Novel Chemical Proteomics Approaches to Study *N*-Myristoylation and N-terminal Methionine Excision

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

by

Andrea Goya Grocin

Department of Chemistry

Imperial College London

80 Wood Lane

London W12 0BZ

September 2020

I, Andrea Goya Grocin, certify that this thesis and the research to which it refers are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: _____

Copyright

Novel Chemical Proteomics Approaches to Study *N*-Myristoylation and N-terminal Methionine Excision

Abstract

N-terminal modifications may constitute the first modifications any protein acquires and can modulate protein function by altering protein stability and 3D structure, promoting or interfering with protein-protein interactions and regulating membrane targeting, localisation or secretion. Here, two interlinked N-terminal protein modifications have been studied: initiator methionine (iMet) excision and co-translational *N*-myristoylation.

N-myristoylation involves de attachment of myristic acid, a 14-carbon saturated fatty acid, onto exposed N-terminal glycines of protein substrates. In humans, this reaction is catalysed by *N*-myristoyltransferases 1 and 2. The two NMT isoforms share a high degree of sequence and structural similarity, which has hindered the development of isoform-specific inhibitors to date and prevented the dissection of isoform-specific substrate pools. Using CRISPR/Cas9 in combination with a previously described metabolic labelling approach and whole proteome profiling, NMT1, and not NMT2, was defined as the main enzyme responsible for *N*-myristoylation of proteins in the cancer cell. In addition, a novel method was designed to accurately assess on-target activity of NMT inhibitors and the fate of *N*-myristoylated substrates across the whole proteome upon NMT inhibition. This new approach relies on post-lysis labelling of exposed N-terminal glycines by *S. aureus* sortase A (SrtA) and no longer relying on metabolic labelling, it can be applied to any type of biological sample.

Methionine aminopeptidases (MetAPs) catalyse initiator methionine (iMet) removal from nascent proteins and are essential to maintain healthy proteome dynamics by priming other N-terminal modifications such as *N*-acetylation or *N*-myristoylation and modulating protein localisation and stability. MetAP2 has been explored by pharmaceutical companies for decades for the treatment of cancer and obesity. However, the links between MetAP2 inhibition and phenotypic effects are still poorly understood. Here, a novel chemical proteomics workflow is proposed to elucidate the substrates of MetAP2 systematically and uncover the missing link between MetAP2 inhibition and phenotype. This new strategy is based on metabolic labelling of cells with the methionine analogue azidohomoalanine (AHA) and in combination with specific pharmacological inhibition of MetAP2, allowed identification of >70 substrates of MetAP2, 94% of which were unknown until reported here.

Together, this work provides fundamental insights into the biological role and importance of *N*-myristoylation and iMet excision in cancer and shapes the path for future steps in the use of NMT and MetAP2 inhibitors for the treatment of human disease.

To my dad,
may this thesis be
the one you never finished.

Acknowledgements

First, I would like to thank my supervisor, Ed Tate, for his guidance when I needed it, as well as the freedom to follow my own intuition most of the rest of the time. Thanks for believing in me even when I doubted myself and for counting on me to take on both internal and external collaborations. Most importantly, thanks for the opportunity to form part of the Tate group, not only a thriving environment for scientific research, innovation and self-improvement, but also for tolerance, respect and support.

I am deeply in debt to Remigiusz Serwa, who introduced me to the world of proteomics and patiently guided me through my first experiments. He was always there to push me to do that extra experiment when I was convinced it would not work. I would also like to thank him for answering every single phone call and discuss my progress even after leaving the Tate group. I can only hope I will pass on my knowledge and influence other student's scientific career one day the same way he impacted my own.

A special mention goes to Nattawadee (Kwang) Panyain, my lab companion along almost the whole of my PhD journey. From my very first day in the 5th floor 'biolab' office to submission of your thesis and then mine. I could not have asked for a better colleague and friend.

My sincere appreciation to Wouter Kallemeijn for the fruitful discussions, for his reliability and organisation skills. He always valued my advice and counted on me. We would have not been able to go through the move to White City without him.

A number of thanks are also due to my co-authors, Remigiusz Serwa, Julia Morales-Sanfrutos and Markus Ritzefeld, for their help bringing the SrtA project forward, which resulted in my first publication.

Many thanks to Cass, who briefly joined the MetAP2 project; to Monica Faronato, for her help with cloning and her patience dealing with me when I arrived with limited lab experience; and to Gregor Lueg, for his useful discussions on pathway analysis and data processing.

Thanks to everyone that helped me write this thesis through the coronavirus lockdown, both scientifically and emotionally. Also thanks to Remigiusz Serwa, Wouter Kallemeijn, Chiara Fabbro, Theodore Flack, Anja Schlott, Cyrille Kounde, Daniel Conole and Ed Tate, for taking the time to reviewing this work and provide excellent feedback that shaped this thesis to the quality it is today.

I would also like to extend my gratitude to every member of the Tate group who I did not have room to mention here. They have made me feel at home (in London, which I think is

quite an achievement!) and have almost become a second family to me.

I am particularly thankful to my parents, for trusting me even through their worries when I decided to take scientific research as my career path. Thanks for always nurturing my curiosity and encouraging me to give the best of myself.

Last and most importantly, I would like to thank the person who shared the core of my PhD journey with me. He was who convinced me to come to London in the first place and who pushed me through every step of it: from indecent housing and scientific setbacks to the best memories of this last four years. We have celebrated each other's victories and supported one another through our failures. For the long discussions on statistics and rigour, and the even longer biology-related explanations; for encouraging me to learn to code, even if clumsily; and for always bringing the best of me. Aitor Muguruza, I would have not been able to do this without you by my side.

List of figures

Chapter 1

Chapter 2

List of tables

Chapter 1

Chapter 6

Appendix A

Contents

Abbreviations

ACN: acetonitrile

AUC: area under the curve

BSA: bovine serum albumin

CETSA: cellular thermal shift assay

CI: confidence interval

CuAAC: copper-catalysed alkyne-azide cycloaddition

FA: formic acid

GNAT: Gcn5-related N-acetyltransferase

GO: Gene Ontology

GOBP: GO Biological Process

GOCC: GO Cellular Compartment

GOMF: GO Molecular Function

Haa: 2-hydroxyacetamide

HUVEC: human umbilical vein endothelial cells

(HMVEC: human microvascular endothelial cells

IGF: in-gel fluorescence

iMet: initiator methionine

indel: insertions and deletions (gene)

ITC: isothermal calorimetry

ITDR: isothermal dose-response

KD: knock down

KO: knockout

MetAP: methionine aminopeptidase

MetAP2i: MetAP2 inhibitor

mRNA: messenger RNA

NAT: *N*-α-acetyltransferase

NMT: *N*-myristoyltransferase

NMTi: NMT inhibitors

PBS: phosphate buffered saline

PDF: peptide deformylase

POEP: protection of eIF2α phosphorylation

PolI, Pol II, Pol III: RNA polymerase I, II and III

PTMs: protein post translational modifications

RPs: ribosomal proteins

RPLs: ribosomal proteins of the large subunit

RPSs: ribosomal proteins of the small subunit

RPBs: ribosome-associated protein biogenesis factors

rRNA: ribosomal RNA

RT: room temperature

SILAC: stable isotope labelling of amino acids in culture

SRB: sulforhodamine B

TBS: Tris buffered saline

TFA: trifluoroacetic acid

WB: Western blot

1

Introduction

1.1 Prevalence and importance of protein post-translational modifications

It is now widely established that complexity of higher organisms is not necessarily dictated by their genomic content. For instance, the estimated total number of genes does not necessarily reflect organism complexity: 6,000 genes in yeast (*S. cerevisiae*), 20,000 genes in humans, 30,000 genes in mouse (*M. musculus*) and 51,000 genes in rice (*O. sativa*) [1]. Instead, it is during the processes laying downstream, during transcription, translation and post-translational protein regulation, where the actual complexity arises, driven by the intricate set of interactions and interplay among the different components of these processes. Thus, the molecular biology dogma has been revised numerous times to include mechanisms by which complexity might emerge in each of these steps. These include alternative start sites during transcription [2], alternative splicing or RNA editing during mRNA processing [3, 4], and different start sites during translation [5]. Regulatory diversity does not end there, however. Living organisms have evolved a much more finely tunable mechanism to control cellular processes beyond regulation of mRNA transcription and protein synthesis rates: protein modifications.

More than 400 different protein modifications have been described to date [6, 7]. While they can occur co- or post-translationally, the term 'post-translational modification' (PTM) is generally used to refer to any of these. This usage is primarily due to the fact that post-translational modifications were identified and characterised before the existence of co-translational modifications was fully appreciated. Recent findings powered by technological advances have revealed a much higher diversity of protein modifications than originally expected, at a remarkably high prevalence in the proteome. Almost all eukaryotic proteins are now believed to bear at least one (if not multiple) co- or post-translational modifications [7]. They not only influence protein function directly (e.g. one phosphorylation can switch one enzyme from its inactive to its active form), but also indirectly, by promoting or blocking associations between proteins, altering protein localisation or protein half-life (Figure 1.1).

Some PTMs are irreversible, affecting protein fate throughout the life-span of a protein, and some are reversible, allowing for fine-tuning upon changing conditions (cell cycle, circadian rhythms, environmental cues, etc.). Furthermore, the functional relevance of these modifications has been shown to depend upon their localisation on the protein sequence, with the same type of PTM (e.g. phosphorylation) capable of exerting activatory or inhibitory roles depending on which region of the protein it modifies [8]. As a result, the proteome is a much larger entity and far more dynamic than the genome [9]. The spatio-temporal dynamism conferred by protein modifications allows for rapid and tight control of essential processes within every cell. Accounting for all the aforementioned sources of variation including protein modifications, it has been estimated that in a given human cell there are 0.6-6 million protein species [10]. With the aim to describe this complexity, a new term was coined by L. Smith

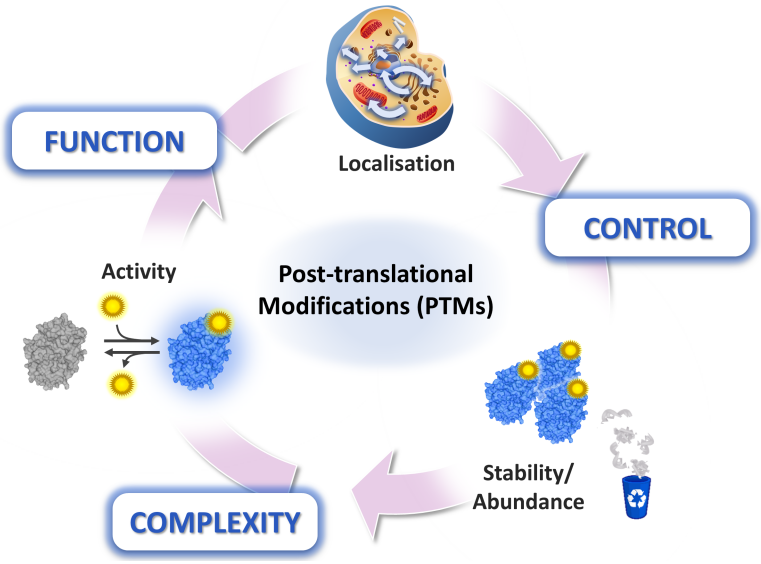


Figure 1.1: By regulating protein activity, localisation and abundance, PTMs modulate protein function in a tightly regulated manner and contribute greatly to proteome complexity.

and N. Kelleher:'*proteoform*', which encompasses the combination of the final specific amino acid sequence of a given protein entity and the set of all its modifications at a given cellular state [11, 12].

1.2 Protein N-terminal modifications

The protein N-terminus is the starting point of the linear polypeptidic sequence; that is, the first amino acids to emerge out of the ribosome during synthesis. As it starts with a free amino group (-NH_2), it is referred to as "N-terminus" (or "amino" terminus). The opposite end of the sequence, the last one to be released by the ribosome, is referred to as "C-terminus", for the free carboxyl group (-COOH) that characterises the end of the polypeptide sequence. When we refer to the N-end of the protein, therefore, it is consensus not to italicise the "N", while it is generally italicised when the "N" is being used to denote the atom to which the modification is being linked to, such as in *N*-acetylation or *N*-myristoylation, where the modification takes place on the nitrogen atom of the α-amino terminus of the protein.

Protein N-terminal modifications comprise the set of modifications that take place at the N-terminus of proteins and as such, they can in theory happen in any protein. While historically less studied in comparison to more popular modifications such as phosphorylation, N-terminal modifications have gained a considerable amount of attention in the last decade, revealing their ubiquity, prevalence and relevance [6].

They often occur co-translationally as the nascent polypeptide emerges from the ribosome, and hence, constitute the first set of modifications any protein undergoes. They are mediated by specific enzymes, which are part of the ribosome-associated protein biogenesis factors

(RBPs) [13]. These are non-ribosomal proteins that act in close contact with the ribosome and recognise their substrates as soon as the nascent peptide emerges from the ribosome exit tunnel. They not only catalyse the addition of co-translational modifications, but are also responsible for ensuring protein folding, serving as quality controls and for the recognition of protein localisation signals and subsequent translocation into the organelles. This way, interacting with the mRNA, the tRNAs, translation factors and modulators and also RBPs, the ribosomes act as a cellular hub to coordinate translation in a spatially and temporally regulated manner [14]. Moreover, altering the chemical properties of the N-terminal α-amine, N-terminal modifications have been shown to play a decisive role on protein fate through the different N-end degron pathways, where protein degradation rates depend on the identity of the N-terminal residue [15](cf. section 1.2.8 for further details on N-degron pathways).

