer to remove detergent.

4.2.4.2 Chemical cleaving of peptides from KSI fusion protein

The KSI fusion protein contains no internal methionine residues, and the linking methionine between KSI and the expression peptide is therefore useful as a cleavage site between the two by use of cyanogen bromide, to which it is uniquely susceptible (Wagstaff, Howard, and Williamson 2010). The sulphur of methionine performs a nucleophilic attack on the carbon of CNBr and is replaced by bromine, and the methionine forms an unstable 5-membered ring which results in cleaving of methionine from the N-terminal of the peptide. There are few limitations to use of the method, except obvious restrictions on inclusion of methionine within the peptide, and lower efficiency of cleavage for serine and threonine residues, which can be optimised (Kaiser and Metzka 1999).

To cleave the peptide from the KSI fusion protein, 5 ml of 80% (or 60% for N-terminal serine peptides) formic acid was added to the pellets and they were magnetically stirred to resuspend. 0.2 g CNBr was added in a fume hood, and the tubes were rapidly sealed and covered in foil. The tubes were stirred overnight.

The following day, 20 ml dH_2O was added to the reactions and the solution was lyophilised. The hydrophobicity of the peptides determined how they were subsequently handled. Dried peptides were resuspended one of two ways:

Hydrophobic peptides:

Hydrophobic peptides were resuspended in 4 ml of a resuspension buffer suitable for hydrophobic peptides:

40% acetonitrile
60% dH_2O
0.1% trifluoroacetic acid

They were left on a rotating shaker at room temperature for 1 hour, then centrifuged 12 000 x g, 10 minutes, 4 °C. The supernatant was filtered with 0.2 µM nylon Corning costar spin-x HPLC filters.

Peptides handled this way were: SIINFEKL, ESIINFEKL, LSIINFEKL, LLSIINFEKL, PSIINFEKL & PPSIINFEKL

Hydrophilic peptides:

Hydrophilic peptides were resuspended in 4 ml of a resuspension buffer suitable for hydrophilic peptides:

20 mM K_2HPO_4
100 mM NaCl

The pH was adjusted with $NaHCO_3$ to pH 7.4 and the solution covered in foil and left overnight on a rotating shaker at room temperature. The following day the mixture was centrifuged 5000 x g, 15 minutes, 4 °C, before filtering with a 0.2 µM filter.

Peptides handled this way were: EESIINFEKL.

4.2.4.3 Purification and analysis of expressed peptides

Peptides were purified by reverse phase HPLC Phenomenex Jupiter C18 column. An aqueous Buffer A (HPLC-grade H_2O with 0.1% trifluoroacetic acid) and organic Buffer B (HPLC-grade acetonitrile with 0.1% trifluoroacetic acid) were used to set up an HPLC gradient. A gradient of 10-40% Buffer B was used.

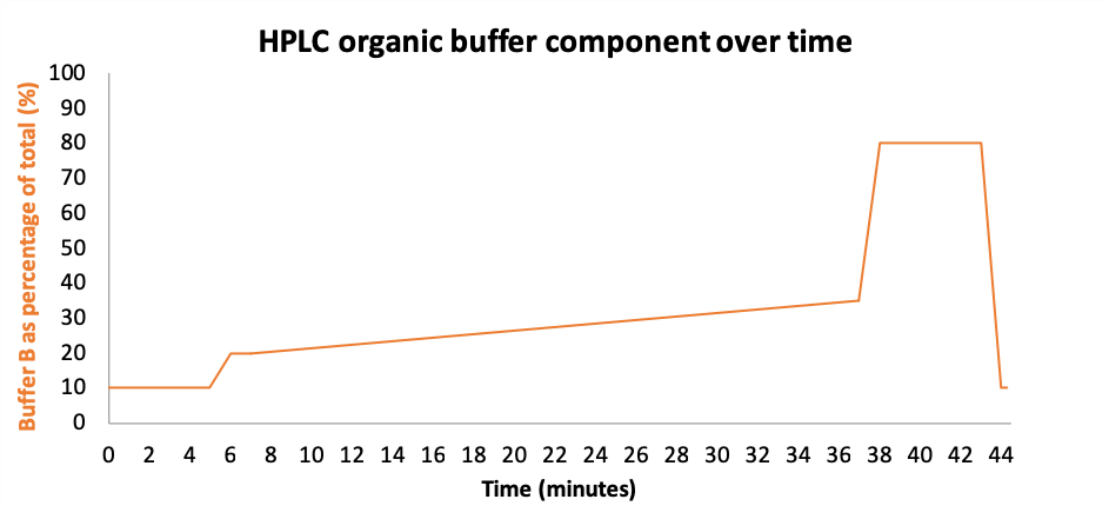


Figure 58 - HPLC separation method for SIINFEKl series peptides given as proportion of Buffer B over time

Peptides were identified by retention time on the column and verified by time of flight positive electrospray mass spectrometry.

4.2.4.4 ^{15}N ^{13}C ^{1}H heteronuclear NMR of labelled peptides

$^{13}C/^{15}N$ isotope labelled PSIINFEKL peptide was prepared for NMR by dissolving in 10% D_2O/90% H_2O, and checked for HSQC quality. The peptide was freeze dried and the lyophilised peptide resuspended in water. This was then dialysed over 1L of 20 mM potassium phosphate (mono/dibasic) buffer at pH 7 using a 100-500 Da cut-off membrane at 4 °C for 2 days.

Data acquisition:

All 2D heteronuclear HSQC experiments used standard pulse sequences from the Varian BioPack software.

Processing and assignment:

The ^{13}C and ^{15}N dimensions were referenced by indirect referencing from the known proton dimension centre position. The gyrometric ratios used were taken from Wishart, Bigam, Yao, et al. (1995). Data was processed in nmrPipe and nmrDraw (Delaglio et al. 1995). Assignments were made in CCPN Analysis (Vranken et al. 2005) from solved ^{1}H assignments.

4.3 Results

4.3.1 Expression

KSI-linked peptides showed good induction upon addition of IPTG during M9 media expression in *E. coli*. Cells were pelleted with about 1 g per 500 ml media after ~16 hours induced growth. After inclusion body purification, pellets were ~0.5 g prior to CNBr cleavage.

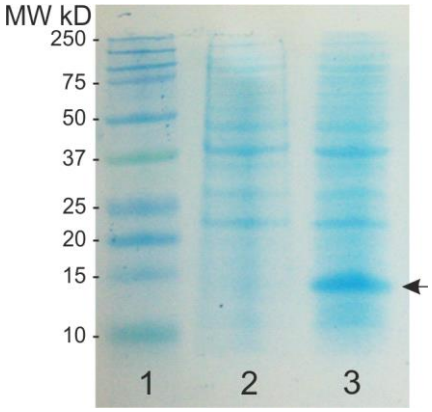


Figure 59 - SDS-PAGE of isotope labelled KSI-fused SIINFEKL expression. Lane1: Bio-Rad Precision Plus Kaleidoscope protein standard. Lane 2: pre IPTG induction sample of peptide expression. Lane 3: post IPTG induction of peptide expression. Arrow denotes band of over-expressed KSI-tagged peptide

4.3.2 Purification and analysis

Expressed and cleaved peptides produced large peaks on the HPLC gradient and fractions collected from these were analysed. Mass spectrometry of unlabelled peptides showed the expected monoisotropic masses for +1 and +2 peaks (Figure 61). Mass spectrometry of labelled peptides showed the expected monoisotropic masses for +1 and +2 peaks (Figure 62). Of the +1 ionised species present close to the expected peptide mass, the majority (over 2/3) had the monoisotropic mass of 1121.78, indicating most of the peptide was fully labelled (Figure 63). 1L of M9 $^{13}C/^{15}N$ labelled expression yielded in the region of 2 mg of HPLC-purified peptide.

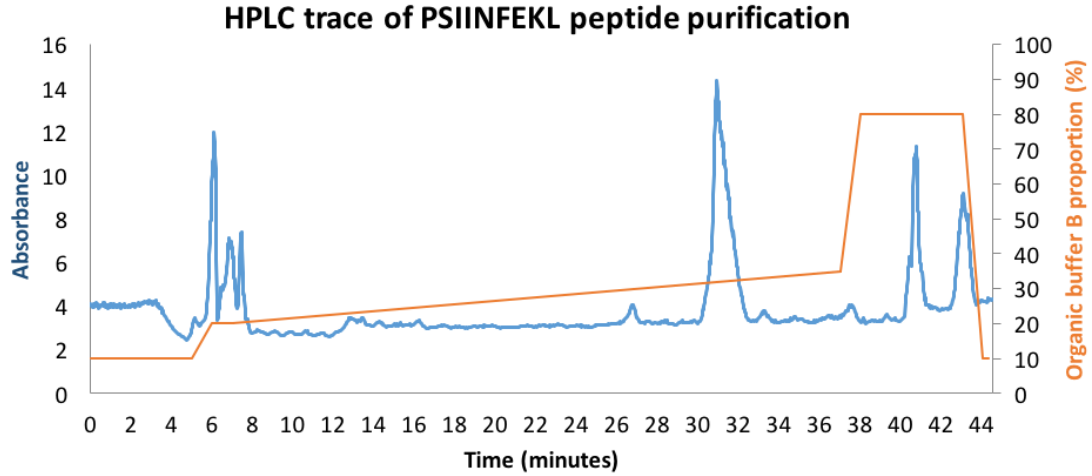


Figure 60 – Reverse phase HPLC trace of PSIINFEKL purification showing detected signal (blue) and organic buffer component (orange) against time. PSIINFEKL is eluted between 30-32 mins

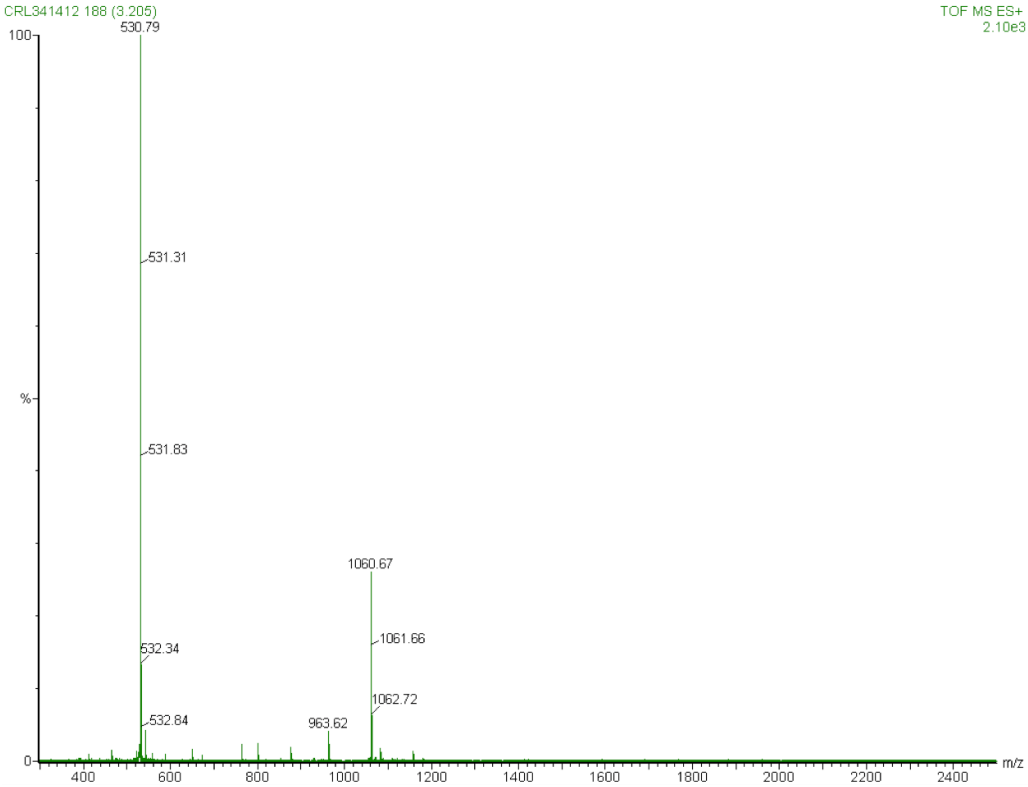


Figure 61 – Time of flight positive electrospray mass spectrometry profile of unlabelled cell-expressed PSIINFEKL sample (post HPLC purification). Expected monoisotropic mass is 1060.24 Da

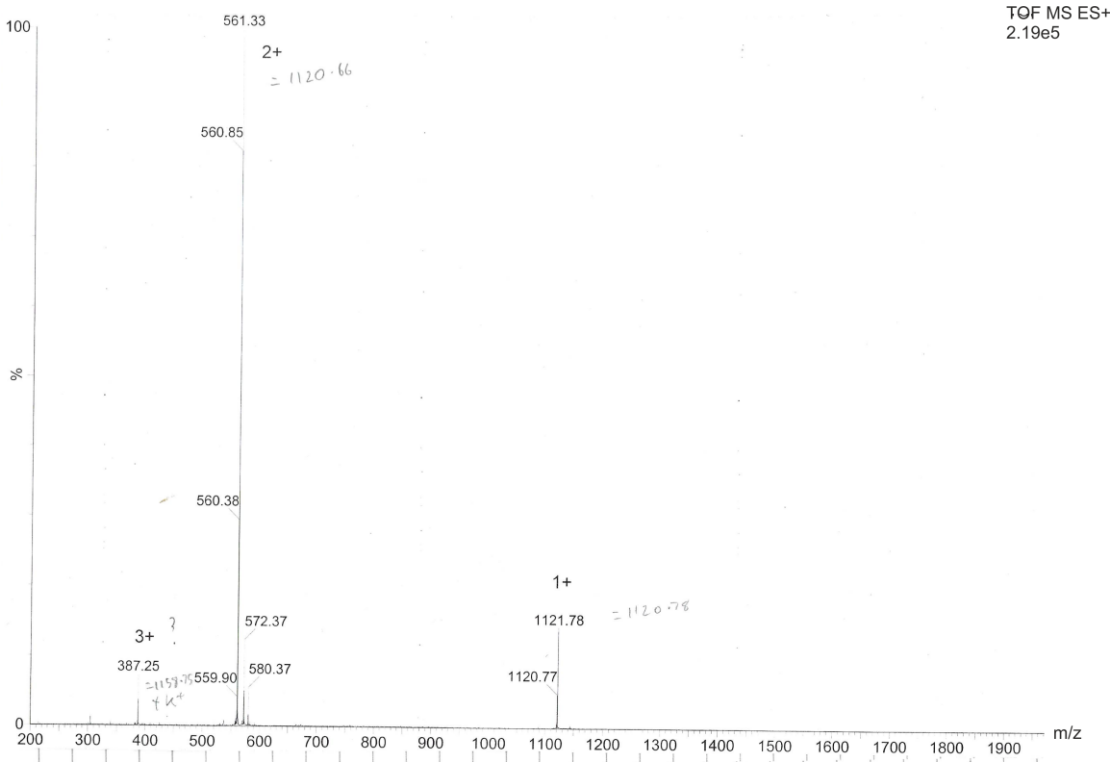


Figure 62 - Positive electrospray time of flight mass spectrometry profile of labelled PSIINFEKL sample (post HPLC purification). Expected monoisotropic mass is 1121.2

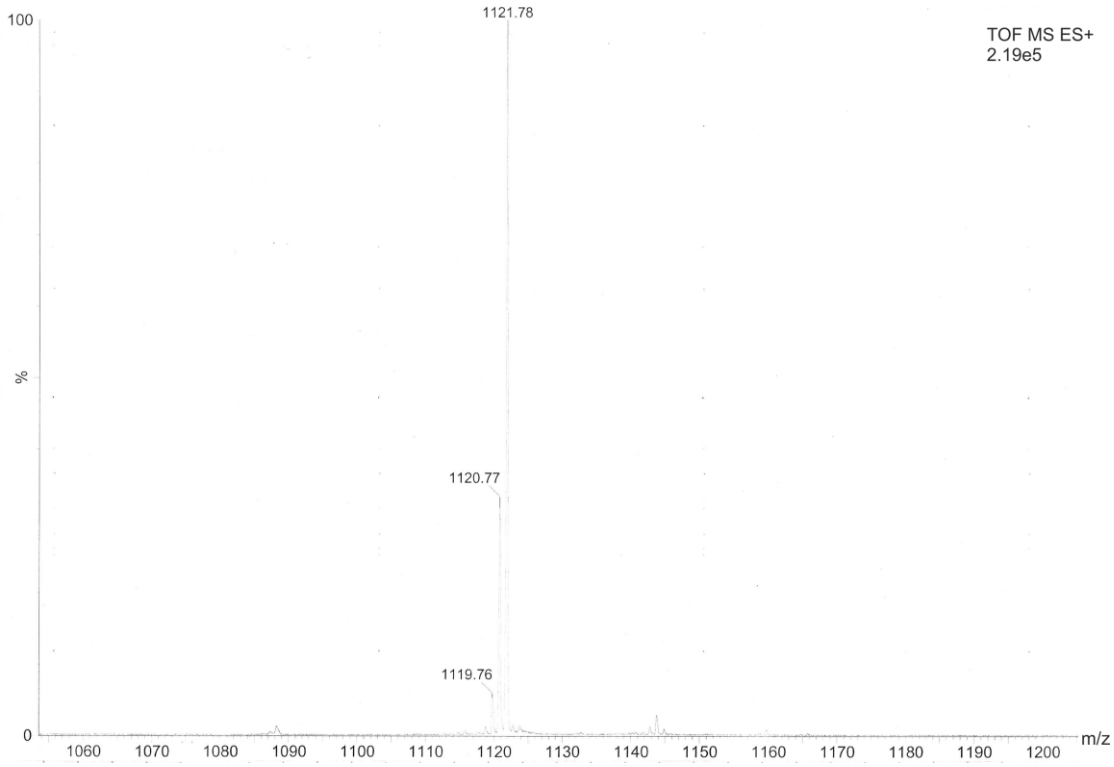


Figure 63 - Positive electrospray time of flight mass spectrometry profile of $^{13}C/^{15}N$ labelled PSIINFEKL showing +1 ionisation peaks

4.3.3 Results of NMR assignment of 13C/15N labelled PSIINFEKL

Where confident assignments for protons are available, 2D heteronuclear experiments offer rapid assignment of other nuclei present, most commonly ^{13}C and ^{15}N. Simple HSQC experiments utilise transfer of magnetisation between connected heteronuclei, and can therefore provide chemical shift assignments in the second dimension based on knowledge of the first, providing significantly more assignment information on a peptide than homonuclear NMR can offer (Cavanagh et al. 1996).

4.3.3.1 ^{1}H-^{15}N HSQC spectra

For the ^{15}N assignments, 2 ^{1}H-^{15}N spectra were collected and the second optimised spectrum (Figure 64) with a higher number of scans and increments was used for ^{15}N assignments.

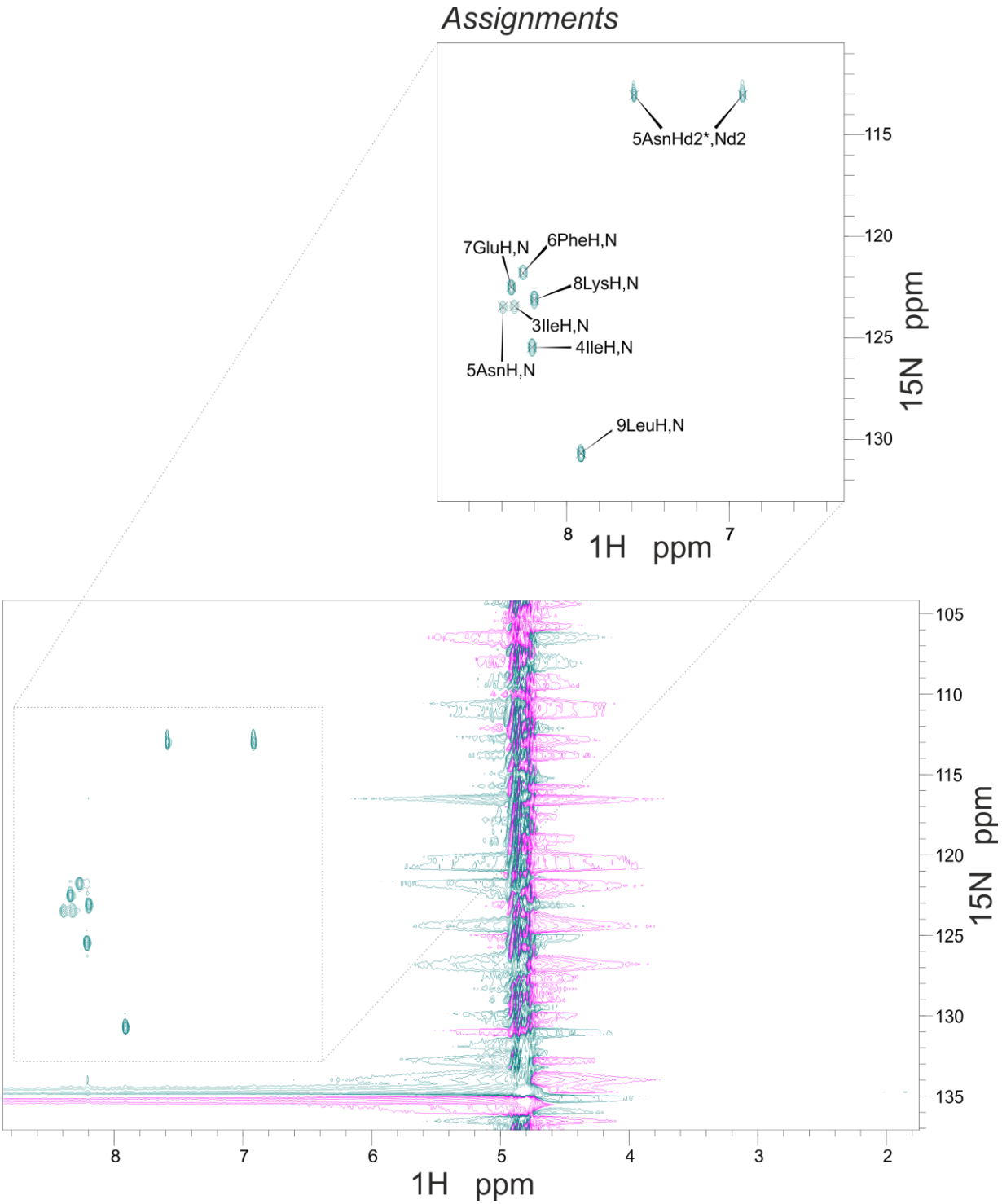


Figure 64 - ^{1}H-^{15}N HSQC spectrum of ^{15}N labelled PSIINFEKL

4.3.3.2 ^{1}H-^{13}C HSQC spectra

Aliphatic CT spectra

For the ^{13}C aliphatic groups, 3 ^{13}C HSQC spectra were collected, of which 2 were constant time (CT) experiments, performed to give additional chemical information about the peptide. The second optimised CT spectrum (Figure 65) was used for aliphatic carbon groups chemical shift assignment.

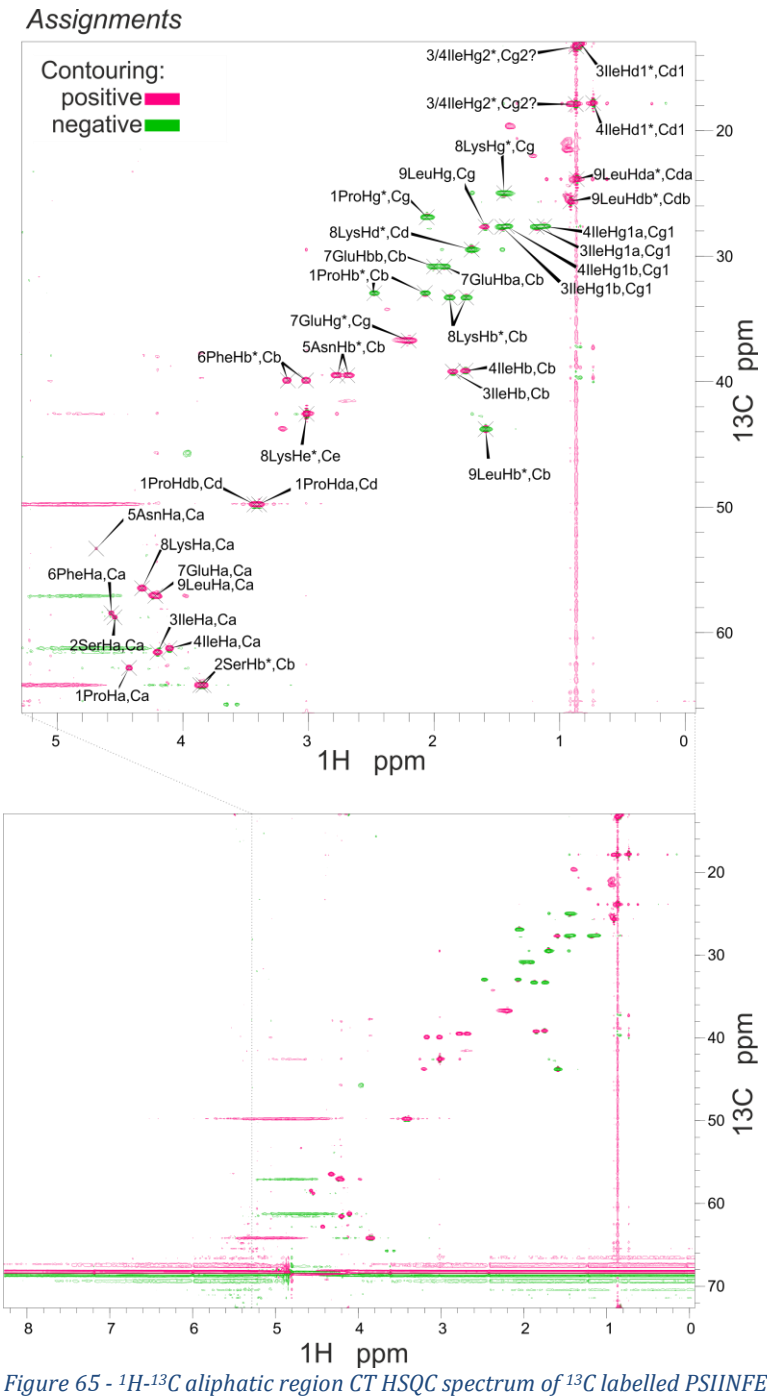


Figure 65 - ^{1}H-^{13}C aliphatic region CT HSQC spectrum of ^{13}C labelled PSIINFEKL

Aromatic CHSQC spectrum

As expected, the ^{1}H-^{13}C aromatic region HSQC (Figure 66) showed only three peaks, corresponding to the 3 chemical environments sampled by 5 CH groups of the phenylalanine ring in PSIINFEKL.

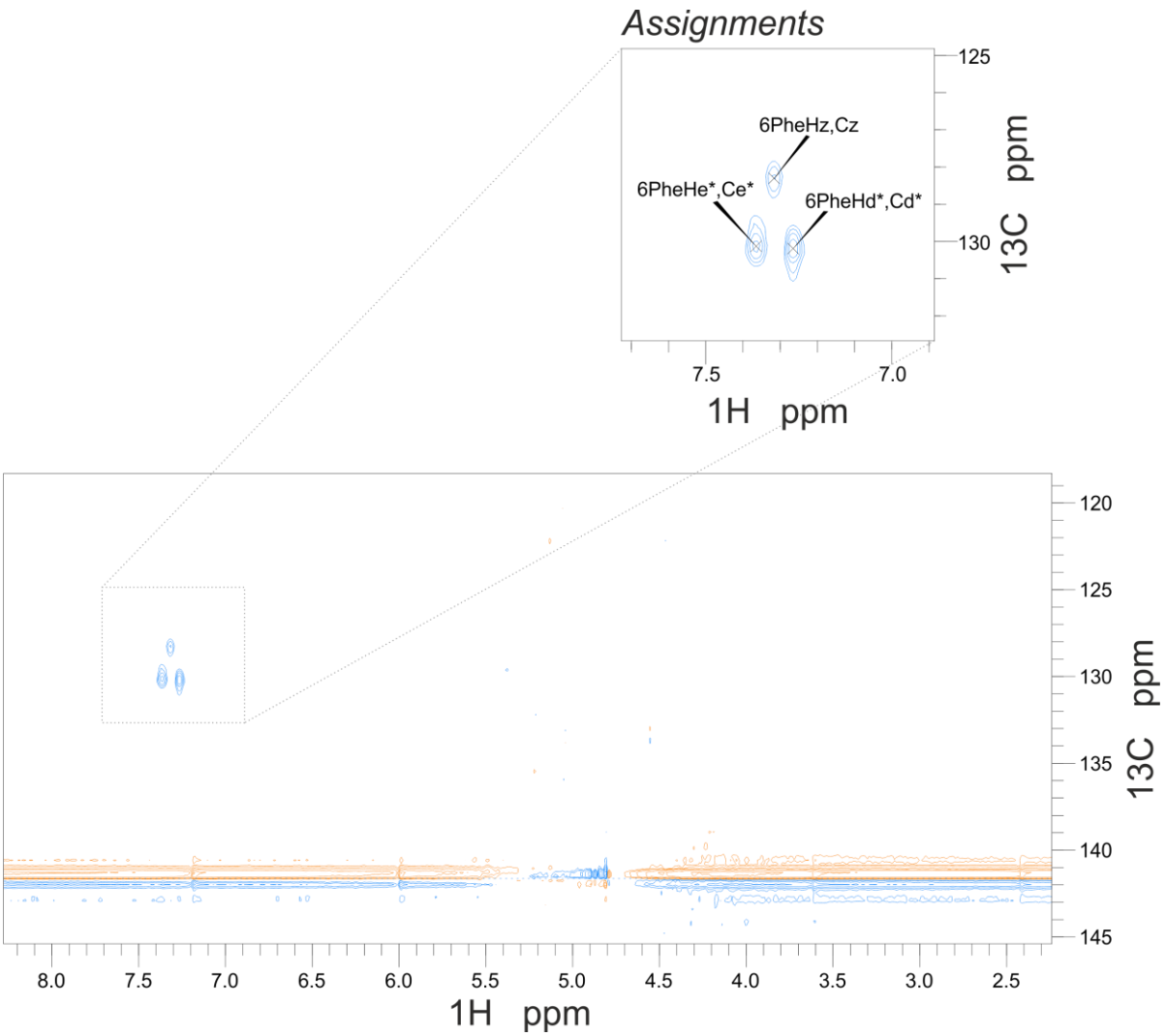


Figure 66 - ^{1}H-^{13}C aromatic region HSQC spectrum of ^{13}C labelled PSIINFEKL

4.3.3.3 Chemical shift assignments

The HSQC spectra (see Figure 64, Figure 65 & Figure 66) obtained provided unambiguous ^{1}H, ^{13}C and ^{15}N assignments in most cases (given Table 8) which matched the existing ^{1}H assignments for PSIINFEKL (See Chapter 3 Proton Resonance Assignments of a Peptide Series & Appendix 13 – Peptide assignments in 90% H2O 10% D2O). No cross-peak for the backbone NH of 1Pro or 2Ser were found in the ^{1}H-^{15}N spectrum. The ^{13}C and ^{15}N nuclei were confidently assigned in all cases other than the Cγ2 nuclei of the 3rd and 4th position isoleucine residues, where the positions on the ^{1}H axis were sufficiently close to pose uncertainty about which methyl group was giving rise to each signal.

Table 8 - HSQC spectral assignments for PSIINFEKL (specific stereochemical assignments e.g. Hba/Hbb have been arbitrarily designated by CCPN Analysis and are undetermined)

1Pro	**H**	-	**Ha**	4.43	**Hba**	2.07	**Hbb**	2.48	**Hg***	2.06	**Hda**	3.39	**Hdb**	3.43
			Ca	62.81	**Cb**	32.95			**Cg**	26.90	**Cd**	49.78		
	N	-												
2Ser	**H**	-	**Ha**	4.55	**Hba**	3.84	**Hbb**	3.86						
			Ca	58.77	**Cb**	64.17								
	N	-												
3Ile	**H**	8.32	**Ha**	4.20	**Hb**	1.85	**Hg1a**	1.18	**Hg1b**	1.46	**Hg2**	0.87	**Hd1***	0.83
			Ca	61.56	**Cb**	39.21	**Cg1**	27.67			**Cg2**	13.30? 17.89?	**Cd1**	13.08
	N	123.45												
4Ile	**H**	8.21	**Ha**	4.11	**Hb**	1.75	**Hg1a**	1.12	**Hg1b**	1.42	**Hg2**	0.87	**Hd1***	0.73
			Ca	61.25	**Cb**	39.21	**Cg1**	27.67			**Cg2**	13.30? 17.89?	**Cd1**	17.80
	N	125.48												
5Asn	**H**	8.39	**Ha**	4.69	**Hba**	2.69	**Hbb**	2.77	**Hd2a**	6.92	**Hd2b**	7.59		
			Ca	53.30	**Cb**	39.50								
	N	123.47							**Nd***	113.05				
6Phe	**H**	8.27	**Ha**	4.57	**Hba**	3.02	**Hbb**	3.16	**Hd***	7.27	**He***	7.37	**Hz**	7.32
			Ca	58.45	**Cb**	39.91			**Cd***	130.19	**Ce***	130.13	**Cz**	128.30
	N	121.79												
7Glu	**H**	8.34	**Ha**	4.22	**Hba**	1.92	**Hbb**	2.00	**Hg***	2.20				
			Ca	57.06	**Cb**	30.83			**Cg**	36.73				
	N	122.50												
8Lys	**H**	8.20	**Ha**	4.33	**Hba**	1.88	**Hbb**	1.75	**Hd***	1.70	**Hg***	1.44	**He***	3.02
			Ca	57.06	**Cb**	33.30			**Cd**	29.48	**Cg**	24.98	**Ce**	42.57
	N	123.13												
9Leu	**H**	7.91	**Ha**	4.22	**Hb***	1.59	**Hg**	1.59	**Hda***	0.86	**Hdb***	0.91		0.87
			Ca	57.06	**Cb**	43.78	**Cg**	27.67	**Cda**	23.87	**Cdb**	25.64		17.89
	N	130.69												

4.4 Discussion

Assignments were straightforward, with proton chemical shifts identified according to the existing homonuclear NMR assignments (see Chapter 3 Proton Resonance Assignments of a Peptide Series) and verified by the positive or negative sign of the peak in the aliphatic CT spectrum. Cross-peaks for the backbone NH of N-terminal 1Pro were not expected, and accordingly did not appear in the ^{1}H-^{15}N spectrum. The lack of cross-peak representing the backbone NH of 2Ser was also unsurprising, given that the second position residue backbone NH was not apparent in the homonuclear experiments performed for SIINFEKL and ESIINFEKL in 90% H_2O (Chapter 3). Some unassigned peaks existed in the aliphatic spectra, presumed due to impurities in the sample because they did not match known proton chemical shifts for PSIINFEKL. This indicates that although the peptide was expressed in large enough quantities for good NMR spectra to be obtained, purity could have been higher, and perhaps a shallower HPLC gradient used to separate impurities.

The Cγ2 nuclei of the 3rd and 4th position isoleucine residues were given ambiguous assignments, but due to the location on the 1H axis and the cross-peaks being positive, it was evident that the chemical group producing these signals was a methyl group. However, despite this uncertainty, the aliphatic spectra presented additional proton assignment details for the ProHβ protons. These were previously assigned chemical shift values around 2.50 ppm, but for which in the ^{1}H-^{13}C spectra an additional peak was seen at 2.07 ppm, overlaying the ProHγ* position. Similarly, the LysHβ protons gave rise to 2 chemical shifts (at 1.88 and 1.75 ppm) where only one (at 1.88 ppm) had previously been recorded as representative of both protons. As the number of nuclei represented by a peak has a bearing on the accuracy of quantitative NMR this is useful for highlighting missteps in a proton assignment from homonuclear ^{1}H-^{1}H TOCSY spectra.

5 Investigating human ERAP1 behaviour with NMR

5.1 Introduction

To date, several methods have been used for in vitro exploration of the activity of ERAP1 against different substrates. Broadly, these are often adept at testing modulation of one aspect of the peptide and how it affects the trimming rate. The simplest experiment is perhaps the use of substrates such as L-leucine-7-amido-4-methylcoumarin (L-AMC) which provides a consistent fluorescent reading when the leucine group is cleaved by ERAP1. A selection of 82 fluorescent substrates was used in a thorough comparison of substrates for ERAP1, ERAP2 and IRAP, suggesting future inhibitor design principles, by Zervoudi et al. (2011). AMC-based fluorescent compounds are small and therefore other assays are more commonly used to assess ERAP1 trimming of longer peptide substrates. A fluorescent substrate used by Evnouchidou, Berardi and Stratikos (2009) was a whole peptide with an N-terminal tryptophan, the fluorescence of the tryptophan being quenched by an internal cysteine/dinitrophenyl, and restored when it was cleaved. This is effective for assessing peptide internal residue preferences of ERAP1, but cannot also compare N-terminal preference. HPLC separation of peptide products is one such method which can be highly effective (Evnouchidou et al. 2008) although the peptides chosen must separate well on the reverse-phase gradient, and multiple reactions started then stopped at timepoints. A lengthy gradient may be required to separate all products making it a time-consuming experiment, but unlike the dinitrophenyl-quenched tryptophan method, it does not require chemical modifications to the peptide which might affect ERAP1 activity.

To examine both N-terminal trimming and internal residue preferences, multiple methods might be required. If the aminopeptidase has such strict 'molecular ruler' behaviour it cannot trim beyond the optimal MHC peptide then no more is required, but ERAP1 is known to degrade shorter peptides (Kanaseki et al. 2006). Given suggestions by Evnouchidou et al. (2008)that ERAP1 exhibits

internal amino acid preference, these assays do not give a full and nuanced picture of trimming by ERAP1, as they monitor only a single trimming step.

Nuclear magnetic resonance is capable of identifying peptides based on chemical shift, and can also be used quantitatively because the size of each peak is proportional to the number of protons (and, by extension, moles of compound) it represents. This presents the opportunity for real-time reaction progress monitoring. Published reactions investigated by this method usually have simple enough products and substrates to quickly identify species from well isolated peaks, even if the substrate in question is a complex polymer (Limtiaco et al. 2011). Utilising reaction progress monitoring NMR for a complex mixture of substrates and products as are present in an ERAP1 reaction is unusual but offers the opportunity to study every step of the reaction, including and beyond cleaving of the first N-terminal residue.

5.1.1 Quantitative NMR

Peak size relates to number of hydrogens present, this can be split into the number of hydrogens present in the represented species, and the number of species represented by the peak, i.e. concentration. Perhaps the simplest way of quantitatively performing 1D proton NMR is by use of a chemical reference. This is a compound which must ideally be stable and have a strong, reliable signal, well separated from all other peaks. It acts as a standard in the sample, with the known reference concentration used to determine sample concentration.

5.1.1.1 Considerations for NMR quantitation

Although the size of a peak on an NMR spectrum corresponds to the number of nuclei it represents, it is not always possible to calculate the number of nuclei from the peak size. Labile peaks in exchange with the solvent will give a low signal and are best discounted altogether. Another consideration is the

relaxation time, T1, of different nuclei. T1 ('longitudinal' or 'spin-lattice') relaxation dictates the period of time taken to return to a relaxed equilibrium state $M_{z,eq}$ from M_z after being subjected to the radio-frequency (R.F.) pulse. It is linked to the tumbling of the molecule within solution and therefore affected by factors such as size of the molecule (Wüthrich 1986). The T1 can often be in the order of several seconds, and if the relaxation delay (D1) of the experiment is of insufficient length, the recorded signal will be a fraction of what it would be if the nuclei were permitted full relaxation.