There are $>$15 different protein N-terminal modifications described to date and their functions are as varied as their chemistries. The most prevalent ones (initiator methionine (iMet) excision [16], N-terminal acetylation [17] and *N*-myristoyation [9]) occur co-translationally, while the protein is still being synthesised in the ribosome and are involved in a myriad of functions including regulation of protein localisation, stability and binding partners. Other less well-studied modifications include some that occur conditionally, as a response to altering environmental cues and are part of signalling cascades in response to a change. These include oxidative-stress mediated N-terminal modifications such as α-ketoamine formation [18], transamination [18], cyclisation [19] or epimerisation [20] and hypoxia-mediated N-cysteine oxidation [21]. Some N-terminal modifications have been described only in orchestration with their role in N-degron pathways (e.g. N-deamidation, N-arginylation) [15]. Others are more rare, with only a few proteins identified with that modification (e.g. *N*-palmitoylation [22]) or have not been that well characterized and their functional impact is still not clear (*N*-methylation [23], *N*-ubiquitination [24], *N*-propionylation [25] or conversion of N-terminal glutamate and glutamine to pyroglutamate [26]) (Figure 1.2).

iMet excision and *N*-myristoylation modifications are discussed in more detail in the next sections of this chapter due to their direct relevance to this work (cf. section 1.3 for iMet excision and section 1.4 for *N*-myristoylation). However, it is worth noting that all N-terminal modifications potentially impact the occurrence of other equally important modifications at this unique site of the protein. For example, blocking the activity of specific enzymes involved in the processing of these modifications (e.g. by NMT or MetAP2 inhibitors), might leave an otherwise blocked or unmodified N-terminus free for other modifications to occur. Therefore, it was deemed relevant to list and review, even if briefly, the most important aspects of these other less known N-terminal modifications as well as to include some insights into their interplay and orchestration. Table 1.1 summarises the main information discussed for each of the N-terminal modifications, which is also described in the following pages in more detail.

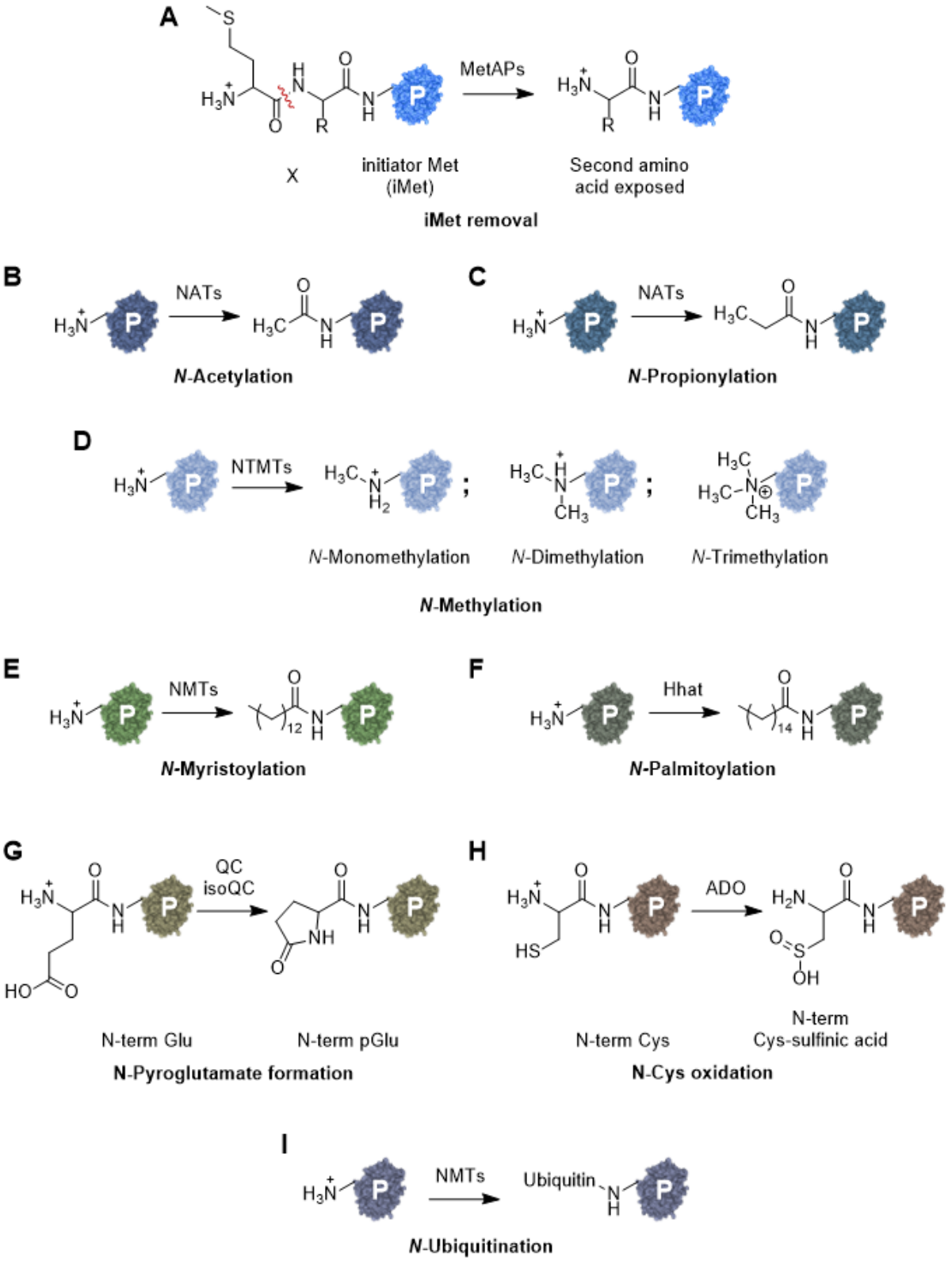


Figure 1.2: Major types of N-terminal modifications: (A) initiator methionine (iMet) removal, **(B)** *N*-acetylation, **(C)** *N*-propionylation, **(D)** *N*-methylation, **(E)** *N*-myristoylation, **(F)** *N*-palmitoylation, **(G)** N-pyroglutamate (pGlu) formation, **(H)** N-cysteine oxidation and **(I)** *N*-ubiquitination. P = rest of the protein.

Table 1.1: Summary of N-terminal modifications: enzymes responsible, number of proteins known to bear these modifications and known functions.

Modification	Enzyme	Enzyme localisation	Co-/Post-translational	Reversible?	Prevalence	Functions
iMet removal	MetAP1 MetAP2	Cytosol Cytosol	Co-	No	>70% of proteome (humans)	Stability (N-end degron) Protein-protein interactions Met metabolism
***N*-acetylation**	NATs (NatA-NatF)	Cytosol (However, some auxiliary subunits seem to have other localisations as well)	Co- (main) Post- (rare)	No	>80% of proteome (humans)	Affiity-enhancing Stability (Ac/N-end degron) Secretion Membrane targeting Protein folding
***N*-propionylation**	NATs	(same as above)	Co- (by similarity)	No	Rare (<20 examples)	Not known
***N*-methylation**	NTMT1 NTMT2	Nucleus Nucleus	Post- (upon translocation into nucleus)	No	Rare (<20 examples)	Protein-protein interactions DNA binding
***N*-myristoylation**	NMT1 NMT2	Cytosol, PM[1] Cytosol, PM[1]	Co- (main) Post- (apoptosis)	No	1-4% of proteome	Reversible membrane binding 3D structure Protein-protein interactions Stability (Gly/N-end degron) Role in apoptosis
***N*-palmitoylation**	Hhat (vertebrates) Rasp (flies)	PM[1]	Post-	No	Very rare (Hh proteins, Spitz)	Multimerisation Signalling upon secretion
Pyroglutamate formation	QC isoQC	Secreted Golgi apparatus	Post-	?	Rare (<20 examples) (hormones and chemokines)	Increased stability Receptor binding Potency modulation
N-Cys oxidation	ADO (humans) PCO1, PCO2 (plants)	Cytosol	Post-	No	Rare (<10 examples)	Arg/N-end degron Hypoxia signalling
***N*-ubiquitination**	Ube2w (E2) HUWE1 (E3)	Nucleus, nucleolus Nucleus	Post-	Yes	>100	Not known

[1] PM, plasma membrane.

1.2.1 *N*-acetylation

In human cells, N-terminal acetylation is probably the most prevalent modification, currently on ca. 80% of all proteins [27]. Its frequency is still high in lower eukaryotes (ca. 50% yeast) but less prevalent in lower organisms like archaea and rare in bacteria [6, 27, 28]. It happens mostly co-translationally, while the polypeptide chain is only about 25-50 amino acids long [29]. It has also been described in plant cell organelles, where it has been detected in more than 30% of the nuclear-encoded chloroplastic proteins [6]. In this case, *N*-acetylation happens post-translationally, upon translocation of the protein into the chloroplast.

N-acetylation is catalysed enzymatically by a set of enzymes called *N*-α-acetyltransferases (NATs) (Figure 1.2B). In eukaryotes these are protein complexes mostly formed by two protein subunits: a catalytic subunit and an auxiliary one. The catalytic subunit binds acetyl-CoA and transfers it onto protein substrates while the auxiliary one has been found to be necessary for the catalytic activity although not taking direct part on the transfer. Auxiliary subunits

Table 1.2: Types of eukaryotic NATs, subunit composition and substrate preferences.

NAT	Catalytic subunit	Auxiliary subunit(s)	Substrates					
NatA	Naa10 (or Naa11[1])	Naa15 (or Naa16[1])	Ser-	Ala-	Thr-	Gly-	Val-	Cys-
NatB	Naa20	Naa25	Met-Asp-	Met-Glu-	Met-Asn-	Met-Gln		
NatC	Naa30	Naa35 + Naa38	Met-Leu-	Met-Phe-	Met-Ile-	Met-Trp-		
NatD	Naa40	-	Ser- (H2A and H4 histones)					
NatE	Naa50	NatA (Naa10 + Naa15)	Met-Leu-	Met-Phe-	Met-Ile-	Met-Trp-		
NatF	Naa60	-	Met-Leu-	Met-Phe-	Met-Ile-	Met-Trp-	Met-Lys-	

[1] Naa11 and Naa16 are orthologs of Naa10 and Naa15 and are found in higher eukaryotes. Their combinations give rise to four different types of NatA in these organisms.

have also been reported to be important for ribosome binding [30, 31]. NATs belong to the Gcn5-related N-acetyltransferase (GNAT) protein superfamily, a large family of enzymes that share a common 3D fold and use acyl-CoAs as donors to acylate their substrates [32]. Six different complexes have been discovered in eukaryotes to date: NatA, NatB, NatC, NatD, NatE and NatF (Table 1.2, adapted from [6]). Most of the differences in their substrate specificities have been generally assigned to the identity of the first two amino acids of their substrates [33], but it is increasingly evident that residues beyond position two might also play a role [17, 34]. NatA, NatB and NatC are responsible for most of the *N*-acetylations in the cell, where NatA (Naa10+Naa15) primarily acetylates proteins after iMet removal and NatB (Naa20+Naa25) and NatC (Naa30+Naa35+Naa38) acetylate the retained iMet when the second residue is acidic (NatB) or hydrophobic (NatC) [6]. NatD (Naa40) has been described to almost exclusively *N*-acetylate histones H2A and H4 [35]. NatE (Naa50+NatA) was discovered when Naa50 was found to associate with NatA without affecting its activity and later found to acetylate iMet-retained substrates on its own [30, 36]. Finally, NatF (Naa60), which is present in humans but not in yeast, acetylates proteins showing an intact iMet [37] and has been reported to localise to the Golgi as well as linked to the *N*-acetylation of transmembrane proteins, probably post-translationally [38]. It is worth noting that the sequence specificities are not absolute and that no complete consensus exists that can accurately predict the presence of *N*-acetylation [6]. Further, there is apparent overlap for the same substrates not only among NATs (Table 1.2), but also with other N-terminal PTMs, where competition for different processing has been proposed [39, 40].

There is no known *N*-α-deacetylase, so *N*-acetylation is believed to be irreversible. However, recent studies have revealed certain proteins that can appear both in the *N*-acetylated and

free N-terminus forms and that these can co-exist within the cell [34]. The origin and functional meaning of these pools of proteins, which involved about 9% of all identified N-termini [34], still requires further study, but some theories have been proposed: (i) NATs only recognise certain substrates weakly, leading to repeated miss of those substrates; (ii) activity of NATs might be dictated by Ac-CoA availability and protein turnover or (iii) there might exist regulation at the level of ribosome binding [6].