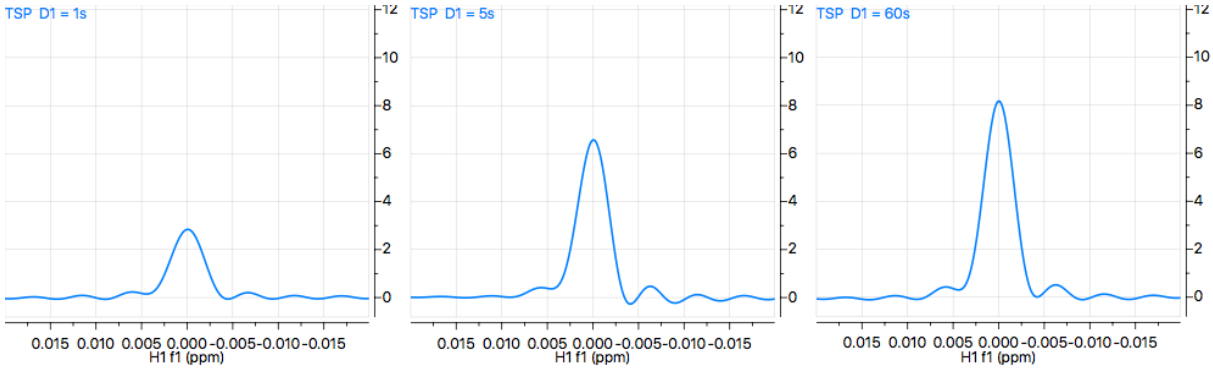


Figure 67 - Partial to full relaxation of TSP reference signal with D1 times from 1 to 60 seconds. Acquisition time: 0.32 s.

The ideal relaxation delay would be 3-5 times longer than the T1 of a given nuclei. This gives a maximum error of around 5% for 3x greater than T1, or 1% if 5x greater (Cookson and Smith 1982). In reaction-monitoring there are constraints imposed by the reaction, if rapid, and a long relaxation delay has a significant impact on experiment time. The D1 must be long enough for the nuclear magnetisation of nuclei of interest to return to equilibrium before the next scan in the experiment. Optimisation of the Ernst flip angle can ensure full relaxation in a shorter T1, giving the maximum possible signal, and this route can reduce experiment times without too much impact. However, this will not yield the good signal-to-noise of large 70-90° flip angles and is inadvisable for quantitation where multiple different T1 values are present in the sample (Cookson and Smith 1982).

Another vital consideration is signal-to-noise. Poor signal-to-noise will generate large errors, and this is particularly a problem for small peaks representing only one or two nuclei, for all low-concentration components of the reaction mix, and must also be considered when the chosen D1 is substantially shorter than T1 relaxation. The lower limit on signal-to-noise is often recommended as a ratio of 250, giving 0.4% error. With an adequate number of scans, the signal-to-noise can be improved until it is acceptable for all peaks. This also lengthens the experiment time, and again a trade-off is required to capture all reaction events while retaining a practicable level of signal-to-noise.

Lineshape of the final spectrum is a key consideration and is affected by sample shimming and acquisition time. Poor shimming can introduce problems where peaks are broad or have shoulders, rather than the expected Lorentzian lineshape. The acquisition period determines, along with spectral (sweep) width, the number of points in in the spectrum, and thereby spectral digitisation and resolution. This is shown by:

$$res = \frac{1}{at} = \frac{2sw}{np}$$

And:

$$at = \frac{np}{2sw}$$

Where 'res' is the resolution, 'at' is acquisition time in seconds, 'sw' is the sweep width in Hz, and 'np' is the number of points in a spectrum.

Digitisation has an effect on integration because data points within an NMR peak will affect fitting, and sampling for integration (see Figure 68). A short acquisition time will lead to truncation of the spectrum and baseline distortion (sinc wiggles) around peaks (this is apparent around the right hand peak of

Figure 68). An additional consideration is that very extensive acquisition times give more noise inclusion, and a poorer signal-to-noise (Becker, Ferretti, and Gambhir 1979). The acquisition time must be sufficient to capture the whole free induction decay signal, otherwise processing can be used to improve the resolution. The spectral width chosen must be sufficient to cover all signals present to prevent aliasing from external signals, which are misleading and difficult to phase correctly during processing (Cavanagh et al. 1996). For proton spectra of random coil peptides, signals are unlikely to fall outside of 0-10 ppm so a spectral width slightly wider than this is sufficient.

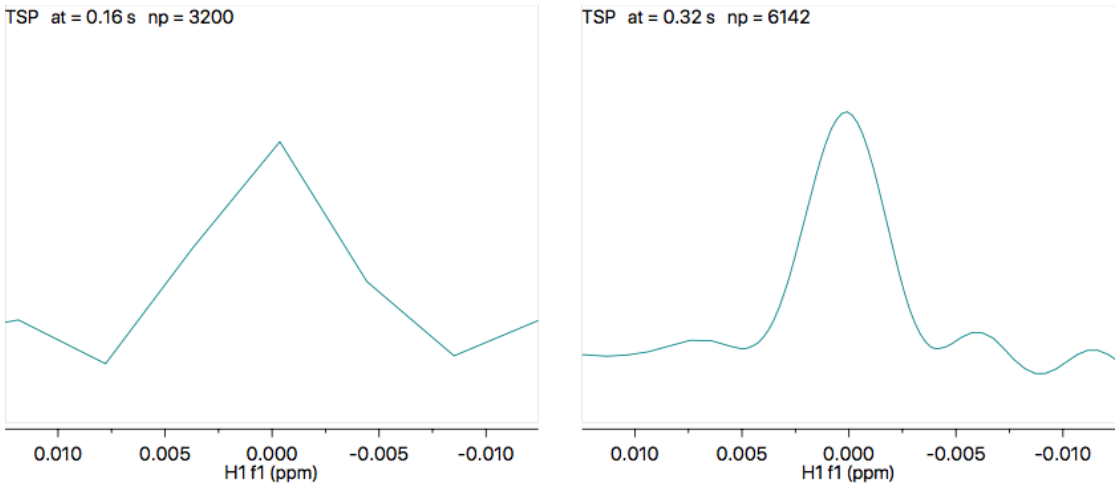


Figure 68 – Improving spectral quality with an increased number of points. Peak represented by a low number of points (left) and a higher number of points (right). The sinc wiggles distorting the base of the peaks (seen to right) is an artefact created by ending acquisition before full signal decay, causing FID truncation. Longer acquisition times can avoid this.

Solvent suppression in solution state NMR is especially critical in sensitive cryoprobes, where the water signal might distort the baseline and obscure small neighbouring signals. Large water signals cause a radiation damping effect, transferring magnetisation to the coil which affects the pulse sequence. Additionally the effects of a very large water peak versus small solute peaks means poor digitisation with the gain value set low to prevent the water signal causing analog-to-digital converter (ADC) receiver overflow (Cavanagh et al. 1996). There are ~110 M protons in water, significantly more than the dissolved

biomolecule of interest, which is likely to be in the μM-mM range. Much of the water signal is removed by solvent deuteration, however any remaining signal can be reduced with a solvent suppression technique included in the pulse sequence. There are a variety of solvent suppression schemes suited to different situations. These include presaturation, where a long R.F. pulse is applied at the water frequency prior to the rest of the pulse sequence, with best results for small, narrow water signals (Hoult 1976) and watergate ('water suppression by gradient-tailored excitation') techniques which can suppress larger signals. The watergate suppression schemes work through dephasing all resonances with gradients before selectively acting on all non-water resonances in a spin echo to produce a spectrum mostly absent of the water peak (Piotto, Saudek, and Sklenář 1992).

5.1.1.2 Processing and analysing reaction series of spectra

The ideal Lorentzian peak to integrate would be well separated from neighbours to incorporate the exponentially widening base fully, thus calculating the integral value from the maximum available area of the peak, but this is not realistic for crowded spectra. In order to cover 99% of the peak, integrals must be taken at 25x the line width of the peak in each direction. This allows the analyst to take in a sufficient amount of the infinitely widening base for quantitation, but is difficult in crowded spectra (Griffiths and Irving 1998). Flat baseline regions without signals either side of the peak also makes the integral bias and slope easily adjusted where necessary. The need for this is significantly reduced if the baseline is very flat, and where required this can be achieved by applying baseline correction.

One of the first points of consideration for processing a spectrum is phasing. Though so routine it could almost be forgotten, good phasing is essential for accurate integrals, and could affect integration significantly if wrong. Phased and symmetrical, Lorentzian-shape lines are particularly essential for deconvolution of a spectrum. Spectral deconvolution is an increasingly powerful method of separating overlapped peaks into constituents (Figure 69). Separation of peaks is

considerably more difficult in the case of a complex peptide proton spectrum where a large number of protons are represented at close chemical shifts, which makes deconvolution necessary.

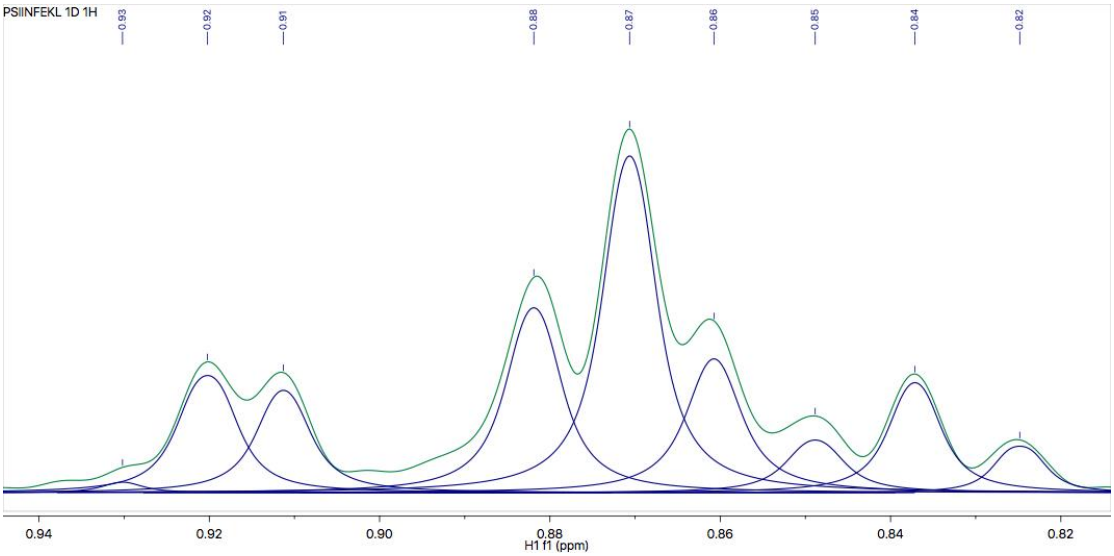


Figure 69 - Global spectral deconvolution applied to region of overlapping peaks with MestReNova 12 software. Spectrum shown in green, deconvoluted peaks in blue.

For deconvolution fitting to work well it requires the expected lineshape, which by default will be Lorentzian although software will often be able to search for Gaussian lines if necessary. For recognition of Lorentz-shapes, no apodisation or window function can be used where it will affect lineshape in a way which prevents effective peak fitting. Exponential line broadening can be used to improve the effective signal-to-noise, though this does not replicate the signal gained by improving the actual signal-to-noise (Cookson and Smith 1982). Because it is exponential and Lorentzian lines, fourier transformed, are exponential, it will not introduce distortions to the peak shape where other apodisation functions might (Becker, Ferretti, and Gambhir 1979). Spectral deconvolution also needs good shimming for the same reason of lineshape. This can be corrected for by reference deconvolution whereby the lineshape of a reference peak is used to correct lineshape across the entire spectrum, but good shimming in the first instance is preferable.

Though often highly effective, deconvolution is limited in particularly complicated regions of spectra. Additionally, it requires more computational power and consequently more time than a simple sum integration of a peak. The MestReNova 12 software offers quantitative global spectral deconvolution (qGSD), which performs additional line fitting to minimise residuals calculated through GSD. It is extremely accurate, nonetheless requiring even more calculation time, and is perhaps excessive for most uses. For the average integration across a reaction series, edited sum integration can be used, which is a sum integration augmented by deconvolution removal of neighbouring peak effects.

As mentioned above in section 5.1.1.1 ('Considerations for NMR quantitation'), the number of points in a spectrum is an important precursor to how the spectrum is digitised. This has a potentially significant impact on how the peak is integrated and accuracy of deconvolution. Beyond increasing the number of points recorded, linear prediction and zero filling can be used in processing to improve spectra, if the acquisition time cannot be increased further.

5.1.1.3 Quantitative NMR and biological reaction monitoring

Quantitative NMR is generally straightforward for pure samples where spectra can be recorded for any required period of time. In reaction progress NMR there is less scope for flexibility. For all parameters in the experiment, where the added dimension of monitoring a reaction adds time constraints to quantitation, a balance must be struck. This is especially true for biological samples and proteins where, in addition to concentrations and temperatures being carefully considered to support the biomolecule being studied, the NMR experiment must be adapted to fit reaction times. A well-designed experiment should incorporate the requirements of the biological sample while maximising possibility of accuracy, for all the points made above.

5.1.2 Designing an NMR experiment by determining T1 values

Since the nuclei must fully return to equilibrium magnetisation after application of the R.F. pulse for maximum signal to be induced in the probe coil, it is crucial to determine T1 values and set appropriate relaxation delay values in the NMR experiment to allow this to happen. Knowing the relaxation time for different species within the reaction is necessary to choose the relaxation delay period of the experiment which balances expediency and accuracy. It also informs which resonances will provide most accurate concentration values from integration.

Inversion recovery experiments can be used to array the relaxation delay time. They consist simply of a 180° pulse followed by the arrayed relaxation time, which determines how closely magnetisation returns towards equilibrium. This is followed by a 90° pulse and acquisition (Wüthrich 1986). It is necessary that the delay time between scans is greater than 3x the estimated maximum T1 for the experiment to give accurate values (Cavanagh et al. 1996) and this must be considered for the resonances of interest for these experiments. A brief investigation into the T1 values of reaction components wil therefore be presented here.

5.1.3 What to look for in an ERAP1 reaction

Because chemical shifts are highly specific to the molecule that provides them, it is possible to identify components in a reaction. This can be used for monitoring the progress of a reaction, with spectra taken at intervals to track the rise and fall of reaction components. shows aminopeptidase activity on EESIINFEKL over time (black traces) from the ring peaks of the phenylalanine residue, where the initial peak at 7.275 ppm for 7PheHδ can be seen to move downfield to the left, to 7.285 ppm as INFEKL is created. This eventually is degraded entirely to phenylalanine, EKL, and the other product amino acids, none of which are seen in this region of the spectra. Aberrant peptide degradation would be seen with the

emergence of unrecognised chemical shift patterns and non-sequential trimming of the peptide.

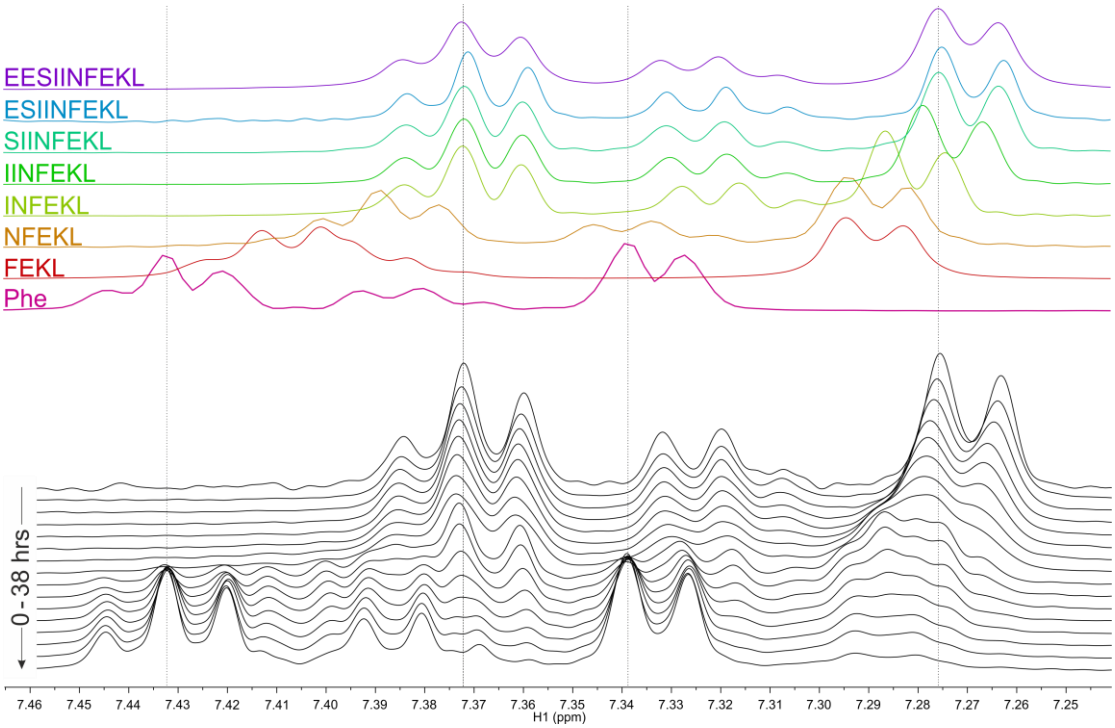


Figure 70 - Following an ERAP1 enzymatic reaction by NMR - the components of the reaction can be seen to change over time (below, black traces), matching the expected chemical shifts of isolated peptide and amino acid (top, colour) for the phenylalanine ring resonances and consistent with aminopeptidase activity

5.2 Methods

5.2.1 ERAP1 expression, purification and experimentation

5.2.1.1 S2 cell expression

Cryo-stored S2 insect cells, previously transfected by Dr Patrick Duriez, were rapidly thawed in a 37 °C water bath and added to 9 ml EX-CELL 420 serum-free media (Sigma Aldrich). This was centrifuged at 200 x g for 3 minutes and the pellet resuspended in 5 ml EX-CELL 420 medium. This was grown in a T25 flask in a 27 °C static incubator (cells did not require CO_2 supply). Trypan blue was used for cell counts. When culture density was at >2 x 10^6 live cells/ml, cells were passaged to 5 x 10^5 live cells/ml and from then on cells were subcultured when culture density was between 6 and 20 x 10^6 live cells/ml, to 1 x 10^6 live cells/ml. After 2-3 days, puromycin was added to 5 μg/ml for selection. Cells were grown statically in T75 flasks when sufficiently expanded, and then into Erlenmeyer conical flasks when at adequate quantities for 4 litre expressions in a shaking incubator. Protein expression was induced with 0.5 mM $CuSO_4$ and the flasks incubated at 27 °C for a week. The cells were centrifuged out of the medium at 3000 rpm for 10 minutes and either discarded or retained for another expression. The supernatant was retained and filtered through a 0.2 μm filter. 5% glycerol was added to supernatants before freezing.

5.2.1.2 Purification

ERAP1 was purified with a 3 kDa C-terminal Twin-Strep-Tag which has high affinity to Strep-Tactin resin, enabling purification directly from culture medium if required (see IBA manual, 2012).

Binding buffer was made like so:

ERAP1 binding buffer:

100 mM Tris-HCl (pH8)
150 mM NaCl
~0.1% w/v Sodium azide

This was added to the expression supernatant, at 10% v/v of supernatant. Further binding buffer was added until the pH was between 7.0-7.5. BioLock Biotin Blocking solution (from IBA Lifesciences) was added at a ratio of 1.5 ml per 500 ml supernatant. The supernatant was incubated at 4 °C for 15-30 minutes. The supernatant was passed through a 0.22 μM vacuum 'Steritop' filter (Millipore) and pH checked again for pH 7-8. Strep-Tactin sepharose resin (IBA Lifesciences) was added at around 300 μl suspension per 50 ml of expression supernatant and incubated at 4 °C for 2 hours on rollers, or overnight for a second purification of the same supernatant. The supernatant was separated from the Strep-Tactin beads with a gravity column. The beads were washed with 5-10 CV of binding buffer. Protein was eluted with 20 ml elution buffer:

ERAP1 elution buffer:

100 mM Tris-HCl (pH8)
150 mM NaCl
5 mM D-desthiobiotin
~0.1% w/v Sodium azide

After elution, the Strep-Tactin was regenerated with 15 column volumes of 1 mM HABA (2-[4'-hydroxy-benzeneazo] benzoic acid) in binding buffer, then cleaned with 30 column volumes of the binding buffer. Eluted protein was run on a SDS-PAGE (see below) to establish purity.

Where purity was considered insufficient, concentrated ERAP1 was further purified by size exclusion chromatrography on a Superdex 75 10/300 GL analytical column (GE Healthcare) in 20 mM potassium phosphate buffer (pH 7). It was then concentrated and buffer exchanged using microcentrifuge spin

columns (10 000 KDa cut-off) into deuterated potassium phosphate buffer (20 mM, pD 7). The ERAP1 was then checked for contaminant with 1D ^{1}H NMR before being flash frozen in aliquots in liquid N_2.

5.2.1.3 SDS-PAGE for analysis of expressed and purified protein

12% SDS polyacrylamide gels were cast in Bio-Rad 0.75 mm mini-PROTEAN plates. A buffer solution for the gel mix was prepared:

Gel buffer:

SDS (0.3%)
Tris (3M), pH 8.45

The running gel was prepared to the following recipe:

12% SDS-PAGE (running):

3.33 ml 30% bis acrylamide (37.5:1)
3.33 ml dH2O
3.33 ml gel buffer
10 µl TEMED

This was mixed before 50 µl ammonium persulphate (100 mg/ml stock) was added. After rapid and thorough mixing, the mixture was distributed between 3 sets of casting plates. Isopropanol was added on top to a few millimetres and the gels were left for ~30 minutes to set. The isopropanol was removed by blotting with filter paper.

A stacking gel was then made to create wells in the polyacrylamide gels. This was mixed to the following recipe:

12% SDS-PAGE (stacking):

660 µl 30% bis acrylamide (37.5:1)
2.9 ml dH2O
1.24 ml gel buffer
10 µl TEMED

This was mixed before 25 µl ammonium persulphate (100 mg/ml stock) was added. The stacking gel mixture was then rapidly and thoroughly mixed again, before being distributed atop the cast running gels. Well combs were inserted (usually 10 wells) and the gels left to set for another 30 minutes.

To prepare samples for running on a gel the samples were mixed with a 4x sample buffer. This was prepared thus:

4x Sample buffer:

200 mM Tris-HCl (pH 6.8)
8% w/v SDS
0.8% v/v bromophenol blue
40% v/v glycerol
400 mM β-mercaptoethanol
ddH_2O to 10 ml

Samples were mixed at a 1:3 ratio with 4x sample buffer and heated at 95 °C in a heat block for 10 minutes. For most samples, 10 µl was loaded into a well. 15 µl was loaded for weak samples (after size exclusion chromatography) or 5 µl for a strong sample (after concentration of the sample). 5 µl ladder (Bio-Rad 'dual-color precision plus protein') was added to the first well.

The gels were run in a Bio-Rad mini-PROTEAN tank according to the manufacturer's instructions. The anode buffer was 200 mM Tris (pH 8.9). The cathode buffer was made to the following recipe:

Cathode buffer:

100 mM Tris (pH 8.2)
100 mM Tricine
0.1% SDS

5.2.1.4 ERAP1 activity verification with L-AMC

Activity was verified at 37 °C using L-Leucine 7-amido-4-methylcoumarin to test 380/460 nm absorbance and fluorescence at minute intervals for 30 minutes.

Deactivated apo-form ERAP1 was prepared by 48 hour buffer exchange using 3000 KDa cutoff membrane at 4 °C against 20 mM potassium phosphate, pH 7, containing 10 mM EDTA and tested for activity at 24 and 48 hours.

5.2.2 NMR strategy

5.2.2.1 Sample preparation

A deuterated peptide buffer was prepared like so:

Deuterated potassium phosphate peptide buffer:

approx. 7.6 mM dibasic potassium phosphate (1M stock)
approx. 12.4 mM monobasic potassium phosphate (1M stock)
mixed to a total concentration of 20 mM potassium phosphate to give a pH of 6.6.

The solution was dehydrated in a vacuum centrifuge and resuspended in the same volume of deuterium oxide, before being dried and reuspended again in the same manner resulting in a buffer at pD 7. Peptides (purchased from ChinaPeptides) were dissolved in deuterated peptide buffer at 1 mg/ml to pD 7. Peptides with a lower pD were buffer exchanged into the deuterated peptide buffer using 100-500 Da cut-off cellulose membranes. These peptide stock solutions were stored at -20 °C

ERAP1 reactions for NMR monitoring were mixed to the following recipe:

ERAP1 reaction monitoring NMR samples:

200 μM peptide
20 μM TSP
0.04% sodium azide (w/v)
2 μM ERAP1
Made up to 180 μl with deuterated potassium phosphate buffer (pD 7)

For these reaction mixes, a peptide, TSP and sodium azide mixture was prepared as given above with a suitable volume of deuterated peptide buffer to make the total volume of 180 μl, minus the volume of ERAP1 which was to be added. This was centrifuged in an Eppendorf. ERAP1 aliquots were thawed and centrifuged alongside the peptide/TSP/azide/buffer mixture. When the spectrometer was ready to accept the sample, the peptide mix was added to the volume of ERAP1 to bring the total volume to 180 μl. A 3 mm NMR tube was used to carry the reaction mix within the spectrometer.

5.2.2.2 Reaction progress monitoring NMR experiments

1D presaturation spectra were collected with a relaxation delay of 5 seconds, 2 steady scans and an acquisition time of 0.32 seconds. The number of scans was varied for different experiment lengths and signal-to-noise, but increased by powers of 2 for straightforward comparison of spectra later. This was carried out at 25 °C on a 600 MHz Varian Inova solution state cryoprobe spectrometer.

5.2.2.3 Processing and analysing reaction series of spectra

Spectra were processed and analysed with the MNova software, for phasing, linear prediction, and an exponential line broadening function of 1. Baseline correction performed with Bernstein polynomial correction, or 'multipoint'

baseline correction with signal-less regions was used to define baseline in MNova.

5.2.2.4 Saturation transfer difference NMR

Samples were prepared as per reaction monitoring NMR samples (see above) and the spectra obtained with the standard Varian BioPack phase-subtracted saturation transfer difference NMR experiment. A final saturation frequency of 3500 Hz was used.

5.2.2.5 Modelling peptide binding from NMR data

The structure of PPSIINFEKL peptide was built into the published ERAP1 crystal structure 2YD0 (Kochan et al. 2011) in UCSF Chimera (Pettersen et al. 2004). The peptide N-terminal was positioned where the bestatin inhibitor, present in the crystal structure, bound to the active site with the 1ProN in position to interact with the catalytic zinc. Dihedral angles and side chain rotamers throughout the rest of the peptide were adjusted manually until the peptide was in proximity to the hydrophobic part of the cavity wall which agreed with the saturation transfer difference data. The Chimera Dock Prep tool was used to add hydrogens to the peptide. 1ProN, 5IleCβ and 10LeuCγ were selected as fixed atoms for an energy minimisation of the peptide and minimisation (using the Minimize Structure tool) carried out with default number of steps and gradient ((Pettersen et al. 2004).

5.3 Results

5.3.1 Establishing T1 values by inversion recovery

The T1 values established by inversion recovery for amino acids (Table 9) gave the longest T1 value of 7.2 seconds and the shortest at 0.6 s. A relaxation delay of 5 seconds was chosen despite two protons (ProHα and SerHα) having longer T1 values, due to the need for both a large number of scans and short enough experiment length to capture events at the start of an ERAP1 reaction. A relaxation delay of 21.6 seconds, 3x the T1 of ProHα, would extend the required time for experiments substantially with little benefit if that resonance is not used during later analysis. Only 15 of the 36 amino acid resonances have a calculated T1 of less than 1.67 s, which would make the relaxation delay >3x T1. The proportion of spins which are not relaxed to $M_{z,eq}$ as a result must be a significant caveat for quantitation.

Table 9 - T1 values for amino acid resonances, collected at 25 ℃, 100 μM, in deuterated potassium phosphate

ID	T1 (s)	error		ID	T1 (s)	error
Pro Ha	7.231	0.5263		**Glu Hb***	1.822	0.4597
Ser Ha	5.244	0.7117		**Ser Hb1**	1.746	0.1317
Glu Ha	4.393	0.923		**Ile Hd***	1.557	0.01427
Phe Hz	3.848	0.6663		**Ile Hb**	1.469	0.2071
Asn Ha	3.813	1.145		**Lys He***	1.464	0.01683
Pro Hd2	3.810	0.1234		**Leu Hb*/Hg**	1.461	0.04192
Pro Hg*	3.689	0.08554		**Glu Hg***	1.357	0.2602
Pro Hd1	3.684	0.125		**Leu He***	1.151	0.01203
Phe Ha	3.602	0.5246		**Lys Hd***	1.101	0.01174
Pro Hb1	3.492	0.2147		**Ile Hg2***	1.043	0.02863
Phe He	3.295	0.2906		**Lys Hg1**	0.9761	0.02744
Ile Ha	3.219	0.06101		**Lys Hg2**	0.9685	0.03235
Leu Ha	3.195	0.1887		**Ile Hg1b**	0.9489	0.1399
Pro Hb2	3.185	0.1611		**Ile Hg1a**	0.9198	0.1466
Phe Hd	3.090	0.1277		**Lys Hb***	0.8524	0.008302
Lys Ha	2.904	0.04181		**Phe Hb2**	0.6453	0.09533
Ser Hb2	2.256	0.2472		**Phe Hb1**	0.596	0.1281
Asn Hb1	2.104	0.2124				
Asn Hb2	2.027	0.5591		**TSP (reference)**	3.505	0.2536

5.3.2 Comparison of aminopeptidase trimming and natural degradation

5.3.2.1 NMR monitoring of peptide degradation in the absence of ERAP1

To establish the markers of degradation expected in the absence of ERAP1, which could, therefore, not be attributed to ERAP1, the isolated peptides were monitored by regular collection of NMR spectra over a period of time. PSIINFEKL, at 200 μM and incubated at 25 °C for 30 hours of experiments (longer than the time required to watch the progress of an ERAP1 reaction) showed little observable change (Figure 72). The predominant change was universal across a spectrum: reduction in lineshape quality as a result of sample shimming worsening over time. Loss in peak size at the methyl region of the spectrum (1-0.65 ppm) suggested a reduction in peptide concentration in the sample to around 94%. It is unclear to what extent this occurs due to peptide being degraded or whether this is partially due to the limitations of the experiment and worsening lineshape, as this percentage seems to descend improbably in the final 28.5 hour spectrum (Figure 71), but poor spectrometer shimming can impair the accuracy of quantitation.

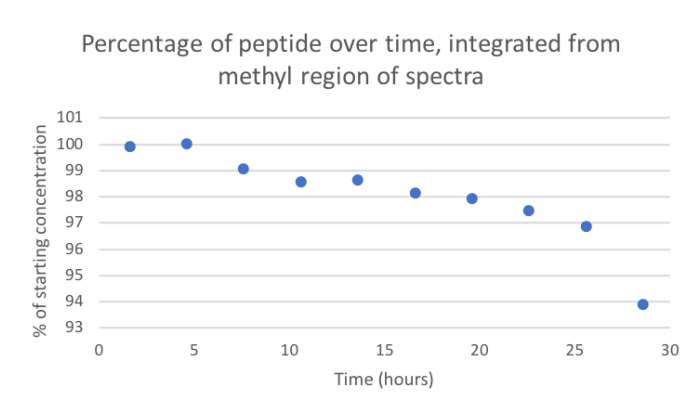


Figure 71 -Measured decrease in peptide concentration over time for a PSIINFEKL control in the absence of ERAP1

No formation of amino acids was observed in this sample during the 30 hours of observation, including the isoleucine Hγ2 peak (a good reporter peak due to the number of protons present) at 1.01 ppm (Figure 72B) or a neighbouring leucine Hδ* peak at 0.97 ppm. Neither were any proline resonances resulting from removal of the N-terminal proline of PSIINFEKL observed.

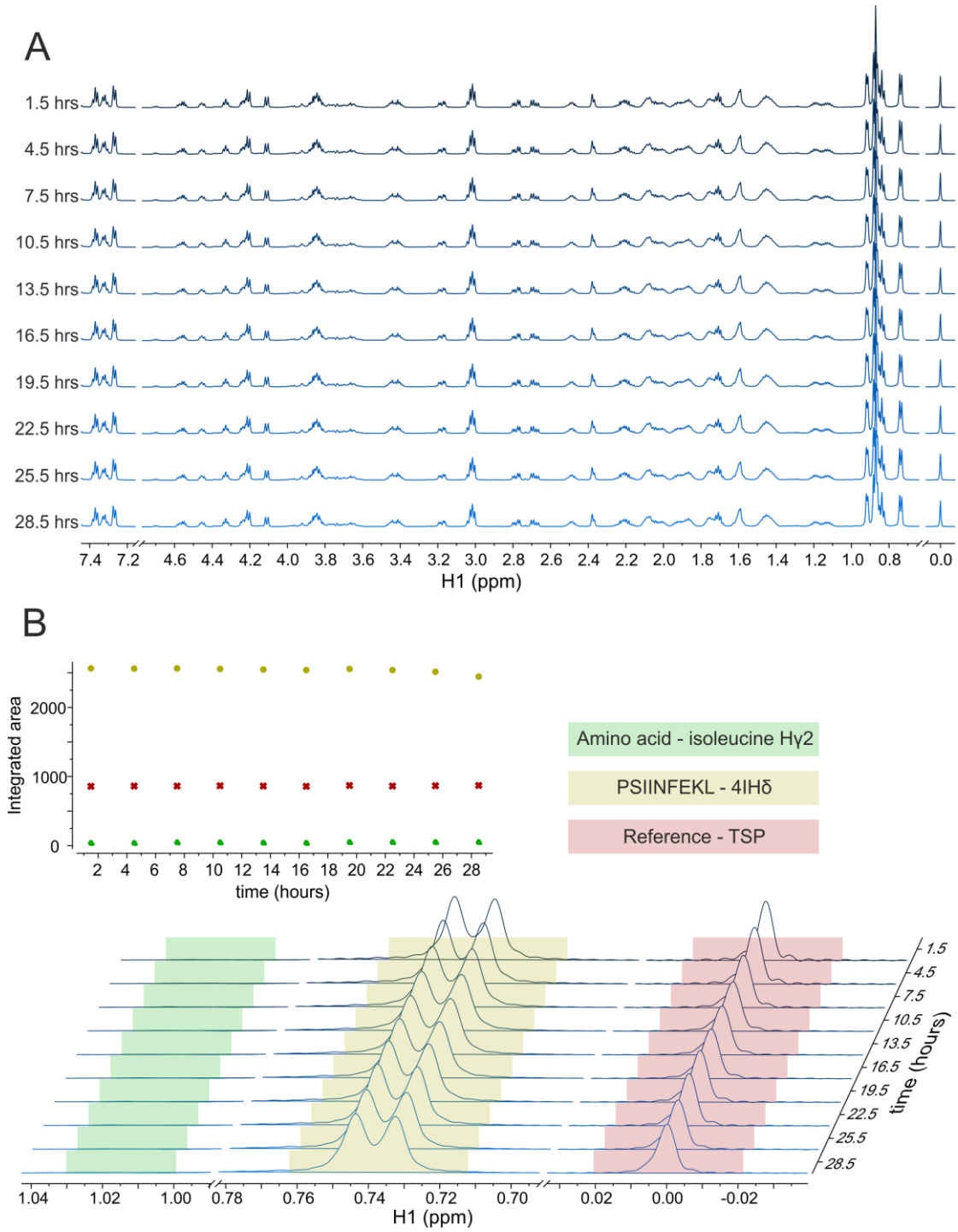


Figure 72 – 200 uM PSIINFEKL (in the absence of ERAP1) with spectra taken over 30 hours at 25 ℃ (times shown are spectral midpoints). A) Stack-displayed timepoint spectra from experiment. B) Integration of peptide/amino acid regions shows little change to peak areas over this time

A much longer monitoring of PPSIINFEKL over weeks was also performed to see the same degradation during a more extensive period of time. Incubation of PPSIINFEKL for this period (24 days) produced noticeable effects on the chemical shifts and peak size changes which were obvious even from casual observation of the set of spectra.

Changes consistent with the N-terminal amino acid being removed from the peptide would show chemical shifts corresponding to PSIINFEKL appearing, and those of PPSIINFEKL disappearing. This is not in evidence from these spectra, as shown in Figure 75. Resonances which should show little effect, at 2.8 and 2.7 ppm, are instead decreasing or changing shape. PPSIINFEKL peaks which would be removed if PPSIINFEKL was converted to PSIINFEKL include 3.74 and 2.58 ppm, both of which remain present in the 24 day spectrum. In addition at 2.5 ppm a peak for PSIINFEKL would be visible if the N-terminal was cleaved, but this is not the case. Also displayed in Figure 75, resonances for the isolated L-proline at 3.35 and 2.25 ppm have not been created. There is no evidence of N-terminus removal in the sample, and instead the chemical shifts present would indicate instead that PPSIINFEKL mostly remains with an intact backbone, but might be undergoing chemical changes to the side chains. This matches personal observations that a decayed peptide (independent of any enzyme being added) has similar to expected chemical shifts, but displays unusual features such as two phenylalanine ring peaks rather than three (see Figure 73). Despite this, there may be some amino acid formation through decay of PPSIINFEKL, as the chemical shift of isoleucine Hγ2 (at 1.01 ppm) looks to be producing a small but measurable peak across the 24 days (Figure 74B).

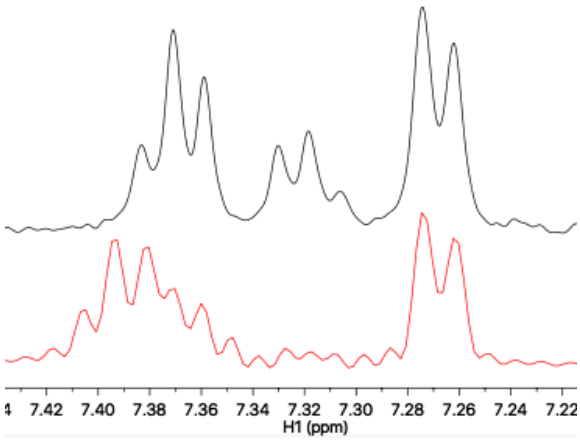


Figure 73 - 6Phe side chain ring peaks of PSIINFEKL (top) and degraded PSIINFEKL (bottom)

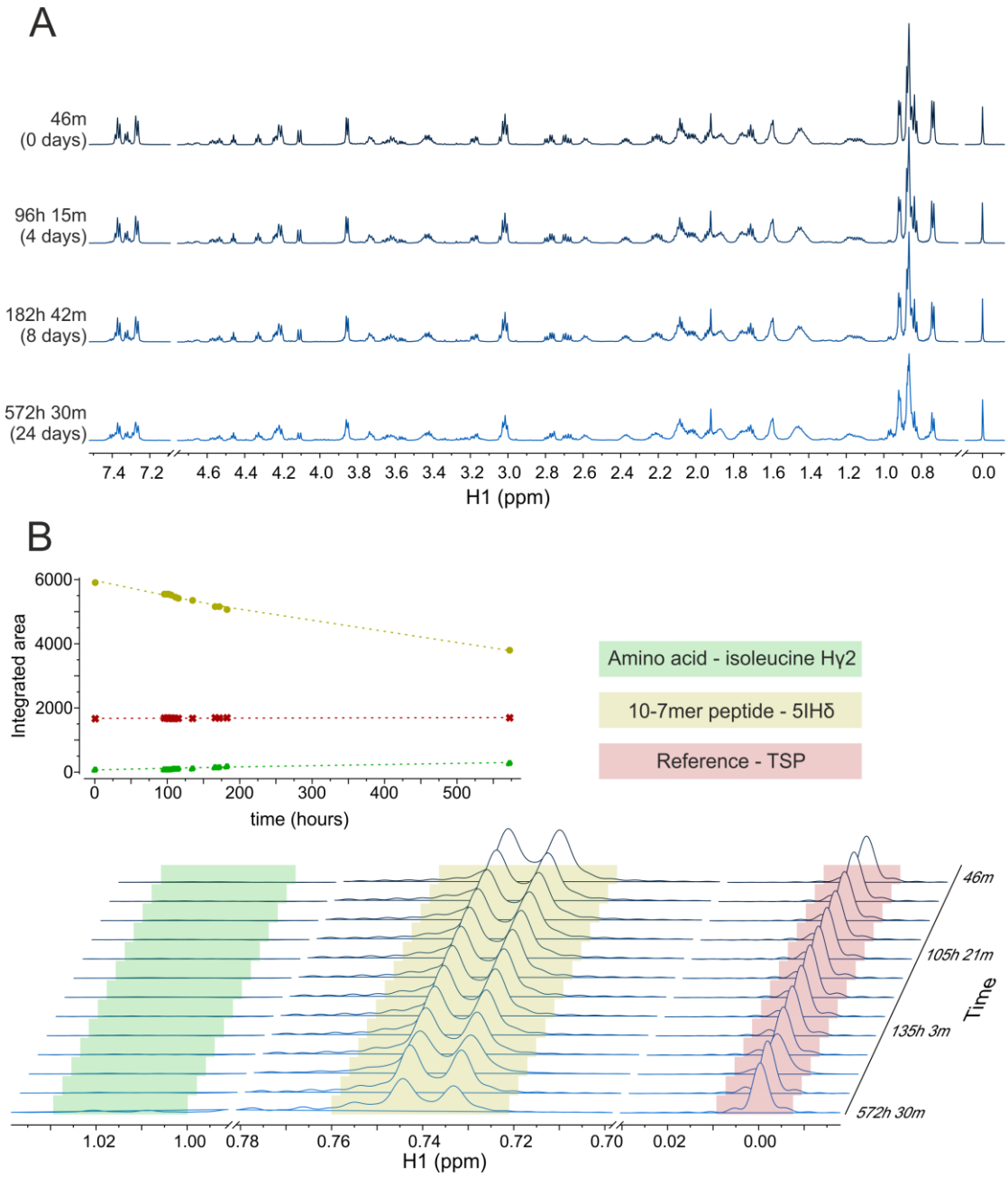


Figure 74 - 200 uM PPSIINFEKL (in the absence of ERAP1) with spectra taken over 24 days at 25 ℃. A) Stack-displayed timepoint spectra from experiment. B) Integration of peptide/amino acid regions shows changes to peak areas over time.

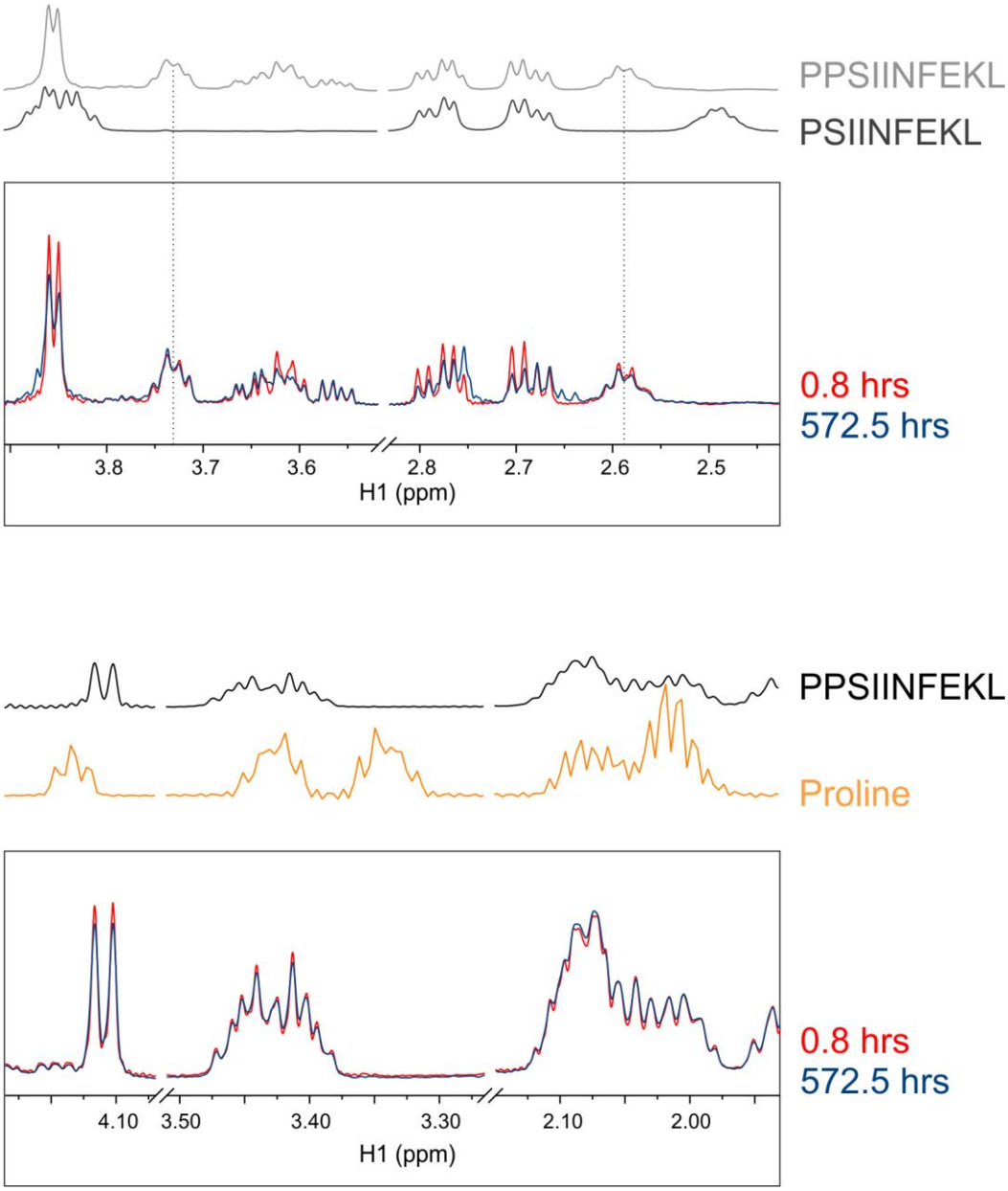


Figure 75 – Sections of spectra of 200 μM PPSIINFEKL-only sample (shown in boxes) at different timepoints, with comparison spectra for isolated peptides/amino acids shown above. Substrate peptide (PPSIINFEKL) peaks do not substantially reduce in size over the period, and likewise product peaks (for PSIINFEKL/proline) do not appear.

5.3.2.2 NMR reaction monitoring of peptide degradation in the presence of ERAP1

As analysis of spectral peaks for 200 μM PSIINFEKL with 2 μM wild-type ERAP1 present over nearly 22 hours, a similar length of time to the 30 hours of isolated PSIINFEKL experiment which yielded no remarkable change to the spectrum (Figure 72), showed substantial changes within the peptide sample. Peaks

corresponding to PSIINFEKL (and product peptides) decreased in size over time and peaks at amino acid chemical shifts were correspondingly created. This is clear both from integrated values and visible within the spectrum in Figure 76, where the peak in yellow representing PSIINFEKL 4IleHδ, SIINFEKL 3IleHδ and IINFEKL 2IleHδ are decreasing as the isoleucine Hγ2 peak appears at 1.01 ppm (green). Other amino acids can also be identified in the spectra. The examples given in Figure 77 show proline appearing at 4.13 ppm as the peptide peaks decrease, serine, phenylalanine and asparagine appearing between 3.95 and 4.03 ppm where they overlap, but can still be identified and, if needed, deconvoluted, and for the ring proton peaks of phenylalanine can be seen emerging downfield of the similar peptide resonances, at 7.43 ppm. PSIINFEKL resonances conversely disappear from the reaction (although not entirely) and this can be seen at 4.11 ppm, 4.46 ppm, around 3.85 ppm, and 7.27, 7.32 and 7.37 ppm. SIINFEKL peaks may be observed at 3.90 ppm, suggesting this is being generated.

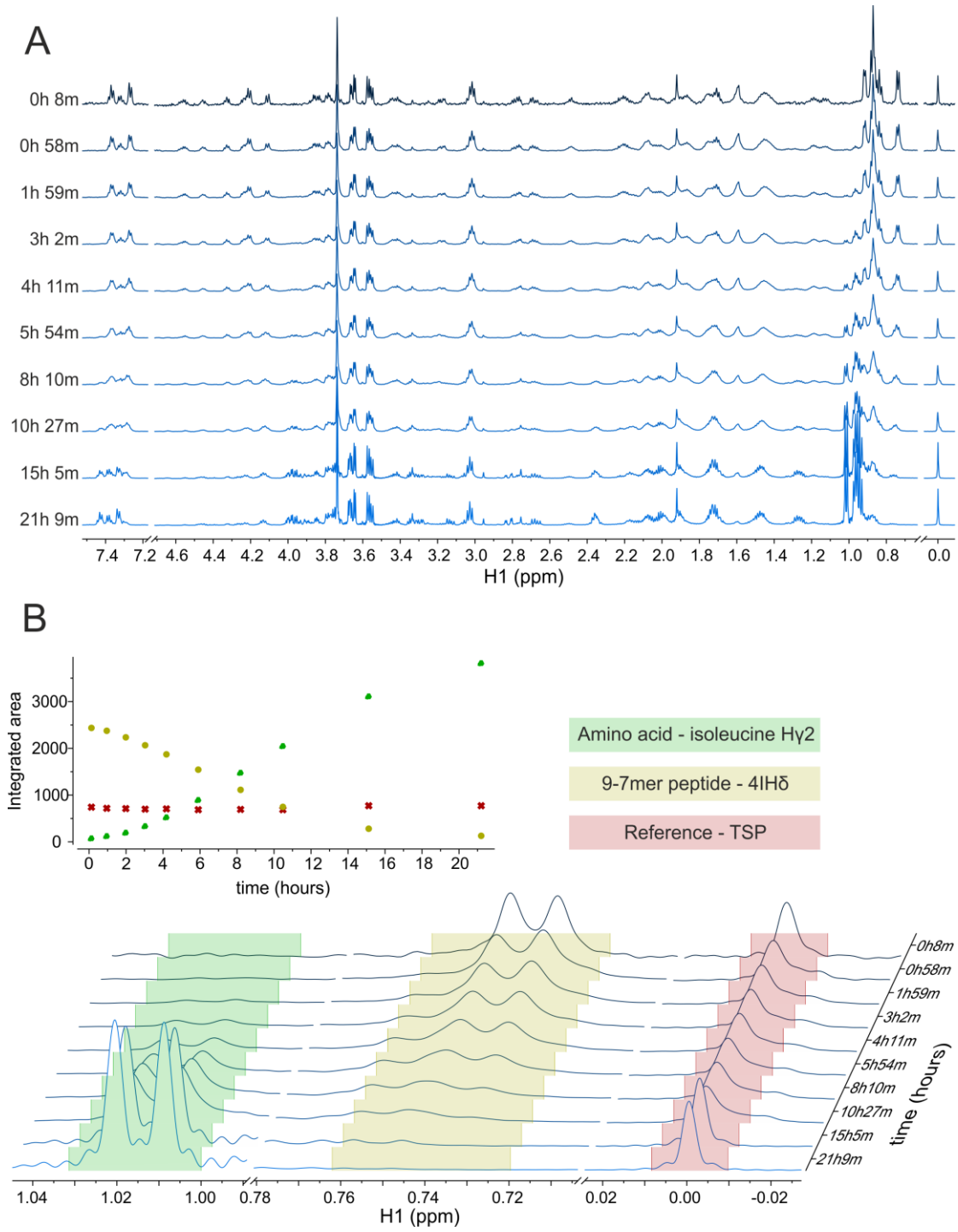


Figure 76 - Degradation of 200 μM PSIINFEKL in the presence of 2 μM wt ERAP1. A) Stack-displayed timepoint spectra from experiment. B) Integration of peptide/amino acid regions shows obvious change to substrate/product peak areas over this time

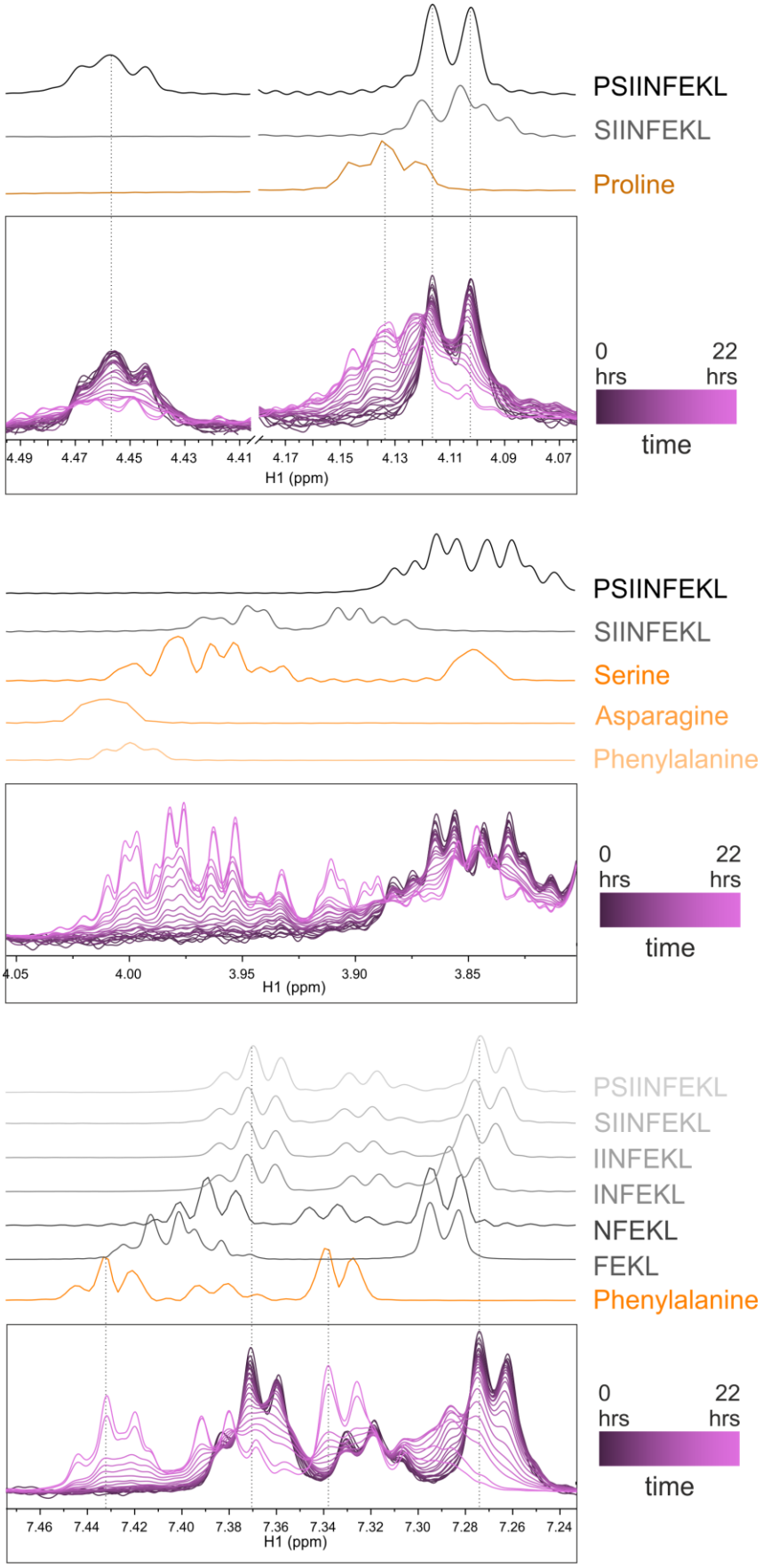


Figure 77 - Sections of PSIINFEKL + wt ERAP1 reaction monitoring spectra (shown in boxes) with comparison spectra for isolated peptides/amino acids shown above.

5.3.2.3 NMR reaction monitoring of peptide degradation in the presence of catalytically impaired ERAP1

Aminopeptidase activity would be expected in impaired or entirely depleted form from an enzyme with significant modification to the active site. The E320A mutant is considered non-functional due to the mutation within the active site GAMEN motif (Reeves et al. 2014). Similarly the zinc ion at the heart of the active site, a feature of ERAP1 as with all other zinc metallopeptidases in the oxytocinase M1 subfamily and most M1 proteins, is critical for activity (Hattori et al. 1999; Serwold et al. 2002; Saric et al. 2002; Rawlings and Barrett 1995).

Deactivation of ERAP1 by buffer exchange against 10 mM EDTA was demonstrated by testing activity on L-AMC against active ERAP1, in which EDTA-incubated ERAP1 showed abrogation of activity (Figure 78).

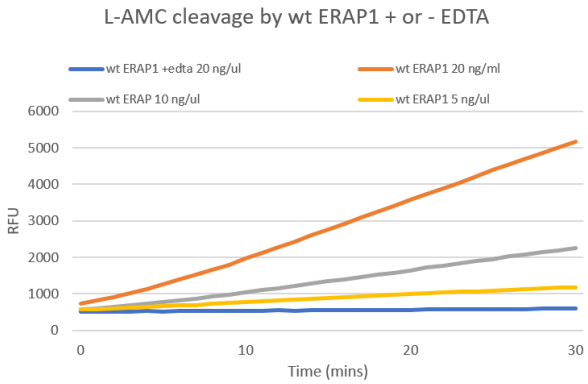


Figure 78 - L-AMC cleavage by ERAP1 (in relative fluorescent units) over time, for 20 ng/ul EDTA-incubated ERAP1 (blue) and active ERAP1 at different concentrations

200 μM PSIINFEKL with 2 μM E320A showed more rapid peptide degradation than PSIINFEKL alone, but of a substantially different sort to the corresponding wt ERAP1 experiment (described above). Comparison of the peak areas for L-isoleucine Hγ2 (1.01 ppm) and the peak indicating all 7-9-mer peptides at 0.77-0.72 ppm shows free isoleucine being created at low concentrations, while removal of peptide is more rapid (Figure 79). Despite some obfuscatory contaminant peaks between 3.7 and 4 ppm, it was largely possible to identify which species were present in the reaction. Peaks which could isolate SIINFEKL, IINFEKL, INFEKL or shorter peptides down to a 3-mer were not visible, indicating none of these were present. PSIINFEKL peaks declined and appeared to be removed throughout the reaction, but this decline was replicated for most

ambiguous peptide peaks with chemical shifts shared between multiple peptide species, when, in an aminopeptidase reaction, these would be expected to stay constant until even the smallest and lower affinity peptides had been removed. Resonances representing isolated amino acids were inconsistent with aminopeptidase activity, with leucine being created along with isoleucine. Signals consistent with lysine generation appeared as well as for proline, but serine, phenylalanine, glutamic acid and asparagine were not visible from the stacked spectra.

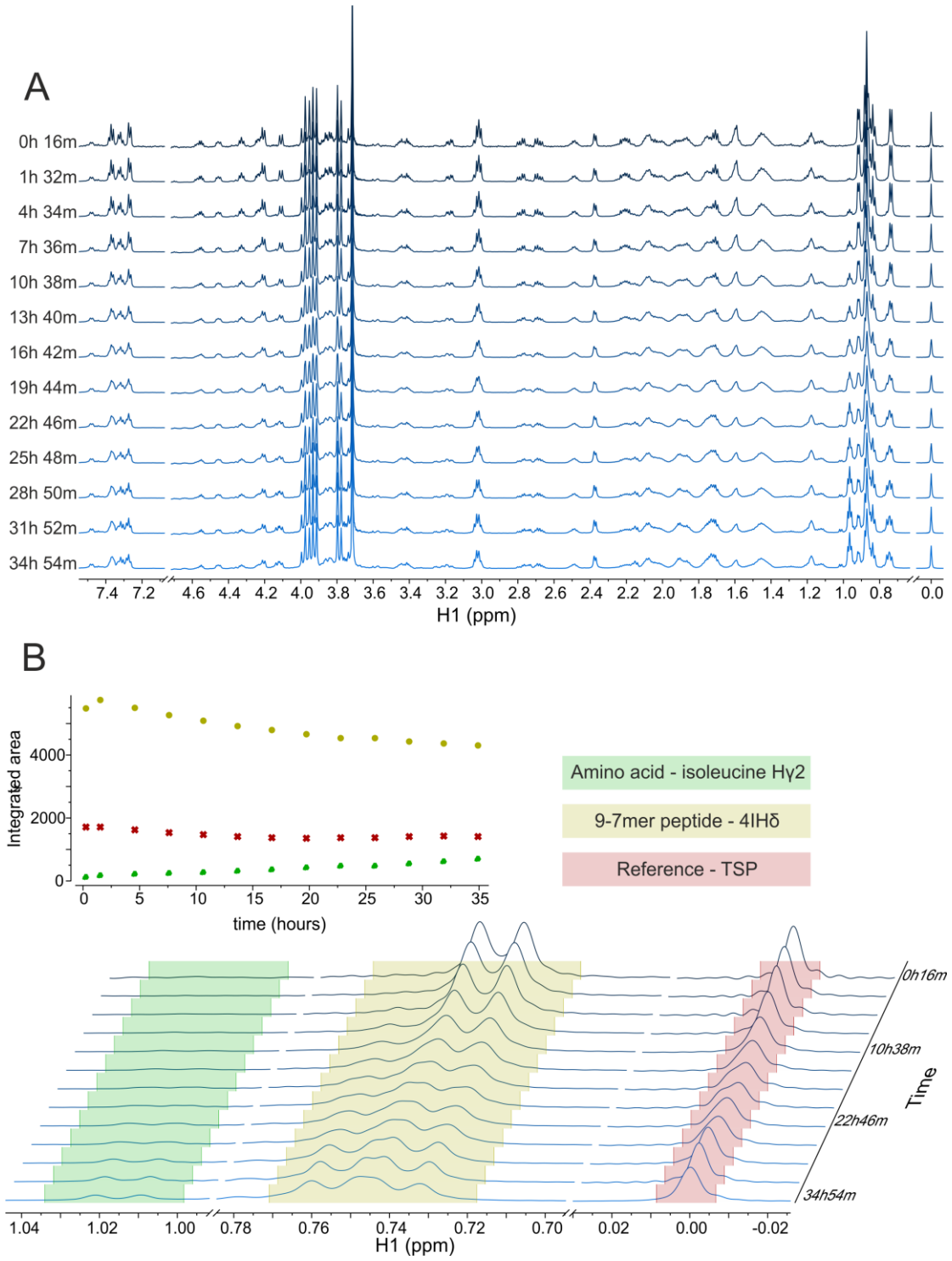


Figure 79 - NMR reaction monitoring of 200 μM PSIINFEKL with 2 μM E320A ERAP1. A) Stack-displayed timepoint spectra from experiment. B) Integration of peptide/amino acid regions shows some change to peak areas over time

The EDTA-incubated wt ERAP1 (-Zn(II)) experiment was performed with a lower (1 μM) concentration of ERAP1 due to significant losses during buffer exchange against 10 uM EDTA and a subsequent necessary buffer exchange to remove EDTA prior to NMR. It was incubated with 200 μM EESIINFEKL for the reaction. Assessment of the spectra suggested similarity overall with the E320A experiment, though not in specifics. EESIINFEKL peaks were depleted (at 4.5-4.4 ppm and at 2.3 ppm) and appeared to be replaced by ESIINFEKL (at 3.85 ppm) (Figure 80). For SIINFEKL it was inconclusive, with chemical shifts appearing close to, but not at, the positions expected between 3.97 and 3.87 ppm (Figure 82) but not decreasing again afterwards as it was used as a substrate. Very small peaks matching IINFEKL chemical shifts also appeared (this can be seen at 4.15 ppm, Figure 80). While peaks consistent with glutamic acid could be seen after 61 hours (Figure 82, 2.1-2.04 ppm) there were no clear peaks for isoleucine, phenylalanine or asparagine, only very small peaks which might result from serine, and nothing apparent for lysine or leucine.

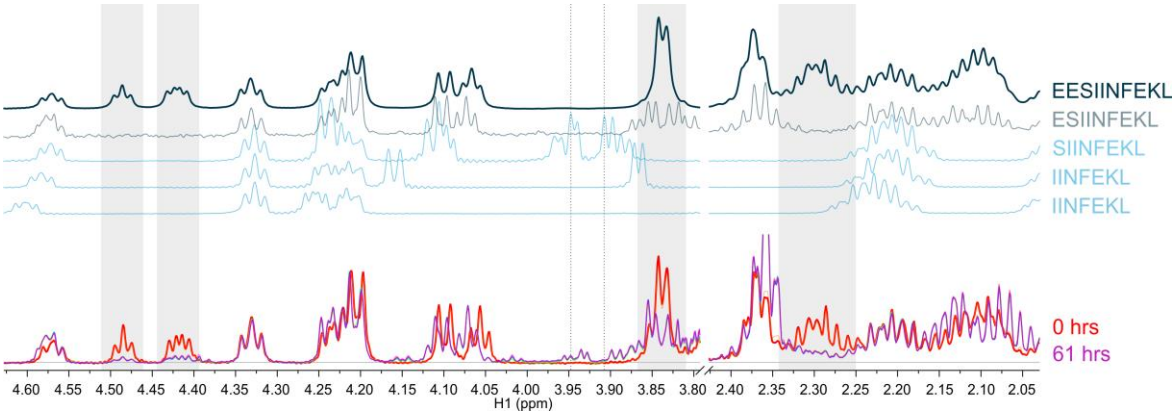


Figure 80 – Reaction monitoring NMR spectra of EESIINFEKL peptide with EDTA-incubated wt ERAP1 (bottom spectra, overlaid) with pure isolated peptide spectra (above) showing removal of EESIINFEKL from reaction (indicated by grey boxes)

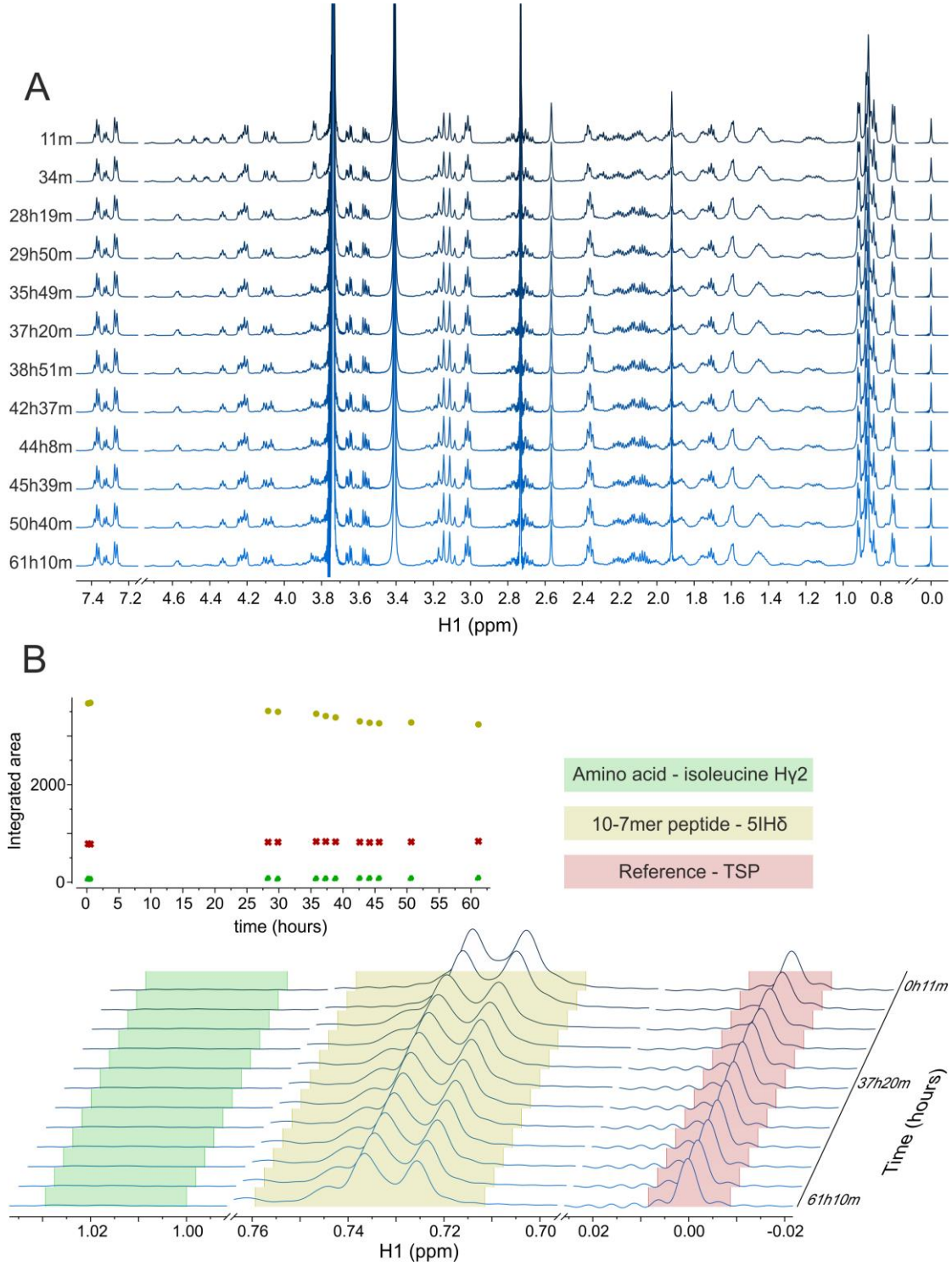


Figure 81 - NMR reaction monitoring of 200 µM EESIINFEKL with 1 µM EDTA-deactivated ERAP1. A) Stack-displayed timepoint spectra from experiment. B) Integration of peptide/amino acid regions shows some change to peaks over time

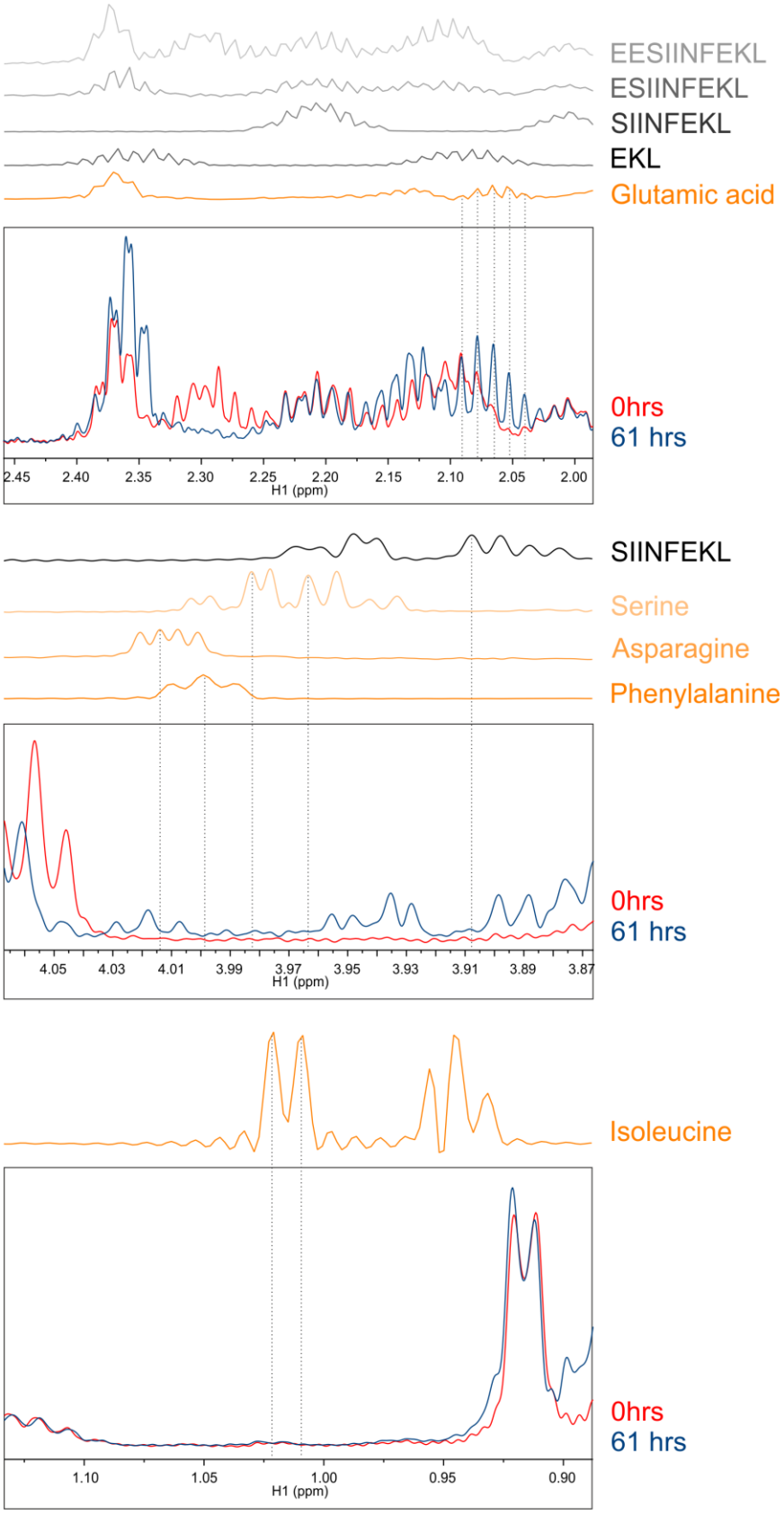


Figure 82 – Reaction monitoring NMR of EESIINFEKL with EDTA-incubated ERAP1 experiment (in boxes) with isolated peptides and amino acid spectra above, showing where amino acid peaks would be expected to appear (dotted lines) in the presence of an active aminopeptidase.

5.3.3 Comparison of variants

The 5SNP variant of ERAP1 (designated *001, see Table 1) possesses 5 single nucleotide polymorphisms. These are M349V (rs2287987), K528R (rs30187), D575N (rs10050860), R725Q (rs17482078) and Q730E (rs27044) which have known associations to disease ((TASC) et al. 2010; D. M. Evans et al. 2011; Kochan et al. 2011). It can be described as a 'hypo-trimmer' ERAP1 variant (Reeves et al. 2014) with ~20% the activity of wild-type ERAP1 (an efficient trimmer, designated *002 – see Table 1) (Reeves et al. 2013). Preferences for different N-terminal residues to wt ERAP1 have also been reported by Reeves et al. (2013) with 5SNP displaying poor activity levels for N-terminal cysteine, aspartic acid, histidine, and serine residues. Though trimming is impaired for all residues, methionine, valine, leucine and glutamine are preferred residues while activity for glutamic acid, alanine, tryptophan, threonine, isoleucine and asparagine appears proportionally reduced across them (Reeves et al. 2013).

The 5SNP and wt ERAP1 used for comparative NMR investigations were tested with the L-AMC assay (Figure 83) and found to be active, with 5SNP having lower activity than wt ERAP1. Reactions with all peptides showed straightforward aminopeptidase activity with successive trimming of newly created N-termini and the chemical shifts for predicted peptide products appearing within the reaction, then subsequently disappearing as they became substrates. The exception to this was PPSIINFEKL, which was degraded slowly and in a non-aminopeptidase fashion. Both the 5SNP and wt ERAP1 PPSIINFEKL reactions showed very little creation of serine (Figure 84) which was present in only negligible concentrations. This was similarly the case for generation of asparagine (Figure 86). Peaks consistent with

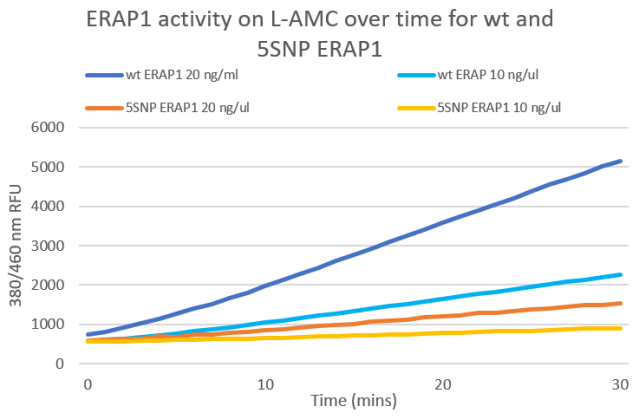


Figure 83 - Comparison of ERAP1 activity on L-AMC over time with wt and 5SNP variants

chemical shifts for phenylalanine and isoleucine were observed, however, and were present in greater concentrations for the wt reaction than for 5SNP. Phenylalanine (Figure 87) was at calculated at 90.9 μM for wt at 20.86 hrs, and at 6.4 μM for 5SNP at 21.54 hrs. Isoleucine (Figure 85) at the same timepoints was calculated at 75.1 μM for wt and 7.6 μM for 5SNP. Several peaks of unknown origin were seen across the spectra, and this, coupled with the way in which amino acids were being generated non-sequentially, suggests endopeptidase activity. For the other five peptides, amino acids were generated in the order of serine, isoleucine, asparagine, phenylalanine (with the exception of wt asparagine reactions, see below) as shown by the lines denoting time at which half maximum concentration being reached (see Figure 84, Figure 85, Figure 86, & Figure 87).

Generation of serine (Figure 84) gave comparatively poor-quality results due to the Hβ peak integrated representing a single proton and being crowded by neighbouring peaks, and the peaks being strongly affected by good or bad shimming. Nonetheless it is clear serine was generated rapidly at the start of the reaction for KSIINFEKL, KKSIINFEKL, ESIINFEKL and EESIINFEKL, and was created more slowly for PSIINFEKL. Though it is difficult to be confident of any difference between wt activity on KSIINFEKL and KKSIINFEKL for wt ERAP1, it is possible that it has reached a half maximum concentration of serine at a later stage in the reaction for 5SNP, at about 1.3 hrs for KSIINFEKL and 2.2 hrs for KKSIINFEKL. Peculiarly, serine was generated more rapidly by 5SNP than wt in both the ESIINFEKL and EESIINFEKL reactions, suggesting 5SNP had a stronger preference for glutamic acid at the peptide N-terminal of 10- and 9-mers than the wt ERAP1 did.