Acetylation not only blocks the N-terminal α-amino group from further post-translational processing, but also neutralises its positive charge [41]. *N*-acetylation has been shown to act as an affinity enhancer, increasing interaction potential with binding partners [42]. It has also been reported to influence membrane targeting and secretion [43–45]. For example, *N*-acetylation of Arl8b small GTPase has been shown to be responsible for lysosomal targeting [43]. Moreover, it has been shown to increase affinity of α-Synuclein for physiological membranes by stabilising the helicity of its N-terminus [44]. In addition, *N*-acetylation has been reported to prevent targeting of certain proteins to the ER and their subsequent sorting to the secretory pathway [45]. Recently, the importance of this modification on forming specific N-terminal degrons has been highlighted [46, 47], with one specialised N-degron pathway (denoted Ac/N-degron pathway) dedicated to the degradation of *N*-acetylated proteins (refer to section 1.2.8 for further information on N-degron pathways). Further, it has been shown to play a role in protein folding and proteostasis of the yeast prion [PSI^+] [48]. When *N*-acteylation was lost, it led to general misfolding of [PSI^+] and a subsequent stress response as well as increased degradation, together reducing the overall size of the prion aggregates [48].

1.2.2 *N*-propionylation

N-propionylation (Figure 1.2C) had gone unnoticed until relatively recently, when in the second half of the 2000s two independent systematic proteomics studies of the human N-terminal proteome revealed the presence of a few proteins modified with a propionyl at their α-amino group [49, 50]. A few years later, in a dedicated study in yeast it was shown that NATs (in particular, NatA) could also attach a propionyl group at the N-terminus and so act as *N*-propionyltransferases in addition to *N*-acetyl transferases both *in vitro* and *in vivo* [25]. The late discovery goes along the supposedly low rates of this modification: it is very rare, with less than 20 examples to date including both human and yeast studies. This could be partially explained by the availability of propionyl-CoA in the cell, which is 2-20-fold less abundant than the more common Acetyl-CoA [25, 51]. Due to the few examples known to date, there is not much known about the function of *N*-propionylation [41]. The additional methylene group as compared to acetyl-CoA results in increased hydrophobicity and bulkiness (Figures 1.2B-C), but it remains to be discovered how this will impact protein function as opposed to *N*-acetylation. Since no "*N*-depropionylase" has been found to date, and similarly to *N*-acetylation, *N*-propionylation is considered irreversible [52].

1.2.3 *N*-methylation

Biochemically, *N*-methylation consists of the enzymatic transfer of a methyl group from *S*-adenosyl methionine (SAM) to the exposed N-terminal α-amino group of certain amino acids [23] (Figure 1.2D). The substrate consensus sequence is X_1-X_2-Lys/Arg, where X_1 can be any amino acid except Asp or Glu and X_2 can be any uncharged amino acid (polar/nonpolar), although Pro is preferred [23, 53]. It is catalysed by two main *N*-methyltransferases: NTMT1 and NTMT2 (also known as NRMT1/2 or METTL11A/B). Both enzymes share both sequence and structural similarity and are localised to the nucleus [54], albeit with different tissue distributions [23, 55]. The main difference between the two enzymes is that NTMT2 seems to preferentially catalyse monomethylation reactions, whereas NTMT1 can catalyse the addition of one, two or three methyl groups [54]. The number of methyl groups affects the chemical properties of the modification. A single methyl group only alters the pK_a of the alpha amino group slightly and has a minor steric impact on binding. In contrast, trimethylation (or dimethylation in case of Pro), not only grants increased bulkiness, but also completely abrogates the nucleophilicity of the α-amino group and confers a constant positive charge that is not dependent on pH [56]. Since no *N*-demethylase is currently known, *N*-methylation is considered irreversible [52].

Although the existence of *N*-methylation has been known for over 40 years, its true relevance hasn't been recognised until the last decade. Early reports include dimethylation (Pro) of cytochrome c557 of the *C. oncopelti* insect parasite [57], trimethylation (Ala) of myosin light chain [58] and monomethylation (Pro) of Drosophila H2B histone [59]. These seemed to indicate a role for N-terminal methylation in regulating protein-protein interactions. More recent studies have established a novel crucial role of *N*-α-methylation in DNA binding and chromatin remodeling. More specifically, in mitosis, due to its role ensuring stable chromatin association of RCC1 [60, 61]; in nucleotide excision repair, through DDB2 [62]; and in centromere assembly, where it ensures efficient binding of CENP-A [63] and CENP-B [64] to centromeric DNA. Moreover, a recent report showed the importance of *N*-methylation of the yeast proteasome ATPase subunit Rpt1 for normal proteasomal function [65].

Recently, a third enzyme capable of methylating the α-amino terminus in humans was discovered, called METTL13 [66]. It is a dual methyltransferase capable of catalysing both N-terminal Gly and Lys55 of human eEF1A. Interestingly this enzyme was found to be present in the cytoplasm and mitochondria in addition to the nucleus, suggesting *N*-methylation might also happen in these two other locations.

1.2.4 *N*-palmitoylation

N-palmitoylation refers to the attachment of palmitate by a covalent amide bond to the N-terminal α-amine of the protein [22] (Figure 1.2F). It is quite a unique modification, as

palmitoylation usually occurs via thioesther linkage in internal cysteines. There are only a handful of proteins known to be modified this way: the hedgehog proteins, Hedgehog (Hh) in *D. melanogaster* [67] and Sonic Hedhehog (Shh) in vertebrates [68], and the *D. melanogaster* protein Spitz [69]. It is believed to extend to the two other members of the hedgehog family in vertebrates, Indian hedgehog (Ihh) and Dessert hedgehog (Dhh) due to sequence similarities (UniProt [70]) although it's presence has not been experimentally confirmed yet.

All hedgehog proteins are synthesised as precursor proteins, with an N-terminal signal peptide that directs them to the ER and that is cleaved upon translocation revealing an N-terminal Cys [22]. Hh proteins then undergo autoproteolytic cleavage at the C-terminus, revealing a newly formed C-terminal Gly, which is then cholesterylated at the C-terminal carboxylate. This anchors Hh proteins to the luminal side of the ER membrane, where hedgehog acyltransferase (Hhat) [71] in vertebrates or Rasp (previously known as Skinny Hedgehog [67]) in flies transfers the palmitic acid group. Hhat and Rasp belong to the membrane bound *O*-acyl transferase (MBOAT) family [22]. Interestingly, Hhat and Rasp first attach palmitate onto the Cys side chain, through regular S-acylation via thioester link, but this then rearranges itself through an intramolecular S-to-N shift to yield the final amide linked palmitate (*N*-pal) [22]. Thus, through this unusual mechanism palmitate is attached covalently and irreversibly to the N-terminal α-amino group of the N-terminal Cys of Hedgehog proteins. Moreover, Rasp has also been shown to *N*-palmitoylate the fly protein Spitz, a ligand for the *D. melanogaster* epidermal growth factor receptor (EGFR) [69].

Shh/Hh are secreted morphogens that are essential for embryogenesis and tissue development, through their important role in proliferation and differentiation [72]. Shh has also been implicated in wound healing, but should otherwise be silent in most adult tissues [73]. However, it has been shown to be aberrantly expressed in a variety of cancers [73, 74], where expression of Shh usually correlates with worse prognosis [75]. Importantly, dual lipidation of hedgehog proteins has been shown to be essential for their function [68]. In particular, N-pal has been shown to be necessary for Hh [76] and Shh signalling [77], allowing for correct multimerisation that enables signalling to distant responsive cells [77]. Due to its essential role for Hh signalling, Hhat has been proposed as a target to inhibit Shh signalling in cancer [74, 78].

An additional protein has been proposed to be *N*-palmitoylated: the G-protein alpha subunit S ($G\alpha_S$) [79]. In this case, the palmitate was shown to be positioned at the α-amine of the N-terminal Gly (Gly2). This unusual position could be explained by the presence of a Cys in the third position (Cys3) (N-terminal sequence (M)GCLGNSK, where M is removed by MetAPs), where the *N*-palmitate could be attached first before following a rearrangement similar to the one in Hh proteins. However, further validation is needed to confirm cellular occurrence and its functional relevance.

1.2.5 N-terminal conversion of Glu and Gln into pyroglutamate

Catalysed by glutamyl cyclase (QC, from the *QPCT* gene) or isoglutamyl cyclase (isoQC, gene *QPCTL*), this modification has been described for a number of proteins and biologically active peptides [80] (Figure 1.2G). Both enzymes seem to have the same substrate selectivity, but different localisation: while QC is mostly secreted, isoQC is almost exclusively localised to the Golgi apparatus [81].

One important function of pyroglutamate (pGlu) is to protect the N-terminal α-amine from aminopeptidases, thus ensuring increased stability and prolonging the life-span of the proteins [26]. In this way, it has been shown to regulate the biological activity of some peptide hormones (Thyrotropin-releasing hormone (TRH), gonadotropin releasing hormone (GnRH) or gastrin), neuropeptides (neurotensin) and chemokines (MCP1/CCL2, MCP2/CCL8, MCP3/CCL7,
MCP4/CCL13 and CX3CL1) [26, 80–82]. In some cases, it has also been shown to affect function directly by modulating potency and receptor binding capacity (TRH) [26] or chemotactic activity (MCP1/CCL2 [83], MCP2/CCL8 [82] and CX3CL1 [81]).

Moreover it has also been linked to some neurodegenerative diseases. In Alzheimer's disease, most of the amyloid-beta (Aβ) peptides in amyloid plaques have a pGlu at their N-terminus, likely accumulated through their resistance to proteolysis [84]. Further, the presence of pGlu can result in accelerated aggregation by increased hydrophobicity and consequent reduction in solubility [85]. In some rare familiar dementias, aggregated peptides also show an N-terminal pGlu [86].

Due to their role in inflammatory processes and disease, inhibition of glutamyl cyclase has been proposed for Alzheimer's disease [87], liver inflammation models [88] as well as proposed as a strategy to prevent thyroid cancer metastasis [89].

1.2.6 N-Cys oxidation

Recent studies have revealed a new mechanism for fast oxygen sensing in both plants and human cells that involves modification of N-terminally exposed Cys into Cys-sulfinic acid [21, 90, 91] (Figure 1.2H). They found that opposite to what was thought, N-terminal oxidation of these hypoxia-sensitive substrates was enzymatically regulated. Plant cysteine oxidase 1 and 2 (PCO1, PCO2) in plants [90, 91] and cysteamine (2-aminoethanethiol) dioxygenase (ADO) in human cells [21] are responsible for catalysing the dioxygenation of the N-terminal Cys side chain of specific substrates under normoxic conditions, leading to degradation of the oxidized substrate through arginylation by Ate1 and the Arg/end N-rule pathway (further explanations on the Arg/N-degron pathway can be found in subsection 1.2.8.2).

In plants, this was shown to lead to stabilisation of group VII ethylene response factors (ERF-

VIIs) upon hypoxia, which promote plant survival under these unfavourable conditions [91]. In human cells, in contrast, ADO was reported to label RGS4/RGS5, two G-protein signalling regulators, and the angiogenic interleukin-32 (IL-32) for degradation. Thus, ADO was shown to control activation of the mitogen-activated protein kinase (MAPK) cascade and of new angiogenic events in response to hypoxia [21]. These differences are likely to reflect the distinct environmental adaptation mechanisms plants and humans must undertake to ensure their survival under adverse conditions like hypoxia.

These are all very recent discoveries, and therefore, it is likely that there are more substrates of plant PCOs and human ADO that are yet to be discovered and that will expand the network of processes that ensure survival under hypoxia. This pathway is not found in yeast, where N-Cys is not a destabilizing residue. Interestingly however, recombinant proteins bearing an N-terminal Cys could be made sensitive to hypoxia in yeast as well, through the introduction of human ADO [21]. This establishes a more widely applicable role for enzymes involved in this pathway that might well be conserved amongst other eukaryotic species.

1.2.7 *N*-ubiquitination

N-terminal ubiquitination refers to the addition of ubiquitin to the N-terminal α-amine of the protein (Figure 1.2I). It should not be confused with ubiquitination through the N-degron pathway, where the addition of ubiquitin happens on the side-chain of internal lysines after recognition of the N-terminal degron sequence.

The position of this modification was discovered through Lys-replacement studies on the myogenic transcriptional switch MyoD, which was found to still be ubiquitinated and degraded when all lysines in its sequence where mutated [92]. It has since been discovered for many more proteins including the lysine-free human papillomavirus 16 oncoprotein E7 [93], the Epstein Barr virus latent membrane proteins LMP1 [94] and LMP2A [95], the differentiator regulator Id2 [96], the cell cycle regulator p21 [97, 98], the signalling kinase ERK3 [98], the small ubiquitin-related modifier SUMO2 and the E3 ligase CHIP [99], the neurodegenerative disease protein Tau and the polyglutamine disease protein ataxin-3 [100]). A recent systematic study increased the list to ca. 100 proteins [24].

Two enzymes of the ubiquitin-proteasome system have been identified to be capable of *N*-ubiquitination: the ubiquitin conjugating (E2) enzyme Ube2w (which has been shown to autoubiquitinate itself as well as mono-ubiquitinate the N-termini of SUMO2, CHIP, Tau and ataxin-3 [99, 100]) and the ubiquitin ligase (E3) HUWE1, described for the *N*-ubiquitination of MyoD but only when all Lys are mutated[101]. Interestingly, Ube2w differs structurally from other known E2 enzymes, and according to the authors, these structural differences would promote interaction with the α-amine rather than a more positively charged Lys-side chain [100].