Isoleucine was integrated from a well separated 3 proton Hγ2 methyl peak and therefore after careful baseline correction gave visibly more accurate integrals than serine (Figure 85). It was produced more rapidly by wt ERAP1 than by 5SNP in all reactions. For KSIINFEKL the trimming rate was initially almost identical between wt and 5SNP, slowing for 5SNP after a concentration of 150 μM was reached, suggesting 5SNP was slower to remove the subsequent N-

terminal isoleucine of INFEKL than of IINFEKL. Isoleucine reaches a half-maximum concentration earlier with KKSIINFEKL than KSIINFEKL for wt ERAP1, but slower for KKSIINFEKL than KSIINFEKL for 5SNP, as with serine generation. This was also the case for phenylalanine generation (Figure 87), and asparagine generation by 5SNP was slower for KKSIINFEKL than KSIINFEKL (Figure 86).

While phenylalanine generation was in line with what was expected, asparagine trimming behaviour was aberrant in the wild-type reactions. In the KSIINFEKL reaction it reached a maximum concentration of approximately 50 μM before apparently dropping. It is unclear why this should be. Though asparagine chemical shifts were present, there were also unexplained chemical shifts positioned nearby the 2.86 ppm resonance. These were not present for the 5SNP experiments and the similar splitting pattern suggested they were also asparagine resonances, but not from the isolated, unadulterated amino acid form.

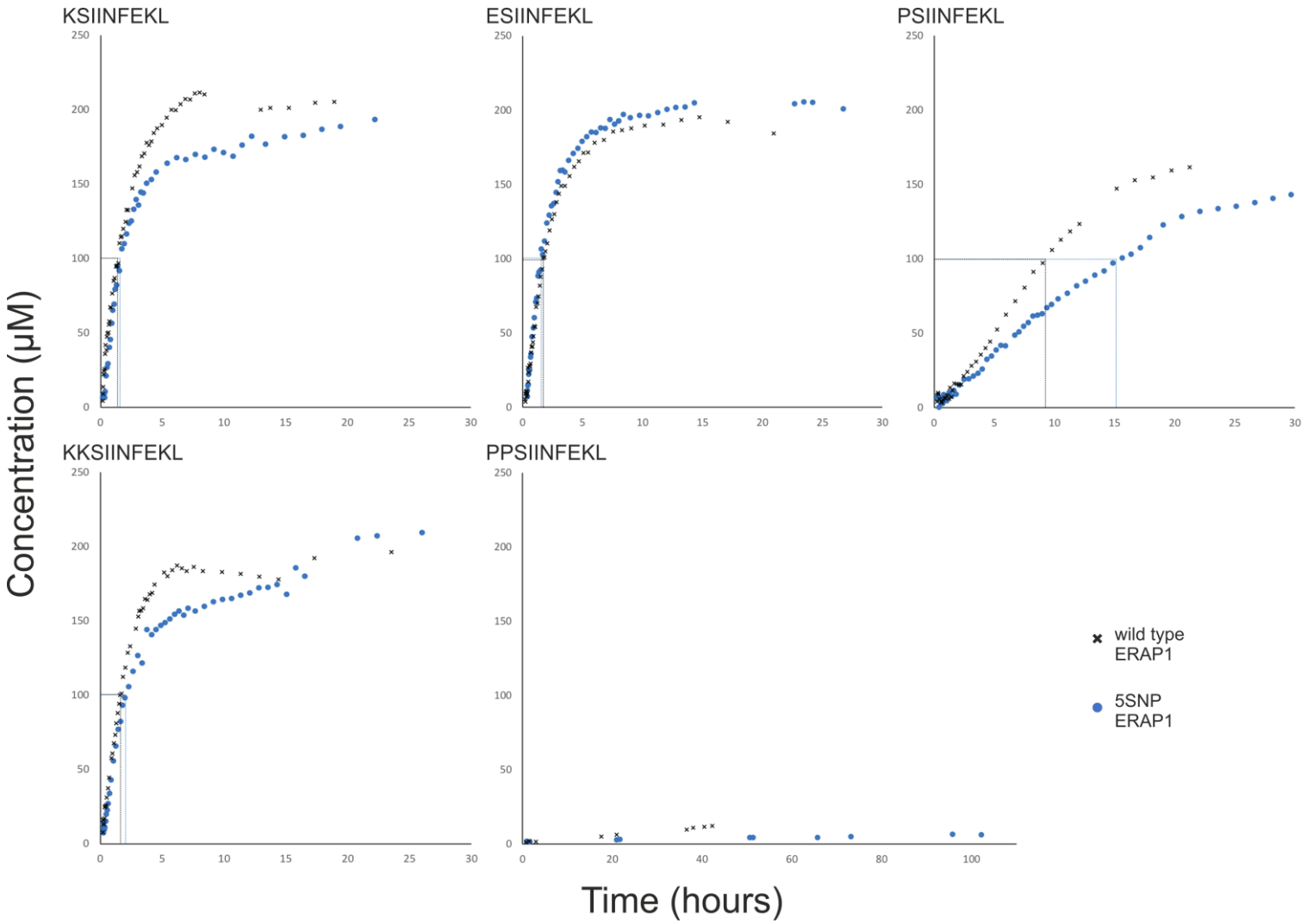


Figure 84 - Concentration of L-serine present in reaction over time for wt and 5SNP ERAP1 reactions of different peptides

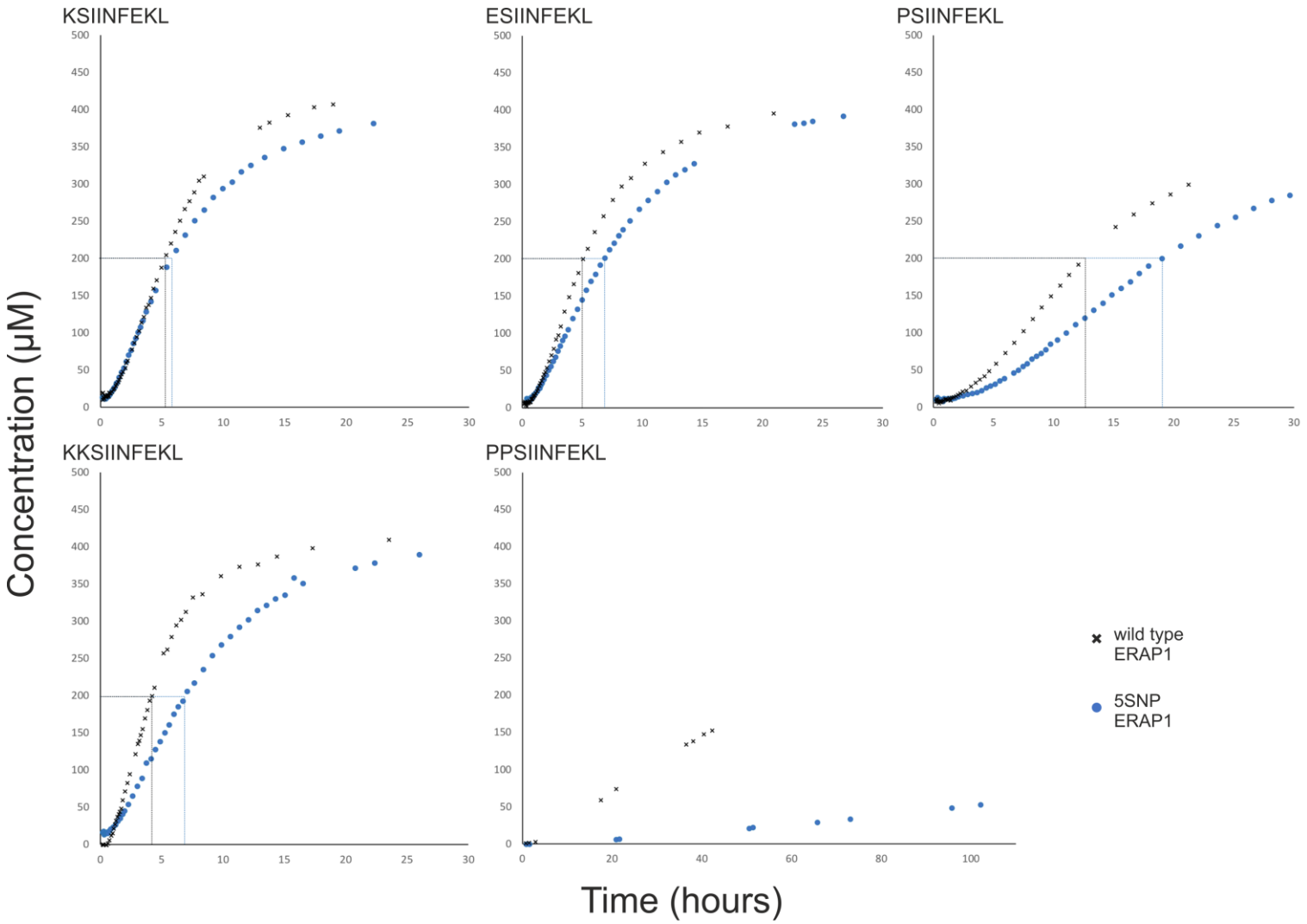


Figure 85 – Concentration of L-isoleucine present in reaction over time, for wt and 5SNP ERAP1 reactions of different peptides

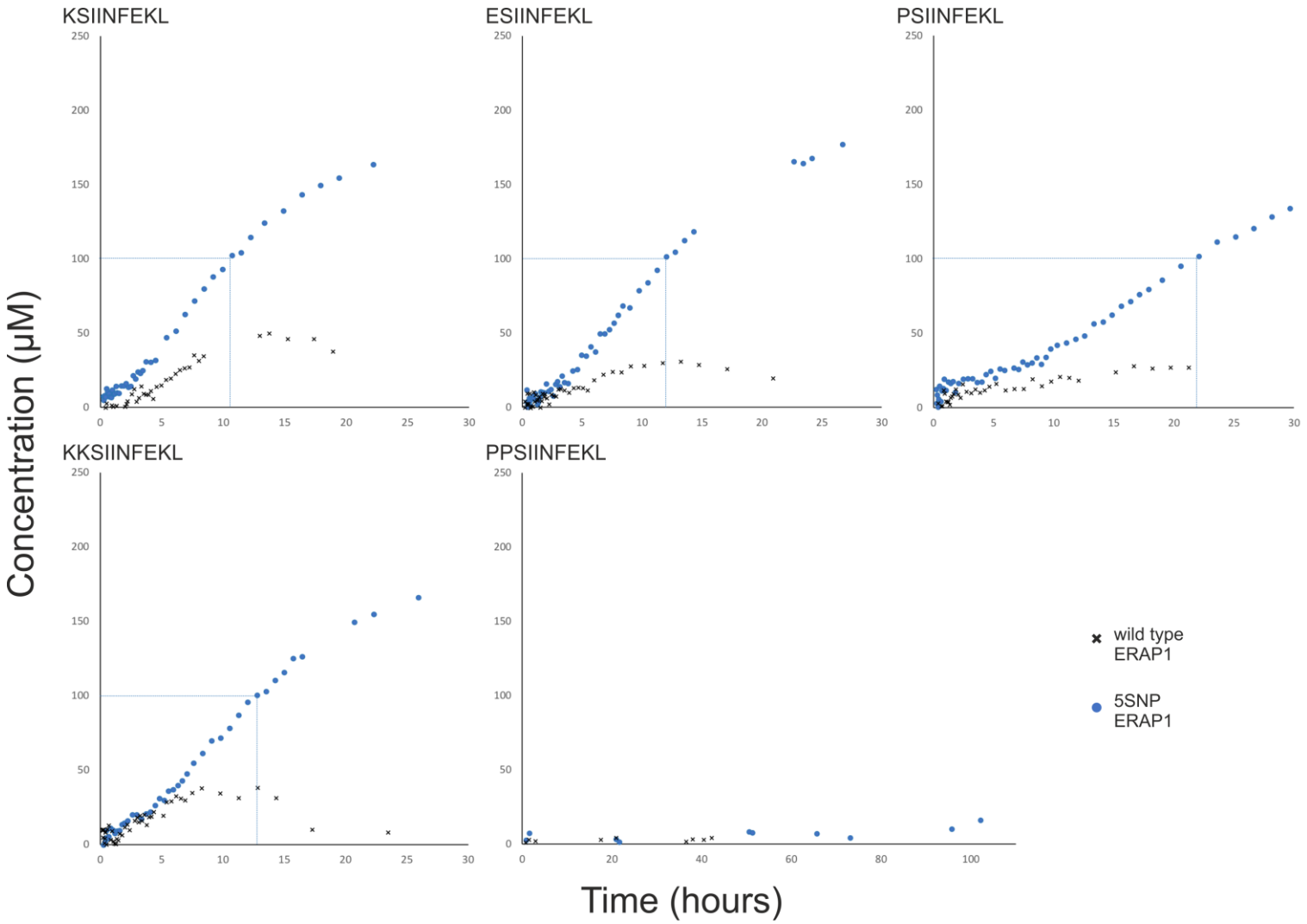


Figure 86 - Concentration of L-asparagine present in reaction over time, for wt and 5SNP ERAP1 reactions of different peptides

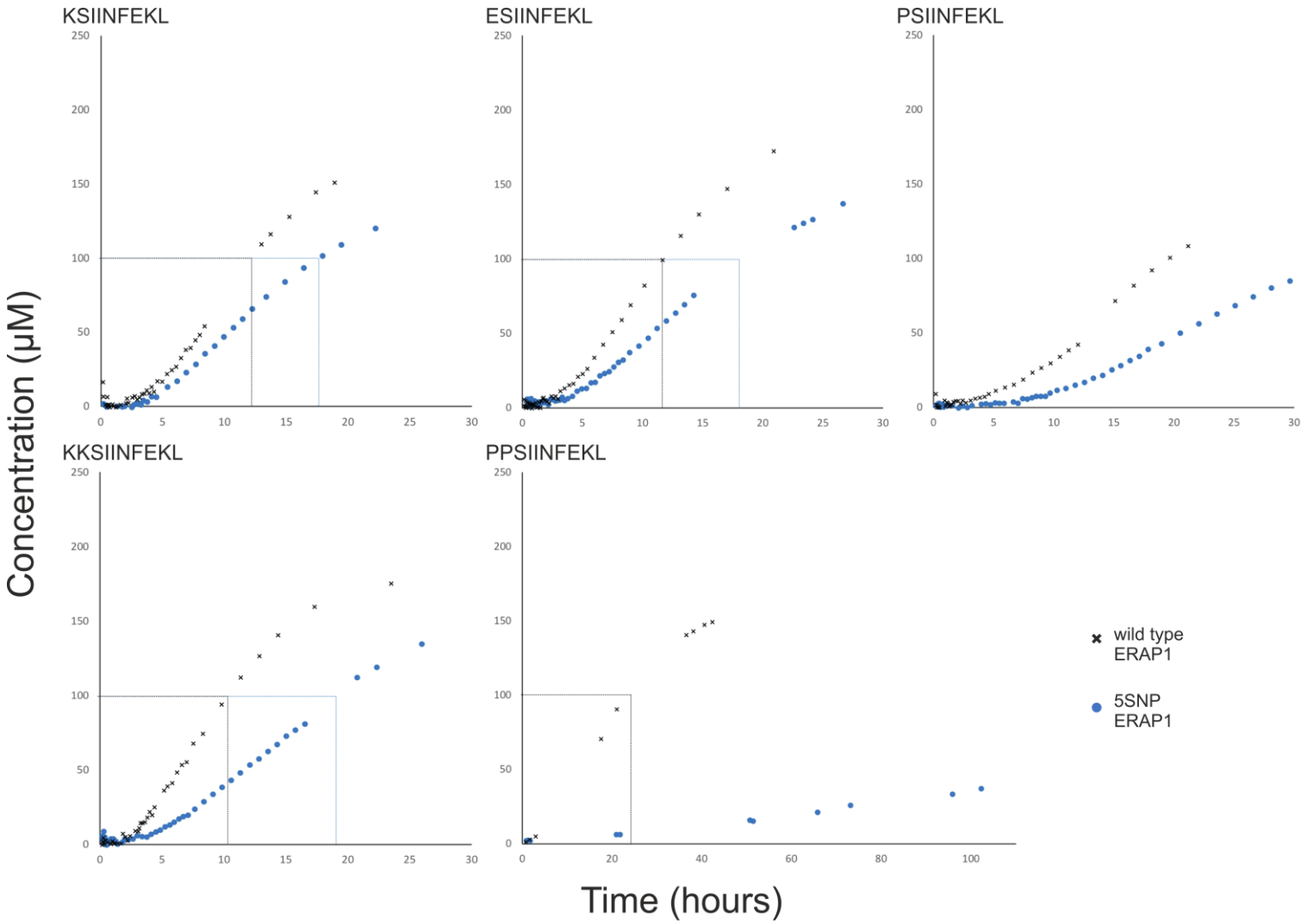


Figure 87 - Concentration of L-phenylalanine present in reaction over time, for wt and 5SNP ERAP1 reactions of different peptides

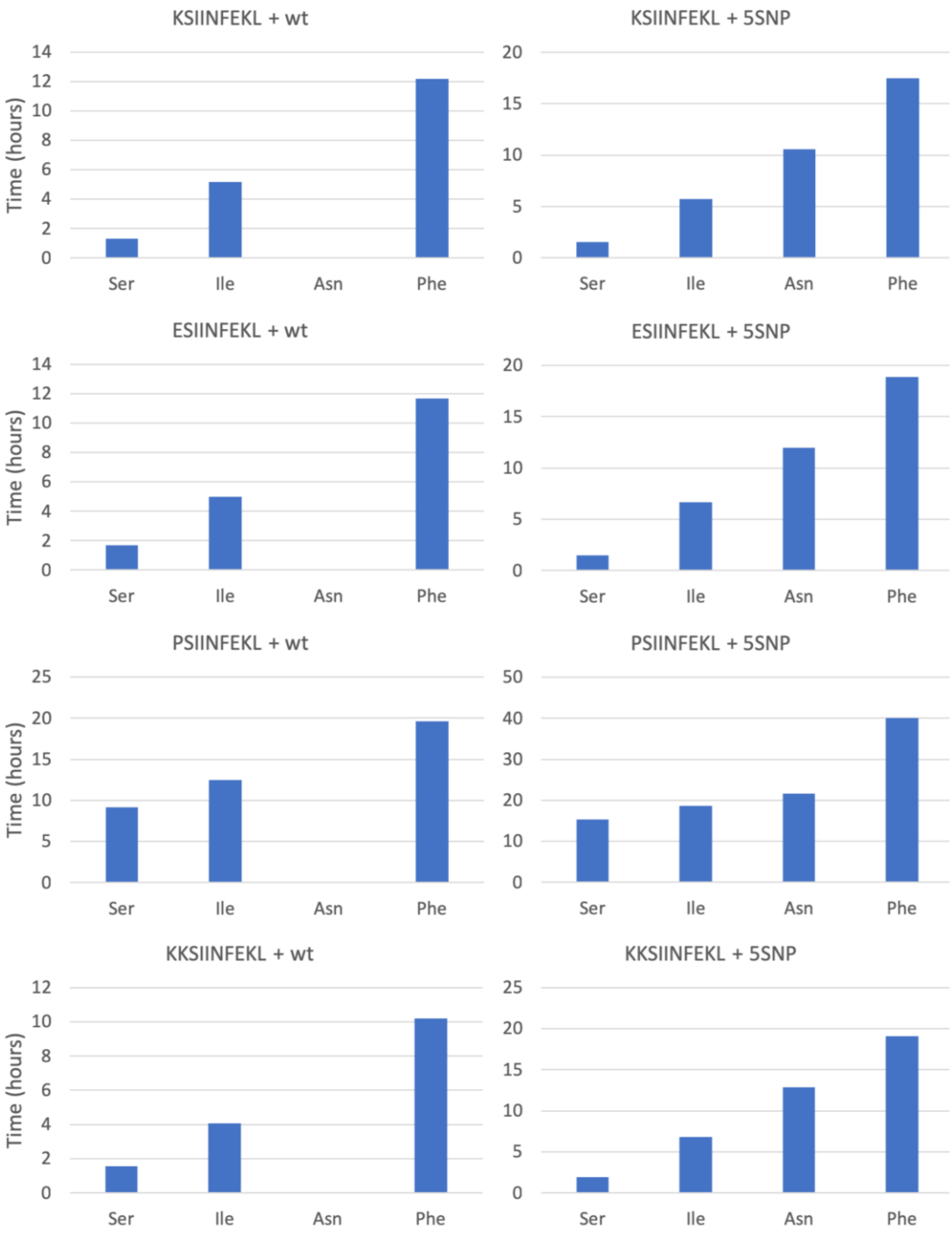


Figure 88 - Graphs of time taken to reach half maximum amino acid concentration in reaction, per each residue. Plotted for each peptide and ERAP1 variant.

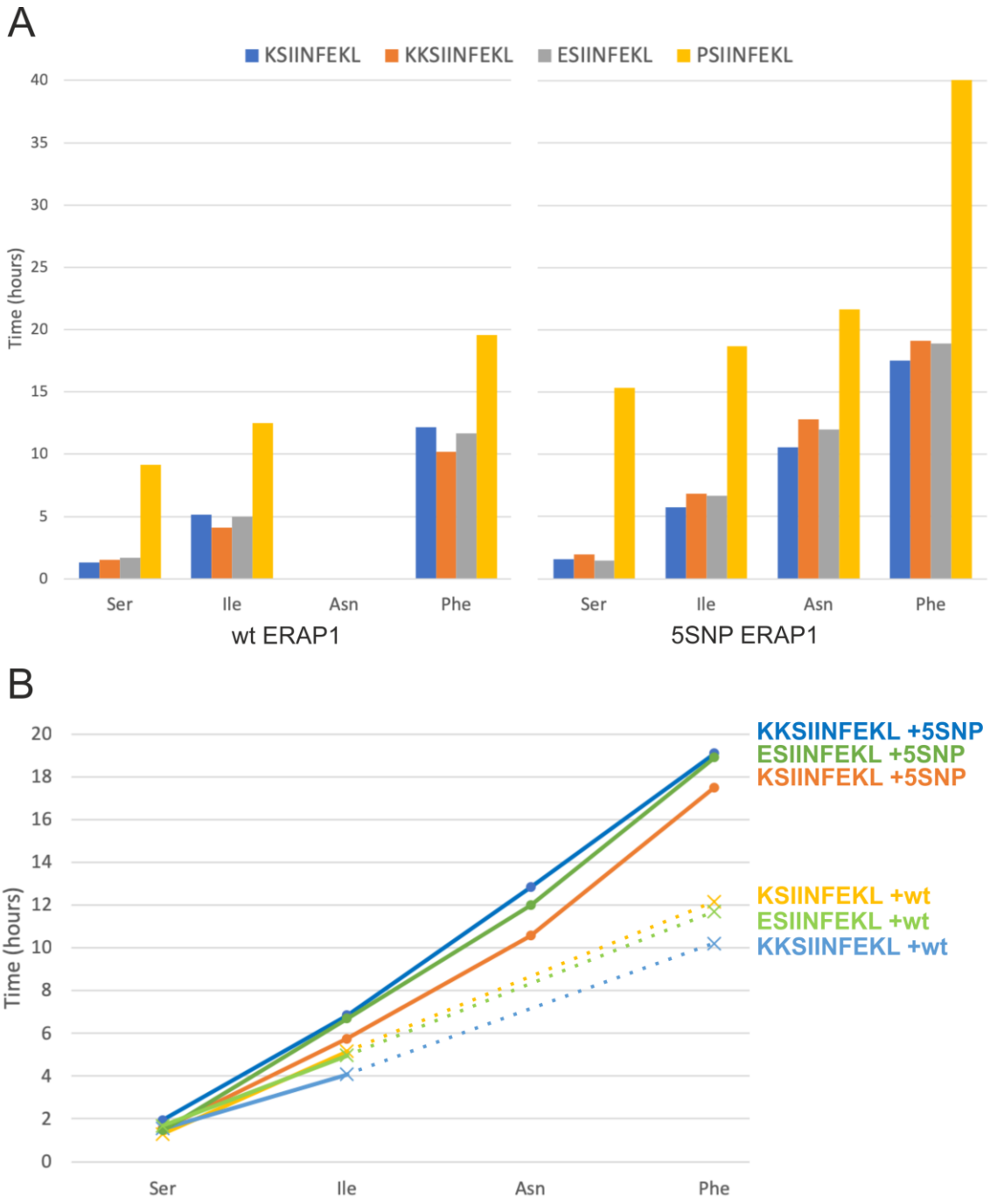


Figure 89 - Time taken to reach 50% of maximum concentration in reaction, per amino acid. A) The half-maximum time for the peptides with wt ERAP1 (left) and 5SNP ERAP1 (right). B) Half maximum concentration time for peptides excluding PSIINFEKL showing difference in activity and preferences between wt and 5SNP ERAP1.

5.3.4 Saturation transfer difference NMR of wt and 5SNP ERAP1 variants

The binding of peptide C-termini to the large internal cavity of ERAP1 has been suggested to take place using hydrophobic side chains, interacting with small pockets in the cavity (Nguyen et al. 2011). This theory arises from the size of the cavity and examination of the crystal structures, as well as the observations of ERAP1 length and C-terminal residue preferences (Chang et al. 2005; Nguyen et al. 2011). However, to date, which parts of a peptide bind to ERAP1is not fully investigated, and no crystal structure of an ERAP1-bound peptide exists. It is possible to examine which parts of the peptide interact with the protein without a solved structure, and this can be done through saturation transfer difference NMR. Presented here are the findings of these experiments.

PPSIINFEKL was used for saturation transfer difference NMR experiments with 5SNP and wt ERAP1, due to its very poor trimming and prolonged retention within the reaction mix. Saturation frequencies were tested on PPSIINFEKL to find one within the range of protein resonances but distant from the peptide chemical shifts. 3500 Hz (close to 10 ppm) was selected for being within the protein envelope, but also causing minimal baseline distortions and no peptide saturation (Figure 90).

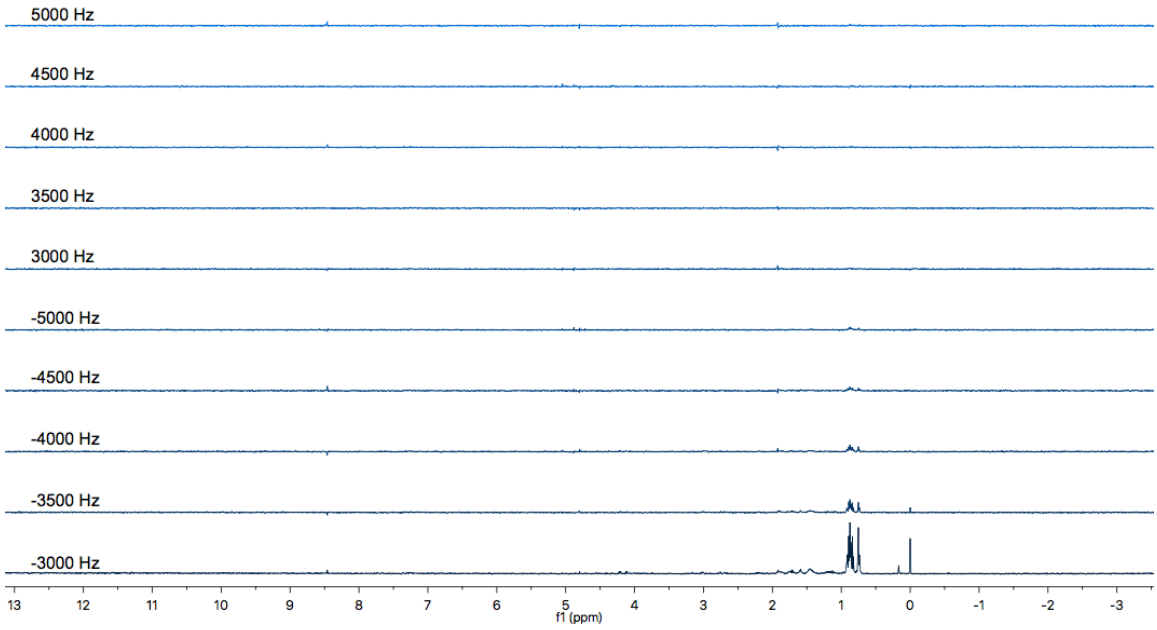


Figure 90 - Testing different saturation frequencies on 200 M PPSIINFEKL with a phase cycled self-subtracting experiment.

The saturation transfer experiments for 5SNP and wt ERAP1 were run for close to 13.5 hours and 1D presaturation experiments run before and after indicated that PPSIINFEKL remained present as a substantial proportion of the reaction contents.

Notably for both experiments, TSP presents itself as an obvious peak at 0 ppm (Figure 91). This is perhaps not especially surprising, as TSP is known to bind to proteins and peptides (Jayawickrama and Larive 1999). Other peaks which are easily identified from both the wt and 5SNP spectra are those of the phenylalanine ring, 7PheHδ (7.27 ppm), 7PheHζ (7.32 ppm) and 7Phe Hε (7.37 ppm), and the methyl groups of the 4th and 5th position isoleucine, and for the leucine at position 10 in the peptide (Figure 91). These fall at 0.74, 0.86 and 0.92 ppm and integration of the peak areas suggests the peak size and thus interaction with ERAP1 may be greater for 5IleHd (0.74 ppm) than for 10LeuHd2 (0.92 ppm) (see Table 10, Figure 92). Another peak obvious enough for automatic peak picking to identify it was the Hα peak for 4Ile or 10Leu (4.21 ppm). This was evident in both the wt and 5SNP ERAP1 spectra (Figure 91). It

was difficult to be certain that there was any resonance for the 8GluHα peak close beside it (4.24 ppm) and no other Hα peaks were apparent.

As the noise level was significant and saturation transfer effects for other peaks was poor, identification of other peaks beyond those already mentioned were significantly less confident. Automatic peak picking with the noise threshold set high conferred slightly more certainty to the designation of peaks as saturation transfer effects, rather than noise, and two further peaks were identified this way. The 6AsnHβb (2.78 ppm) was picked automatically by the MNova software for the 5SNP, but not wt variant, and a neighbouring peak at 2.68 ppm for 6AsnHβa could also be seen (Figure 91). For the wt variant, conversely, there appeared to be a peak for 6AsnHβa but nothing for 6AsnHβb, and this was not selected by automatic peak picking. The other peak identified by the program was 2ProHδb (3.74 ppm) for wt ERAP1, but not for 5SNP, although there does appear to be a small peak here for both variants. No peak is obvious for 2ProHδa (3.62 ppm). 1ProHγ (2.09 ppm), 8GluHβb (2.03 ppm) and 10LeuHβ/γ (1.60 ppm) were cautiously identified in both spectra, but with weak signals which could potentially result from noise.

The 5SNP ERAP1 signals appear to be much stronger than those for the wt, for TSP as well as the peaks resulting from peptide. Comparison of peak areas between saturation transfer experiments and a PPSIINFEKL presaturation experiment shows this (Figure 92) although the peak areas for the PPSIINFEKL-only STD test experiment (blue bars in Figure 92) show that minor noise and baseline distortion make integration of most of the peaks highly unconvincing. It is more suited to the methyl peaks which still have a poor signal-to-noise (for the wt ERAP1 spectrum it is 1.95, 6.28 and 2.52 for 0.91, 0.86 and 0.74 ppm respectively, for 5SNP a SNR of 2.69, 8.2 and 2.03 for 0.91, 0.86 and 0.74 ppm) but have stronger signals. The peak area comparison for these would suggest more STD-signal-to-presat-signal for the 5IleHδ peak (0.74 ppm) than 10LeuHδ2 (0.92 ppm) for both variants (Figure 92). Interestingly 5SNP also has a much stronger TSP signal than wt ERAP1 (Figure 91).

Overall, for both ERAP1 variants, binding appears to be strongest to hydrophobic residues, with clear signals from 4Ile, 5Ile, 7Phe and 10Leu. Most of these peaks in the saturation transfer difference spectra are for the hydrophobic side chain, encompassing the methyl groups of 4Ile, 5Ile and 10Leu, and the ring structure of 7Phe. However, it is not exclusively side chain groups which are clearly seen, but also the Hα of either 4Ile, 10Leu or both, since they are so closely positioned on a 1H spectrum that they are indiscernible from each other.

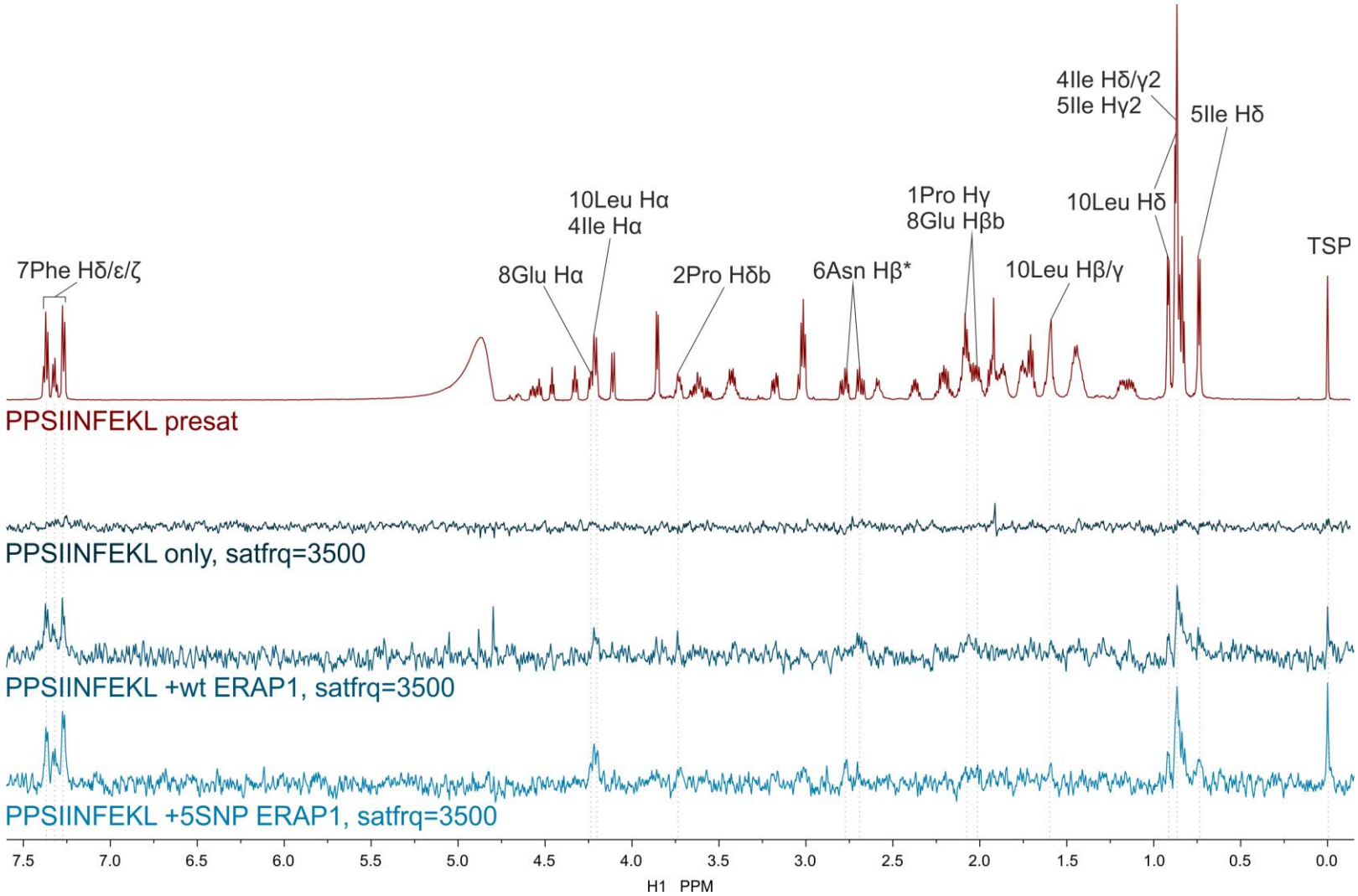


Figure 91 - Spectra of PPSIINFEKL, as (top) a presaturation 1D experiment, (2nd) a saturation transfer experiment in the absence of protein (3rd) a saturation transfer difference spectrum with wt ERAP1, (bottom) a saturation transfer difference spectrum with 5SNP ERAP1

Table 10 - Peaks in saturation transfer difference spectra of PPSIINFEKL with either wt or 5SNP ERAP1 variants

Position (ppm)	Proton ID	Feature visible in wt spectrum	Feature visble in 5SNP spectrum	Feature selected by automatic peak picking in MNova	wt absolute integral value	5SNP absolute integral value
Peaks clearly visible in spectra						
0.74	5IleHd	✓	✓	✓	19.18	27.34
0.86	4IleHg2/ 4IleHd/ 5IleHg2/ 10LeuHd1	✓	✓	✓	79.16	111.66
0.92	10LeuHd2	✓	✓	✓	7.48	18.25
4.21	10Leu/4Ile Ha	✓	✓	✓	22.23	37.24
7.27	7PheHd	✓	✓	✓	38.87	58.98
7.32	7PheHz	✓	✓		27.09	32.77
7.37	7PheHe	✓	✓	✓	46.67	43.43
Ambiguous peaks – may result from noise						
1.60	10LeuHb/g	✓	✓		12.24	15.57
2.03	8GluHbb	✓	✓		15.16	12.88
2.09	1ProHg	✓	✓		27.35	14.38
2.68	6AsnHba	✓			27.77	9.48
2.78	6AsnHbb		✓	✓	7.16	15.5
3.74	2ProHdb	✓		✓	3.67	15.74
4.24	8GluHa		✓		-1.43	11.24
Non-peptide peaks						
0.00	TSP	✓	✓	✓	12.99	35.17

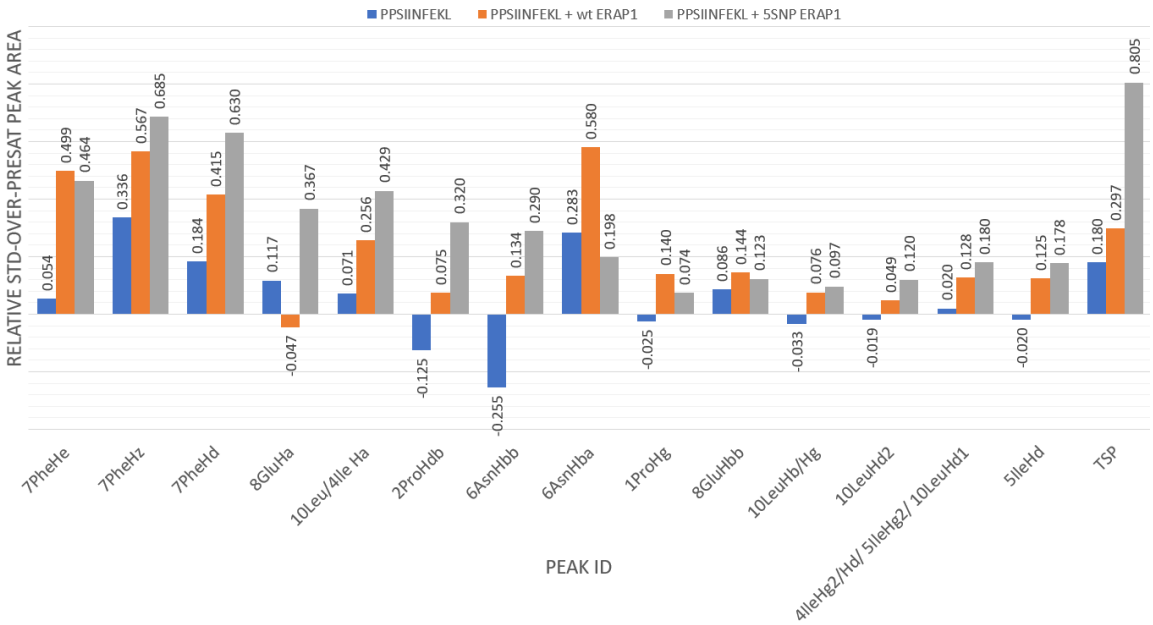


Figure 92 - Comparison of integrated peak area values between PPSIINFEKL STD spectra (peptide only, with wt ERAP1, and with 5SNP ERAP1) at 3500 Hz saturation frequency, and peak areas from the 'reference' PPSIINFEKL presaturation spectrum

5.3.4.1 Modelling the position of the peptide from NMR data

Tentative attempts to use the saturation transfer difference results to model a possible binding position for PPSIINFEKL within the ERAP1 cavity showed a plausible binding location for the peptide. The hydrophobic region of the cavity wall formed in the closed form of ERAP1 by the meeting of domains II and IV provided possible binding sites for the hydrophobic side chains, as suggested by Nguyen et al. (2011). These hydrophobic cavities (see Figure 93B) form possible interactions with 5Ile, 7Phe and 10Leu which would be in close proximity to the wall if this binding model is correct. Interactions between 4Ile and the cavity are less clear, although it could interact with the nearby hydrophobic patch. The side chain amine of 9Lys might interact with the hydrophilic regions of the cavity, as might 8Glu, although these groups do not show up in the deuterated spectrum of PPSIINFEKL and it is unknown whether they contributed to binding. Unlike the other side chains, 6Asn points into the solvent within the cavity. It is not clear

how it could have contributed to binding, suggesting either wrong identification of 6AsnHb* in the saturation transfer difference NMR data, or the peptide has alternative binding conformations. The modelled position of the peptide lying flush with the curved cavity wall of ERAP1, with the N-terminal extending upwards to domains I and II, and the C-terminal extending down into the cavity from domains II to IV, is shown in Figure 94.

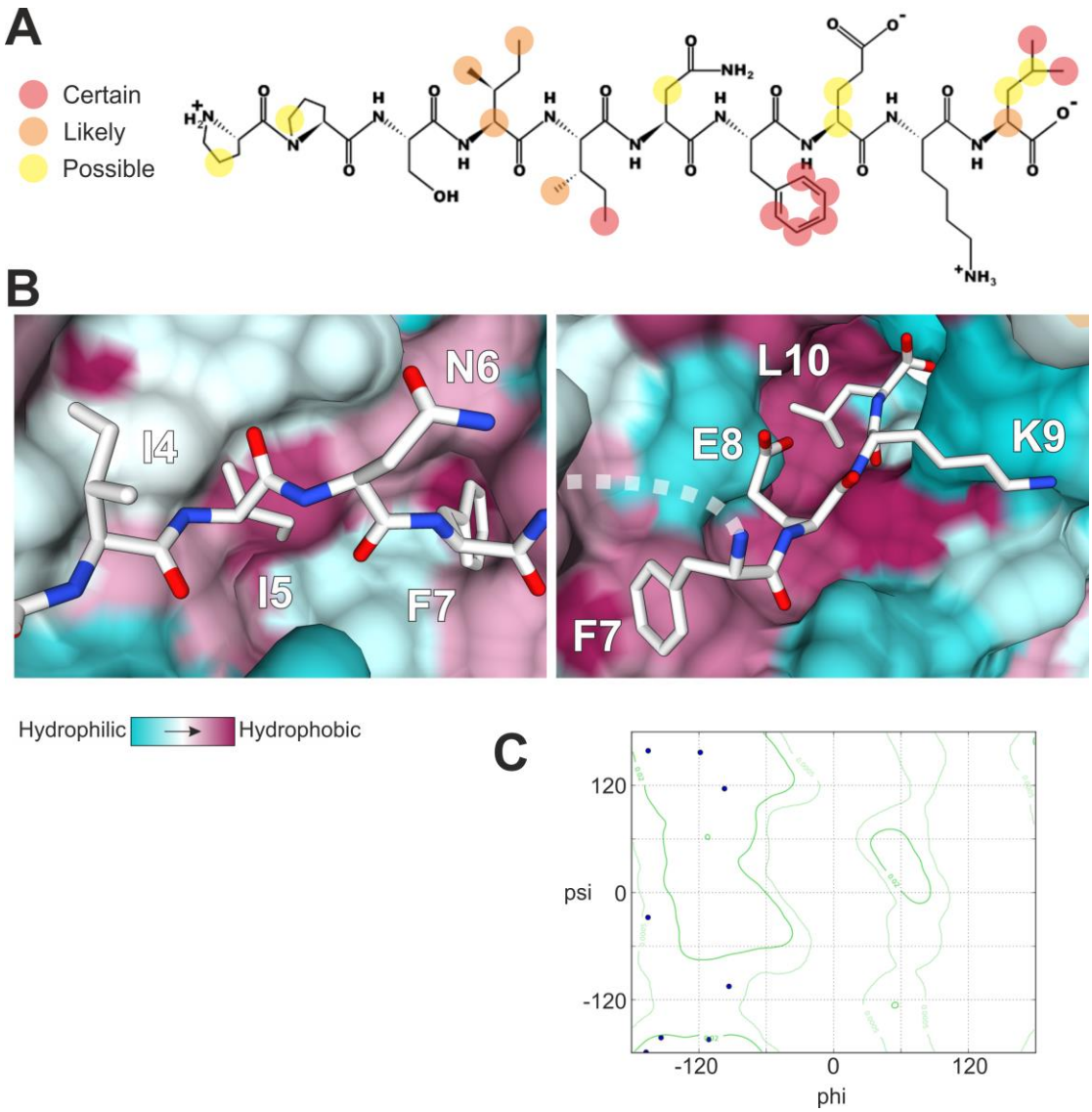


Figure 93 - Modelling the possible position of PPSIINFEKL within the ERAP1 binding cavity. A) saturation transfer difference NMR results from both wt and 5SNP ERAP1 marked on the structure of PPSIINFEKL. B) Stick model of PPSIINFEKL showing side chains of I5, F7 and L10 directed towards hydrophobic pockets. C) Ramachandran plot of modelled PPSIINFEKL with acceptable dihedral angles for all residues. PDB ID: 2YD0, modelling performed in UCSF Chimera (Kochan et al. 2011; Pettersen et al. 2004).

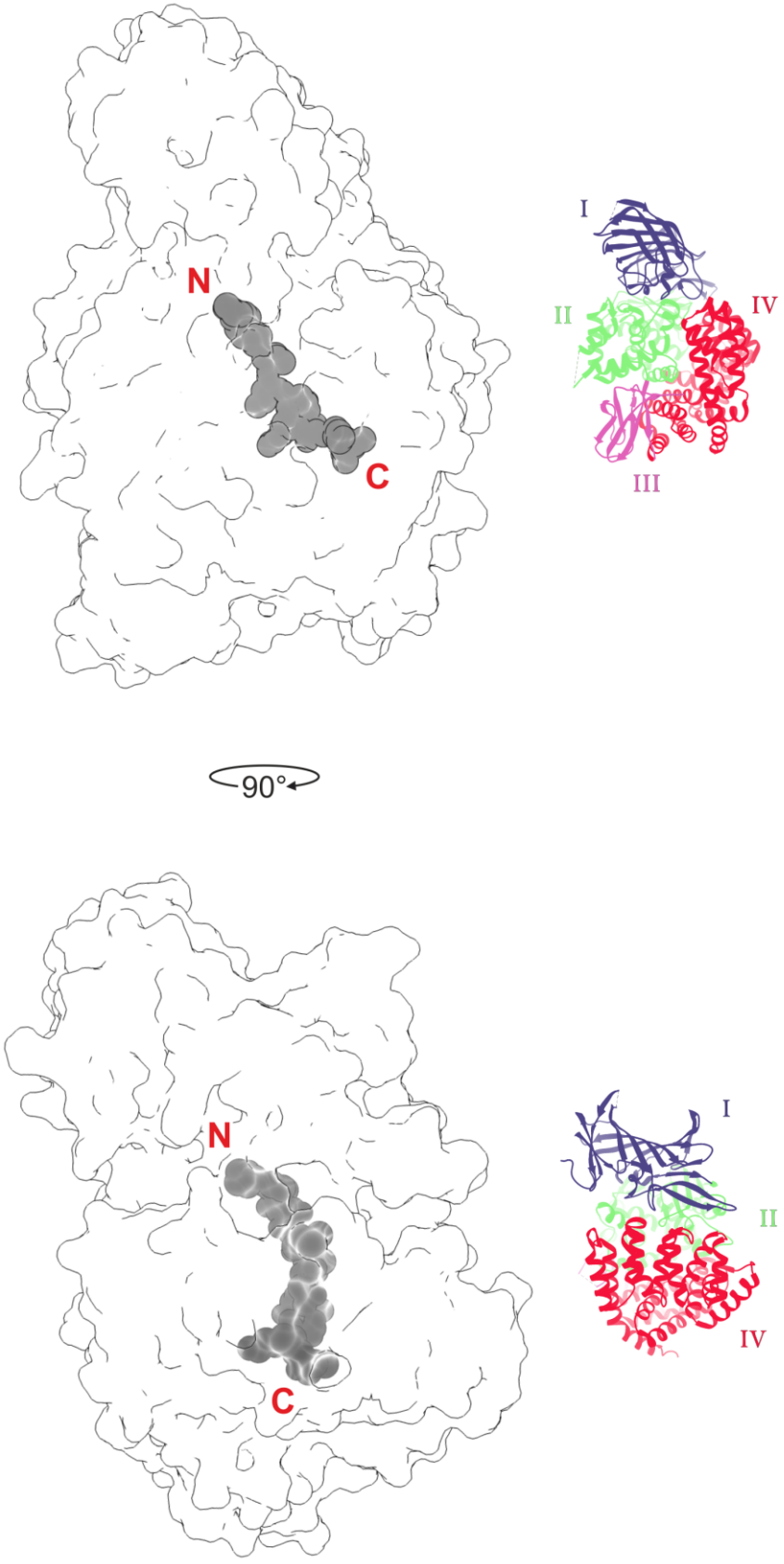


Figure 94 – Modelled position of PPSIINFEKL (dark grey) within a closed form of ERAP1 (ribbon representation of same shown to right). PDB ID: 2YD0, modelling performed in UCSF Chimera (Kochan et al. 2011; Pettersen et al. 2004).

5.4 Discussion

Degradation of peptides by peptidases can be seen using NMR to monitor the reaction. The difference between samples containing peptide alone and peptide with active ERAP1 can be seen even when looking at a small number of key resonances (compare Figure 72B with Figure 76B) where peaks representing amino acids can be seen to appear in the spectrum, substrate peptide peaks correspondingly decrease, and product peptides are created then removed. The sequential generation of amino acids and product peptides is in accordance with what you would expect from an aminopeptidase for all the peptides shown here, with the exception of PPSIINFEKL.

PPSIINFEKL and PSIINFEKL were both predicted to be non-trimmed peptides due to the N-terminal proline. Surprisingly for PSIINFEKL it was not only degraded but appeared to be degraded in an aminopeptidase specific manner, despite N-terminal prolines being variously described as poorly cleaved by ERAP1 since early work was done on the protein (Serwold, Gaw, and Shastri 2001) and impossible due to the nature of the active site (Efstratios Stratikos, personal communications). The fact that PSIINFEKL is degraded, albeit at a slower rate than the other nonameric peptides, could be an effect of the relatively high concentrations of ERAP1 in the reaction, at a hundredth of the initial peptide concentration, although the *in vivo* study published by Reeves et al. (2013) found evidence of a small amount of trimming for proline N-terminal extended SIINFEHL. Hearn, York, and Rock (2009) similarly suggested a dramatically lower trimming ability for 1Pro compared to other N-termini, but did not state a complete inability to cleave it. Certainly, it would be more difficult for the reaction to occur, due to the pyrrolidine ring offering a more rigid N-terminal than the amine which is normally bound by E320 at the start of the reaction, and a secondary amine rather than $-NH_3^+$ (or deprotonated form), but it is 2Pro which ERAP1 is known to not trim, matching TAP preferences, due to the tertiary amide of 2Pro (Serwold, Gaw, and Shastri 2001).

PPSIINFEKL also displayed a profile of degradation unlike the gradual decomposition of PPSIINFEKL alone, especially for the wild-type variant which generated around 150 μM of both phenylalanine and isoleucine over the course of the experiment. This is surprising, and though EDTA-deactivated ERAP1 appears to create glutamic acid and ESIINFEKL from the EESIINFEKL substrate and some unexplained chemical shifts are present, there is nothing to suggest generation of amino acids from internal trimming. The conclusion to this is either that ERAP1 can trim in an endopeptidase specific manner when N-terminal residues are so unfavourable, or (despite streptactin affinity purification followed by size exclusion chromatography) a small quantity of some other endopeptidase was present and degrading the peptide. This may also explain the low (rather than non-existent) activity of the E320A mutant, which is described as 'non-functional' (Reeves et al. 2013). Feasibly, another active site residue could coordinate both the water molecule and the peptide N-terminal, though this would be incredibly unfavourable and might no longer restrict ERAP1 to aminopeptidase activity, it seems unlikely.

For the PPSIINFEKL/wt ERAP1 experiment, the question arises why phenylalanine and isoleucine appear to be made at approximately equal amounts. If proportional amounts of each are present then this cannot be due to PPSIINFEKL being cleaved to PPS and IINFEKL because the isoleucine molarity would be twice that of phenylalanine when ERAP1 had trimmed IINFEKL. Neither can it be cleaved to PPSI and INFEKL because asparagine is not present. It is plausible that the resonances for one or both of these amino acids is in this case representing a different molecule, perhaps IN or INF, of insufficient length for ERAP1 to trim rapidly.

If PPSIINFEKL is being degraded in an endopeptidase fashion by ERAP1, this could have implications for the saturation transfer difference results, which indicated that ERAP1 was predominantly binding the hydrophobic methyl groups of the 4^{th} and 5^{th} position isoleucines, as well as the C-terminal leucine, and the phenylalanine side chain ring. Transfer of saturation was also seen at 4.21 ppm, a chemical shift where both the Hα of 10Leu and 4Ile sit. It is unclear

whether this particular peak arises because of interaction between 4IleHα and ERAP1, or between 10LeuHα and ERAP1. Alternatively, it arises from both, and other Hα resonances are present but not visible because they represent a single proton, and are therefore insignificant amid the spectral noise. Interactions of the hydrophobic regions with ERAP1 fits with the reported observations of ERAP1 preferring 9-mer and 10-mer substrates with hydrophobic C-termini (Chang et al. 2005; Evnouchidou et al. 2008). Crystal structures of ERAP1 indicate shallow pockets for these hydrophobic side chains to fit into within the binding cavity (Nguyen et al. 2011; Kochan et al. 2011) and these were used to create a model of PPSIINFEKL binding which fitted saturation transfer data. These saturation transfer experiments would therefore agree with that, but it is uncertain whether the N-terminal is binding as it would with a non-proline N-terminal. There was no obvious difference between the results for 5SNP and wt ERAP1, other than smaller peaks for wt ERAP1. The wild-type variant also had a smaller peak for TSP, relative to the other peaks present, which might indicate a stronger peptide affinity but lower concentration of ERAP1 than thought.

The TSP reference also had saturation transfer from ERAP1, and this could have implications for quantitation accuracy. TSP is a reliable reference compound and fairly stable, but as well as binding to biopolymers, is known to adsorb to glass surfaces (Jayawickrama and Larive 1999). A more significant issue for accurate calculation of reaction components is the size and crowding of peaks. For resonances which stand alone and represent several protons (such as the Hγ2 of L-isoleucine) quantitation is readily done. For small Hα peaks this is more difficult, especially where the water signal and solvent suppression encroach. Another significant issue can be baseline changes due to large neighbouring peaks, though this problem was negated by careful baseline correction with the MNova software, where necessary selecting blank regions in all spectra for 'multipoint' manual baseline correction. The ERAP1 signal could have contributed to peak area in some parts of the spectra, being at 1% of the initial substrate peptide concentration, and a trim pulse might improve this.

Table 11 - Time taken (hours) to reach half of maximum concentration of amino acid product, for each peptide and ERAP1 variant. Dashes are given where a half maximum concentration was not reached throughout the whole experiment period.

	KSIINFEKL		KKSIINFEKL		ESIINFEKL		PSIINFEKL		PPSIINFEKL	
	5SNP	wt	5SNP	wt	5SNP	wt	5SNP	wt	5SNP	wt
Ser	1.6	1.3	1.9	1.5	1.5	1.7	15.3	9.1	-	-
Ile	5.8	5.2	6.8	4.1	6.7	5.0	18.7	12.5	-	-
Asn	10.6	-	12.8	-	12.0	-	21.6	-	-	-
Phe	17.5	12.2	19.1	10.2	18.9	11.7	40.1	19.6	-	-

Despite these limitations, the experiments were sufficient to show the differences between wt and 5SNP for the different peptides. There was little difference between wild-type trimming of KSIINFEKL and ESIINFEKL, but much slower activity for PSIINFEKL, which is expected. It is unclear why KSIINFEKL and ESIINFEKL show such similar trimming; perhaps the experiments are not optimised for maximising the difference between ERAP1 trimming preferences (for extensions other than the very poorly trimmed, like proline). It would be of interest to see if a hydrophobic N-terminal showed any significant increase in ERAP1 activity. Assignments of LSIINFEKL and LLSIINFEKL in 90% H_2O have already been made, and so this is a plausible avenue for future expansion of ERAP1 monitoring by NMR.

The half maximum concentration for the isoleucine, asparagine and phenylalanine was reached after a longer period of time for ESIINFEKL than KSIINFEKL in the 5SNP experiments, which would suggest 5SNP prefers K- extensions to E- extensions, but for serine the generation appeared more rapid from ESIINFEKL than for the wild-type. This is hard to explain: it seems unlikely that the cleaved N-termini have any further impact on the activity of ERAP1, so perhaps the calculated serine concentrations for the ESIINFEKL/5SNP experiment is incorrect. The finding that KSIINFEKL appears more rapidly trimmed than ESIINFEKL by 5SNP is surprising, given that this contradicts published observations that E- is a preferred N-terminal residue to K- for by wt and 5SNP variants (Reeves et al. 2013).

Interestingly wt ERAP1 showed an obvious length preference for 10-mer KKSIINFEKL over 9-mer KSIINFEKL, whereas for 5SNP, a slow trimmer, the additional N-terminal residue appeared to slow generation of later amino acids. This could be due to different preferences for internal residues between wt and 5SNP but could also be due to length preferences. 5SNP evidently processes the K/S cleavage at the start of KSIINFEKL rapidly and it is possible that it would struggle with a lysine-followed lysine for KKSIINFEKL. Work by Hearn et al. (2009) has previously suggested that the second position can dictate peptide trimming, for example VLSIINFEKL appearing to be trimmed more rapidly by wild-type ERAP1 than LVSIINFEKL. Prior to that publication, internal peptide sequence was found to affect ERAP1 trimming (Evnouchidou et al. 2008). Different length preferences also seem plausible. In the study by Reeves et al. (2013) it was demonstrated that M349V/K528R and K528R/Q730E haplotypes processed X6-SHL8 efficiently but X5-SHL8 poorly, in contrast to the wild-type which generated SHL8 very readily from X5-SHL8. By contrast M349V and M349V/D575N/R725Q trimmed X5-SHL8 well and X6-SHL8 poorly. This was stated to be due to substrate sequence preferences, which is a practical conclusion to have drawn since the peptides used were derived from naturally occurring sequences (X5 was AIVMKSIINFEHL whereas X6 was LEQLEKSIINFEHL) rather than N-terminal repeats of the same amino acids (Reeves et al. 2013).

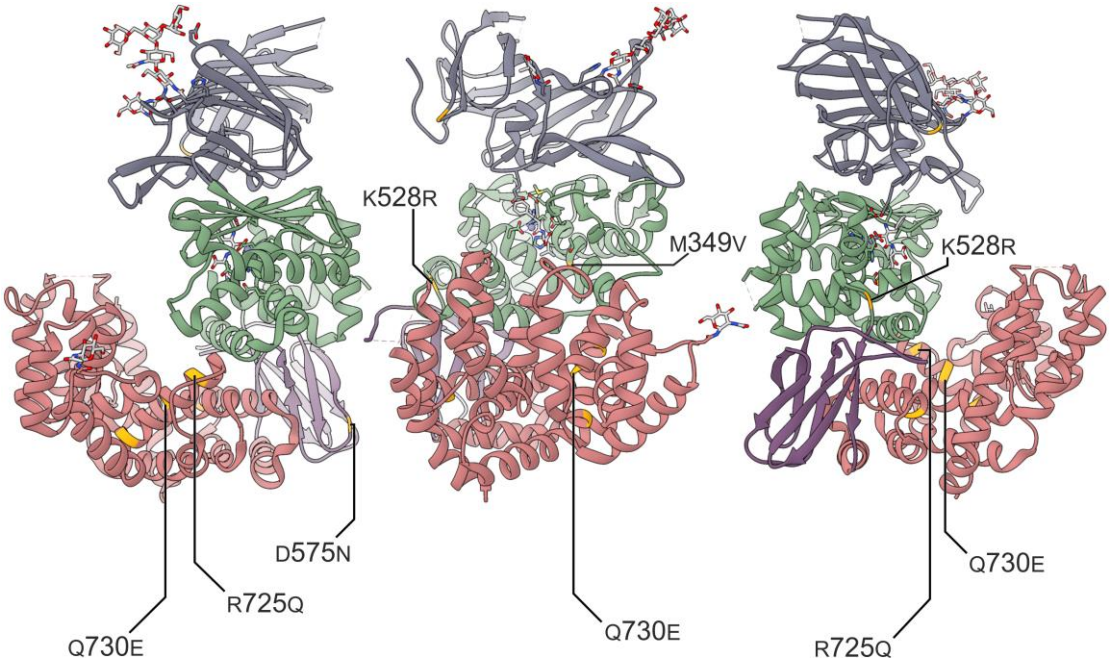


Figure 95 - The 5 single nucleotide polymorphisms in the '5SNP' ERAP1 variant structure

It is difficult to suggest the effects of mutations such as those in 5SNP with any conviction since looking at a static wild-type protein crystal structure does not give a full picture of how they will affect behaviour. Nonetheless, the SNPs in 5SNP are not all close to the active site, but R725Q and Q730E are in domain IV (Figure 95) where they could come into contact with residues towards the C-terminal and affect length specificity. How mutations that convert the position from positive to uncharged, and uncharged to negative residues would make a difference here is unclear. They may have less impact on the characteristics of 5SNP than the other SNPs, since the one with strongest linkage to AS is R528 at the hinge between domains II and III (D. M. Evans et al. 2011). This residue might be involved in closing of the protein (see Figure 6) and a bulky arginine side chain substituted for lysine might feasibly alter how this occurs. Since the closed form of ERAP1 is suggested to be active and the open state inactive (Kochan et al. 2011) this would presumably have an impact on peptide trimming. Observations of pMHC I complexes being edited by ERAP1 hyper-trimmers, even when the peptide is an optimal length for MHC I (Reeves et al. 2013) might suggest a disadvantage for a hypo-trimmer here, which could struggle to compete with MHC I for the peptide and possibly result in sub-optimal pMHC I complexes.

6 Conclusions

This study aimed to further understanding of the operation of ERAP1, with a focus on its activity on free peptides and into potential action on MHC I-bound peptides. ERAP1 is known to have sequence specific 'molecular ruler' behaviour which modulates its behaviour towards different N-terminal and internal residues (Evnouchidou et al. 2008) and different peptide lengths (Chang et al. 2005) to produce a selection of peptides available to MHC class I. MHC I-bound (or partially bound) peptides have also been demonstrated to undergo trimming by ERAP1 (Reeves et al. 2013) and ERAP1/2 dimers (Chen et al. 2016) and it was posited that impaired binding of peptide in the MHC I binding groove would enable ERAP1 to trim the peptide.

In Chapter 2 the crystal structures of four SIINFEKL/H-2K^b/β_2m disulphide trap single chain trimers with A pocket mutations of key residues were presented. Peptides were bound in canonical fashion and there was no evidence of major structural impairment or aberrant peptide binding. A pocket mutations resulted in changes to the hydrogen bonding network, differing orientations of crucial side chains (such as K66 in the E63A mutant structure) and additional or positionally shifted water molecules in the peptide N-terminal binding region. These structures were used for MD simulations by Dr Athanasios Papakyriakou and published alongside data showing ERAP1 trimming of the peptide N-terminal was increased in the three mutant structures compared to the 'wild-type' SL8-SCT, thus suggesting reduced MHC I/peptide affinity could modulate trimming by ERAP1 (Papakyriakou et al. 2018). In future, reproducing trimming of extended-peptide dtSCTs and/or pMHC I complexes by isolated ERAP1 would be an interesting insight into the relationship between ERAP1 and MHC I. This potential reaction could be monitored by NMR.

Proton assignments of a small library of previously uncharacterised peptides, based on SIINFEKL, with 1 or 2 residues added or subtracted at the N-terminal, were determined from 2D homonuclear experiments. The combination of

chemical shifts were determined to be sufficiently unique to each peptide for identification by 1D proton NMR. ^{13}C and ^{15}N assignments were also made for PSIINFEKL. For this the peptide was expressed in *E. coli* first in unlabelled and then in isotope labelled form, with sufficient quality and quantity for the assignment from HSQC spectra.

The proton assignments were used for ERAP1 experimentation in Chapter 5. It was demonstrated that ERAP1 reactions could be monitored with sequential collection of NMR spectra. Isolated peptide showed patterns of slow decay which were unlike N-terminal removal. By contrast, peptides in the presence of 5SNP or wild-type ERAP1 were rapidly degraded, resulting in significant changes to chemical shifts and peak heights within minutes for high affinity peptides such as KKSIINFEKL. Amino acids were created consecutively in order of peptide sequence for all peptides but PPSIINFEKL. PSIINFEKL, expected to be trimmed extremely poorly, was degraded more slowly than the K- or E- N-terminal extensions, but still within a similar experimental timeframe and was less suited to being a control peptide than expected. PPSIINFEKL also degraded, though in a non-aminopeptidase specific fashion, with approximately equivalent concentrations of species with chemical shifts matching phenylalanine Hε and isoleucine Hγ2 but little serine or asparagine. This suggests a possible endopeptidase activity for ERAP1 where the peptide N-terminal does not have the chemical groups required for active site binding and bond hydrolysis.

Saturation transfer difference NMR with PPSIINFEKL indicated both 5SNP and wt ERAP1 were binding methyl side chains of the 4^{th}, 5^{th}, and 10^{th} position residues, as well as the side chain ring of the phenylalanine at position 7. This is largely in line with previous findings that ERAP1 prefers hydrophobic C-termini (Chang et al. 2005; Evnouchidou et al. 2008) and may have binding pockets for hydrophobic side chains such as these (Nguyen et al. 2011). Wild-type ERAP1 was demonstrated to degrade a 10-mer more rapidly than a 9-mer, whereas the 5SNP variant trimmed the same 10-mer more slowly than the 9-mer.

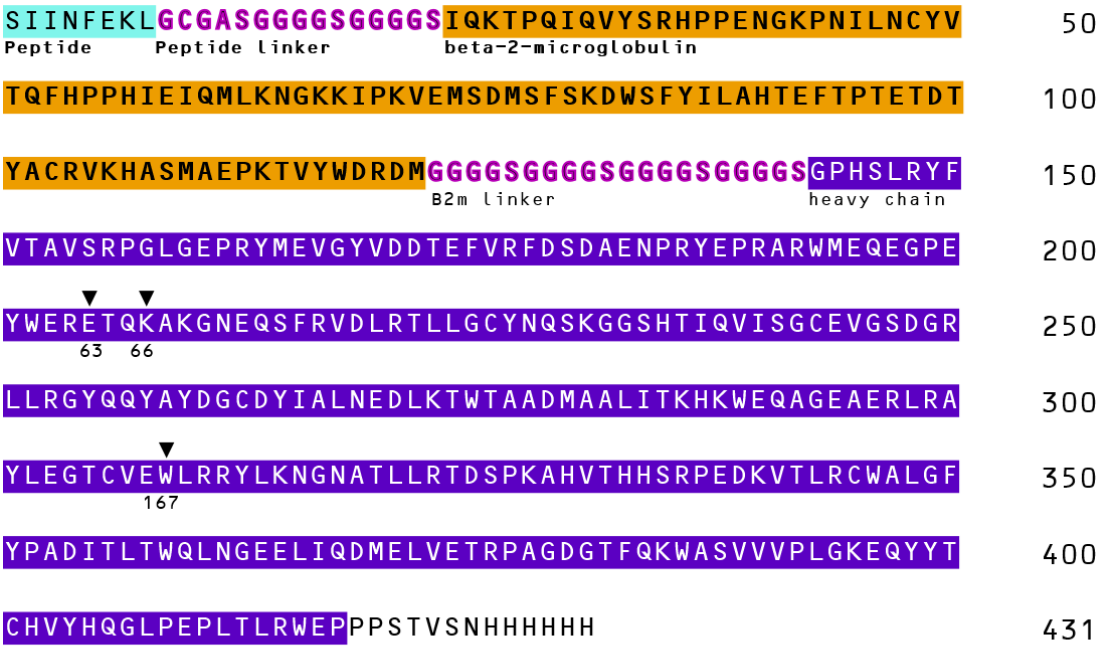
The insights gained from this experiment might be improved with optimisation of the experiment, possibly by using different or shorter peptides to improve isolation of monitored peaks within a spectrum and by running longer experiments to monitor formation of all products to completion. The complexity of the spectra being analysed posed challenges, nonetheless this thesis shows the potential for using a technique as simple as 1D proton NMR to view every step of an ERAP1 reaction, which has not previously been tested.

Future directions for this work could include refinement of the NMR reaction monitoring by using different ERAP1 variants, and also different peptides as stated above, or alternatively by using isotope-labelled peptides. The isotope-labelled peptides could have an additional use for testing MHC I peptide binding: because the peptides show random coil chemical shifts in solution, it would be of interest whether these were different for an MHC-bound peptide. Further from this, MHC I could be refolded with the N-terminally extended peptides in order to see whether any alternative binding conformations (e.g. extruded and buldged in the middle) were present. The dtSCT mutants trimming by ERAP1 has been examined in vivo (Papakyriakou et al. 2018) and an interesting extension of this work could be in vitro binding and interaction experiments, perhaps with a technique such as surface plasmon resonance. Additionally, further attempts to monitor ERAP1 trimming a dtSCT by proton NMR could provide interesting insight into how the two proteins optimise the peptide pool.

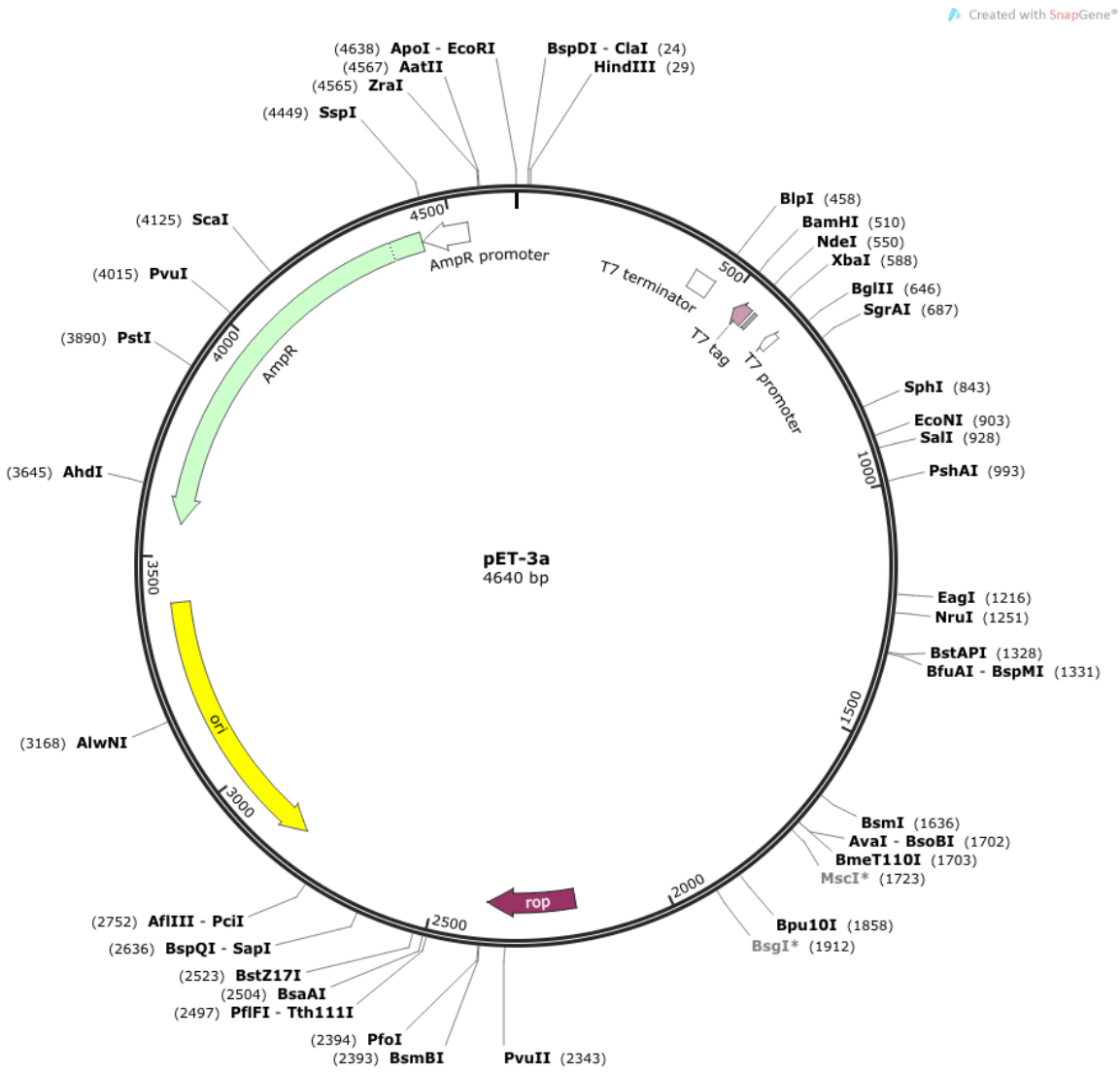
Appendices

Appendix 1 - Construct sequence for H-2K^{b} dtSCT

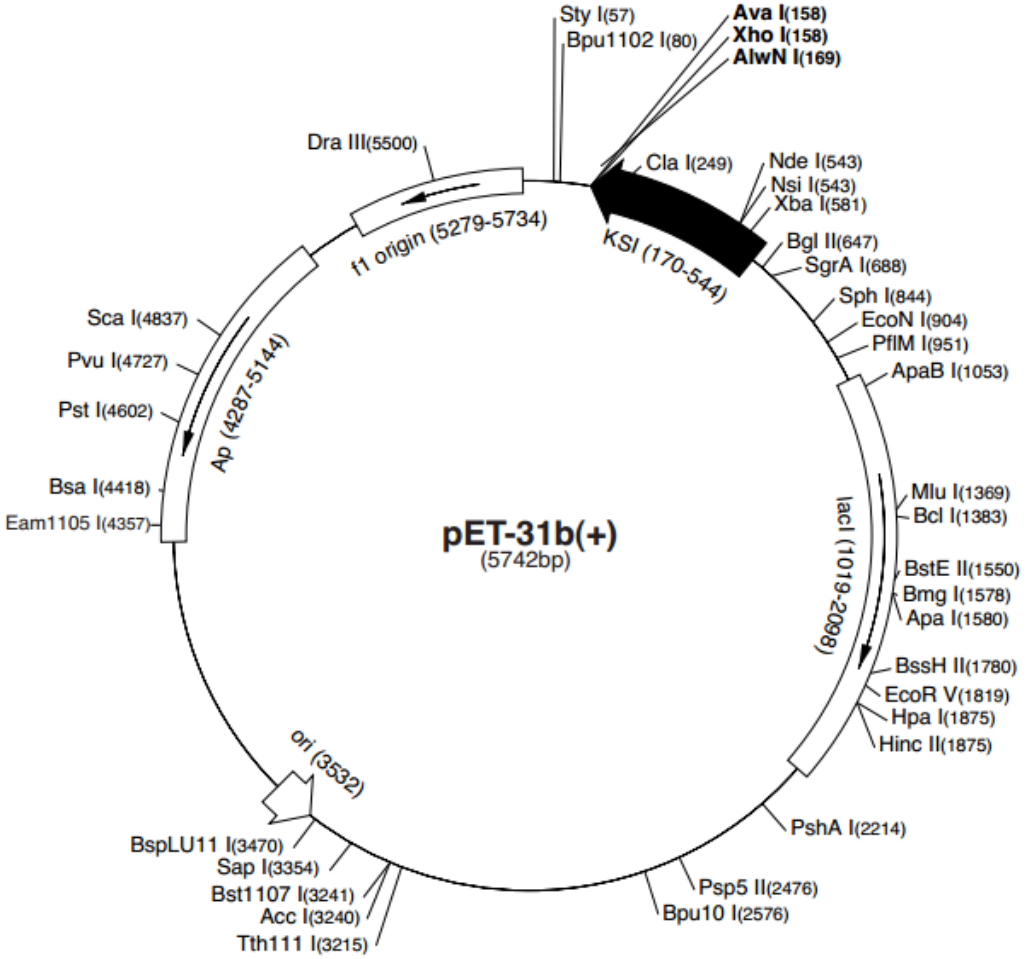
Amino acid sequence of expressed SL8-SCT 'wild type' protein. Regions (e.g. peptide, heavy chain) are marked by coloured boxes. Arrow markers indicate mutation sites for E63A, K66A and W167A.