The study of N-terminally ubiquitinated proteins has been limited to date because most widely used approaches for the systematic study of ubiquitination are not able to detect it. On the one hand, proteomics approaches that rely on enrichment through ubiquitin binding domains (UBDs) [102] or epitope tagged ubiquitin expression [103–105] are not capable of providing site-specific information. On the other hand, approaches that depend on immunoenrichment with K-ϵ-GG antibody [106] are limited to detection of Lys ubiquitination. A recent report using a new type of antibody capable of recognising a longer fragment of the Ub remnant after cleavage with the endoproteinase LysC revealed new insights into this type of N-terminal modification [24]. LysC cleaves ubiquitin leaving a 13 amino acid remnant covalently attached to proteins, which are recognised by this new ubiquitin antibody. By not relying in the specific recognition of an ubiquitinated Lysine, it was shown to also serve for the discovery of *N*-ubiquitination events[24]. In this study Akimov and colleagues showed that *N*-ubiquitinated proteins were distributed across all protein expression levels and that expression levels of these proteins did not change upon proteasome inhibition, which suggested a broader role for this PTM beyond degradation, in contrast to the focus of most previous studies [24]. Expansion of the list of proteins known to be N-ubiquitinated also revealed that it can happen on the iMet or on the exposed second amino acid residue after iMet removal, when this is a small hydrophobic amino acid, especially when Pro or Val are present in the second position [24]. This contrasts with previous reports on Ube2w specificity, where it was found to preferentially recognise 'intrinsically disordered' N-termini and to be inhibited by Pro residues on positions 2-4 [107]. However, this could simply indicate that there is more than one enzyme involved in N-terminal ubiquitination. Remarkably, the patterns for *N*-ubiquitination showed a striking negative correlation with N-terminal acetylation [24], highlighting the temporal control of N-terminal modifications, in which post translational modifications like *N*-ubiquitination can only happen on free N-termini left unmodified by preceding co-translational modifications.

1.2.8 N-degron pathways

The N-degron pathway (previously called N-end rule pathway) was discovered for the first time in 1986 and initially described the relationship between the half-life of a protein and the identity of its N-terminal residue [108]. Since then, all 20 amino acids have been shown to be destabilizing in specific sequence contexts [15]. Moreover, C-degrons have also been described, which act analogously to N-degrons, but at the opposite end of the protein [109]. N-degrons are recognised by a set of substrate-specific E3 ligases or other proteins called N-recognins, which tag them for proteasomal degradation (or the proteasome-like ClpAP protease in bacteria) by ubiquitination of nearby internal lysines [110].

Interestingly, although N-degron pathways exist in all kingdoms, they are not conserved across prokaryotes and eukaryotes. In bacteria, there are two main N-degron pathways: the formyl-methyonine (fMet)/N-degron pathway and Leu/N-degron pathways [15]. In eukaryotes, six types have been described so far, which include the fMet/N-degron pathway in organella

(mitochondria, choroplasts), the fMet/N-degron pathway in the cytosol of the yeast *S. cerevisiae* and the Pro/N-, Ac/N-, Arg/N- and Gly/N-degron pathways in eukaryotic cytoplasms [15]. Below are described the four N-degron pathways occurring in the cytoplasm of higher eukaryotes (Figure 1.3).

1.2.8.1 The Ac/N-degron pathway

This pathway targets most acetylated N-terminal residues, which can be subclassified according to their propensity to being acetylated [15]: (i) usually *N*-acetylated (iMet, Ala, Ser, Thr), (ii) rarely *N*-acetylated (Val and Cys) and (iii) very rarely *N*-acetylated (Gly and Pro) (Figure 1.3A).

Known Ac/N-recognins include Doa10 in yeast and Teb4 in mammals (both ER membrane-embedded) and Not4, the E3 subunit of the multifunctional cytosolic/nuclear complex Ccr4-Not [46]. N-terminal acetylation is a very a common modification in eukaryotes, happening in ca. 80% of the proteome in humans (cf. ref [27] and section 1.2.1 for further information). Being such a common modification, it has been proposed that this N-degron depends on exposure for efficient recognition by the Ac/N-degron pathway [47]. Acetylation has been proposed as an affinity-enhancing modification, playing an important role in protein-protein interactions [42]. Hence, it has been proposed that the Ac/N-degron pathway might act as a complex regulator through subunit quality control and stoichiometry control mechanisms. Thus, subunits displaying weaker binding affinities for their binding partners or no longer needed are recycled to maintain exquisitely regulated complex dynamics across changing conditions [47].

1.2.8.2 The Arg/N-degron pathway

This pathway targets most unacetylated N-terminal residues, which can be subclassified into three categories [15, 111, 112] (Figure 1.3B):

- Primary, residues directly recognised by Arg/N-recognins: Arg, Lys, His, Leu, Phe, Tyr, Trp, Ile and iMet-Ile, iMet-Leu, iMet-Arg, iMet-Lys. In yeast, iMet followed by other bulky hydrophobic residues like Trp, Phe or Ile were also reported [111] but these did not translate into human cells [112].
- Secondary, residues recognised upon arginylation by Ate1 R-transferase: Asp, Glu and oxidized Cys.
- Tertiary, residues recognised upon deamidation by Nta1 (yeast) and specific Ntan1 and Ntanq1 (animals and plants) N-terminal amidases followed by arginylation by Ate1: Asn and Gln.

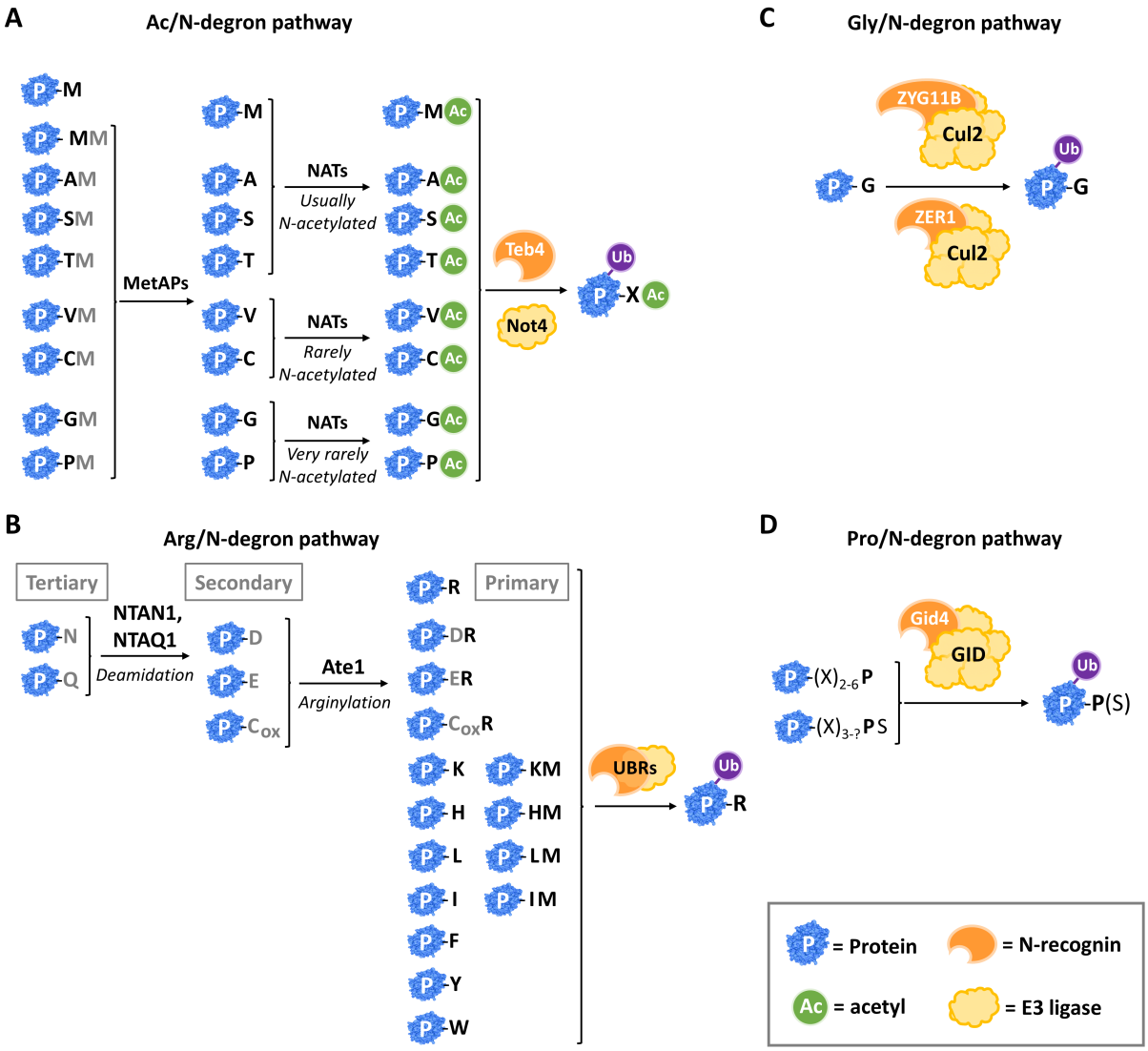


Figure 1.3: Main N-degron pathways on the cytoplasm of higher eukaryotes: (A) the Ac/N-degron pathway, **(B)** the Arg/N-degron pathway, **(C)** the Gly/N-degron pathway and the **(D)** Pro/N-degron pathway. Where N-recognin and E3 ligases responsible for these pathways are different for yeast and mammals, the mammal ones are displayed. P = rest of the protein.

Ubr1 is the only known Arg/N-recognin in yeast but it acts in complex with another E3 ligase, Ufd4, which increases its polyubiquitination capability [113]. In mammals, four Arg/N-recognins have been described: Ubr1, Ubr2, Ubr4 and Ubr5 [114].

Regulation of protein degradation through the Arg/N-degron pathway has been proposed to be similar to that of the Ac/N-degron pathway: proteins that are either misfolded or that fail to rapidly bind to either binding partners or chaperones (e.g. Hsp90 system) have been shown to serve as targets for this pathway [115, 116].

Moreover, the Arg/N-degron pathway has been proposed to play a role as a sensor of several molecules allowing the cells to adapt to their environments. Through Cys oxydation, it has been proposed as a sensor of NO and O_2 (cf. section 1.2.6) [21]. It has also been reported as

a sensor of short peptide availability, regulating their uptake in yeast [117]. The Arg/N-degron pathway also plays a role in the sensing of hemin (the Fe^{3+}-heme), which down-regulates Ate1 R-transferase both as a low micromolar inhibitor and by promoting its degradation [118].

The Arg/N-degron pathway has also been linked to autophagy through p62/Sqstm1, a component of the autophagy-lysosome system that was also found to function as a non-E3 Arg/N-recognin [119]. Stresses such as heat shock and the unfolded protein response where shown to induce translocation of ER proteins back into the cytosol, where they are N-terminally arginylated followed by lysosomal degradation [119].

1.2.8.3 The Gly/N-degron pathway

A recent study by Timms and colleagues discovered a new N-degron pathway that targeted proteins bearing an N-terminal Gly for degradation [112] (Figure 1.3C). The Cullin-RING ligase Cul2, together with ZYG11B and ZER1 adaptors where shown to be the N-recognins of this N-degron pathway. Both $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$ act complementarily, with most substrates being shared between the two of them but showing some differences on the N-degrons they recognise [112]. In intact proteins, Gly/N-degrons appeared to be depleted of the proteome [112]. Instead, this pathway was shown to lead to rapid degradation of newly exposed N-terminal glycines when *N*-myristoylation fails or upon caspase cleavage during apoptosis [112]. Thus, this N-degron pathway was proposed to act as a quality control mechanism for *N*-myristoylation and to ensure no longer needed fragments are degraded during the apoptotic cascade.

1.2.8.4 The Pro/N-degron pathway

There is only one example known of this N-degron pathway and it mediates the degradation of gluconeogenic enzymes when the yeast *S. cerevisiae* is moved into a medium containing glucose. This is mediated by the multisubunit GID Ub ligase, of which the Gid4 subunit is the N-recognin that recognises Pro at either the N-terminus or at the second position (Figure 1.3D). This pathway acts on an array of gluconeogenic enzymes including fructose-1,6-biphosphatase (Fbp1), isocitrate lyase (Icl1), malate dehydrogenase (Mdh2) and phosphoenolpyruvate carboxykinase (Pck1), all of which are targeted for degradation when glucose becomes available in the medium [120].

1.3 N-terminal methionine excision

All start codons code for either an *N*-formylated methionine (formylmethyonine, fMet) in prokaryotes and eukaryotic organelles (mitochondria and chloroplasts) or for a 'free' methionine in archaea and the eukaryotic cytoplasm [121]. Hence, all proteins are initially synthesised with either a 'protected' fMet or a 'free', amino-exposed Met. In bacteria and organelles, peptide deformylases (PDFs) remove the formyl moiety from the N-terminus of the initiator fMet in the majority of the cases [122], leaving an amino-exposed methionine identical to that of eukaryotes and archaea. This first 'free' methionine of the sequence is often referred to as initiator methionine (iMet) to distinguish it from other methionines in the sequence and highlight its specific role in initiating protein synthesis. In about 50-80% of the cases depending on the organism, this is further processed by methionine aminopeptidases (cf. section 1.3.1), which cleave iMet off exposing the second amino acid residue of the sequence at the N-terminus [6]. These first protein modifications happen very fast and while the nascent peptide is still bound to the ribosome. It has been estimated that both PDFs and MetAPs bind and process their substrates when the polypeptide chain is only about 45-50 amino acids long, only sufficient for 10-15 amino acids to be out of the ribosomal exit tunnel [123].