Appendix 2 – pET3a vector map



Appendix 3 – pET-31b vector map

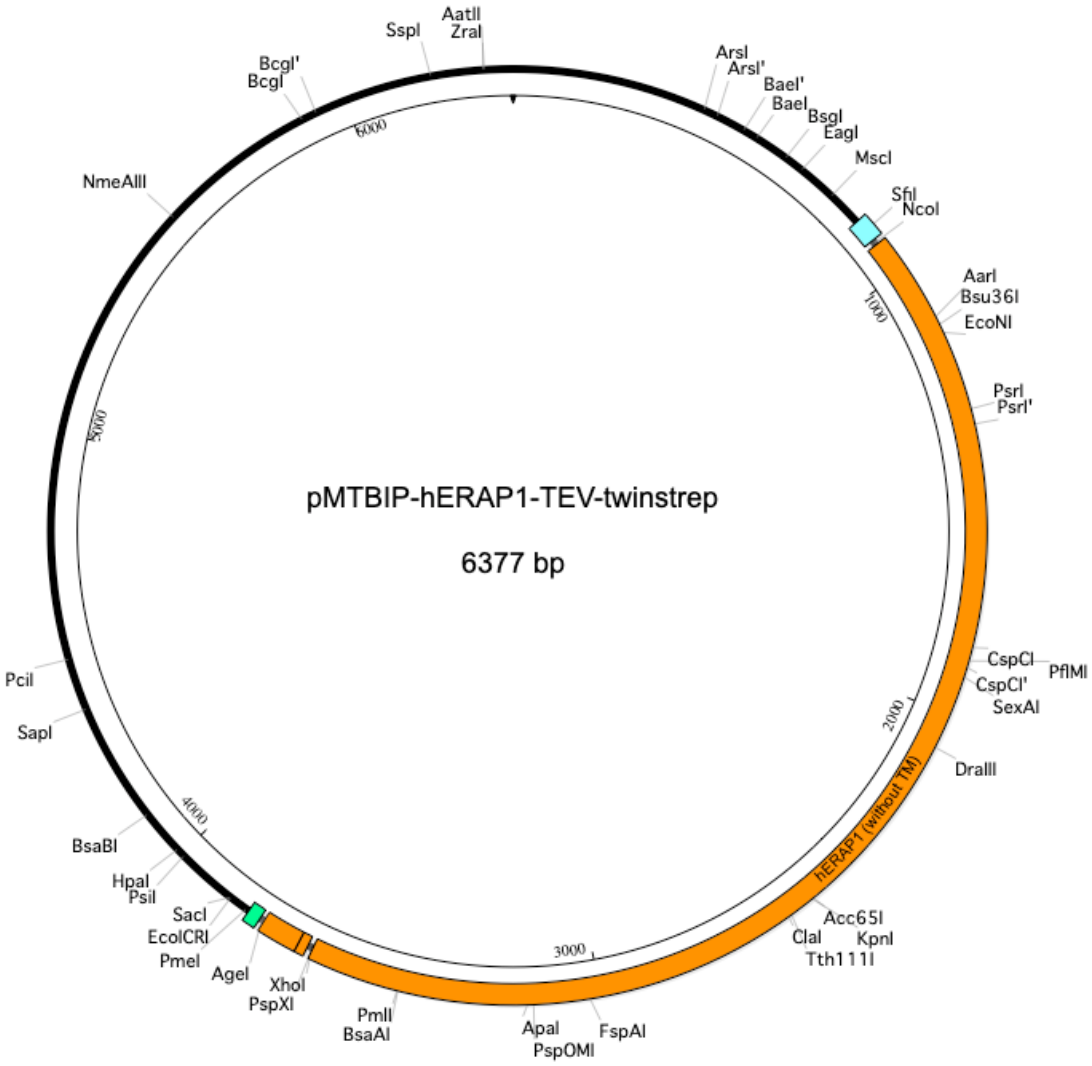


Appendix 4 – Peptide cloning oligos DNA sequences

Peptide cloning oligo DNA sequences (all shown 5'-3').

Name	Sequence	Tm (°C)	CG (%)	No. nt	Molecular weight (g/mol)
SIINFEKL Upper	agcattattaactttgaaaaactgtaaatg	62.7	23.3	30	9236.1
SIINFEKL Lower	ttacagtttttcaaagttaataatgctcat	63.1	23.3	30	9169.1
ESIINFEKL Upper	gaaagcattattaactttgaaaaactgtaaatg	65.2	24.2	33	10191.8
ESIINFEKL Lower	ttacagtttttcaaagttaataatgctttccat	66.6	24.2	33	10066.7
EESIINFEKL Upper	gaagaaagcattattaactttgaaaaactgtaaatg	67.2	25	36	11147.4
EESIINFEKL Lower	ttacagtttttcaaagttaataatgctttcttccat	68.5	25	36	10964.3
PSIINFEKL Upper	ccgagcattattaactttgaaaaactgtaaatg	68.6	30.3	33	10143.7
PSIINFEKL Lower	ttacagtttttcaaagttaataatgctcggcat	70	30.3	33	10116.7
PPSIINFEKL Upper	ccgccgagcattattaactttgaaaaactgtaaatg	74.3	36.1	36	11051.3
PPSIINFEKL Lower	ttacagtttttcaaagttaataatgctcggcggcat	75.6	36.1	36	11064.3
LSIINFEKL Upper	ctgagcattattaactttgaaaaactgtaaatg	66	27.3	33	10158.7
LSIINFEKL Lower	ttacagtttttcaaagttaataatgctcagcat	67.5	27.3	33	10100.7
LLSIINFEKL Upper	ctgctgagcattattaactttgaaaaactgtaaatg	69.9	30.6	36	11081.3
LLSIINFEKL Lower	ttacagtttttcaaagttaataatgctcagcagcat	71.2	30.6	36	11032.3
KSIINFEKL Upper	aaaagcattattaactttgaaaaactgtaaatg	64.8	21.2	33	10175.8
KSIINFEKL Lower	ttacagtttttcaaagttaataatgcttttcat	65.2	21.2	33	10081.7
KKSIINFEKL Upper	aaaaaaagcattattaactttgaaaaactgtaaatg	66.6	19.4	36	11115.4
KKSIINFEKL Lower	ttacagtttttcaaagttaataatgctttttttcat	67	19.4	36	10994.3

Appendix 5 – ERAP1 plasmid map



Appendix 6 – ERAP1 amino acid sequence

human ERAP1
RSPWQSTEAS PKRSDGTPFP WNKIRLPEYV IPVHYDLLIH ANLTTLTFWG TTKIEITASQ 60
human ERAP1
PTSTIILHSH HLQLSRATLR KGAGERPSEE PLQVLEHPRQ EQIALLAPEP LLVGLPYTVV 120
human ERAP1
IHYAGNLSET FHGFYKSTYR TKEGELRILA STQFEPTAAR MAFPCFDEPA SKASFSIKIR 180
human ERAP1
REPRHLAISN MPLVKSVTVA EGLIEDHFDV TVKMSTYLVA FIISDFESVS KITKSGVKVS 240
human ERAP1
VYAVPDKINQ ADYALDAAVT LLEFYEDYFS IPYPLPKQDL AAIPDFQSGA MENWGLTTYR 300
human ERAP1
ESALLFDAEK SSASSKLGIT MTVAHELAHQ WFGNLVTMEW WNDLWLNEGF AKFMEFVSVS 360
human ERAP1
VTHPELKVGD YFFGKCFDAM EVDALNSSHP VSTPVENPAQ IREMFDDVSY DKGACILNML 420
human ERAP1
REYLSADAFK SGIVQYLQKH SYKNTKNEDL WDSMASICPT DGVKGMDGFC SRSQHSSSSS 480
human ERAP1
HWHQEGVDVK TMMNTWTLQK GFPLITITVR GRNVHMKQEH YMKGSDGAPD TGYLWHVPLT 540
human ERAP1
FITSKSDMVH RFSLKTKTDV LILPEXVEWI KFNVGMNGYY IVHYEDDGWD SLTGLLKGTH 600
human ERAP1
TAVSSNDRAS LINNAFQLVS IGKLSIEKAL DLSLYLKHET EIMPVFQGLN ELIPMYKLME 660
human ERAP1
KRDMNEVETQ FKXFLIRLLR DLIDKQTWTD EGSVSERMLR SQLLLLACAH NYQPCVQRAE 720
human ERAP1
GYFRKWKESN GNLSLPVDVT LAVFAVGAQS TEGWDFLYSK YQFSLSSTEK SQIEFALCRT 780
human ERAP1
QNKEKLQWLL DESFKGDKIK TQEFPQILTL IGRNPVGYPL AWQFLRKNWN KLVQKFELGS 840
human ERAP1
SSIAHVVMGT TNQFSTRTRL EEVKGFFSSL KENGSQLRCV QQTIETIEEN IGWMDKNFDK 900
human ERAP1 TEV site Twin-Strep tag
IRVWLQSEKL ERMLEENLYF QGSAWSHPQF EKGGGSGGGS GGSAWSHPQF EK 952

Appendix 7 - TOPS screen (Bulek et al. 2012)

	1	2	3	4	5	6	7	8	9	10	11	12
A	PEG 4K 15 w/v Glycerol 15 v/v cacodylat..0.1 M (6)	PEG 4K 20 w/v cacodylat..0.1 M (6) Glycerol 15 v/v	PEG 4K 25 w/v cacodylat..0.1 M (6) Glycerol 15 v/v	PEG 8K 15 w/v Glycerol 15 v/v cacodylat..0.1 M (6)	PEG 8K 20 w/v Glycerol 15 v/v cacodylat..0.1 M (6)	PEG 8K 25 w/v cacodylat..0.1 M (6) Glycerol 15 v/v	PEG 4K 15 w/v cacodylat..0.1 M (6) AS 0.2 M	PEG 4K 20 w/v AS 0.2 M cacodylat..0.1 M (6)	PEG 4K 25 w/v AS 0.2 M cacodylat..0.1 M (6)	PEG 8K 15 w/v cacodylat..0.1 M (6) AS 0.2 M	PEG 8K 20 w/v AS 0.2 M cacodylat..0.1 M (6)	PEG 8K 25 w/v cacodylat..0.1 M (6) AS 0.2 M
B	PEG 4K 15 w/v Glycerol 15 v/v cacodyl..0.1 M (6.5)	PEG 4K 20 w/v cacodyl..0.1 M (6.5) Glycerol 15 v/v	PEG 4K 25 w/v cacodyl..0.1 M (6.5) Glycerol 15 v/v	PEG 8K 15 w/v cacodyl..0.1 M (6.5) Glycerol 15 v/v	PEG 8K 20 w/v cacodyl..0.1 M (6.5) Glycerol 15 v/v	PEG 8K 25 w/v cacodyl..0.1 M (6.5) Glycerol 15 v/v	PEG 4K 15 w/v cacodyl..0.1 M (6.5) AS 0.2 M	PEG 4K 20 w/v cacodyl..0.1 M (6.5) AS 0.2 M	PEG 4K 25 w/v cacodyl..0.1 M (6.5) AS 0.2 M	PEG 8K 15 w/v cacodyl..0.1 M (6.5) AS 0.2 M	PEG 8K 20 w/v cacodyl..0.1 M (6.5) AS 0.2 M	PEG 8K 25 w/v cacodyl..0.1 M (6.5) AS 0.2 M
C	PEG 4K 15 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 4K 20 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 4K 25 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 8K 15 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 8K 20 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 8K 25 w/v MES 0.1 M (7) Glycerol 15 v/v	PEG 4K 15 w/v MES 0.1 M (7) AS 0.2 M	PEG 4K 20 w/v MES 0.1 M (7) AS 0.2 M	PEG 4K 25 w/v MES 0.1 M (7) AS 0.2 M	PEG 8K 15 w/v MES 0.1 M (7) AS 0.2 M	PEG 8K 20 w/v MES 0.1 M (7) AS 0.2 M	PEG 8K 25 w/v MES 0.1 M (7) AS 0.2 M
D	PEG 4K 15 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 4K 20 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 4K 25 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 8K 15 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 8K 20 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 8K 25 w/v HEPES 0.1 M (7) Glycerol 15 v/v	PEG 4K 15 w/v HEPES 0.1 M (7) AS 0.2 M	PEG 4K 20 w/v HEPES 0.1 M (7) AS 0.2 M	PEG 4K 25 w/v HEPES 0.1 M (7) AS 0.2 M	PEG 8K 15 w/v HEPES 0.1 M (7) AS 0.2 M	PEG 8K 20 w/v HEPES 0.1 M (7) AS 0.2 M	PEG 8K 25 w/v HEPES 0.1 M (7) AS 0.2 M
E	PEG 4K 15 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 20 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 25 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 15 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 20 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 25 w/v HEPES 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 15 w/v HEPES 0.1 M (7.5) AS 0.2 M	PEG 4K 20 w/v HEPES 0.1 M (7.5) AS 0.2 M	PEG 4K 25 w/v HEPES 0.1 M (7.5) AS 0.2 M	PEG 8K 15 w/v HEPES 0.1 M (7.5) AS 0.2 M	PEG 8K 20 w/v HEPES 0.1 M (7.5) AS 0.2 M	PEG 8K 25 w/v HEPES 0.1 M (7.5) AS 0.2 M
F	PEG 4K 15 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 20 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 25 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 15 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 20 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 8K 25 w/v Tris 0.1 M (7.5) Glycerol 15 v/v	PEG 4K 15 w/v Tris 0.1 M (7.5) AS 0.2 M	PEG 4K 20 w/v Tris 0.1 M (7.5) AS 0.2 M	PEG 4K 25 w/v Tris 0.1 M (7.5) AS 0.2 M	PEG 8K 15 w/v Tris 0.1 M (7.5) AS 0.2 M	AS 0.2 M PEG 8K 20 w/v Tris 0.1 M (7.5)	AS 0.2 M PEG 8K 25 w/v Tris 0.1 M (7.5)
G	PEG 4K 15 w/v Tris 0.1 M (8) Glycerol 15 v/v	PEG 4K 20 w/v Tris 0.1 M (8) Glycerol 15 v/v	PEG 4K 25 w/v Tris 0.1 M (8) Glycerol 15 v/v	PEG 8K 15 w/v Tris 0.1 M (8) Glycerol 15 v/v	PEG 8K 20 w/v Tris 0.1 M (8) Glycerol 15 v/v	PEG 8K 25 w/v Tris 0.1 M (8) Glycerol 15 v/v	AS 0.2 M PEG 4K 15 w/v Tris 0.1 M (8)	AS 0.2 M PEG 4K 20 w/v Tris 0.1 M (8)	AS 0.2 M PEG 4K 25 w/v Tris 0.1 M (8)	AS 0.2 M PEG 8K 15 w/v Tris 0.1 M (8)	AS 0.2 M PEG 8K 20 w/v Tris 0.1 M (8)	AS 0.2 M PEG 8K 25 w/v Tris 0.1 M (8)
H	PEG 4K 15 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	PEG 4K 20 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	PEG 4K 25 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	PEG 8K 15 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	PEG 8K 20 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	PEG 8K 25 w/v Tris 0.1 M (8.5) Glycerol 15 v/v	AS 0.2 M PEG 4K 15 w/v Tris 0.1 M (8.5)	AS 0.2 M PEG 4K 20 w/v Tris 0.1 M (8.5)	AS 0.2 M PEG 4K 25 w/v Tris 0.1 M (8.5)	AS 0.2 M PEG 8K 15 w/v Tris 0.1 M (8.5)	AS 0.2 M PEG 8K 20 w/v Tris 0.1 M (8.5)	AS 0.2 M PEG 8K 25 w/v Tris 0.1 M (8.5)

Appendix 8 - TOPS2 screen (Bulek et al. 2012)

	1	2	3	4	5	6	7	8	9	10	11	12
A	cacodyl..0.1 M (6.5) PEG 4K 10 w/v AS 0.2 M	cacodyl..0.1 M (6.5) PEG 4K 12.5 w/v AS 0.2 M	cacodyl..0.1 M (6.5) PEG 4K 15 w/v AS 0.2 M	cacodyl..0.1 M (6.5) PEG 4K 17.5 w/v AS 0.2 M	cacodyl..0.1 M (6.5) PEG 4K 20 w/v AS 0.2 M	cacodyl..0.1 M (6.5) PEG 4K 22.5 w/v AS 0.2 M	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 10 w/v	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 12.5 w/v	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 15 w/v	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 17.5 w/v	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 20 w/v	cacodyl..0.1 M (6.5) AS 0.2 M PEG 8K 22.5 w/v
B	cacodyl..0.1 M (6.5) PEG 4K 10 w/v AS 0.1 M	cacodyl..0.1 M (6.5) PEG 4K 12.5 w/v AS 0.1 M	cacodyl..0.1 M (6.5) PEG 4K 15 w/v AS 0.1 M	cacodyl..0.1 M (6.5) PEG 4K 17.5 w/v AS 0.1 M	cacodyl..0.1 M (6.5) PEG 4K 20 w/v AS 0.1 M	cacodyl..0.1 M (6.5) PEG 4K 22.5 w/v AS 0.1 M	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 10 w/v	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 12.5 w/v	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 15 w/v	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 17.5 w/v	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 20 w/v	cacodyl..0.1 M (6.5) AS 0.1 M PEG 8K 22.5 w/v
C	MES 0.1 M (6.7) PEG 4K 10 w/v AS 0.2 M	MES 0.1 M (6.7) PEG 4K 12.5 w/v AS 0.2 M	MES 0.1 M (6.7) PEG 4K 15 w/v AS 0.2 M	MES 0.1 M (6.7) PEG 4K 17.5 w/v AS 0.2 M	MES 0.1 M (6.7) PEG 4K 20 w/v AS 0.2 M	MES 0.1 M (6.7) PEG 4K 22.5 w/v AS 0.2 M	MES 0.1 M (6.7) AS 0.2 M PEG 8K 10 w/v	MES 0.1 M (6.7) AS 0.2 M PEG 8K 12.5 w/v	MES 0.1 M (6.7) AS 0.2 M PEG 8K 15 w/v	MES 0.1 M (6.7) AS 0.2 M PEG 8K 17.5 w/v	MES 0.1 M (6.7) AS 0.2 M PEG 8K 20 w/v	MES 0.1 M (6.7) AS 0.2 M PEG 8K 22.5 w/v
D	HEPES 0.1 M (7) PEG 4K 10 w/v AS 0.2 M	HEPES 0.1 M (7) PEG 4K 12.5 w/v AS 0.2 M	HEPES 0.1 M (7) PEG 4K 15 w/v AS 0.2 M	HEPES 0.1 M (7) PEG 4K 17.5 w/v AS 0.2 M	HEPES 0.1 M (7) PEG 4K 20 w/v AS 0.2 M	HEPES 0.1 M (7) PEG 4K 22.5 w/v AS 0.2 M	HEPES 0.1 M (7) AS 0.2 M PEG 8K 10 w/v	HEPES 0.1 M (7) AS 0.2 M PEG 8K 12.5 w/v	HEPES 0.1 M (7) AS 0.2 M PEG 8K 15 w/v	HEPES 0.1 M (7) AS 0.2 M PEG 8K 17.5 w/v	HEPES 0.1 M (7) AS 0.2 M PEG 8K 20 w/v	HEPES 0.1 M (7) AS 0.2 M PEG 8K 22.5 w/v
E	HEPES 0.1 M (7.5) PEG 4K 10 w/v AS 0.2 M	HEPES 0.1 M (7.5) PEG 4K 12.5 w/v AS 0.2 M	HEPES 0.1 M (7.5) PEG 4K 15 w/v AS 0.2 M	HEPES 0.1 M (7.5) PEG 4K 17.5 w/v AS 0.2 M	HEPES 0.1 M (7.5) PEG 4K 20 w/v AS 0.2 M	HEPES 0.1 M (7.5) PEG 4K 22.5 w/v AS 0.2 M	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 10 w/v	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 12.5 w/v	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 15 w/v	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 17.5 w/v	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 20 w/v	HEPES 0.1 M (7.5) AS 0.2 M PEG 8K 22.5 w/v
F	PEG 4K 10 w/v AS 0.2 M bis-tris 0.1 M (6)	bis-tris 0.1 M (6) PEG 4K 12.5 w/v AS 0.2 M	bis-tris 0.1 M (6) PEG 4K 15 w/v AS 0.2 M	bis-tris 0.1 M (6) PEG 4K 17.5 w/v AS 0.2 M	bis-tris 0.1 M (6) PEG 4K 20 w/v AS 0.2 M	bis-tris 0.1 M (6) PEG 4K 22.5 w/v AS 0.2 M	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 10 w/v	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 12.5 w/v	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 15 w/v	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 17.5 w/v	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 20 w/v	bis-tris 0.1 M (6) AS 0.2 M PEG 8K 22.5 w/v
G	PEG 4K 10 w/v AS 0.2 M bis-tris 0.1 M (6.5)	PEG 4K 12.5 w/v AS 0.2 M bis-tris 0.1 M (6.5)	PEG 4K 15 w/v AS 0.2 M bis-tris 0.1 M (6.5)	PEG 4K 17.5 w/v AS 0.2 M bis-tris 0.1 M (6.5)	PEG 4K 20 w/v AS 0.2 M bis-tris 0.1 M (6.5)	PEG 4K 22.5 w/v AS 0.2 M bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 10 w/v bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 12.5 w/v bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 15 w/v bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 17.5 w/v bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 20 w/v bis-tris 0.1 M (6.5)	AS 0.2 M PEG 8K 22.5 w/v bis-tris 0.1 M (6.5)
H	bis-tris 0.1 M (7) PEG 4K 10 w/v AS 0.2 M	bis-tris 0.1 M (7) PEG 4K 12.5 w/v AS 0.2 M	bis-tris 0.1 M (7) PEG 4K 15 w/v AS 0.2 M	bis-tris 0.1 M (7) PEG 4K 17.5 w/v AS 0.2 M	bis-tris 0.1 M (7) PEG 4K 20 w/v AS 0.2 M	bis-tris 0.1 M (7) PEG 4K 22.5 w/v AS 0.2 M	bis-tris 0.1 M (7) AS 0.2 M PEG 8K 10 w/v	bis-tris 0.1 M (7) AS 0.2 M PEG 8K 12.5 w/v	bis-tris 0.1 M (7) AS 0.2 M PEG 8K 15 w/v	bis-tris 0.1 M (7) AS 0.2 M PEG 8K 17.5 w/v	bis-tris 0.1 M (7) AS 0.2 M PEG 8K 20 w/v	AS 0.2 M PEG 8K 22.5 w/v bis-tris 0.1 M (7)

Appendix 9 - SL8-SCT optimisation screen (Optimising TOPS screen hits A12 and B12)

Ammonium sulphate (mM)	PEG 8000 (%)			
	23	25	23	25
0	460µl PEG	500µl PEG	460µl PEG	500µl PEG
	75µl NaCa (pH 6.5)	75µl NaCa (pH 6.5)	100µl NaCa	100µl NaCa 0 µl
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	0 µl AS 440µl H_2O	AS 400µl H_2O
	0 µl AS 440µl H_2O	0 µl AS 400µl H_2O		
100	460µl PEG	500µl PEG	460µl PEG	500µl PEG
	75µl NaCa (pH 6.5)	75µl NaCa (pH 6.5)	100µl NaCa	100µl NaCa
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	25 µl AS 415µl H_2O	25 µl AS 375µl H_2O
	25 µl AS 415µl H_2O	25 µl AS 375µl H_2O		
200	460µl PEG	500µl PEG	460µl PEG	500µl PEG
	75µl NaCa (pH 6.5)	75µl NaCa (pH 6.5)	100µl NaCa	100µl NaCa
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	50 µl AS 390µl H_2O	50 µl AS 350µl H_2O
	50 µl AS 390µl H_2O	50 µl AS 350µl H_2O		
	6.3	6.3	6.5	6.5
	100 mM sodium cacodylate (pH)			

Appendix 10 - SL8-SCT optimisation screen (Optimising TOPS screen hits A12 and B12)

Ammonium sulphate (mM)	PEG 8000 (%) 27	25	27	25
50	540µl PEG	500µl PEG	540µl PEG	500µl PEG
	75µl NaCa (pH 6.5)	75µl NaCa (pH 6.5)	100µl NaCa	100µl NaCa 12 µl AS 388µl H_20
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	12 µl AS 348µl H_20	
	12 µl AS 348µl H_20	12 µl AS 388µl H_20		
100	540µl PEG	500µl PEG	540µl PEG	500µl PEG
	75µl NaCa (pH 6.5)	75µl NaCa (pH 6.5)	100µl NaCa	100µl NaCa
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	25 µl AS 335µl H_20	25 µl AS 375µl H_20
	25 µl AS 335µl H_20	25 µl AS 375µl H_20		
200	540µl PEG	500µl PEG	540µl PEG	500µl PEG
	75µl NaAc (pH 6.5)	75µl NaAc (pH 6.5)	100µl NaAc	100µl NaAc
	25µl NaCa (pH 5.5)	25µl NaCa (pH 5.5)	50 µl AS 310µl H_20	50 µl AS 350µl H_20
	50 µl AS 310µl H_20	50 µl AS 350µl H_20		
100	540µl PEG	500µl PEG	540µl PEG	500µl PEG
	100µl Bis Tris	100µl Bis Tris	100µl sodium HEPES	100µl sodium HEPES
	25 µl AS 335µl H_20	25 µl AS 375µl H_20	25 µl AS 335µl H_20	25 µl AS 375µl H_20
	pH 6.8	pH 6.8	pH 6.8	pH 6.8
	6.3	6.3	6.5	6.5
	100 mM sodium cacodylate (pH)			

Vary buffers – sodium HEPES (pH 6.8) and Bis Tris (pH 6.8)

Appendix 11 – Potential hydrogen bonds of dtSCT mutants (calculated by WhatIf)

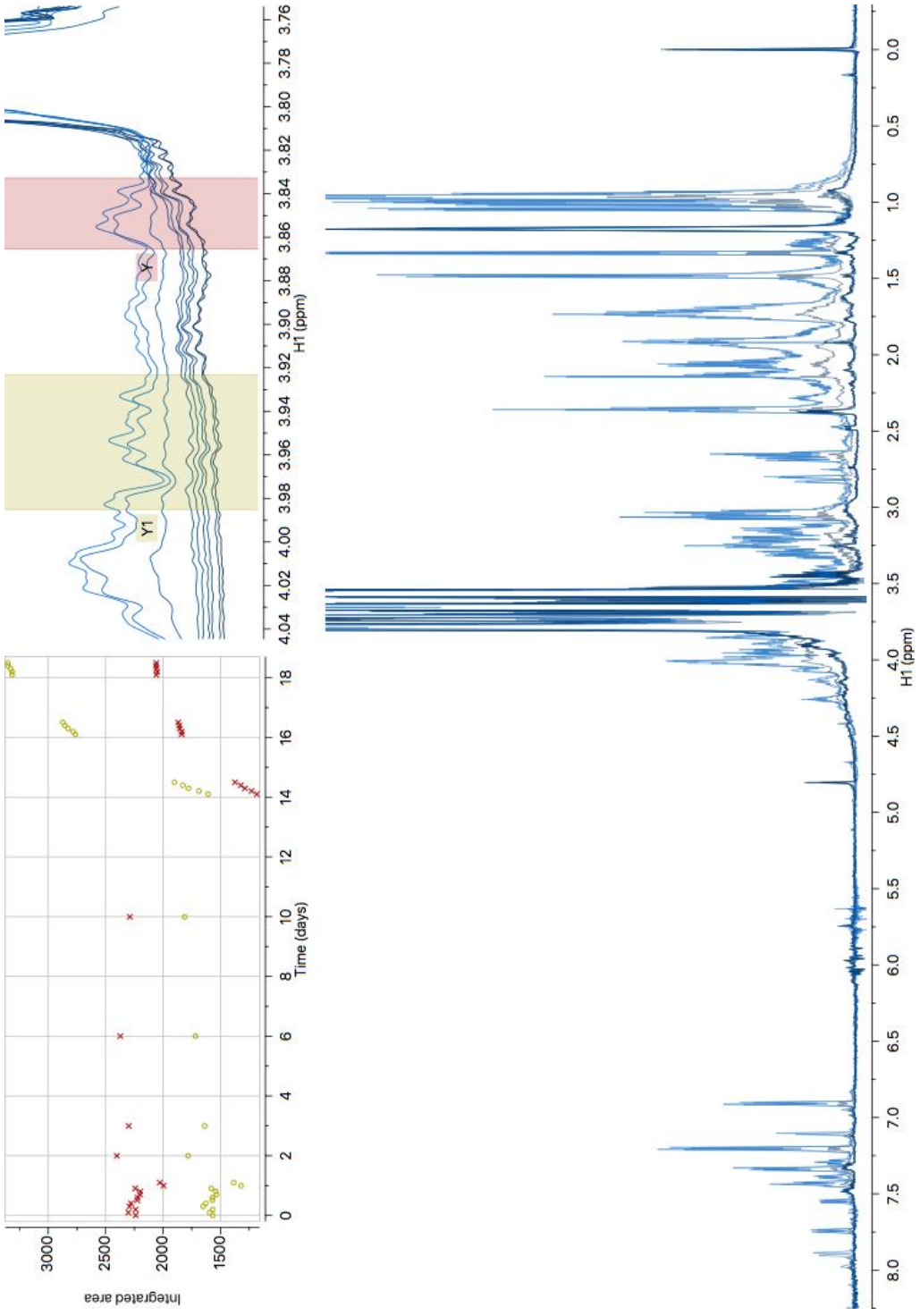
construct	Donor residue	Donor ID	Chain	Donor	Acceptor Residue	Acceptor ID	Chain	Acceptor	D->A distance	H->A distance	DHA max. angular error	HA max. angular error	Hydrogen coordinates
SLB-SCT	1 SER	1	A	N	283 TYR	171	A	OH	2.83	1.87	20	60.7	2.44 15.69 34.50
	271 TYR	159	A	OH	1 SER	1	A	O	2.73	1.75	15.8	53.6	2.42 20.81 35.30
	1 SER	1	A	OG	175 GLU	63	A	OE2	2.93	1.93	4	74.9	-1.61 16.32 34.23
	1 SER	1	A	OG	175 GLU	63	A	OE2	2.93	2.15	45.6	43.9	-2.63 15.90 33.71
	178 LYS	66	A	NZ	1 SER	1	A	OG	3.05	2.05	3.3	48.6	-1.97 18.27 33.50
	1 SER	1	A	OG	776 HOH	403	C	O	3.45	2.47	14.5	0	-1.71 17.65 35.33
	1 SER	1	A	OG	776 HOH	403	C	O	3.45	2.45	0	70.3	-2.15 18.71 36.26
	389 SER	1	B	N	670 TYR	171	B	OH	2.87	2.01	37.4	62.9	-12.27 -28.18 35.84
	658 TYR	159	B	OH	389 SER	1	B	O	2.29	1.29	3.9	35.8	-12.34 -23.24 36.19
	389 SER	1	B	OG	562 GLU	63	B	OE2	2.33	1.36	18.6	61.2	-16.07 -27.70 35.03
	389 SER	1	B	OG	562 GLU	63	B	OE2	2.33	1.51	43.9	32.2	-16.77 -27.63 34.70
	565 LYS	66	B	NZ	389 SER	1	B	OG	3.23	2.32	30.6	60.3	-16.30 -25.60 33.86
	389 SER	1	B	OG	776 HOH	366	C	O	2.79	1.95	40.1	0	-16.64 -26.75 36.23
	389 SER	1	B	OG	776 HOH	366	C	O	2.79	1.79	0	86.7	-16.65 -26.49 37.20
	2 ILE	2	A	N	776 HOH	1	C	O	3.26	2.26	6.6	0	0.76 18.68 32.15
	178 LYS	66	A	NZ	2 ILE	2	A	O	2.39	1.4	6.4	42.8	-1.89 19.77 33.18
	390 ILE	2	B	N	776 HOH	33	C	O	3.13	2.19	23.6	0	-13.99 -25.29 33.14
	565 LYS	66	B	NZ	390 ILE	2	B	O	2.51	1.54	16	48.4	-16.37 -24.34 33.60
	3 ILE	3	A	N	776 HOH	2	C	O	3.05	2.08	17.2	0	0.99 23.10 33.52
	182 ASN	70	A	ND2	3 ILE	3	A	O	3.13	2.17	19.1	39.3	-2.71 23.48 30.18
	391 ILE	3	B	N	776 HOH	277	C	O	3.08	2.09	9.5	0	-13.54 -20.92 34.58
	569 ASN	70	B	ND2	391 ILE	3	B	O	3.19	2.23	19.4	34.5	-17.12 -20.76 30.83

	1 SER	1	A	N	283 TYR	171	A	OH	2.84	1.89	24.2	66.9	2.32 38.08 34.44
	271 TYR	159	A	OH	1 SER	1	A	O	2.83	1.85	14.2	46	2.53 43.28 35.15
	389 SER	1	B	N	671 TYR	171	B	OH	2.76	1.97	46.5	53.9	-12.52 -5.88 35.65
	659 TYR	159	B	OH	389 SER	1	B	O	2.51	1.52	2.8	29.7	-12.17 -0.80 36.03
	2 ILE	2	A	N	777 HOH	166	S	O	2.96	1.98	14.9	0	0.32 41.07 32.09
	178 LYS	66	A	NZ	2 ILE	2	A	O	2.72	1.89	41.4	19.7	-3.16 43.14 34.18
	390 ILE	2	B	N	777 HOH	173	S	O	2.76	1.83	27	0	-14.34 -2.95 33.04
	566 LYS	66	B	NZ	390 ILE	2	B	O	2.92	2.21	53.2	32.3	-18.07 -0.54 34.77
	3 ILE	3	A	N	777 HOH	70	S	O	3.18	2.21	17.8	0	0.90 45.48 33.41
	182 ASN	70	A	ND2	3 ILE	3	A	O	3.02	2.07	22.4	38.3	-2.78 45.97 30.21
	391 ILE	3	B	N	777 HOH	104	S	O	3.11	2.13	15	0	-13.65 1.50 34.18
E63A	570 ASN	70	B	ND2	391 ILE	3	B	O	3.07	2.07	7.6	21.6	-16.77 2.11 30.71
	1 SER	1P	A	N	283 TYR	171H	A	OH	2.8	1.83	17	62.6	2.37 15.40 34.47
	271 TYR	159H	A	OH	1 SER	1P	A	O	2.63	1.64	5.7	38.2	2.42 20.44 35.29
	1 SER	1P	A	OG	175 GLU	63H	A	OE2	2.71	1.78	25.8	63.3	-1.82 16.61 34.29
	1 SER	1P	A	OG	175 GLU	63H	A	OE2	2.71	1.85	37.3	63.8	-2.31 16.11 33.80
	1 SER	1P	A	OG	777 HOH	83	S	O	3.04	2.16	34.9	0	-1.35 17.54 34.04
	1 SER	1P	A	OG	777 HOH	83	S	O	3.04	2.04	0	35.9	-2.28 18.11 33.54
	1 SER	1P	A	OG	777 HOH	108	S	O	3.12	2.12	1.5	0	-1.42 17.95 35.32
	1 SER	1P	A	OG	777 HOH	108	S	O	3.12	2.12	0	61	-1.66 18.93 35.80
	389 SER	1P	B	N	671 TYR	171H	B	OH	2.78	1.86	29.2	62.4	-12.41 -28.29 35.60
	659 TYR	159H	B	OH	389 SER	1P	B	O	2.55	1.55	1.3	34.1	-12.44 -23.19 36.22
	389 SER	1P	B	OG	563 GLU	63H	B	OE2	2.54	1.54	5.1	60.8	-16.41 -27.49 35.05
	389 SER	1P	B	OG	563 GLU	63H	B	OE2	2.54	1.7	41.4	48.1	-17.00 -27.73 34.65
K66A	389 SER	1P	B	OG	777 HOH	477	S	O	3.02	2.13	33.5	0	-16.14 -26.13 35.03

K66A	2 ILE	2P	A	N	777 HOH	2	S	O	2.91	1.95	20	0	0.44 18.36 32.06
	777 HOH	83	S	O	2 ILE	2P	A	O	2.69	1.69	0	42.5	-2.23 19.37 33.18
	777 HOH	108	S	O	2 ILE	2P	A	O	2.75	1.75	0	53.5	-1.65 20.11 35.30
	390 ILE	2P	B	N	777 HOH	70	S	O	2.97	2.03	25.1	0	-14.32 -25.39 33.07
	777 HOH	477	S	O	390 ILE	2P	B	O	2.65	1.65	0	38	-16.98 -24.33 34.06
	3 ILE	3P	A	N	777 HOH	1	S	O	2.93	1.97	21.1	0	1.02 22.78 33.36
	182 ASN	70H	A	ND2	3 ILE	3P	A	O	3.18	2.24	22.5	35	-2.62 23.24 30.07
	391 ILE	3P	B	N	777 HOH	73	S	O	2.96	1.96	5.9	0	-13.66 -20.98 34.36
	570 ASN	70H	B	ND2	391 ILE	3P	B	O	3.19	2.24	23.5	33.5	-17.39 -20.68 30.84
W167A	1 SER	1p	A	N	283 TYR	171H	A	OH	2.82	1.85	18.5	62.2	2.44 15.28 34.54
	271 TYR	159H	A	OH	1 SER	1p	A	O	2.51	1.52	12.8	44.2	2.46 20.26 35.20
	1 SER	1p	A	OG	175 GLU	63H	A	OE2	2.96	1.99	18	67.3	-1.67 16.36 34.25
	1 SER	1p	A	OG	175 GLU	63H	A	OE2	2.96	2.06	31.7	62.3	-2.26 15.68 33.68
	1 SER	1p	A	OG	777 HOH	256S		O	3.18	2.18	4.3	0	-1.43 17.80 35.19
	1 SER	1p	A	OG	777 HOH	256S		O	3.18	2.18	0	62.9	-1.71 18.86 35.62
	1 SER	1p	A	OG	777 HOH	795S		O	3.1	2.35	49	0	-1.89 16.71 35.47
	389 SER	1p	B	N	671 TYR	171H	B	OH	2.87	1.91	21.9	68.4	-11.95 -28.21 35.74
	659 TYR	159H	B	OH	389 SER	1p	B	O	2.4	1.4	1	36.8	-12.25 -23.35 36.24
	563 GLU	63H	B	OE1	389 SER	1p	B	OG	3.13	2.48	57.8	58.1	-16.06 -26.69 33.18
	389 SER	1p	B	OG	563 GLU	63H	B	OE2	2.59	1.59	4.3	72.4	-16.09 -27.66 35.03
	389 SER	1p	B	OG	563 GLU	63H	B	OE2	2.59	1.74	39.2	50.5	-16.72 -28.05 34.77
	566 LYS	66H	B	NZ	389 SER	1p	B	OG	3.31	2.46	38.1	55.5	-16.13 -25.59 33.63
	389 SER	1p	B	OG	777 HOH	327S		O	2.74	1.75	13	0	-16.39 -26.74 36.23
	389 SER	1p	B	OG	777 HOH	327S		O	2.74	1.74	0	68.3	-16.77 -26.59 36.87

W167A	2 ILE	2p	A	N	777 HOH	1S		O	3	2.01	8.6	0	0.72 18.23 32.07
	178 LYS	66H	A	NZ	2 ILE	2p	A	O	2.62	1.68	25.8	36.3	-1.94 18.94 33.58
	777 HOH	256S		O	2 ILE	2p	A	O	2.63	1.63	0	53.4	-1.67 20.01 35.05
	390 ILE	2p	B	N	777 HOH	96S		O	2.93	1.95	12.8	0	-13.76 -25.43 33.01
	566 LYS	66H	B	NZ	390 ILE	2p	B	O	2.57	1.64	27.4	60.8	-16.17 -24.44 33.49
	777 HOH	385S		O	390 ILE	2p	B	O	2.61	1.61	0	49.4	-16.33 -23.48 35.86
	3 ILE	3p	A	N	777 HOH	3S		O	3.05	2.13	27.6	0	0.96 22.71 33.24
	182 ASN	70H	A	ND2	3 ILE	3p	A	O	3.13	2.2	25.9	37.6	-2.82 22.99 30.06
	391 ILE	3p	B	N	777 HOH	97S		O	2.99	2.02	17.1	0	-13.46 -20.96 34.25
	570 ASN	70H	B	ND2	391 ILE	3p	B	O	3.2	2.26	25.3	35.2	-17.18 -20.86 30.82

Appendix 12 – NMR reaction monitoring of E63A dtSCT degradation in presence of wt ERAP1



Monitoring for degradation of dtSCT mutant E63A by wild-type ERAP1. The reaction was tracked over 18 days (shown from dark to light blue) by proton NMR. Trimming of N-terminal serine from E63A was monitored. No L-serine was detected until peaks began to appear across the spectrum, indicating significant protein

Peptides in 90% H_2O, 10% D_2O

SIINFEKL	**S1**	**NH**		**Hα**		**Hβ1**	*3.90?*	**Hβ2**	*3.78?*	**Hγ**												
	I2	**NH**		**Hα**	4.15	**Hβ**	1.79	**Hγ1a**	1.11	**Hγ1b**	1.40	**Hγ2**	0.81	**Hδ**	0.80							
	I3	**NH**	8.28	**Hα**	4.02	**Hβ**	1.67	**Hγ1a**	1.06	**Hγ1b**	1.37	**Hγ2**	0.76	**Hδ**	0.65							
	N4	**NH**	8.44	**Hα**	4.62	**Hβ1**	2.65	**Hβ2**	2.65	**Hδ1**	6.92	**Hδ2**	7.61									
	F5	**NH**	8.32	**Hα**	4.49	**Hβ1**	2.94	**Hβ2**	3.09	**Hδ**	7.19	**Hε**	7.29	**Hζ**	7.24							
	E6	**NH**	8.33	**Hα**	4.15	**Hβ1**	1.92	**Hβ2**	1.85	**Hγ1**	2.13	**Hγ2**	2.13									
	K7	**NH**	8.23	**Hα**	4.24	**Hβ1**	1.79	**Hβ2**	1.79	**Hγ1**	1.37	**Hγ2**	1.37	**Hδ1**	1.64	**Hδ2**	1.64	**Hε1**	2.94	**Hε2**	2.94	**Hζ**	
	L8	**NH**	7.96	**Hα**	4.13	**Hβ1**	1.52	**Hβ2**	1.52	**Hγ**	1.52	**Hδ1**	0.81	**Hδ2**	0.81							
ESIINFEKL	**E1**	**NH**		**Hα**	*4.00?*	**Hβ1**	*2.04?*	**Hβ2**	*2.04?*	**Hγ1**	*2.31?*	**Hγ2**	*2.31?*									
	S2	**NH**		**Hα**	4.50	**Hβ1**	3.78	**Hβ2**	3.78	**Hγ**												
	I3	**NH**	8.46	**Hα**	4.15	**Hβ**	1.80	**Hγ1a**	1.13	**Hγ1b**	1.43	**Hγ2**	0.83	**Hδ**	0.82							
	I4	**NH**	8.31	**Hα**	4.04	**Hβ**	1.69	**Hγ1a**	1.07	**Hγ1b**	1.38	**Hγ2**	0.78	**Hδ**	0.67							
	N5	**NH**	8.47	**Hα**	4.65	**Hβ1**	2.61	**Hβ2**	2.68	**Hδ1**	6.95	**Hδ2**	7.63									
	F6	**NH**	8.34	**Hα**	4.51	**Hβ1**	2.97	**Hβ2**	3.12	**Hδ**	7.22	**Hε**	7.32	**Hζ**	7.27							
	E7	**NH**	8.36	**Hα**	4.17	**Hβ1**	1.87	**Hβ2**	1.87	**Hγ1**	2.15	**Hγ2**	2.15									
	K8	**NH**	8.25	**Hα**	4.27	**Hβ1**	1.83	**Hβ2**	1.83	**Hγ1**	1.40	**Hγ2**	1.40	**Hδ1**	1.68	**Hδ2**	1.68	**Hε1**	2.96	**Hε2**	2.96	**Hζ**	
	L9	**NH**	7.98	**Hα**	4.15	**Hβ1**	1.54	**Hβ2**	1.54	**Hγ**	1.54	**Hδ1**	0.84	**Hδ2**	0.84							
KSIINFEKL	**K1**	**NH**		**Hα**	4.06	**Hβ1**	1.92	**Hβ2**	1.92	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.70	**Hδ2**	1.70	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.61
	S2	**NH**	8.87	**Hα**	4.56	**Hβ1**	3.84	**Hβ2**	3.84	**Hγ**												
	I3	**NH**	8.55	**Hα**	4.19	**Hβ**	1.86	**Hγ1a**	1.19	**Hγ1b**	1.46	**Hγ2**	0.87	**Hδ**	0.86							
	I4	**NH**	8.28	**Hα**	4.09	**Hβ**	1.74	**Hγ1a**	1.12	**Hγ1b**	1.42	**Hγ2**	0.83	**Hδ**	0.73							
	N5	**NH**	8.53	**Hα**	4.68	**Hβ1**	2.68	**Hβ2**	2.76	**Hδ1**	6.94	**Hδ2**	7.65									

KSIINFEKL	**F6**	**NH**	8.36	**Hα**	4.54	**Hβ1**	3.02	**Hβ2**	3.13	**Hδ**	7.24	**Hε**	7.35	**Hζ**	7.30									
	E7	**NH**	8.25	**Hα**	4.27	**Hβ1**	1.91	**Hβ2**	2.01	**Hγ1**	2.34	**Hγ2**	2.34											
	K8	**NH**	8.31	**Hα**	4.27	**Hβ1**	1.85	**Hβ2**	1.85	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.59
	L9	**NH**	8.20	**Hα**	4.26	**Hβ1**	1.61	**Hβ2**	1.61	**Hγ**	1.61	**Hδ1**	0.89	**Hδ2**	0.89								
PSIINFEKL	**P1**	**NH**		**Hα**	4.46	**Hβ1**	2.49	**Hβ2**	2.49	**Hγ1**	2.08	**Hγ2**	2.08	**Hδ1**	3.43	**Hδ2**	3.43							
	S2	**NH**	8.87	**Hα**	4.54	**Hβ1**	3.85	**Hβ2**	3.85	**Hγ**														
	I3	**NH**	8.46	**Hα**	4.19	**Hβ**	1.85	**Hγ1a**	1.19	**Hγ1b**	1.48	**Hγ2**	0.88	**Hδ**	0.78									
	I4	**NH**	8.34	**Hα**	4.10	**Hβ**	1.75	**Hγ1a**	1.12	**Hγ1b**	1.43	**Hγ2**	0.84	**Hδ**	0.73									
	N5	**NH**	8.52	**Hα**	4.69	**Hβ1**	2.69	**Hβ2**	2.78	**Hδ1**	7.68	**Hδ2**	7.00											
	F6	**NH**	8.39	**Hα**	4.57	**Hβ1**	3.02	**Hβ2**	3.17	**Hδ**	7.27	**Hε**	7.36	**Hζ**	7.32									
	E7	**NH**	8.40	**Hα**	4.23	**Hβ1**	2.00	**Hβ2**	1.92	**Hγ1**	2.21	**Hγ2**	2.21											
	K8	**NH**	8.31	**Hα**	4.32	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.74	**Hδ2**	1.74	**Hε1**	3.02	**Hε2**	3.02	**Hζ**	
	L9	**NH**	8.04	**Hα**	4.21	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.87	**Hδ2**	0.92									
EESIINFEKL	**E1**	**NH**		**Hα**	4.10	**Hβ1**	2.16	**Hβ2**	2.16	**Hγ1**	2.53	**Hγ2**	2.53											
	E2	**NH**	8.39	**Hα**	4.46	**Hβ1**	2.10	**Hβ2**	2.01	**Hγ1**	2.48	**Hγ2**	2.48											
	S3	**NH**	8.65	**Hα**	4.47	**Hβ1**	3.84	**Hβ2**	3.84	**Hγ**														
	I4	**NH**	8.38	**Hα**	4.19	**Hβ**	1.86	**Hγ1a**	1.45	**Hγ1b**	1.18	**Hγ2**	0.87	**Hδ**	0.86									
	I5	**NH**	8.25	**Hα**	4.07	**Hβ**	1.75	**Hγ1a**	1.43	**Hγ1b**	1.14	**Hγ2**	0.83	**Hδ**	0.74									
	N6	**NH**	8.45	**Hα**	4.67	**Hβ1**	2.76	**Hβ2**	2.68	**Hδ1**	6.94	**Hδ2**	7.65											
	F7	**NH**	8.30	**Hα**	4.53	**Hβ1**	3.12	**Hβ2**	3.04	**Hδ**	7.34	**Hε**	7.30	**Hζ**	7.25									
	E8	**NH**	8.21	**Hα**	4.27	**Hβ1**	1.92	**Hβ2**	2.03	**Hγ1**	2.38	**Hγ2**	2.38											
	K9	**NH**	8.28	**Hα**	4.27	**Hβ1**	1.85	**Hβ2**	1.85	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.58
	L10	**NH**	8.36	**Hα**	4.34	**Hβ1**	1.65	**Hβ2**	1.65	**Hγ**	1.65	**Hδ1**	0.86	**Hδ2**	0.92									

KKSIINFEKL	**K1**	**NH**		**Hα**	4.02	**Hβ1**	1.90	**Hβ2**	1.90	**Hγ1**	1.44	**Hγ2**	1.44	**Hδ1**	1.69	**Hδ2**	1.69	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.59
	K2	**NH**	8.83	**Hα**	4.38	**Hβ1**	1.79	**Hβ2**	1.79	**Hγ1**	1.44	**Hγ2**	1.44	**Hδ1**	1.69	**Hδ2**	1.69	**Hε1**	2.99	**Hε2**	2.99	**Hζ**	7.59
	S3	**NH**	8.68	**Hα**	4.48	**Hβ1**	3.83	**Hβ2**	3.83	**Hγ**													
	I4	**NH**	8.42	**Hα**	4.20	**Hβ**	1.86	**Hγ1a**	1.18	**Hγ1b**	1.44	**Hγ2**	0.86	**Hδ**	0.86								
	I5	**NH**	8.28	**Hα**	4.10	**Hβ**	1.74	**Hγ1a**	1.12	**Hγ1b**	1.42	**Hγ2**	0.83	**Hδ**	0.73								
	N6	**NH**	8.52	**Hα**	4.68	**Hβ1**	2.68	**Hβ2**	2.76	**Hδ1**	6.95	**Hδ2**	7.65										
	F7	**NH**	8.36	**Hα**	4.55	**Hβ1**	3.13	**Hβ2**	3.03	**Hδ**	7.24	**Hε**	7.35	**Hζ**	7.30								
	E8	**NH**	8.25	**Hα**	4.27	**Hβ1**	2.01	**Hβ2**	1.91	**Hγ1**	2.33	**Hγ2**	2.33										
	K9	**NH**	8.32	**Hα**	4.27	**Hβ1**	1.86	**Hβ2**	1.85	**Hγ1**	1.44	**Hγ2**	1.45	**Hδ1**	1.72	**Hδ2**	1.73	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.59
	L10	**NH**	8.18	**Hα**	4.25	**Hβ1**	1.61	**Hβ2**	1.61	**Hγ**	1.61	**Hδ1**	0.88	**Hδ2**	0.90								
PPSIINFEKL	**P1**	***NH***	*9.12*	**Hα**	4.64	**Hβ1**	2.57	**Hβ2**	2.57	**Hγ1**	2.06	**Hγ2**	2.06	**Hδ1**	3.41	**Hδ2**	3.41						
		NH	*8.33*																				
	P2			**Hα**	4.52	**Hβ1**	2.36	**Hβ2**	2.36	**Hγ1**	1.91	**Hγ2**	2.03	**Hδ1**	3.60	**Hδ2**	3.72						
	S3	**NH**	8.65	**Hα**	4.44	**Hβ1**	3.84	**Hβ2**	3.84	**Hγ**													
	I4	**NH**	8.36	**Hα**	4.19	**Hβ**	1.85	**Hγ1a**	1.17	**Hγ1b**	1.44	**Hγ2**	0.86	**Hδ**	0.86								
	I5	**NH**	8.29	**Hα**	4.08	**Hβ**	1.74	**Hγ1a**	1.13	**Hγ1b**	1.43	**Hγ2**	0.84	**Hδ**	0.73								
	N6	**NH**	8.49	**Hα**	4.68	**Hβ1**	2.68	**Hβ2**	2.77	**Hδ1**	6.94	**Hδ2**	7.65										
	F7	**NH**	8.32	**Hα**	4.54	**Hβ1**	3.13	**Hβ2**	3.03	**Hδ**	7.35	**Hε**	7.31	**Hζ**	7.24								
	E8	**NH**	8.23	**Hα**	4.27	**Hβ1**	2.02	**Hβ2**	1.92	**Hγ1**	2.36	**Hγ2**	2.36										
	K9	**NH**	8.29	**Hα**	4.27	**Hβ1**	1.85	**Hβ2**	1.85	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.00	**Hε2**	3.00	**Hζ**	7.58
	L10	**NH**	8.29	**Hα**	4.29	**Hβ1**	1.63	**Hβ2**	1.63	**Hγ**	1.63	**Hδ1**	0.90	**Hδ2**	0.90								

Appendix 14 – Peptide assignments in 100% D_2O

Peptides in 20 mM potassium phosphate, pD 7, 100% D_2O

SIINFEKL	**S1**	**Hα**	4.15	**Hβ1**	3.91	**Hβ2**	3.97														
	I2	**Hα**	4.23	**Hβ**	1.85	**Hγ1a**	1.19	**Hγ1b**	1.47	**Hγ2**	0.89	**Hδ**	0.89								
	I3	**Hα**	4.07	**Hβ**	1.73	**Hγ1a**	1.11	**Hγ1b**	1.43	**Hγ2**	0.73	**Hδ**	0.72								
	N4	**Hα**	4.65	**Hβ1**	2.70	**Hβ2**	2.78														
	F5	**Hα**	4.53	**Hβ1**	3.02	**Hβ2**	3.17	**Hδ**	7.26	**Hε**	7.36	**Hζ**	7.31								
	E6	**Hα**	4.19	**Hβ1**	1.91	**Hβ2**	1.99	**Hγ1**	2.22	**Hγ2**	2.22										
	K7	**Hα**	4.28	**Hβ1**	1.87	**Hβ2**	1.87	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.00	**Hε2**	3.00		
	L8	**Hα**	4.18	**Hβ1**	1.58	**Hβ2**	1.58	**Hγ**	1.58	**Hδ1**	0.89	**Hδ2**	0.89								
ESIINFEKL	**E1**	**Hα**	4.07	**Hβ1**	2.13	**Hβ2**	2.10	**Hγ1**	2.37	**Hγ2**	2.37										
	S2	**Hα**	4.57	**Hβ1**	3.85	**Hβ2**	3.95														
	I3	**Hα**	4.21	**Hβ**	1.87	**Hγ1a**	1.20	**Hγ1b**	1.45	**Hγ2**	0.87	**Hδ**	0.87								
	I4	**Hα**	4.10	**Hβ**	1.75	**Hγ1a**	1.14	**Hγ1b**	1.42	**Hγ2**	0.83	**Hδ**	0.74								
	N5	**Hα**	4.70	**Hβ1**	2.69	**Hβ2**	2.79														
	F6	**Hα**	4.57	**Hβ1**	3.18	**Hβ2**	3.02	**Hδ**	7.27	**Hε**	7.37	**Hζ**	7.32								
	E7	**Hα**	4.23	**Hβ1**	2.01	**Hβ2**	1.92	**Hγ1**	2.21	**Hγ2**	2.21										
	K8	**Hα**	4.33	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.02	**Hε2**	3.02		
	L9	**Hα**	4.21	**Hβ1**	1.59	**Hβ2**	1.59	**Hγ**	1.59	**Hδ1**	0.91	**Hδ2**	0.91								
KSIINFEKL	**K1**	**Hα**	3.99	**Hβ1**	1.92	**Hβ2**	1.92	**Hγ1**	1.47	**Hγ2**	1.47	**Hδ1**	1.72	**Hδ2**	1.72	**Hε1**	3.00	**Hε2**	3.00		
	S2	**Hα**	4.57	**Hβ1**	3.85	**Hβ2**	3.85														
	I3	**Hα**	4.21	**Hβ**	1.89	**Hγ1a**	1.18	**Hγ1b**	1.45	**Hγ2**	0.89	**Hδ**	0.89								
	I4	**Hα**	4.12	**Hβ**	1.76	**Hγ1a**	1.12	**Hγ1b**	1.43	**Hγ2**	0.84	**Hδ**	0.73								
	N5	**Hα**	4.69	**Hβ1**	2.69	**Hβ2**	2.77														
	F6	**Hα**	4.57	**Hβ1**	3.03	**Hβ2**	3.18	**Hδ**	7.27	**Hε**	7.37	**Hζ**	7.32								
	E7	**Hα**	4.23	**Hβ1**	1.89	**Hβ2**	1.99	**Hγ1**	2.21	**Hγ2**	2.21										

KSIINFEKL	**K8**	**Hα**	4.32	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.74	**Hδ2**	1.74	**Hε1**	3.01	**Hε2**	3.01
	L9	**Hα**	4.20	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.89	**Hδ2**	0.89						
PSIINFEKL	**P1**	**Hα**	4.46	**Hβ1**	2.50	**Hβ2**	2.50	**Hγ1**	2.08	**Hγ2**	2.09	**Hδ1**	3.44	**Hδ2**	3.42				
	S2	**Hα**	4.55	**Hβ1**	3.87	**Hβ2**	3.83												
	I3	**Hα**	4.23	**Hβ**	1.88	**Hγ1a**	1.48	**Hγ1b**	1.19	**Hγ2**	0.91	**Hδ**	0.88						
	I4	**Hα**	4.12	**Hβ**	1.77	**Hγ1a**	1.41	**Hγ1b**	1.12	**Hγ2**	0.84	**Hδ**	0.74						
	N5	**Hα**	4.49	**Hβ1**	2.67	**Hβ2**	2.77												
	F6	**Hα**	4.24	**Hβ1**	3.16	**Hβ2**	3.22	**Hδ**	7.28	**Hε**	7.41	**Hζ**	7.37						
	E7	***Hα***	*4.34*	***Hβ1***	*1.88*	**Hβ2**	2.01	**Hγ1**	2.21	**Hγ2**	2.21								
		Hα	*4.24*	***Hβ1***	*1.91*														
	K8	***Hα***	*4.32*	**Hβ1**	1.88	**Hβ2**	1.88	***Hγ1***	*1.48*	***Hγ2***	*1.48*	***Hδ1***	*1.73*	***Hδ2***	*1.73*	**Hε1**	3.02	**Hε2**	3.02
		Hα	*4.34*					***Hγ1***	*1.46*	***Hγ2***	*1.46*	***Hδ1***	*1.72*	***Hδ2***	*1.72*				
	L9	**Hα**	4.20	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.88	**Hδ2**	0.88						
EESIINFEKL	**E1**	**Hα**	4.07	**Hβ1**	2.11	**Hβ2**	2.11	**Hγ1**	2.37	**Hγ2**	2.37								
	E2	**Hα**	4.41	**Hβ1**	1.94	**Hβ2**	2.09	**Hγ1**	2.29	**Hγ2**	2.29								
	S3	**Hα**	4.47	**Hβ1**	3.84	**Hβ2**	3.84												
	I4	**Hα**	4.20	**Hβ**	1.86	**Hγ1a**	1.18	**Hγ1b**	1.45	**Hγ2**	0.87	**Hδ**	0.87						
	I5	**Hα**	4.09	**Hβ**	1.75	**Hγ1a**	1.13	**Hγ1b**	1.42	**Hγ2**	0.83	**Hδ**	0.73						
	N6	**Hα**	4.70	**Hβ1**	2.69	**Hβ2**	2.78												
	F7	**Hα**	4.58	**Hβ1**	3.02	**Hβ2**	3.17	**Hδ**	7.27	**Hε**	7.37	**Hζ**	7.32						
	E8	**Hα**	4.24	**Hβ1**	1.92	**Hβ2**	2.00	**Hγ1**	2.20	**Hγ2**	2.20								
	K9	**Hα**	4.33	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.44	**Hγ2**	1.44	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	3.01	**Hε2**	3.01
	L10	**Hα**	4.20	**Hβ1**	1.59	**Hβ2**	1.59	**Hγ**	1.59	**Hδ1**	0.87	**Hδ2**	0.87						
KKSIINFEKL	**K1**	**Hα**	3.96	**Hβ1**	1.89	**Hβ2**	1.89	**Hγ1**	1.46	**Hγ2**	1.46	**Hδ1**	1.72	**Hδ2**	1.72	**Hε1**	3.01	**Hε2**	3.01

KKSIINFEKL	**K2**	**Hα**	4.40	**Hβ1**	1.82	**Hβ2**	1.82	**Hγ1**	1.47	**Hγ2**	1.47	**Hδ1**	1.71	**Hδ2**	1.71	**Hε1**	2.99	**Hε2**	2.99
	S3	**Hα**	4.48	**Hβ1**	3.84	**Hβ2**	3.84												
	I4	**Hα**	4.21	**Hβ**	1.88	**Hγ1a**	1.19	**Hγ1b**	1.46	**Hγ2**	0.88	**Hδ**	0.88						
	I5	**Hα**	4.11	**Hβ**	1.77	**Hγ1a**	1.14	**Hγ1b**	1.42	**Hγ2**	0.85	**Hδ**	0.75						
	N6	**Hα**	4.69	**Hβ1**	2.78	**Hβ2**	2.69												
	F7	**Hα**	4.58	**Hβ1**	3.18	**Hβ2**	3.03	**Hδ**	7.24	**Hε**	7.35	**Hζ**	7.30						
	E8	**Hα**	4.24	**Hβ1**	2.01	**Hβ2**	1.92	**Hγ1**	2.21	**Hγ2**	2.21								
	K9	**Hα**	4.32	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.72	**Hδ2**	1.72	**Hε1**	3.01	**Hε2**	3.01
	L10	**Hα**	4.20	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.88	**Hδ2**	0.88						
PPSIINFEKL	**P1**	**Hα**	4.65	**Hβ1**	2.59	**Hβ2**	2.59	**Hγ1**	2.09	**Hγ2**	2.09	**Hδ1**	3.44	**Hδ2**	3.44				
	P2	**Hα**	4.54	**Hβ1**	2.38	**Hβ2**	2.38	**Hγ1**	2.05	**Hγ2**	1.94	**Hδ1**	3.62	**Hδ2**	3.73				
	S3	**Hα**	4.46	**Hβ1**	3.85	**Hβ2**	3.86												
	I4	**Hα**	4.21	**Hβ**	1.86	**Hγ1a**	1.19	**Hγ1b**	1.46	**Hγ2**	0.88	**Hδ**	0.88						
	I5	**Hα**	4.11	**Hβ**	1.76	**Hγ1a**	1.12	**Hγ1b**	1.42	**Hγ2**	0.83	**Hδ**	0.75						
	N6	**Hα**	4.70	**Hβ1**	2.78	**Hβ2**	2.69												
	F7	**Hα**	4.58	**Hβ1**	3.17	**Hβ2**	3.03	**Hδ**	7.27	**Hε**	7.37	**Hζ**	7.32						
	E8	**Hα**	4.24	**Hβ1**	2.00	**Hβ2**	1.93	**Hγ1**	2.21	**Hγ2**	2.21								
	K9	**Hα**	4.33	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.46	**Hγ2**	1.46	**Hδ1**	1.72	**Hδ2**	1.72	**Hε1**	3.02	**Hε2**	3.02
	L10	**Hα**	4.22	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.91	**Hδ2**	0.92						
IINFEKL	***I1/2***	***Hα***	*3.87*	***Hβ***	*1.94*	***Hγ1a***	*1.20*	***Hγ1b***	*1.48*	***Hγ2***	*0.97*	***Hδ***	*0.97*						
	I1/2	***Hα***	*4.17*	***Hβ***	*1.75*	***Hγ1a***	*1.13*	***Hγ1b***	*1.45*	***Hγ2***	*0.85*	***Hδ***	*0.74*						
	N3	**Hα**	4.68	**Hβ1**	2.69	**Hβ2**	2.77												
	F4	**Hα**	4.57	**Hβ1**	3.01	**Hβ2**	3.18	**Hδ**	7.28	**Hε**	7.37	**Hζ**	7.32						
	E5	**Hα**	4.24	**Hβ1**	1.92	**Hβ2**	2.01	**Hγ1**	2.20	**Hγ2**	2.20								
	K6	**Hα**	4.33	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.44	**Hγ2**	1.44	**Hδ1**	1.72	**Hδ2**	1.72	**Hε1**	3.01	**Hε2**	3.01

IINFEKL	**L7**	**Hα**	4.22	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.87	**Hδ2**	0.91					
INFEKL	**I1**	**Hα**	3.76	**Hβ**	1.85	**Hγ1a**	1.12	**Hγ1b**	1.40	**Hγ2**	0.89	**Hδ**	0.84					
	N2	**Hα**	4.73	**Hβ1**	2.71	**Hβ2**	2.80											
	F3	**Hα**	4.59	**Hβ1**	3.01	**Hβ2**	3.19	**Hδ**	7.28	**Hε**	7.37	**Hζ**	7.32					
	E4	**Hα**	4.25	**Hβ1**	2.01	**Hβ2**	1.93	**Hγ1**	2.21	**Hγ2**	2.24							
	K5	**Hα**	4.33	**Hβ1**	1.88	**Hβ2**	1.88	**Hγ1**	1.45	**Hγ2**	1.45	**Hδ1**	1.72	**Hδ2**	1.73	**Hε1**	3.02	**Hε2**	3.02
	L6	**Hα**	4.22	**Hβ1**	1.60	**Hβ2**	1.60	**Hγ**	1.60	**Hδ1**	0.88	**Hδ2**	0.90					

Amino acids in 20 mM potassium phosphate, pD 7, 100% D_2O. Labile protons not shown.

Proline	**Hα**	4.13	**Hβ1**	2.08	**Hβ2**	2.35	**Hγ1**	2.02	**Hγ2**	2.02	**Hδ1**	3.35	**Hδ2**	3.42			
Serine	**Hα**	3.85	**Hβ1**	3.95	**Hβ2**	3.98											
Isoleucine	**Hα**	3.67	**Hβ**	1.99	**Hγ1a**	1.27	**Hγ1b**	1.48	**Hγ2**	1.02	**Hδ**	0.94					
Asparagine	**Hα**	4.01	**Hβ1**	2.86	**Hβ2**	2.96											
Phenylalanine	**Hα**	4.00	**Hβ1**	3.14	**Hβ2**	3.28	**Hδ**	7.34	**Hε**	7.43	**Hζ**	7.38					
Glutamic acid	**Hα**	3.76	**Hβ1**	2.07	**Hβ2**	2.13	**Hγ1**	2.37	**Hγ2**	2.37							
Lysine	**Hα**	3.76	**Hβ1**	1.91	**Hβ2**	1.91	**Hγ1**	1.46	**Hγ2**	1.51	**Hδ1**	1.73	**Hδ2**	1.73	**Hε1**	3.03	**Hε2**	3.03
Leucine	**Hα**	3.74	**Hβ1**	1.71	**Hβ2**	1.71	**Hγ**	1.71	**Hδ1**	0.97	**Hδ2**	0.97					

Appendix 16 – T1 values of peptide resonances

Peptide	Peak ID	Freq. (ppm)	Intensity	T1 (s)	error
ESIINFEKL	4IleHd1*	0.736804	39.1453	0.5978	0.02181
ESIINFEKL	4IleHg2*	0.836415	43.1666	0.9219	0.02161
ESIINFEKL	9LeuHda/3IleHd1*	0.867666	88.3631	0.729	0.013
ESIINFEKL	9LeuHdb	0.921378	41.2313	0.6813	0.0353
ESIINFEKL	3/4IleHg1b/8LysHg*	1.45361	44.518	0.6233	0.09142
ESIINFEKL	9LeuHg/Hb*	1.58838	63.1289	0.6274	0.04326
ESIINFEKL	8LysHd*	1.70753	54.2631	0.5667	0.02942
ESIINFEKL	4IleHb	1.7212	45.3322	0.6284	0.08046
ESIINFEKL	7GluHba/8LysHba	1.9214	123.172	5.46	0.4521
ESIINFEKL	1GluHb*	2.09425	56.9621	0.4748	0.06829
ESIINFEKL	7GluHg*	2.23293	40.3285	0.4938	0.1019
ESIINFEKL	1GluHg*	2.36086	76.9217	0.6009	0.05878
ESIINFEKL	5AsnHba	2.6929	44.91	0.6828	0.1144
ESIINFEKL	5AsnHbb	2.77688	46.1722	0.6872	0.07384
ESIINFEKL	6PheHba/8LysHe*	3.01419	117.507	0.8618	0.0468
ESIINFEKL	6PheHbb	3.17142	44.5903	0.4576	0.08537
ESIINFEKL	2SerHba	3.8189	59.4869	0.6122	0.07778
ESIINFEKL	2SerHbb	3.85503	57.501	0.7164	0.08604
ESIINFEKL	GluHa	4.07671	55.779	2.026	0.2948
ESIINFEKL	4IleHa	4.09624	74.1834	1.48	0.1492
ESIINFEKL	3IleHa/9LeuHa	4.21343	110.377	1.428	0.1144
ESIINFEKL	7GluHa	4.23199	45.5981	1.346	0.2765
ESIINFEKL	8LysHa	4.32965	42.6431	1.333	0.2493
ESIINFEKL	2SerHa/6PheHa	4.57868	41.9098	1.631	0.4494
ESIINFEKL	6PheHd*	7.27404	85.9757	1.955	0.08979
ESIINFEKL	6PheHz	7.31896	45.3945	2.627	0.2872
ESIINFEKL	6PheHe*	7.3717	83.1081	2.116	0.1409
KSIINFEKL	4IleHd1	0.752944	41.1669	0.6466	0.02151
KSIINFEKL	4IleHg2	0.83986	46.3934	0.9054	0.03522
KSIINFEKL	9LeuHd/3IleHd1	0.875994	61.7644	0.8739	0.02547
KSIINFEKL	I3 Hg2	0.92287	35.064	0.5918	0.01473

KSIINFEKL	6KHg/3/4IleHg1b/1LysHg	1.45706	23.4837	0.6116	0.06527
KSIINFEKL	9LeuHg/Hb	1.59085	24.0714	0.632	0.07398
KSIINFEKL	1/8LysHd	1.70999	32.7611	0.5935	0.03241
KSIINFEKL	8LysHb/3IleHb/7GluHb	1.88773	26.0612	0.5875	0.07349
KSIINFEKL	1LysHb	1.92094	75.6013	5.001	0.4234
KSIINFEKL	7GluHg	2.20707	32.2035	0.6792	0.1847
KSIINFEKL	5AsnHba	2.69048	41.2066	0.6064	0.1169
KSIINFEKL	5AsnHbb	2.77154	37.4096	0.7468	0.08511
KSIINFEKL	1/8LysHe	3.00885	104.301	0.8868	0.04663
KSIINFEKL	6PheHba	3.02154	87.0933	0.816	0.04909
KSIINFEKL	6PheHbb	3.17291	24.1512	0.6477	0.1617
KSIINFEKL	2SerHb	3.84968	38.3588	0.6794	0.04111
KSIINFEKL	1LysHa	3.99129	28.0336	1.451	0.1562
KSIINFEKL	4IleHa	4.11629	39.2207	1.614	0.1412
KSIINFEKL	9LeuHa	4.20418	48.6237	1.351	0.05461
KSIINFEKL	8LysHa	4.3243	27.2292	1.309	0.2693
KSIINFEKL	6Phe/2SerHa	4.5665	26.4225	1.533	0.3847
KSIINFEKL	6PheHd	7.27456	45.4095	1.969	0.1216
KSIINFEKL	6PheHz	7.31948	23.7148	2.108	0.2073
KSIINFEKL	6PheHe	7.37221	50.6676	2.192	0.1384
PSIINFEKL	4IleHd1*	0.744155	44.3071	0.6259	0.03982
PSIINFEKL	4IleHg2*	0.837907	46.4141	0.8165	0.03283
PSIINFEKL	9LeuHda*/3IleHd1*	0.872087	120.58	0.7128	0.01227
PSIINFEKL	9LeuHd*/3IleHg2*	0.912127	36.4369	0.6044	0.01194
PSIINFEKL	3/4IleHg1b/8LysHg*	1.45315	44.4535	0.5801	0.1013
PSIINFEKL	9LeuHb*/Hg	1.58987	60.0853	0.6658	0.06549
PSIINFEKL	8LysHd*	1.70804	49.6881	0.6635	0.04692
PSIINFEKL	1ProHg*	2.07231	62.0914	1.83	0.2923
PSIINFEKL	7GluHg*	2.20805	43.7245	0.4613	0.03738
PSIINFEKL	1ProHb*	2.48735	37.404	0.9881	0.291
PSIINFEKL	5AsnHba	2.70318	48.6987	0.6192	0.08439
PSIINFEKL	5AsnHbb	2.77447	62.253	0.5437	0.08197
PSIINFEKL	8LysHe*	3.01568	121.029	0.7828	0.06407
PSIINFEKL	6PheHba	3.02838	97.1898	0.6061	0.05365
PSIINFEKL	6PheHbb	3.16998	40.3443	0.76	0.3104

PSIINFEKL	1ProHda	3.41413	50.7981	1.798	0.2433
PSIINFEKL	1ProHdb	3.45319	49.2956	1.738	0.613
PSIINFEKL	2SerHba	3.83113	68.5379	0.6204	0.1109
PSIINFEKL	2SerHbb	3.85554	61.1799	0.6306	0.08587
PSIINFEKL	4IleHa	4.10262	77.002	1.475	0.2116
PSIINFEKL	9LeuHa	4.20028	92.6491	1.65	0.2298
PSIINFEKL	6Phe/7Glu/3IleHa	4.21493	111.9	1.648	0.1971
PSIINFEKL	8LysHa	4.32821	53.2553	0.6591	0.13
PSIINFEKL	1ProHa	4.44247	50.0783	2.331	1.335
PSIINFEKL	6PheHd*	7.27553	116.236	1.769	0.1921
PSIINFEKL	6PheHz	7.3185	67.4237	2.231	0.171
PSIINFEKL	6PheHe*	7.37124	121.681	2.16	0.1624
EESIINFEKL	5IleHd*	0.724177	41.2745	0.5871	0.0379
EESIINFEKL	5IleHg2*	0.835507	51.7526	0.9879	0.02232
EESIINFEKL	4IleHd*/Hg2*/10LeuHdb*	0.865781	111.684	0.7631	0.02018
EESIINFEKL	10LeuHda*	0.921446	102.117	0.586	0.02333
EESIINFEKL	9LysHg*/4IleHg1b	1.44001	37.2554	0.5978	0.1111
EESIINFEKL	10LeuHb*/Hg	1.5904	51.5247	0.6012	0.0931
EESIINFEKL	9LysHd*	1.70759	44.4828	0.6406	0.08864
EESIINFEKL	2/8Glu Hba	1.92147	508.383	5.956	0.1949
EESIINFEKL	1/2GluHbb	2.09432	30.1525	0.5637	0.1208
EESIINFEKL	8GluHg*	2.18124	28.9415	0.5712	0.1214
EESIINFEKL	2GluHg*	2.28671	42.097	0.6866	0.1735
EESIINFEKL	1GluHg*	2.3746	43.0991	0.6415	0.05362
EESIINFEKL	6AsnHba	2.69297	31.3288	0.7002	0.1275
EESIINFEKL	6AsnHbb	2.7789	27.209	0.487	0.1056
EESIINFEKL	9LysHe*	3.00254	39.9812	0.6855	0.07185
EESIINFEKL	7PheHba	3.01426	73.4731	0.8535	0.05068
EESIINFEKL	7PheHbb	3.16368	33.2129	0.6906	0.165
EESIINFEKL	3SerHb*	3.84338	59.9501	0.4974	0.06362
EESIINFEKL	1GluHa	4.06213	29.9081	2.036	0.5414
EESIINFEKL	5IleHa	4.09241	43.7959	1.344	0.1206
EESIINFEKL	4IleHa	4.1969	49.4471	1.63	0.1459
EESIINFEKL	10LeuHa	4.2135	46.6448	1.305	0.3366
EESIINFEKL	8GluHa	4.23206	27.0125	0.7315	0.3005

EESIINFEKL	9LysHa	4.33069	26.7055	1.126	0.3701
EESIINFEKL	3SerHa	4.48597	25.965	0.9625	0.3585
EESIINFEKL	7PheHd*	7.27606	55.5255	1.937	0.1415
EESIINFEKL	7PheHz	7.32099	32.3369	2.42	0.2415
EESIINFEKL	7PheHe*	7.37274	57.7955	2.248	0.2455
KKSIINFEKL	5IleHd*	0.746108	40.1416	0.558	0.09686
KKSIINFEKL	5IleHg2*	0.840837	63.2618	1.003	0.08969
KKSIINFEKL	4IleHd*/10LeuHdb	0.873064	167.86	0.7318	0.03038
KKSIINFEKL	10LeuHda*	0.920916	162.91	0.6631	0.02653
KKSIINFEKL	4IleHg1a	1.18655	40.4452	2.355	0.4602
KKSIINFEKL	2LysHga	1.46194	84.6691	0.5926	0.04822
KKSIINFEKL	10LeuHb*Hg	1.59573	63.592	0.7824	0.1112
KKSIINFEKL	1/2Lys Hd*	1.71781	125.188	0.696	0.04057
KKSIINFEKL	1/2/9LysHe*	3.01568	221.031	0.9706	0.0259
KKSIINFEKL	3SerHb*	3.84968	63.0656	0.5171	0.08602
KKSIINFEKL	4Ile/8Glu/10LeuHa	4.22079	75.6262	2.075	0.3761
KKSIINFEKL	7PheHd*	7.27651	65.6889	2.045	0.3121
KKSIINFEKL	7PheHz	7.33803	46.7561	3.818	0.9267
KKSIINFEKL	7PheHe*	7.37124	52.7868	2.225	0.1882
PPSIINFEKL	5IleHd	0.745745	28.8867	0.6034	0.0263
PPSIINFEKL	5IleHg2	0.83852	33.476	0.9716	0.04123
PPSIINFEKL	4IleHg2/4IleHd	0.867817	72.3153	0.7389	0.0315
PPSIINFEKL	10LeuHd1	0.91274	29.8653	0.642	0.02104
PPSIINFEKL	10LeuHd2	0.922506	32.2915	0.6179	0.02284
PPSIINFEKL	5IleHg1b/9LysHg*	1.45279	34.809	0.6029	0.04956
PPSIINFEKL	10LeuHb*/Hg	1.59049	47.3042	0.616	0.08292
PPSIINFEKL	9LysHd*	1.70865	41.7374	0.5501	0.05277
PPSIINFEKL	9LysHb*	1.87858	23.6222	0.5266	0.07599
PPSIINFEKL	8GluHb*/2ProHg2	1.92253	158.218	5.35	0.2787
PPSIINFEKL	8GluHb1	2.01628	48.9809	0.569	0.1264
PPSIINFEKL	2ProHg1/1ProHg*	2.08659	99.9866	1.16	0.1118
PPSIINFEKL	8GluHg*	2.20671	48.4677	0.6061	0.1072
PPSIINFEKL	2ProHb*	2.3571	31.144	0.2594	0.07777
PPSIINFEKL	1ProHb*	2.59539	34.5057	1.254	0.3168

PPSIINFEKL	6AsnHb2	2.69305	39.1511	0.5476	0.06897
PPSIINFEKL	6AsnHb1	2.77703	39.6174	0.4958	0.1203
PPSIINFEKL	9LysHe*	3.0163	106.284	0.8713	0.07386
PPSIINFEKL	7PheHb2	3.02899	84.5342	0.7592	0.07139
PPSIINFEKL	7PheHb1	3.17353	32.7687	0.6411	0.2549
PPSIINFEKL	1ProHd*	3.44013	39.3126	1.173	0.1622
PPSIINFEKL	2ProHd1	3.60811	32.3683	0.3552	0.1379
PPSIINFEKL	2ProHd2	3.73897	38.6355	0.5003	0.1054
PPSIINFEKL	S2Hb*	3.86104	106.633	0.5257	0.03137
PPSIINFEKL	5IleHa	4.10323	64.3217	1.574	0.319
PPSIINFEKL	4IleHa	4.20577	72.8573	1.615	0.1781
PPSIINFEKL	10LeuHa	4.21945	75.7626	1.352	0.185
PPSIINFEKL	9LysHa	4.33078	38.3262	1.029	0.289
PPSIINFEKL	3SerHa	4.46066	53.6033	2.17	0.7007
PPSIINFEKL	2ProHa	4.54855	51.0059	494.2	123
PPSIINFEKL	7PheHd	7.27517	121.079	1.967	0.09535
PPSIINFEKL	7PheHz	7.31912	68.3592	2.063	0.2147
PPSIINFEKL	7PheHe	7.37185	112.559	2.023	0.1101

Bibliography

(TASC), John D Reveille, Anne-Marie Sims, Patrick Danoy, David M Evans, Paul Leo, Jennifer J Pointon, et al. 2010. “Genome-Wide Association Study of Ankylosing Spondylitis Identifies Non-MHC Susceptibility Loci.” *Nature Genetics* 42 (2): 123–27. https://doi.org/10.1038/ng.513.