1.3.1 Prevalence of iMet removal

iMet removal is conserved and essential in all kingdoms, from bacteria to metazoa [16]. In bacteria, it is estimated that PDFs deformylate >90% of the newly synthesised proteins and that MetAPs remove iMet from >50% of the proteome [6, 16]. In archaea, the estimations indicate a reduced amount of processing, with around 40% of the proteome having iMet excised. In eukaryotes, >70% of the cytoplasmic proteins are estimated to be processed by MetAPs [6]. iMet excision is also thought to be prevalent in chloroplasts, in which at least 40% of the chloroplast-produced proteins have been reported to be processed by local MetAP after deformylation by PDFs (estimated to occur in 90% of the cases [16]). The few studies in mitochondria seem to indicate a lower prevalence of iMet processing in mitochondria. From the 13 proteins encoded in mammals and 33 in plants, only 10-50% of the proteins seem to be deformylated and only one (Cox III) is thought to have its iMet removed [6]. However, further studies will be needed to have a complete picture.

1.3.2 Methionine aminopeptidases

1.3.2.1 3D structure

MetAPs are dinuclear metalloenzymes and can be activated by different divalent metal cations *in vitro* such as Co(II) [124], Zn(II) [125], Fe(II) [126] or Mn(II) [127], while the physiological ion is still a matter of debate. They belong to the pita bread fold protease family and are

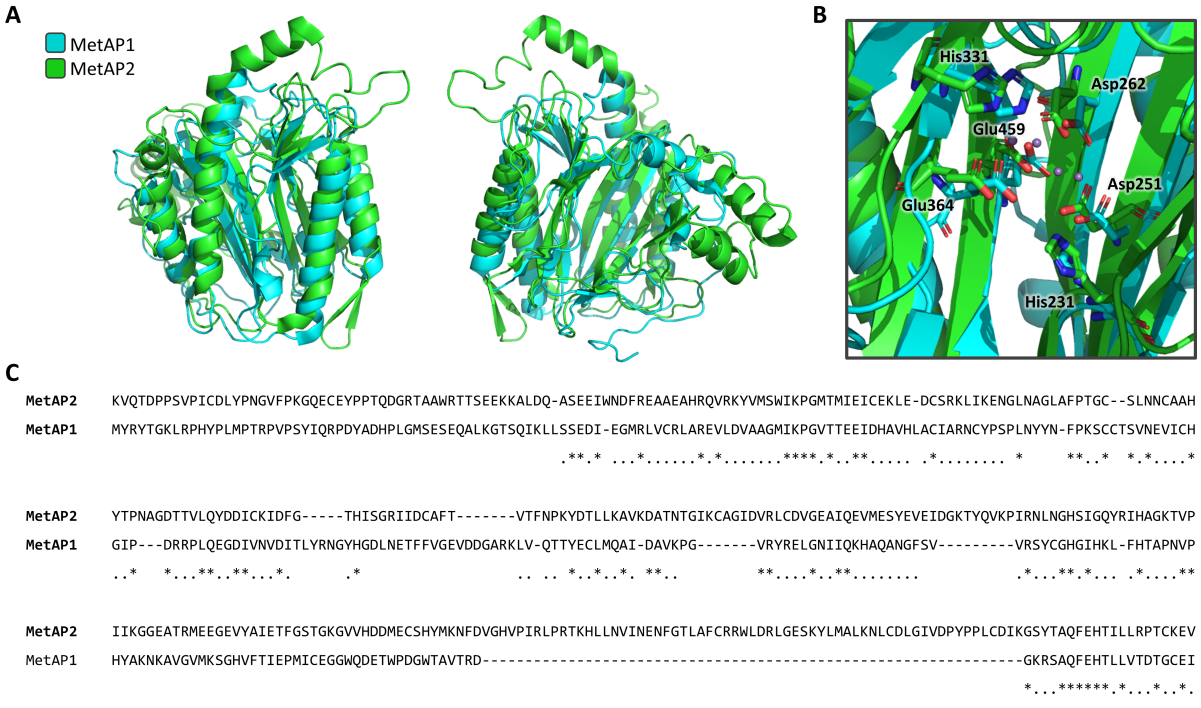


Figure 1.4: MetAP1 and MetAP2 show high degree of structural similarity despite a poor sequence identity. (A) Front and back view of MetAP1 (PDB accession 1R58) and MetAP2 (PDB accession 4FLJ) structural alignment. **(B)** Zoom into the active site of the structural alignment performed in (A). Six residues are highlighted in the metal- and peptide substrate-binding site: the five catalytic residues that are responsible for metal binding (Asp251, Asp262, His331, Glu364 and Glu459) and the His residue (His231) that is covalently modified by fumagillin and other fumagillin-like irreversible MetAP2 inhibitors. Residue numbering is according to their position on the MetAP2 sequence. Purple balls represent manganese ions of each 3D structure. **(C)** Sequence alignment of MetAP2 and MetAP1. An asterisk (*) indicates a fully conserved residue between the two sequences whereas a dot (.) indicates partial conservation of residue properties at that position. Sequence and structural alignment were performed with PyMOL (version 2.4.0).

evolutionary close to creatinase and aminopeptidase P [128]. The overall 3D structure is well conserved among different types of MetAPs (Figure 1.4A) [16]. The 5-residue catalytic and metal ion binding site composed of 2 Asp, 1 His and 2 Glu, is located at the interface between domains, with these five key residues being particularly well conserved across species and MetAP types [124] (Figure 1.4B).

1.3.2.2 Types of MetAPs

There are two main classes of MetAPs: type I MetAPs (or MetAP1) and type II MetAPs (or MetAP2) [129, 130]. Prokaryotes only contain MetAP1 and archea only MetAP2, while eukaryotes have both MetAP1 and MetAP2 [129, 130]. Both enzymes share only a low level of sequence identity (ca. 20%) [131], but this is compensated by the high degree of structural similarity [130] (Figure 1.4). The main difference between the two enzymes resides on two additional compact domains in MetAP2s, a small two-turn beta-sheet domain and a larger helical domain, inserted towards the C-terminus of the catalytic domain [6, 132].

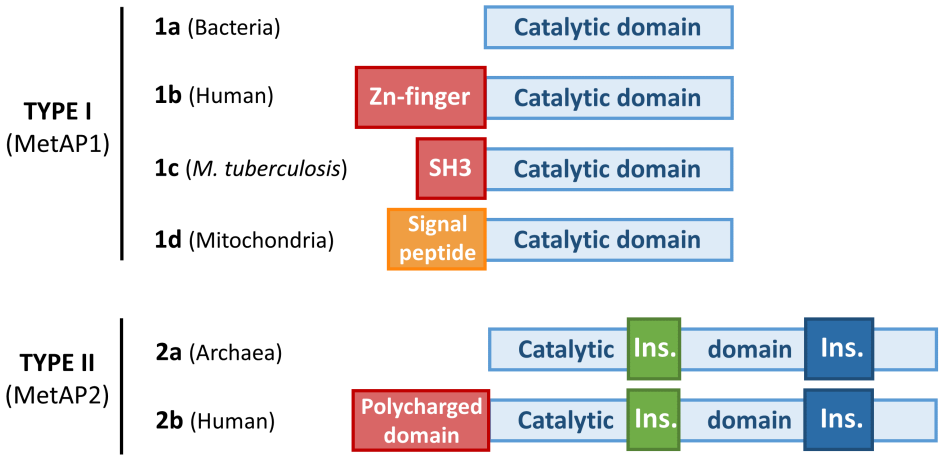


Figure 1.5: Types of MetAPs and their domain structures. There are four subtypes of Type I MetAPs (MetAP1a-d) and two subtypes of Type II MetAPs (MetAP2a-b). Their main differences lie on the N-terminal extensions and the two insertions displayed by Type II MetAPs.

Moreover, MetAP2 has also a wider substrate binding pocket than MetAP1 [133, 134]. To differentiate prokaryotic and eukaryotic MetAPs, they are further subclassified into subtypes a-d for MetAP1 and a-b for MetAP2 [6]. Prokaryotic MetAPs, consisting of subtypes MetAP1a (bacteria) and MetAP2a (archaea) are the simplest forms and only possess a catalytic domain, whereas the rest of the subtypes show different additional N-terminal domains [135] (Figure 1.5). Eukaryotic MetAP1b contains two zinc-finger motifs in its N-terminal domain [136], while eukaryotic MetAP2b has a polycharged region composed of both polybasic and polyacidic spans of residues at its N-terminus [137]. MetAP1c was added when it was found that some bacterial MetAP1s (e.g. *Mycobacterium tuberculosis*) contain an additional SH3 binding domain at their N-termini, thus differentiating them from the main bacterial MetAP1a [138]. MetAP1d consists of the type of MetAPs targeted to eukaryotic organelles, and so, contains a targeting signal sequence at its N-terminus [139]. None of these additional N-terminal domains are required for catalytic activity, but have been shown to have great impact on MetAP function in cells and have been suggested to be implicated in regulation and/or ribosomal binding [136, 140, 141] and are further discussed in the next subsection 1.3.2.3.

1.3.2.3 The roles of the different N-terminal domains of MetAPs

While the N-terminal domains of type I MetAPs (MetAP1) were originally thought to be the main mediators of ribosomal binding, increasing evidence suggests otherwise. The zinc-finger N-terminal domain of MetAP1b from *S. cerevisiae* was shown to be necessary for MetAP function [136, 140] but not crucial for ribosomal association, as mutants showing deletion of the zinc-finger domain still showed the same ribosome profile distribution [136]. However, point mutations at the zinc finger domain did alter MetAP1b binding to the 60S subunit of the active 80S ribosome complex, indicating that it might serve as an additional ribosome-

docking site and be necessary for correct alignment [136]. The SH3 binding domain of *M. tuberculosis* MetAP1c could also mediate some interactions with the ribosome RNA, as it does in other ribosomal proteins [138]. However, it is unclear why some of these organisms possess two MetAPs (MetAP1a and MetAP1c) that are structurally almost identical except for the additional SH3 binding domain in MetAP1c. Moreover, a study of *E. coli* MetAP1a, which is devoid of any additional domains at its N-terminus, showed that it interacted with ribosome through a positively charged loop formed by four Lys residues: K211, K218, K224, K226 [123].

The N-terminal polycharged domain of metazoan MetAP2b has been shown to have a role beyond ribosome binding. It mediates binding of MetAP2b (previously known as p67) to the eukaryotic initiation factor 2 alpha (eIF2α) and protects it from inhibitory phosphorylation [137]. This activity, often referred to as protection of eIF2α phosphorylation (POEP), is crucial to promote translation and thus regulates global protein synthesis within the cell. This is mediated by the first Lys-rich span (residues 41-46) of the N-terminal polycharged domain and an *O*-glycosylation site in residues 60-63 [137, 142]. Interestingly, this POEP seems to be reserved for the MetAP2 of animals, since both fungi or plant MetAP2s seem to lack one of those two interaction points [16]. MetAP2 has been shown to have some autoproteolytic activity resulting in two stable fragments: an N-terminal 26 kDa fragment and a C-terminal 52 kDa fragment, both of which seem to retain their POEP and aminopeptidase functions, respectively [143]. Moreover, it has been shown to be cleaved by calpain activity to release similar-sized fragments [144]. While the role of these proteolytic-fragments is still poorly understood, it has been suggested that it might provide a means for separate regulation for each activity under certain biological conditions [6].

Regarding the ribosomal part involved in the interaction, a recent study showed that in yeast both MetAPs, MetAP1b and MetAP2b, as well as the NatA catalytic subunit Naa10 (Ard1) bind to the ribosome through the b1-4 loops of the ES27L expansion segment [145]. Interestingly, these sections of the yeast expansion segment are not present in bacteria, where the ES27L is much smaller [145]. Importantly, according to this study, ES27L-bound MetAP1 (but not MetAP2 or Naa10) plays a role not only in iMet excision but also in translation fidelity [145].

Together, the data suggests all MetAPs are likely to associate with ribosomes, probably through direct interactions within the catalytic domain, and that in some MetAPs, this is further stabilised through the N-terminal domains. Since longer segments seem to have appeared through evolution, these might have a function in fine-tuning the alignment towards the ribosome and the nascent peptide or in specific regulation that might be necessary in higher organisms.

1.3.2.4 Substrate specificities of MetAPs

MetAPs have been shown to cleave iMet preferentially when the second residue is a small, uncharged amino acid such as Gly, Ala, Ser, Cys, Val, Thr or Pro [131, 146, 147]. The few studies that compared MetAP1 and MetAP2 substrate specificities, indicated similar substrate specificities both for bacterial MetAP1a and archaeal MetAP2a [148], yeast MetAP1b and MetAP2b [131], and human MetAP1b and MetAP2b [147].