Achour, Adnane, Jakob Michaëlsson, Robert A Harris, Jacob Odeberg, Per Grufman, Johan K Sandberg, Victor Levitsky, Klas Kärre, Tatyana Sandalova, and Gunter Schneider. 2002. “A Structural Basis for LCMV Immune Evasion: Subversion of H-2D^{b} and H-2K^{b} Presentation of Gp33 Revealed by Comparative Crystal Structure Analyses.” *Immunity* 17 (6): 757–68. https://doi.org/10.1016/S1074-7613(02)00478-8.

Adams, Paul D, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, et al. 2010. “PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution.” *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 2): 213–21. https://doi.org/10.1107/S0907444909052925.

———. 2012. “Dictionary of Common Terms Used in PHENIX.” 2012. https://www.phenix-online.org/documentation/dictionary.html.

Arunan, Elangannan, Gautam R Desiraju, Roger A Klein, Joanna Sadlej, Steve Scheiner, Ibon Alkorta, David C Clary, et al. 2011. “Definition of the Hydrogen Bond (IUPAC Recommendations 2011).” *Pure Appl. Chem.* 83 (8): 1637–41. https://doi.org/10.1351/PAC-REC-10-01-02.

Becker, Edwin D, James A Ferretti, and Prem N Gambhir. 1979. “Nuclear Magnetic Resonance.” *Analytical Chemistry* 51 (9): 1413–20. https://doi.org/10.1021/ac50045a016.

Berman, Helen, Kim Henrick, and Haruki Nakamura. 2003. “Announcing the Worldwide Protein Data Bank.” *Nature Structural Biology* 10 (December): 980. http://dx.doi.org/10.1038/nsb1203-980.

Berman, Helen M, John Westbrook, Zukang Feng, Gary Gilliland, T N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. 2000. “The Protein Data Bank.” *Nucleic Acids Research* 28 (1): 235–42. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC102472/.

Bessonov, Kyrylo, Elena S. Gusareva, and Kristel Van Steen. 2015. “A Cautionary Note on the Impact of Protocol Changes for Genome-Wide Association SNP × SNP Interaction Studies: An Example on Ankylosing Spondylitis.” *Human Genetics*. https://doi.org/10.1007/s00439-015-1560-7.

Blees, Andreas, Dovile Januliene, Tommy Hofmann, Nicole Koller, Carla Schmidt, Simon Trowitzsch, Arne Moeller, and Robert Tampé. 2017. “Structure of the Human MHC-I Peptide-Loading Complex.” *Nature* 551 (7681): 525–28. https://doi.org/10.1038/nature24627.

Blum, J.S., P.A. Wearsch, and P Cresswell. 2013. “Pathways of Antigen Processing.” *Annual Review of Immunology*, 443–73. https://doi.org/10.1146/annurev-immunol-032712-095910.Pathways.

Bouvier, M, and D C Wiley. 1994. “Importance of Peptide Amino and Carboxyl Termini to the Stability of MHC Class I Molecules.” *Science (New York, N.Y.)* 265 (5170): 398–402. http://www.ncbi.nlm.nih.gov/pubmed/8023162.

Bulek, Anna M, Florian Madura, Anna Fuller, Christopher J Holland, Andrea J a Schauenburg, Andrew K Sewell, Pierre J Rizkallah, and David K Cole. 2012. “TCR/PMHC Optimized Protein Crystallization Screen.” *Journal of Immunological Methods* 382 (1–2): 203–10. https://doi.org/10.1016/j.jim.2012.06.007.

Cascio, Paolo, Craig Hilton, Alexei F. Kisselev, Kenneth L. Rock, and Alfred L. Goldberg. 2001. “26S Proteasomes and Immunoproteasomes Produce Mainly N-Extended Versions of an Antigenic Peptide.” *EMBO Journal* 20 (10): 2357–66. https://doi.org/10.1093/emboj/20.10.2357.

Cavanagh, J, W.J. Fairbrother, A.G. Palmer III, and N.J. Skelton. 1996. *Protein NMR Spectroscopy: Principles and Practice*. Academic Press. https://books.google.co.uk/books?id=_FccoEnOIXwC.

Chang, Shih-Chung, Frank Momburg, Nidhi Bhutani, and Alfred L Goldberg. 2005. “The ER Aminopeptidase, ERAP1, Trims Precursors to Lengths of MHC Class I Peptides by a ‘Molecular Ruler’ Mechanism.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (47): 17107–12. https://doi.org/10.1073/pnas.0500721102.

Chen, Hanna, Lenong Li, Mirjana Weimershaus, Irini Evnouchidou, Peter van Endert, and Marlene Bouvier. 2016. “ERAP1-ERAP2 Dimers Trim MHC I-

Bound Precursor Peptides; Implications for Understanding Peptide Editing." *Scientific Reports* 6 (August): 28902. https://doi.org/10.1038/srep28902.

Cookson, David J, and Brian E Smith. 1982. "Optimal Experimental Parameters for Quantitative Pulse Fourier Transform Proton Nuclear Magnetic Resonance Spectrometry." *Analytical Chemistry* 54 (14): 2591–93. https://doi.org/10.1021/ac00251a042.

Cresswell, Peter, Anne L Ackerman, Alessandra Giodini, David R Peaper, and Pamela a Wearsch. 2005. "Mechanisms of MHC Class I-Restricted Antigen Processing and Cross-Presentation." *Immunological Reviews* 207 (October): 145–57. https://doi.org/10.1111/j.0105-2896.2005.00316.x.

Cui, Xinle, Feras Hawari, Sura Alsaaty, Marion Lawrence, Christian A Combs, Weidong Geng, Farshid N Rouhani, Dianne Miskinis, and Stewart J Levine. 2002. "Identification of ARTS-1 as a Novel TNFR1-Binding Protein That Promotes TNFR1 Ectodomain Shedding." *Journal of Clinical Investigation* 110 (4): 515–26. https://doi.org/10.1172/JCI13847.cDNA.

Dam, Julie, Rongjin Guan, Kannan Natarajan, Nazzareno Dimasi, Lukasz K Chlewicki, David M Kranz, Peter Schuck, David H Margulies, and Roy A Mariuzza. 2003. "Variable MHC Class I Engagement by Ly49 Natural Killer Cell Receptors Demonstrated by the Crystal Structure of Ly49C Bound to H-2Kb." *Nat Immunol* 4 (12): 1213–22. http://dx.doi.org/10.1038/ni1006.

Delaglio, Frank, Stephan Grzesiek, Geerten W Vuister, Guang Zhu, John Pfeifer, and Ad Bax. 1995. "NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes." *Journal of Biomolecular NMR* 6 (3): 277–93. https://doi.org/10.1007/BF00197809.

Deng, Lu, Sangwoo Cho, Emilio L Malchiodi, Melissa C Kerzic, Julie Dam, and Roy A Mariuzza. 2008. "Molecular Architecture of the Major Histocompatibility Complex Class I-Binding Site of Ly49 Natural Killer Cell Receptors." *The Journal of Biological Chemistry* 283 (24): 16840–49. https://doi.org/10.1074/jbc.M801526200.

Denton, Alice E, Robb Wesselingh, Stephanie Gras, Matthew R Olson, Justine D Mintern, Weiguang Zeng, David C Jackson, et al. 2011. "Affinity Thresholds for Naive CD8 + CTL Activation by Peptides and Engineered Influenza A Viruses." *The Journal of Immunology* 187: 5733–44.

https://doi.org/10.4049/jimmunol.1003937.

Dick, Tobias P, Naveen Bangia, David R Peaper, Peter Cresswell, and New Haven. 2002. “Disulfide Bond Isomerization and the Assembly of MHC Class I-Peptide Complexes.” *Immunity* 16: 87–98.

Dong, Gang, Pamela a Wearsch, David R Peaper, Peter Cresswell, and Karin M Reinisch. 2009. “Insights into MHC Class I Peptide Loading from the Structure of the Tapasin-ERp57 Thiol Oxidoreductase Heterodimer.” *Immunity* 30 (1): 21–32. https://doi.org/10.1016/j.immuni.2008.10.018.

Elliott, Tim, and Anthony Williams. 2005. “The Optimization of Peptide Cargo Bound to MHC Class I Molecules by the Peptide-Loading Complex.” *Immunological Reviews* 207 (October): 89–99. https://doi.org/10.1111/j.0105-2896.2005.00311.x.

Emsley, P, B Lohkamp, W G Scott, and K Cowtan. 2010. “Features and Development of Coot.” *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 4): 486–501. https://doi.org/10.1107/S0907444910007493.

Emsley, Paul, and Kevin Cowtan. 2004. “Coot : Model-Building Tools for Molecular Graphics Research Papers.” *Acta Crystallographica Section D*, 2126–32. https://doi.org/10.1107/S0907444904019158.

Evans, David M, Chris C a Spencer, Jennifer J Pointon, Zhan Su, David Harvey, Grazyna Kochan, Udo Oppermann, et al. 2011. “Interaction between ERAP1 and HLA-B27 in Ankylosing Spondylitis Implicates Peptide Handling in the Mechanism for HLA-B27 in Disease Susceptibility.” *Nature Genetics* 43 (8): 761–67. https://doi.org/10.1038/ng.873.

Evans, Philip R, and Garib N Murshudov. 2013. “How Good Are My Data and What Is the Resolution?” *Acta Crystallographica Section D* 69 (7): 1204–14. https://doi.org/10.1107/S0907444913000061.

Evnouchidou, Irini, Marcelo J. Berardi, and Efstratios Stratikos. 2009. “A Continuous Fluorigenic Assay for the Measurement of the Activity of Endoplasmic Reticulum Aminopeptidase 1: Competition Kinetics as a Tool for Enzyme Specificity Investigation.” *Analytical Biochemistry* 395 (1): 33–40. https://doi.org/10.1016/j.ab.2009.07.032.

Evnouchidou, Irini, Frank Momburg, Athanasios Papakyriakou, Angeliki Chroni,

Leondios Leondiadis, Shih-Chung Chang, Alfred L Goldberg, and Efstratios Stratikos. 2008. “The Internal Sequence of the Peptide-Substrate Determines Its N-Terminus Trimming by ERAP1.” *PloS One* 3 (11): e3658. https://doi.org/10.1371/journal.pone.0003658.

Evnouchidou, Irini, Mirjana Weimershaus, Loredana Saveanu, and Peter van Endert. 2014. “ERAP1-ERAP2 Dimerization Increases Peptide-Trimming Efficiency.” *Journal of Immunology (Baltimore, Md. : 1950)*, June. https://doi.org/10.4049/jimmunol.1302855.

Falk, K, O Rötzschke, and HG Rammensee. 1990. “Cellular Peptide Composition Governed by Major Histocompatibility Complex Class I Molecules.” *Nature* 348: 248–51. http://www.nature.com/nature/journal/v348/n6298/abs/348248a0.html.

Fremont, D. H., E. A. Stura, M Matsumura, P A Peterson, and I A Wilson. 1995. “Crystal Structure of an H-2Kb-Ovalbumin Peptide Complex Reveals the Interplay of Primary and Secondary Anchor Positions in the Major Histocompatibility Complex Binding Groove.” *Proceedings of the National Academy of Sciences* 92 (March): 2479–83.

Fremont, D H, M Matsumura, E a Stura, P a Peterson, and I a Wilson. 1992. “Crystal Structures of Two Viral Peptides in Complex with Murine MHC Class I H-2Kb.” *Science (New York, N.Y.)* 257 (5072): 919–27. https://doi.org/10.1126/science.1323877.

Gandhi, Amit, Damodharan Lakshminarasimhan, Yixin Sun, and Hwai-Chen Guo. 2011. “Structural Insights into the Molecular Ruler Mechanism of the Endoplasmic Reticulum Aminopeptidase ERAP1.” *Scientific Reports* 1 (January): 186. https://doi.org/10.1038/srep00186.

Garboczi, David N, Partho Ghosh, Ursula Utz, Qing R Fan, William E Biddison, and Don C Wiley. 1996. “Structure of the Complex Between Human T-Cell Receptor, Viral Peptide and HLA-A2.” *Nature* 384: 134–41.

García-medel, Noel, Alejandro Sanz-Bravo, Dung Van Nguyen, Patricia Gomez-Molina, Carlos Alvarez-navarro, and Jose A Lopez de Castro. 2012. “Functional Interaction of the Ankylosing Spondylitis-Associated Endoplasmic Reticulum Aminopeptidase 1 Polymorphism and HLA-B27 in Vivo.” *Molecular & Cellular Proteomics*, 1416–29.

https://doi.org/10.1074/mcp.M112.019588.

Garman, Elspeth. 1999. “Cool Data: Quantity AND Quality.” *Acta Crystallographica Section D Biological Crystallography* 55 (10): 1641–53. https://doi.org/10.1107/S0907444999008653.

Grandea, Andres G, Tatiana N Golovina, Sara E Hamilton, Venkataraman Sriram, Thomas Spies, Randy R Brutkiewicz, John T Harty, Laurence C Eisenlohr, and Luc Van Kaer. 2000. “Impaired Assembly yet Normal Trafficking of MHC Class I Molecules in Tapasin Mutant Mice.” *Immunity* 13: 213–22.

Griffiths, Lee, and Alan Irving. 1998. “Assay by Nuclear Magnetic Resonance Spectroscopy: Quantification Limits.” *Analyst* 123 (5): 1061–68. https://doi.org/10.1039/A800625C.

Grossmann, Nina, Ahmet S Vakkasoglu, Sabine Hulpke, Rupert Abele, Rachelle Gaudet, and Robert Tampe. 2014. “Mechanistic Determinants of the Directionality and Energetics of Active Export by a Heterodimeric ABC Transporter.” *Nature Communications*. https://doi.org/10.1038/ncomms6419.

Harvey, David, Jennifer J Pointon, David M Evans, Tugce Karaderi, Claire Farrar, Louise H Appleton, Roger D Sturrock, et al. 2009. “Investigating the Genetic Association between ERAP1 and Ankylosing Spondylitis.” *Human Molecular Genetics* 18 (21): 4204–12. https://doi.org/10.1093/hmg/ddp371.

Hattori, a., H. Matsumoto, S. Mizutani, and M. Tsujimoto. 1999. “Molecular Cloning of Adipocyte-Derived Leucine Aminopeptidase Highly Related to Placental Leucine Aminopeptidase/Oxytocinase.” *Journal of Biochemistry* 125 (5): 931–38. https://doi.org/10.1093/oxfordjournals.jbchem.a022371.

Hearn, Arron, Ian a York, and Kenneth L Rock. 2009. “The Specificity of Trimming of MHC Class I-Presented Peptides in the Endoplasmic Reticulum.” *Journal of Immunology (Baltimore, Md. : 1950)* 183 (9): 5526–36. https://doi.org/10.4049/jimmunol.0803663.

Hekkelman, M L, T A H Beek, S R Pettifer, D Thorne, T K Attwood, and G Vriend. 2010. “WIWS : A Protein Structure Bioinformatics Web Service Collection.” *Nucleic Acids Research* 38 (May): 719–23. https://doi.org/10.1093/nar/gkq453.

Hermann, C., J. Trowsdale, and L. H. Boyle. 2015. “TAPBPR: A New Player in the

MHC Class I Presentation Pathway." *Tissue Antigens* 85 (3): 155–66. https://doi.org/10.1111/tan.12538.

Hinz, Andreas, Johanna Jedamzick, Valentina Herbring, Hanna Fischbach, Jessica Hartmann, David Parcej, Joachim Koch, and Robert Tampé. 2014. "Assembly and Function of the MHC I Peptide-Loading Complex Are Conserved across Higher Vertebrates." *Journal of Biological Chemistry* 289 (48): 33109–17. https://doi.org/10.1074/jbc.M114.609263.

Hooft, Rob W W, Chris Sander, and Gerrit Vriend. 1996. "Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures." *PROTEINS* 26: 363–76.

Hoult, D I. 1976. "Solvent Peak Saturation with Single Phase and Quadrature Fourier Transformation." *Journal of Magnetic Resonance (1969)* 21 (2): 337–47. https://doi.org/https://doi.org/10.1016/0022-2364(76)90081-0.

IBA. 2012. "Expression and Purification of Proteins Using Strep-Tag® or Twin-Strep-Tag® A Comprehensive Manual." www.strep-tag.comhttp://www.iba-lifesciences.com/patents.htmloruponinquiryatinfo@iba-lifesciences.comor.

Infantes, Susana, Yolanda Samino, Elena Lorente, Mercedes Jiménez, Ruth García, Margarita Del Val, and Daniel López. 2010. "H-2Ld Class I Molecule Protects an HIV N-Extended Epitope from in Vitro Trimming by Endoplasmic Reticulum Aminopeptidase Associated with Antigen Processing." *Journal of Immunology (Baltimore, Md. : 1950)* 184 (7): 3351–55. https://doi.org/10.4049/jimmunol.0901560.

Ito, Kiyoshi, Yoshitaka Nakajima, Yuko Onohara, Masahide Takeo, Kanako Nakashima, Futoshi Matsubara, Takashi Ito, and Tadashi Yoshimoto. 2006. "Crystal Structure of Aminopeptidase N (Proteobacteria Alanyl Aminopeptidase) from Escherichia Coli and Conformational Change of Methionine 260 Involved in Substrate Recognition." *The Journal of Biological Chemistry* 281 (44): 33664–76. https://doi.org/10.1074/jbc.M605203200.

Jayawickrama, Dimuthu, and Cynthia Larive. 1999. *Analysis of the (Trimethylsilyl)Propionic Acid–β(12–28) Peptide Binding Equilibrium with NMR Spectroscopy*. *Analytical Chemistry*. Vol. 71. https://doi.org/10.1021/ac980989w.

Jones, Peter M, Mark W Robinson, John P Dalton, and Anthony M George. 2011.

"The Plasmodium Falciparum Malaria M1 Alanyl Aminopeptidase (PfA-M1): Insights of Catalytic Mechanism and Function from MD Simulations." *PloS One* 6 (12): e28589. https://doi.org/10.1371/journal.pone.0028589.

Kaiser, Raymond, and Lorraine Metzka. 1999. "Enhancement of Cyanogen Bromide Cleavage Yields for Methionyl-Serine and Methionyl-Threonine Peptide Bonds." *Analytical Biochemistry* 266 (1): 1–8. https://doi.org/10.1006/ABIO.1998.2945.

Kanaseki, Takayuki, Nicolas Blanchard, Gianna Elena Hammer, Federico Gonzalez, and Nilabh Shastri. 2006. "ERAAP Synergizes with MHC Class I Molecules to Make the Final Cut in the Antigenic Peptide Precursors in the Endoplasmic Reticulum." *Immunity* 25 (5): 795–806. https://doi.org/10.1016/j.immuni.2006.09.012.

Kochan, Grazyna, Tobias Krojer, David Harvey, Roman Fischer, Liye Chen, Melanie Vollmar, Frank von Delft, et al. 2011. "Crystal Structures of the Endoplasmic Reticulum Aminopeptidase-1 (ERAP1) Reveal the Molecular Basis for N-Terminal Peptide Trimming." *Proceedings of the National Academy of Sciences of the United States of America* 108 (19): 7745–50. https://doi.org/10.1073/pnas.1101262108.

Krumeich, F. 2015. "Bragg's Law of Diffraction." ETH Zurich Electron Microscopy. 2015. http://www.microscopy.ethz.ch/bragg.htm.

Kyte, Jack, and Russell F Doolittle. 1982. "A Simple Method for Displaying the Hydropathic Character of a Protein." *Journal of Molecular Biology* 157 (1): 105–32. https://doi.org/https://doi.org/10.1016/0022-2836(82)90515-0.

Leonhardt, Ralf M, Kirstin Keusekotten, Cemalettin Bekpen, Michael R Knittler, Ralf M Leonhardt, Kirstin Keusekotten, Cemalettin Bekpen, and Michael R Knittler. 2005. "Critical Role for the Tapasin-Docking Site of TAP2 in the Functional Integrity of the MHC Class I-Peptide-Loading Complex." *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.175.8.5104.

Levitt, M H. 2008. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley. https://books.google.co.uk/books?id=gC3dngEACAAJ.

Limtiaco, John F K, Szabolcs Beni, Christopher J Jones, Derek J Langeslay, and Cynthia K Larive. 2011. "NMR Methods to Monitor the Enzymatic Depolymerization of Heparin." *Analytical and Bioanalytical Chemistry* 399

(2): 593–603. https://doi.org/10.1007/s00216-010-4132-7.

Lorente, Elena, Alejandro Barriga, Carolina Johnstone, Carmen Mir, Mercedes Jiménez, and Daniel López. 2013. "Concerted In Vitro Trimming of Viral HLA-B27-Restricted Ligands by Human ERAP1 and ERAP2 Aminopeptidases." *PloS One* 8 (11): e79596. https://doi.org/10.1371/journal.pone.0079596.

Lybarger, Lonnie, Y Y Lawrence Yu, Michael J Miley, Daved H Fremont, Nancy Myers, Tina Primeau, Steven M Truscott, Janet M Connolly, and Ted H Hansen. 2003. "Enhanced Immune Presentation of a Single-Chain Major Histocompatibility Complex Class I Molecule Engineered to Optimize Linkage of a C-Terminally Extended Peptide." *Journal of Biological Chemistry* 278 (29): 27105–11. https://doi.org/10.1074/jbc.M303716200.

Mareeva, Tatiana, Erik Martinez-hackert, and Yuri Sykulev. 2008. "How a T Cell Receptor-like Antibody Recognizes Major." *The Journal of Biological Chemistry* 283 (43): 29053–59. https://doi.org/10.1074/jbc.M804996200.

Matsumura, M, D H Fremont, P a Peterson, and I a Wilson. 1992. "Emerging Principles for the Recognition of Peptide Antigens by MHC Class I Molecules." *Science (New York, N.Y.)* 257 (5072): 927–34. http://www.ncbi.nlm.nih.gov/pubmed/1323878.

Mcgowan, Sheena, Corrine J Porter, Jonathan Lowther, Colin M Stack, Sarah J Golding, Tina S Skinner-adams, Katharine R Trenholme, et al. 2009. "Structural Basis for the Inhibition of the Essential Plasmodium Falciparum M1 Neutral Aminopeptidase." *Proceedings of the National Academy of Sciences* 106 (8): 2537–42.

Mcpherson, Alexander. 2004. "Introduction to Protein Crystallization." *Methods 34* 34: 254–65. https://doi.org/10.1016/j.ymeth.2004.03.019.

Meier, B.H., and R.R. Ernst. 1979. "Eucidation of Chemical Exchange Networks by Two-Dimensional NMR Spectroscopy: The Heptamethylbenzenonium Ion." *Journal of the American Chemical Society* 101 (21). https://doi.org/10.1021/ja00515a053.

Mitaksov, Vesselin, and Daved H Fremont. 2006. "Structural Definition of the H-2K d Peptide-Binding Motif *." *The Journal of Biological Chemistry* 281 (15): 10618–25. https://doi.org/10.1074/jbc.M510511200.

Mitaksov, Vesselin, Steven M. Truscott, Lonnie Lybarger, Janet M. Connolly, Ted H. Hansen, and Daved H. Fremont. 2007. "Structural Engineering of PMHC Reagents for T Cell Vaccines and Diagnostics." *Chemistry & Biology* 14 (8): 909–22. https://doi.org/10.1016/j.chembiol.2007.07.010.

Neisig, A, J Roelse, and AJ Sijts. 1995. "Major Differences in Transporter Associated with Antigen Presentation (TAP)-Dependent Translocation of MHC Class I-Presentable Peptides and the Effect of Flanking." *The Journal of Immunology*. http://www.jimmunol.org/content/154/3/1273.short.

Nguyen, TT, SC Chang, Irini Evnouchidou, Ian a York, Christos Zikos, Kenneth L Rock, Alfred L Goldberg, Efstratios Stratikos, and Lawrence J Stern. 2011. "Structural Basis For Antigenic Peptide Precursor Processing by the Endoplasmic Reticulum Aminopeptidase ERAP1." *Nature Structural & Molecular Biology* 18 (5): 604–13. https://doi.org/10.1038/nsmb.2021.Structural.

Papakyriakou, Athanasios, Emma Reeves, Mary Beton, Halina Mikolajek, Leon Douglas, Grace Cooper, Tim Elliott, Jörn M. Werner, and Edward James. 2018. "The Partial Dissociation of MHC Class I– Bound Peptides Exposes Their N Terminus to Trimming by Endoplasmic Reticulum Aminopeptidase 1." *Journal of Biological Chemistry* 293 (20): 7538–48. https://doi.org/10.1074/jbc.RA117.000313.

Parcej, David, and Robert Tampe. 2010. "ABC Proteins in Antigen Translocation and Viral Inhibition." *Nat Chem Biol* 6 (8): 572–80. http://dx.doi.org/10.1038/nchembio.410.

Pettersen, Eric F, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. 2004. "UCSF Chimera — A Visualization System for Exploratory Research and Analysis." *J. Comput. Chem.* 25: 1605–12. https://doi.org/10.1002/jcc.20084.

Piotto, Martial, Vladimir Saudek, and Vladimir Sklenář. 1992. "Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions." *Journal of Biomolecular NMR* 2 (6): 661–65. https://doi.org/10.1007/BF02192855.

Raghuraman, Gayatri, Philip Edward Lapinski, and Malini Raghavan. 2002. "Tapasin Interacts with the Membrane-Spanning Domains of Both TAP

Subunits and Enhances the Structural Stability Of." *The Journal of Biological Chemistry* 277 (44): 41786–94. https://doi.org/10.1074/jbc.M207128200.

Rawlings, N D, and A J Barrett. 1995. "Evolutionary Families of Metallopeptidases." *Methods in Enzymology* 248: 183–228. http://www.ncbi.nlm.nih.gov/pubmed/7674922.

Read, R. J. 1986. "Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors." *Acta Crystallographica Section A Foundations of Crystallography* 42 (3): 140–49. https://doi.org/10.1107/S0108767386099622.

Reeves, Emma, Alexandra Colebatch-Bourn, Tim Elliott, Christopher J Edwards, and Edward James. 2014. "Functionally Distinct ERAP1 Allotype Combinations Distinguish Individuals with Ankylosing Spondylitis." *Proceedings of the National Academy of Sciences of the United States of America*, November, 1–6. https://doi.org/10.1073/pnas.1408882111.

Reeves, Emma, Christopher J Edwards, Tim Elliott, and Edward James. 2013. "Naturally Occurring ERAP1 Haplotypes Encode Functionally Distinct Alleles with Fine Substrate Specificity." *Journal of Immunology (Baltimore, Md. : 1950)* 191 (1): 35–43. https://doi.org/10.4049/jimmunol.1300598.

Reiser, Jean-baptiste, Claudine Darnault, Annick Guimezanes, Claude Grégoire, Juan Carlos Fontecilla-camps, Bernard Malissen, Dominique Housset, and Gilbert Mazza. 2000. "Crystal Structure of a T Cell Receptor Bound to an Allogeneic MHC Molecule." *Nature Immunology* 1 (4): 291–97.

Rhodes, Gale. 2010. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Complementary Science. Elsevier Science. https://books.google.co.uk/books?id=rwnR6qvaWgkC.

Rosman, KJR, and PDP Taylor. 1998. "Isotopic Compositions of the Elements (1997)." *Pure and Applied Chemistry* 70 (January): 217–35.

Rossjohn, Jamie, Stephanie Gras, John J. Miles, Stephen J. Turner, Dale I. Godfrey, and James McCluskey. 2015. "T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules." *Annual Review of Immunology* 33 (1): 169–200. https://doi.org/10.1146/annurev-immunol-032414-112334.

Sadasivan, B, P J Lehner, B Ortmann, T Spies, and P Cresswell. 1996. "Roles for Calreticulin and a Novel Glycoprotein, Tapasin, in the Interaction of MHC

Class I Molecules with TAP." *Immunity* 5 (2): 103–14. http://www.ncbi.nlm.nih.gov/pubmed/8769474.

Saric, Tomo, Shih-Chung Chang, Akira Hattori, Ian a York, Shirley Markant, Kenneth L Rock, Masafumi Tsujimoto, and Alfred L Goldberg. 2002. "An IFN-Gamma-Induced Aminopeptidase in the ER, ERAP1, Trims Precursors to MHC Class I-Presented Peptides." *Nature Immunology* 3 (12): 1169–76. https://doi.org/10.1038/ni859.

Saveanu, Loredana, Oliver Carroll, Vivian Lindo, Margarita Del Val, Daniel Lopez, Yves Lepelletier, Fiona Greer, et al. 2005. "Concerted Peptide Trimming by Human ERAP1 and ERAP2 Aminopeptidase Complexes in the Endoplasmic Reticulum." *Nature Immunology* 6 (7): 689–97. https://doi.org/10.1038/ni1208.

Schomburg, Lutz, Heike Kollmus, Sönke Friedrichsen, and Karl Bauer. 2000. "Molecular Characterization of a Puromycin-Insensitive Leucyl-Specific Aminopeptidase, PILS-AP." *European Journal of Biochemistry* 267 (11): 3198–3207. https://doi.org/10.1046/j.1432-1327.2000.01348.x.

Seregin, Sergey S, David P W Rastall, Irini Evnouchidou, F Charles, Dionisia Quiroga, Ram P Kamal, Sarah Godbehere-roosa, et al. 2013. "Associated with Increased Risk of Ankylosing Spondylitis Reduce HLA-B27 Mediated Presentation of Multiple Antigens." *Autoimmunity* 46 (8): 497–508. https://doi.org/10.3109/08916934.2013.819855.

Serwold, Thomas, Stephanie Gaw, and Nilabh Shastri. 2001. "ER Aminopeptidases Generate a Unique Pool of Peptides for MHC Class I Molecules." *Nature Immunology* 2 (7). http://www.nature.com/ni/journal/v2/n7/abs/ni0701_644.html.

Serwold, Thomas, Federico Gonzalez, Jennifer Kim, Richard Jacob, and Nilabh Shastri. 2002. "ERAAP Customizes Peptides for MHC Class I Molecules in the Endoplasmic Reticulum." *Nature* 419 (6906): 480–83. https://doi.org/10.1038/nature01074.

Sijts, E. J A M, and P. M. Kloetzel. 2011. "The Role of the Proteasome in the Generation of MHC Class I Ligands and Immune Responses." *Cellular and Molecular Life Sciences* 68 (9): 1491–1502. https://doi.org/10.1007/s00018-011-0657-y.

Spiliotis, Elias T., H Manley, Manuel Osorio, Martha C. Zúñiga, and Michael Edidin. 2000. “Selective Export of MHC Class I Molecules from the ER after Their Dissociation from TAP.” *Immunity* 13 (6): 841–51. https://doi.org/10.1016/S1074-7613(00)00081-9.

Steiner, Thomas. 2002. “The Whole Palette of Hydrogen Bonds REVIEWS The Hydrogen Bond in the Solid State.” *Angew. Chem. Int. Ed.* 41: 48–76.

Stratikos, Efstratios, Athanasios Stamogiannos, Efthalia Zervoudi, and Doriana Fruci. 2014. “A Role for Naturally Occurring Alleles of Endoplasmic Reticulum Aminopeptidases in Tumor Immunity and Cancer.” *Frontiers in Oncology* 4: 1–10. https://doi.org/10.3389/fonc.2014.00363.

Stratikos, Efstratios, and Lawrence J Stern. 2013. “Antigenic Peptide Trimming by ER Aminopeptidases--Insights from Structural Studies.” *Molecular Immunology* 55 (3–4): 212–19. https://doi.org/10.1016/j.molimm.2013.03.002.

Tholander, Fredrik, Ayumo Muroya, Bernard-Pierre Roques, Marie-Claude Fournie-Zaluski, Marjolein M G M Thunnissen, and Jesper Z Haeggstrom. 2008. “Structure-Based Dissection of the Active Site Chemistry of Leukotriene A4 Hydrolase: Implications for M1 Aminopeptidases and Inhibitor Design.” *Chemistry & Biology* 15: 920–29. https://doi.org/10.1016/j.chembiol.2008.07.018.

Thunnissen, Marjolein M G M, Par Nordlund, and Jesper Z Haeggstrom. 2001. “Crystal Structure of Human Leukotriene A4 Hydrolase, a Bifunctional Enzyme in Inflammation.” *Nature Structural Biology* 8 (2): 131–35.

Truscott, Steven M, Lonnie Lybarger, John M Martinko, Vesselin E Mitaksov, David M Kranz, Janet M Connolly, Daved H Fremont, et al. 2007. “Disulfide Bond Engineering to Trap Peptides in the MHC Class I Binding Groove.” *J. Immunol.* 178: 6280–89. https://doi.org/10.4049/jimmunol.178.10.6280.

Tsujimoto, Masafumi, and Akira Hattori. 2005. “The Oxytocinase Subfamily of M1 Aminopeptidases.” *Biochimica et Biophysica Acta* 1751 (1): 9–18. https://doi.org/10.1016/j.bbapap.2004.09.011.

Ulrich, Eldon L, Hideo Akutsu, Jurgen F Doreleijers, Yoko Harano, Yannis E Ioannidis, Jundong Lin, Miron Livny, et al. 2008. “BioMagResBank.” *Nucleic Acids Research* 36 (suppl_1): D402–8.

http://dx.doi.org/10.1093/nar/gkm957.

Vagin, A, and A Teplyakov. 1997. “MOLREP : An Automated Program for Molecular Replacement.” *Journal of Applied Crystallography* 30 (6): 1022–25. https://doi.org/10.1107/S0021889897006766.

Vranken, Wim F, Wayne Boucher, Tim J Stevens, Rasmus H Fogh, Anne Pajon, Miguel Llinas, Eldon L Ulrich, John L Markley, John Ionides, and Ernest D Laue. 2005. “The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline.” *Proteins: Structure, Function, and Bioinformatics* 59 (4): 687–96. https://doi.org/10.1002/prot.20449.

Wagstaff, Jane L, Mark J Howard, and Richard a Williamson. 2010. “Production of Recombinant Isotopically Labelled Peptide by Fusion to an Insoluble Partner Protein: Generation of Integrin Avβ6 Binding Peptides for NMR.” *Molecular BioSystems* 6 (12): 2380–85. https://doi.org/10.1039/c0mb00105h.

Wearsch, Pamela A, and Peter Cresswell. 2007. “Selective Loading of High-Affinity Peptides onto Major Histocompatibility Complex Class I Molecules by the Tapasin-ERp57 Heterodimer.” *Nature Immunology* 8 (8): 873–82. https://doi.org/10.1038/ni1485.

Williams, A, C Au Peh, and T Elliott. 2002. “The Cell Biology of MHC Class I Antigen Presentation.” *Tissue Antigens* 59 (1): 3–17. http://www.ncbi.nlm.nih.gov/pubmed/11972873.

Williams, Anthony P, Chen Au Peh, Anthony W Purcell, James Mccluskey, and Tim Elliott. 2002. “Optimization of the MHC Class I Peptide Cargo Is Dependent on Tapasin.” *Immunity* 16: 509–20.

Williamson, Michael P, and Tetsuo Asakura. 1997. “Protein Chemical Shifts.” In *Protein NMR Techniques*, edited by David G Reid, 53–69. Totowa, NJ: Humana Press. https://doi.org/10.1385/0-89603-309-0:53.

Winter, G. 2010. “{\it Xia2}: An Expert System for Macromolecular Crystallography Data Reduction.” *Journal of Applied Crystallography* 43 (1). https://doi.org/10.1107/S0021889809045701.

Wishart, David S., Colin G. Bigam, Arne Holm, Robert S. Hodges, and Brian D. Sykes. 1995. “1H, 13C and 15N Random Coil NMR Chemical Shifts of the Common Amino Acids. Investigations of Nearest-Neighbor Effects.” *Journal*

of Biomolecular NMR 5: 67–81.

Wishart, David S, Colin G Bigam, Jian Yao, Frits Abildgaard, H D Jane, Eric Oldfield, John L Markley, and Brian D Sykes. 1995. "1H, 13C and 15N Chemical Shift Referencing in Biomolecular NMR" 6: 135–40.

Wong, Alan H M, Dongxia Zhou, and James M Rini. 2012. "The X-Ray Crystal Structure of Human Aminopeptidase N Reveals a Novel Dimer and the Basis for Peptide Processing." *The Journal of Biological Chemistry* 287 (44): 36804–13. https://doi.org/10.1074/jbc.M112.398842.

Wüthrich, Kurt. 1986. *NMR of Proteins and Nucleic Acids*. A Wiley-Interscience Publication. Wiley. https://books.google.co.uk/books?id=zfBqAAAAMAAJ.

York, Ian a, Shih-Chung Chang, Tomo Saric, Jennifer a Keys, Janice M Favreau, Alfred L Goldberg, and Kenneth L Rock. 2002. "The ER Aminopeptidase ERAP1 Enhances or Limits Antigen Presentation by Trimming Epitopes to 8-9 Residues." *Nature Immunology* 3 (12): 1177–84. https://doi.org/10.1038/ni860.

Young, AC, SG Nathenson, and JC Sacchettini. 1995. "Structural Studies of Class I Major Histocompatibility Complex Proteins: Insights into Antigen Presentation." *The FASEB Journal*, no. 5: 26–36. http://www.fasebj.org/content/9/1/26.short.

Zervoudi, Efthalia, Athanasios Papakyriakou, Dimitra Georgiadis, Irini Evnouchidou, Anna Gajda, Marcin Poreba, Guy S. Salvesen, et al. 2011. "Probing the S1 Specificity Pocket of the Aminopeptidases That Generate Antigenic Peptides: S1 Specificity of ERAP1, ERAP2 and IRAP." *Biochem J.* 435 (2): 411–20. https://doi.org/10.1042/BJ20102049.Probing.