For yeast MetAPs, endogenous levels of MetAP1 and overexpressed MetAP2 cleaved iMet efficiently when the second amino acid (position 2, P2) was Ser, Ala or Pro and slightly less efficiently when the P2 residue was Val [131]. This experiment was performed using recombinant GST in which the amino acid at the P2 position (WT = Ser) was mutated to Ala, Cys, Gly, Pro, Thr, Val, Gln, Glu or Leu. Plasmids coding for each GST mutant were transfected into WT, MetAP1 or MetAP2 knockout (KO) yeast strains, and allowed to grow before purification of recombinant GST protein on glutathione agarose resin. iMet processing status of the purified recombinant GST was assessed by its overall mass, by electrospray mass spectrometry (ESI-MS). MetAP2, even when overexpressed, was less efficient than MetAP1 when the second amino acid was Cys, Thr or Gly [131]. Of note, endogenous levels of MetAP2 were shown not to be sufficient to efficiently cleave iMet before any of the tested amino acids at the second position, indicating that MetAP1 is the main enzyme responsible for driving cellular iMet excision in yeast [131]. An important point to consider is that GST contains a Pro in the third position, which has been shown to partially prevent iMet cleavage when the second amino acid has an intermediate radius of gyration [146, 147], which could explain why limiting levels of MetAP2 could not efficiently cleave any of the GST mutants.

In contrast, a synthetic peptide array study of both human MetAPs suggested MetAP2 as the main driver of iMet processing when the second amino acid was Val or Thr [147]. It also showed that MetAP2 was able to accept larger functional groups at the side chain of the second residue than MetAP1 [147], consistent with the fact that MetAP2 has a wider peptide substrate pocket than MetAP1 [133, 134]. A systematic proteomics study that assessed N-terminal peptide status of mammalian cells upon fumagillin treatment further supported the idea of MetAP2 being the main enzyme responsible for processing of proteins containing Val or Thr, and to a lesser extent Gly, in the P2 position [149]. However, expression of MetAP2 in the absence of MetAP1 (HAP1 MetAP1 KO) was unable to restore processing of MS-, MP-, and MA- N-termini, indicating MetAP1 is the main processing enzyme at these P2 positions [150]. Sequence preferences of human MetAPs are summarised in Figure 1.6.

1.3.3 Functional implications of iMet removal

The importance of N-terminal methionine excision for general cellular and organism homeostasis has been supported by decades of mounting evidence. MetAPs are conserved in all

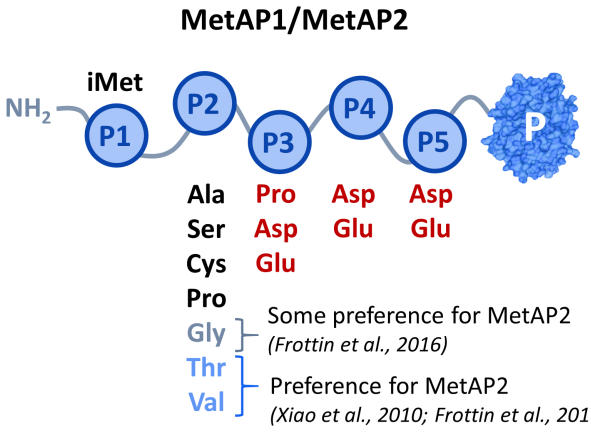


Figure 1.6: Sequence preferences of human MetAPs. iMet is generally cleaved when the second amino acid (P2 position) is Ala, Ser, Cys, Pro, Gly, Val or Thr. Both human MetAPs disfavour Pro at the P3 position and acidic residues at positions P3-P5. MetAP2 has a larger substrate binding pocket and thus accomodates Thr and Val residues better at the P2 position [147, 149]. Some substrates containing a Gly in the P2 position have also been shown to be only cleaved by MetAP2 [149].

kingdoms of life and genetic studies have shown they are also essential in every organism studied [132, 151–155]. Moreover, this process has been shown to be tightly regulated during different stages of life as well as in tumourigenesis [6]. However, the specific roles of iMet retention have remained elusive for decades and are still poorly understood. Why over 2/3 of the proteome would undergo this early modification, with all the energy costs involved is still a mystery to a great extent. Especially, when iMet excision is an irreversible process and does not allow for dynamic post-translational regulation without the need for new protein synthesis. This lack of knowledge might be explained in part by the fact that N-terminal methionine removal or retention might affect each protein differently, which combined with the lack of particular examples makes it difficult to decipher the specific roles of methionine excision. Furthermore, separately investigating each MetAP type in higher eukaryotes is difficult, especially in non-genetically mutated organisms that have not already adapted to the genetic pressure imposed by the lack of the counterpart MetAP. In this section, the key findings on the roles of MetAPs are summarised and discussed.

1.3.3.1 Genetic studies to elucidate the biological role of MetAP1 and MetAP2

Genetic or chemical ablation of the only enzyme responsible for N-terminal methionine excision in bacteria [151] and archaea [152] as well as double deletion in eukaryotes are lethal [132, 153]. This suggests an universal and essential role for N-terminal excision in all living organisms. Genetic deletion of one of the two types of MetAPs in yeast, however, where both types are expressed, renders viable mutants, albeit at lower growth rates [132, 156]. This indicates a clear but partial overlap between the two types of MetAPs, where the second one is able to partially rescue the loss of the first one. Moreover, it opens up a new path to explore

the potential benefits of therapeutically targeting one of the two MetAP enzymes in specific conditions. In plants, where there is one gene for MetAP1 and two for MetAP2 [157], partial genetic knock down of either type of MetAP (MetAP1 or both MetAP2s at once) did not have any visible effects on plant development or growth [153]. Full MetAP1 KOs were still phenotypically identical to the WT plant, whereas full MetAP2 could not be generated [153]. In human cells, genetic knock down of either MetAP1 or MetAP2 had strong impacts on growth, although the extent of inhibition depended on the cell type in both cases with human umbilical vein endothelial cells (HUVEC) showing 70-80% inhibition but A549 only 20-30% [154, 155].

In yeast, MetAP1 was suggested to be the main driver of N-terminal methionine excision, and MetAP2 was thought to be of little contribution to general proteome processing [131]. MetAP2 seems to have a more important role in metazoa than it does in lower eukaryotes. This coincides with the presence of the additional Lys-rich domain at the N-terminus of animal MetAP2s and could therefore be explained by the second role of MetAP2 in POEP (cf. subsection 1.3.2.3 for further details), which is not found in yeast or plants [16, 137, 142]. MetAP2 expression has been shown to positively correlate with actively proliferating mammalian cells, while it was downregulated in starved mammalian cell cultures [158]. Moreover, MetAP2 downregulation coincided with higher levels of eIF2α phosphorylation [158]. In flies, MetAP2 is required for normal patterning during embryonic development, with weak mutations on MetAP2 inducing strong developmental defects and stronger mutations completely inhibiting tissue growth [159]. Conditional Cre-LoxP MetAP2 KO in mice resulted in early gastrulation defects and targeted deletion of MetAP2 only in the hemangioblast lineage resulted in abnormal vascular development and embryonic death [155]. Importantly, the WT phenotype was recovered by double p53/MetAP2 KO[155], indicating the effects caused by MetAP2 deficiency are mediated by p53 signalling.

1.3.3.2 Interconnection with other cellular processes

MetAP activity has also been linked to other important cellular processes. For instance, MetAP1 and MetAP2 activities have been linked to Met metabolism in yeast [160]. L-Met is capable of inhibiting both MetAPs through product inhibition [160]. This was shown to be much greater in the case of MetAP2 (binding constant of 150 µM *vs* 5 mM for MetAP1), which translated in general growth inhibition of yeast cells lacking MetAP1 in the presence of excess Met [160]. In turn, MetAP1 was shown to be responsible for preventing expression of genes involved in Met biosynthesis and sulphur uptake in the presence of high levels of Met [160]. Moreover, N-terminal methionine excision has also been shown to participate in controlling glutathione redox homeostasis in archaea, yeast, plants and human cells, although it is not clear yet how MetAP2 mediates such regulatory activity [149, 152]. Finally, there is also evidence that MetAP2 activity can be regulated by the cellular redox status itself [161].

1.3.3.3 Impact on other N-terminal modifications

iMet removal precedes the addition of other N-terminal modification, either co- or post-translational [6]. Therefore, defects in correct iMet excision will strongly impact the N-terminal modification status of the protein, thereby affecting protein regulation in multiple ways (see Table 1.1 for a summary of the main functions of different N-terminal modifications).

1.3.3.4 Protein half-life and N-degron pathways

Both *N*-acetylated iMet and free iMet have been shown to act both as protective or degradation signals depending on their exposure [46, 47, 112, 162]. Overall, retained iMet has been shown to be protective of protein degradation, especially in protein sequences that naturally do not usually undergo iMet excision, in which retention of iMet has a significant positive impact in the protein lifespan [112]. However, retained iMet, which is usually *N*-acetylated has also been shown to serve as a degradation signal for N-recognins of the Ac/N-degron pathway [46]. Moreover, unacetylated iMet followed by a hydrophobic residue such as Ile or Leu or by positively charged amino acids such as Arg or Lys can also be recognised by the Arg/N-degron pathway system [15, 112]. More details on the N-degron pathways can be found in the relevant section 1.2.8.

1.3.3.5 Mediator of protein binding

There are not many examples of iMet specifically participating and mediating protein-protein recognition and binding, but one such case is the Golgi SNARE protein Ykt6. In Ykt6, iMet (Met1) has been shown to participate directly in the interaction with PTAR1/GGTase-III, allowing for Ykt6 second prenylation, which is important for correct Golgi SNARE complex assembly [163]. The direct implication of the iMet was shown both by structural analysis of the crystal structure of the complex as well as by the fact that the addition of an N-terminal His_6 tag prevented processing of Ykt6 by GGTase-III [163].

Together, through its MetAP and POEP activities, MetAP2 plays a crucial role in maintaining normal cell homeostasis by controlling general protein synthesis, affecting Met metabolism and regulating several aspects of protein function such as protein stability, the presence or absence of other N-terminal modifications and protein-protein interactions (Figure 1.7).

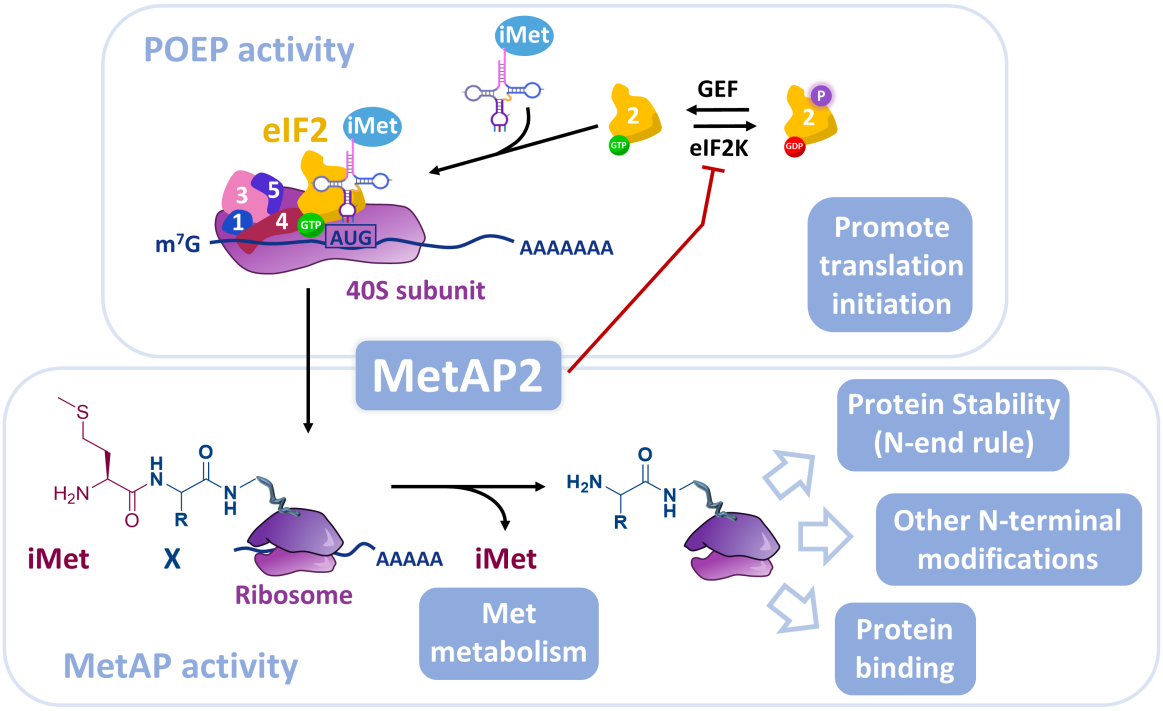


Figure 1.7: Enzymatic and non-enzymatic roles of MetAP2 and funtional implications. In addition to its methionine aminopeptidase catalytic activity, MetAP2 has been shown to possess protection of eIF2α phosphorylation (POEP) activity. Binding to eIF2α is mediated through the second Lys-rich span and an *O*-glycosylation site at its N-terminal domain. Both biological activities are thought to act independently of one another. Through these two independent activities, MetAP2 has an impact on overall protein translation as well as in regulating protein function by affecting protein stability through the N-end rule pathway, protein N-terminal modification status or binding partners. Moreover, MetAPs have been shown to play a role in cellular L-Met metabolism.

1.3.4 MetAP2 inhibitors, relevance of MetAP2 in disease and new insights into the biological role of MetAP2

1.3.4.1 Fumagillin and fumagillin-like irrevesible MetAP2 inhibitors for the treatment of cancer

MetAP2 has been a sought-after target for the treatment of cancer for decades and more recently, also for metabolic diseases. The interest in MetAP2 as a target started with the serendipitous discovery and isolation of a fungal mycotoxin from an *Apergillus fumigatus* infection of a culture of human endothelial cells in the 1990s [164]. This compound, called fumagillin (**1**), had been discovered 40 years before and studied in relation to its potent anti-amoebacidal properties [165, 166]. Fumagillin was shown to exert strong anti-proliferative effects at low micromolar potency in endothelial cells and to inhibit angiogenesis in a chick chorioallantoic membrane model [164]. Due to its anti-proliferative effects in endothelial cells and subsequent anti-angiogenic activity, it was further pursued for its potential benefits for the treatment of cancer. In mice, it was shown to suppress tumour-induced neovascularisation and consequent tumour growth in a dorsal air sac tumour model [164]. However, Ingber and col-

leagues from Takeda rapidly realised that fumagillin had a flaw as a potential anticancer agent: it induced severe weight loss in mice [164]. Therefore, they set out to synthesize fumagillin-like molecules that would preserve the anti-tumour activity with reduced weight-loss effects. From the reported fumagillin analogues, TNP-470 (*O*-(chloroacetylcarbamoyl)fumagillol, initially reported as AGM-1470) (**3**) was the most promising, showing an endothelial cell proliferation inhibition EC_{50} of ca. 40 pM [164, 167, 168]. Importantly, the cytostatic effect was preserved for a wide-window of concentrations and cytotoxicity was only observed for higher concentrations of 75 μM and above [168]. Treatment with TNP-470 in mice exerted strong anti-tumoural effects in a wide range of solid tumours, which was assumed to be driven by the potent anti-angiogenic effects of the compound rather than by direct effects on tumour cells [164]. This substantial trail of evidence drove Takeda to start clinical trials with their lead compound TNP-470 for the treatment of solid tumours in 1992 [169]. Independent preclinical studies also showed promise for fumagillin and TNP-470 for the treatment of other cancers such as hemangioendothelioma [170], mesothelioma [171], neuroblastoma [172] and cholangiocarcinoma[173], as well as for joint inflammation and arthritis[174].

In 1997 and 1998, several breakthrough studies came to light over a short period of time. MetAP2 was shown to be the biological target of both TNP-470 and ovalicin (**2**) by two independent studies [175, 176]. Ovalicin, a structurally-related compound had been isolated in the 1960s from the fungus *Pseudeurotium ovale* and had been studied for its immunosuppressive properties [177, 178]. New studies supported both immunosuppressive and anti-angiogenic effects of both TNP-470 and ovalicin through inhibition of their common molecular target, MetAP2 [176, 179]. Interestingly, binding of these compounds to MetAP2 was shown to inhibit MetAP2 catalytic activity, without interferring with its POEP activity [176]. Soon after, the crystal structure of MetAP2 bound to fumagillin was published, which revealed a covalent bond between His231 of human MetAP2 and the carbon of the spirocyclic epoxide ring of fumagillin [133, 180]. Moreover, these structural studies revealed the reason behind the exquisite selectivity of fumagillin and its analogues for MetAP2 over MetAP1: the binding pocket of MetAP1 did not allow for this interaction both by steric clashes and excessive distancing of key residues to make the necessary contacts for binding [133]. Following these structural studies, TNP-470-induced inhibition of proliferation was shown to correlate well with MetAP activity in a cell culture assay using a biotinylated form of fumagillin [181], further supporting the idea that the sought-after effects of these inhibitors were indeed due to MetAP2 inhibition.

TNP-470 had been taken into clinical trials by Takeda for the treatment of solid cancers in 1992 and had shown great promise in Phases I and II both on its own and in combination with first-line therapies [169, 184–187]. However, clinical development of TNP-470 was abandoned in the early 2000s after strong neurotoxic adverse effects appeared in phase III [188]. At this stage, other pharmaceutical companies and independent academic laboratories started to develop either novel types of irreversible inhibitors still based on the fumagillin

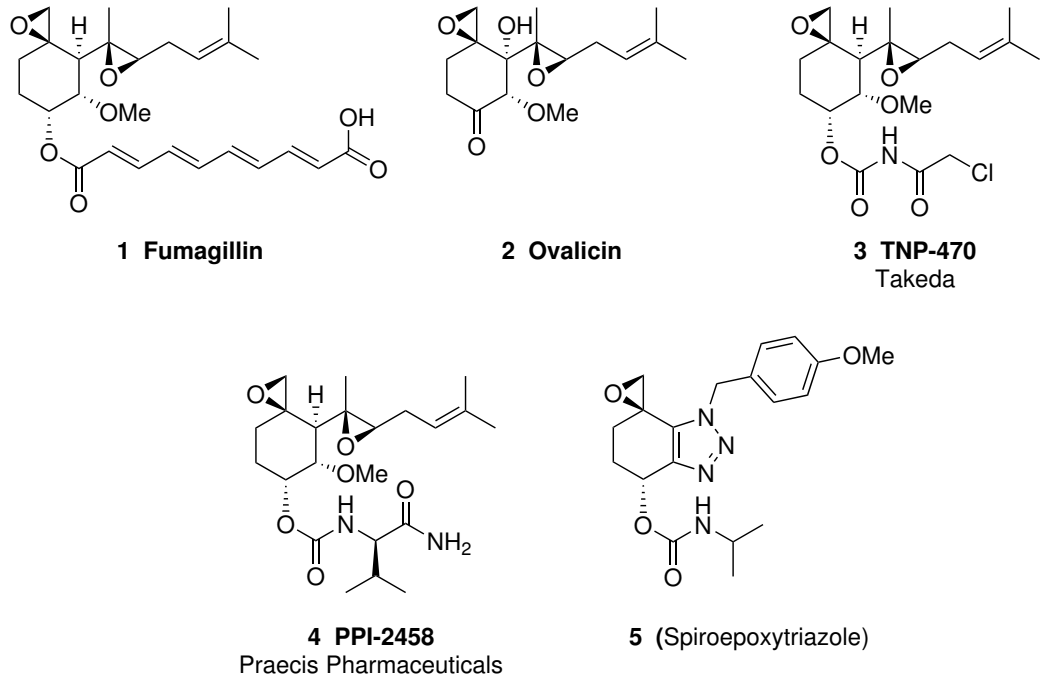


Figure 1.8: Irreversible MetAP2 inhibitors for the treatment of cancer: fumagillin and fumagillin derivatives. The natural compounds fumagillin (**1**) and ovalicin (**2**), the semi-synthetic analogues TNP-470 (**3**) [164] and PPI-2458 (**4**) [182], and the more drug-like spiroepoxytriazole (**5**) [183].

molecule or started to look for drug-like small molecules that were potent reversible inhibitors of MetAP2 (reversible inhibitors are discussed in the next section 1.3.4.2). A new formulation (Caplostatin) in which TNP-470 was conjugated into a *N*-(2-hydroxypropyl)methacrylamide (HPMA) copolymer was able to reduce the neurotoxic adverse effects by preventing its permeation through the blood brain barrier [188, 189]. However, conjugation of TNP-470 to the HPMA copolymer did not solve the poor oral bioavailability issue and still required parenteral administration [189]. Therefore, novel orally available formulations were pursued such as Lodamin (by conjugation to another copolymer) [188] or a pro-drug version of TNP-470, PPI-2458 (**4**) developed by Praecis Pharmaceuticals (later bought by GSK) [182]. PPI-2458 was another fumagillin analogue that went into clinical trials for the treatment of solid cancers and non-Hodgkin's lymphoma [190]. Similar to TNP-470, PPI-2458 was also investigated for the treatment of joint inflammation [191, 192]. More 'drug-like' fumagillin analogues were also developed, such as a spiroepoxytriazole series (**5**) [183].

1.3.4.2 Reversible MetAP2 inhibitors

Novartis was the first pharmaceutical company to report a reversible MetAP inhibitor in 2003. LAF389 (**6**), a synthetic pro-drug analogue of naturally ocurring bengamides, was shown to be a potent dual MetAP1 and MetAP2 inhibitor [193]. Since it inhibited both MetAPs, all cell types tested responded to this inhibitor *in vitro*. However, the authors showed that it selectively affected tumour cells *in vivo* [193]. With promising results in preclinical studies,

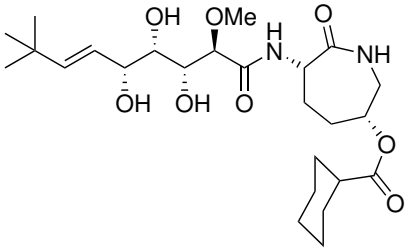
Novartis took this inhibitor to phase I clinical trials for the treatment of high grade solid tumours [194]. However, the compound failed due to cardiovascular toxicity and a lack of consistent response among patients [194].

In the meantime, Abbott (now known as AbbVie) reported the first rationally designed, reversible and selective MetAP2 inhibitor [195, 196]. Based on a 2-amino-2-hydroxyamine core, the bestatin-like compound A-357300 (**7**) was shown to display both MetAP inhibition activity as well as inhibition of proliferation of human microvascular endothelial cells (HMVEC) [196]. It was also shown to display antitumour activity on carcinoma, sarcoma and neuroblastoma murine models [196, 197]. Based on a set of hits derived from a mass spectrometry-based affinity selection screening, Abbott reported another series of novel reversible MetAP2 inhibitors: the anthranilic acid aryl sulfonamide series. This new series showed improved potency and selectivity as compared to A357300 [198, 199]. Rational optimization of these hits lead to development of A-800141 (**8**), a novel reversible inhibitor that was not only remarkably potent ($IC_{50} = 12$ nM, comparable to TNP-470 in their assay) but also orally active [200]. A-800141 also showed anti-angiogenic effects and anticancer activity in a variety of tumour xenografts including B cell lymphoma, neuroblastoma, and prostate and colon carcinomas, either as a single agent or in combination with cytotoxic agents [201].

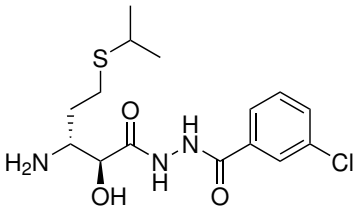
Johnson and Johnson also ran a high-throughput screening program to identify novel reversible small molecule inhibitors of MetAP2. However, the most potent series from the screen, exemplified by the triazole sulfonamide **9**, showed no activity in inhibiting proliferation of HUVECs nor did it exert any anti-angiogenic activity in a rat aortic ring explant culture [202]. The screen had been performed using cobalt as the metal cation and when switching to manganese, another proposed metal ion for MetAP2,a 40-fold reduction in potency was obtained. This was assumed to be the reason behind the lack of response in cells and *ex vivo* culture and consequently, Johnson and Johnson rapidly abandoned the effort.

Another company to run a high-throughput screen for MetAP2 inhibitors was Glaxo SmithKline (GSK). The hits identified were based on a 1,2,4-triazole core (compound **10**) and potently inhibited endothelial cell proliferation as well as inhibited angiogenesis in an aortic ring tissue model in a dose-dependent manner [203]. GSK was also the first to describe a carboxamide (compound **11**) as a MetAP2 inhibitor [204], although they have not made their results with this compound available yet.

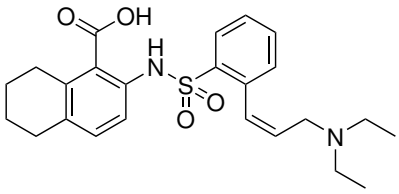
The weight loss effect of TNP-470 and other fumagillin analogues had been considered an undesired effect in the past, and subsequent inhibitor designs managed to mitigate it while maintaining anticancer activity. However, the non-toxic and sustained weight loss exerted by some MetAP2 inhibitors could well serve as potential candidates for the treatment of metabolic disorders, a growing problem in both developed and developing economies. Takeda, well aware of this potential and closely following the progress in this field made by Zafgen (discussed in subsection 1.3.4.3 below), launched a MetAP2 programme for the discovery of



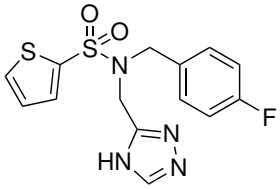
6 LAF389 (Bengamide)
Novartis



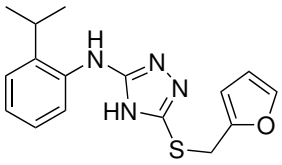
7 A-357300 (Bestatin-like)
Abbott



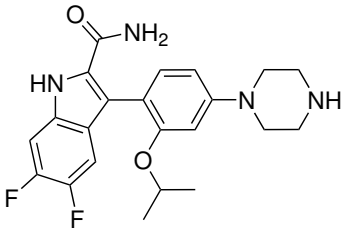
8 A-800141
(Anthranilic acid sulfonamide)
Abbott



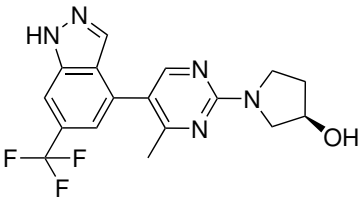
9 JNJ-4929821 (triazole sulfonamide)
Johnson & Johnson



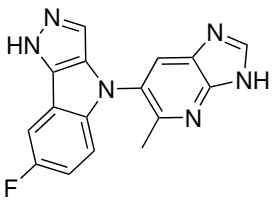
10 (triazole)
GSK



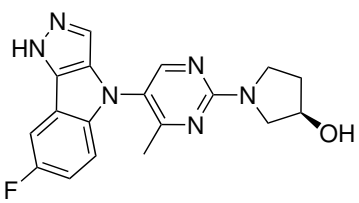
11 2229238 (carboxamide)
GSK



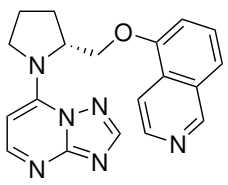
12 (indazole)
Takeda



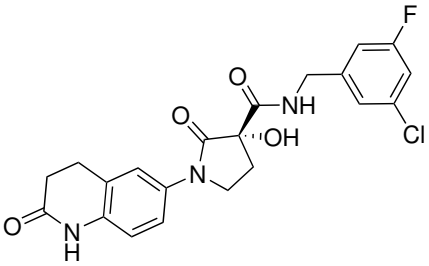
13 (pyrazolo[4,3-b]indole)
Takeda



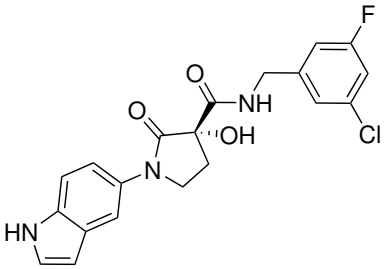
14 (pyrimido(pyrazolo[4,3-b]indole))
Takeda



15 (triazolo pyrimidine)
Merck KGaA



16 (cyclic tartronic diamide)
Merck KGaA



17 M8891
Merck KGaA

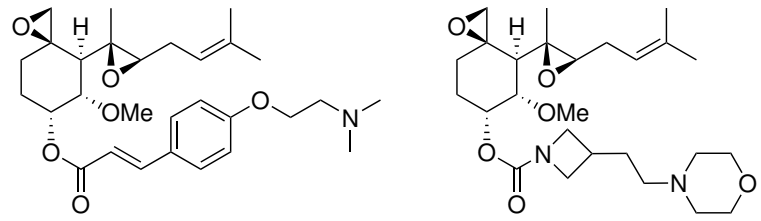
Figure 1.9: Diversity across reversible MetAP2 inhibitors. LAF389 bengamide (**6**) [193], A-357300 bestatin-like compound (**7**) [195], anthranilic acid sulfonamide A-800141 (**8**) [200], JNJ-492821 (**9**) [202], triazole (**10**) [203], 2229238A carboxamide (**11**) [204], indazole (**12**) [205], pyrazolo[4,3-b]indole (**13**) [206], pyrimido(pyrazolo[4,3-b]indole) (**14**) [207], triazolo pyrimidine (**15**) [208], cyclic tartronic acid (**16**) [204] and M8891 clinical compound from Merck KGaA (**17**) [209]

compounds with anti-obesity activity. Combining a fragment-based drug discovery approach and optimisation iterations based on structural studies, Takeda managed to deliver two new series of reversible MetAP2 inhibitors. Lead compounds from both series, compound **12** (indazole) and compound **13** (pyrazolo[4,3-b]indole), showed significant and sustained weight loss in a mouse model of obesity when dosed orally [205, 206]. More recently, they published two new studies in which they use a new compound, arisen from the combination of the halves of the two previous series (Compound **14**) [207, 210]. They have recently shown that this new compound **14** leads to sustained weight loss in obese rodents as well as to increased energy expenditure in human adipocytes derived from obese individuals [210]. Importantly, they showed that these effets are primarily due to the effect of compound **14** directly on brown adipocytes [207], in contrast to previous hypotheses on the weight loss activity being an indirect effect of inhibition of angiogenesis in adipose tissue [211].

Merck KGaA also launched a MetAP2 programme from their German headquarters and in 2017 they published the development of a new series of MetAP2 reversible inhibitors based on a pyrimidine bicyclic core [208]. Based on rational optimization of a hit from a high throughput screen, their most potent compound **15** had an IC_{50} of 38 nM against MetAP2 while displaying favourable liphophilic ligand efficiency (LLE) values [208]. However, HUVEC growth inhibition assays did not show a response as favourable as expected and unable to optimise the structure of this series further, they turned to another high throughput screen hit [204]. This compound, based on a cyclic tartronic diamide scaffold, was found to have undergone an oxidation event during storage, which was what actually conferred the necessary affinity for MetAP2. Further isolation of the *S*-enantiomer and subsequent optimization steps revealed a compound with robust and potent activity in HUVEC proliferation assays (**16**, IC_{50}=74 nM, EC_{50}=15 nM) [204]. Moreover, it showed favourable pharmacokinetic characteristics and reduced tumour growth in a glioblastoma model in mice with subcutaneous injection of human U87-MG cells [204]. Unfortunately, it was shown to suffer from enterohepatic circulation [209]. Efforts to prevent enterohepatic circulation while maintaining potency led them to their latest clinical compound M8891 (**17**), which is currently in phase I trials [209].

1.3.4.3 Second generation of MetAP2 irreversible inhibitors for the treatment of metabolic diseases

Beloranib (**18**) was initially developed by Chong Kun Dang Co. (CKD) Pharmaceuticals under the name CKD-732 and initially intended for the treatment of cancer [212]. It had undergone phase I clinical trials for the treatment of refractory solid tumours either alone or in combination with promising results when it was outlicensed to Zafgen in 2009 [213, 214]. Zafgen further developed this compound, now renamed to Beloranib (ZGN-433), and progressed it into phase I clinical trials, this time for the treatment of obesity [215]. A previous series of reports had demonstrated the potential of irreversible MetAP2 inhibitors, and in particular CKD-732, in inducing non-toxic weight loss [216–218]. Interestingly, the effects seen on adipose tissue



18 Beloranib (CKD-732, ZGN-433)
Zafgen

19 ZGN-1061
Zafgen

Figure 1.10: Second generation fumagillin derivatives for the treatment of obesity and metabolic diseases. Beloranib (also known as CKD-732 or ZGN-433) (**18**) [212] and ZGN-1061 (**19**) [221]

and food intake required lower doses of the compound than its antitumour effects, a dose at which vascularisation was not affected [215, 217, 218]. Confident of the safety profile of Beloranib at these lower doses and with adverse effects being only mild nausea and dizziness [215], Beloranib was taken to phase II [219]. While impact on body weight loss and apetite reduction were promising and most common adverse effects were mild and sleep-related, three cases of venous thromboembolism events caused the trial to be terminated early [219]. Despite these warnings, the company decided to push the clinical candidate forward once more into phase III clinical trials, this time for the treatment of Prader-Willi Syndrome, a rare genetic disorder that leads to obsesive food intake that aggravates the condition and contributes to morbidity [220]. However, two fatal events of pulmonary embolism and two events of deep vein thrombosis halted the trial early, definitively forcing Zafgen to cease clinical development of Beloranib[220].

Somehow not discouraged by their previous failure, Zafgen announced another clinical candidate for obesity and diabetes: ZGN-1061 (**19**). This new synthetic fumagillin-derivative was shown to have a better safety profile than Beloranib, while maintaining its efficacy. According to Zafgen, this was explained by the reduced ability of ZGN-1061 to inhibit endothelial cell proliferation *in vitro*, owed to lower intracellular concentrations, shorter half-life and reduced cellular MetAP2 inhibition achieved with ZGN-1061 as compared with Beloranib [221]. With compelling results in preclinical studies, they rapidly moved it into phase I [222]. ZFN-1061 was put on hold by the Food and Drug Administration (FDA) while it was being tested in phase II trials in 2018 owing to cardiovascular safety risk [223, 224]. No progress on it has been published since.

1.3.5 Digging into the mode of action of MetAP2 inhibitors

1.3.5.1 Inhibition of cell proliferation

Most studies using MetAP2 inhibitors have focused on their effect inhibiting endothelial cell proliferation (cf. sections 1.3.4.1 and 1.3.4.2). Studies with the irreversible MetAP2 inhibitor

TNP-470 showed this was driven by a G0/G1 arrest and inhibition of DNA synthesis [168]. It also reduced the expression of cyclin D1 and inhibited retinoblastoma (RB) protein phosphorylation, a typical marker of cell proliferation [225, 226]. Interestingly, this effect on cellular proliferation was maintained after compound washout [168], suggesting long-lasting effects and that cells struggle to revert back to a proliferating phenotype once MetAP2 has been efficiently targeted. In addition, PPI-2458 was shown to decrease the expression of proliferating cell nuclear antigen (PCNA), another cell cycle component, in a manner proportional to MetAP2 activity [227]. The reversible inhibitors A-357300 also displayed the characteristic inhibition of proliferation of HMVEC through G1 arrest [196], indicating this effect is not limited to irreversible inhibitors. In contrast to previous studies, however, MetAP2 inhibition with A-357300 did not alter cyclin D1 levels but instead decreased Cyclin A levels in a dose-dependent manner [196]. Downregulating MetAP2 levels by shRNA in glioblastoma cells was also shown to cause G0/G1 arrest [228]. Further, actively proliferating cells were shown to correlate with higher MetAP2 levels than arrested cells [158]. These different lines of evidence support an important role for MetAP2 in sustaining cellular proliferation through direct or indirect action on the cell cycle.

1.3.5.2 Apparent cell type selectivity

An important aspect of MetAP2 inhibition that was noticed early on was that there was a significant disparity in the responses that different cell types showed to fumagillin and its analogues. While HUVECs were the most responsive, some other tumour cell lines needed doses at least 10-times higher to exhibit the same response [164, 168]. If MetAP2 has been shown to be essential in higher eukaryotes such as flies or mice [155, 159] and has an important role in maintaining cellular proliferation, how could some cells show almost no effect with MetAP2 inhibitors while others completely stalled at extremely low doses? Probably the simplest hypothesis for this would be possible differences on the expression levels of MetAP2, and due to their suggested partial overlap in substrates, potentially expression levels of MetAP1 as well. That is, cells that have more of either or both isoforms need higher doses of inhibitor to ensure enough occupancy to exert a phenotypic effect. However, a study showed that there was no significant difference in MetAP2 levels between sensitive and resistant cell lines, neither at the mRNA or protein levels [229].This has been posteriorly validated multiple times, including the cell and tissue protein atlas [55]. Moreover, while it was known that cells increased MetAP2 levels when actively proliferating but decreased its levels when arrested, incubation of proliferating cells with fumagillin specifically increased MetAP2 levels in all cell lines tested independently of their responsiveness rates [229]. This indicates that in all proliferating cells loss of MetAP2 activity was equally compensated for, but only certain cell lines could not cope with the downstream consequences of MetAP2 inhibition. These questions were partially addressed by a study that showed that the G1/G0 cell cycle arrest exerted by TNP-470 was mediated by p53 and p21 and that cell lines lacking functional

p53 or p21 were resistant while those displaying functional p53 and p21 were sensitive [230].

1.3.5.3 Cytotoxicity

While the most commonly reported response to MetAP2 inhibitors was cytostaticity, several reports indicated that they could also induce cytotoxicity in certain cell types. Fumagillin was shown to cause apoptosis in mesothelioma cells, which were reported to have greater MetAP2 mRNA levels (and so hypothesised they would be more reliant on the activity of the enzyme) than their normal counterparts. This was shown to be due to mitochondrial damage, nucleosome formation and activation of caspases and was characterised by a significant decrease in the expression of telomerase and the anti-apoptotic gene Bcl-2 [171]. Moreover, treatment of cholangiocarcinoma cell lines with TNP-470 was also reported to induce apoptosis, this time through activation of p38-phosphorylation, up-regulation of Bax and down-regulation of Bcl-xL. TNP-470 also enhanced the antitumour activity of 5-fluorouracil, cisplatin, doxorubicin, and gemcitabine [173]. The underlying mechanism for cytotoxicity needs to be fundamentally distinct from the molecular pathways that drive the cytostatic effect, since induction of G0/1 arrest and entry into quiescence should protect cells from broad spectrum chemotherapeutics.

1.3.5.4 Immunosuppressive effects

Another known effect of fumagillin and its analogues, as well as of the structurally-related ovalicin, is their immunomodulatory capabilities [179]. Indeed, MetAP2 expression was shown to be higher in germinal centre B cells, specifically in dark zone B cells, and corresponding cancerous cell lines, such as B-cell lymphoma but not in other non-B cell lymphomas such as Hodgkin's or T-cell lymphomas [231]. A more recent study identified a role for MetAP2 in differentiation of B cells into plasma cells and antibody production [232]. Interestingly, the specific pattern of MetAP2 expression seen in human spleen samples, could not be recapitulated in mouse spleen, indicating species-specific differences in MetAP2 activities [232].

1.3.5.5 Effect of MetAP2 inhibition on embryonic development through non-canonical Wnt signalling

The lab of Craig Crews, which had carried out important research on MetAP2 including solving the first structure of the complex with fumagillin [133] and published one of the first PROTACs against MetAP2 [233], reported an interesting alternative for the role of MetAP2 in zebrafish embryos. They described an important role for MetAP2 in signalling through the non-canonical Wnt planar cell polarity (PCP) pathway. Treatment of zebrafish embryos with TNP-470, fumagillin, the dual MetAP1/2 inhibitor bengamide E or antisense MetAP2 olionucleotide blocked Wnt signalling downstream of the Wnt receptor Frizzled but upstream of calmodulin-dependent Kinase II (CamK-II), RhoA small G-protein and c-Jun N-terminal

kinase (JNK) [234]. Interestingly, JNK activity has been previously linked to endothelial cell proliferation and both c-jun and RhoA are considered critical factors for angiogenesis [235, 236]. They further showed this was driven by an accumulation of Rab37, since induction of Rab37 accumulation alone by point mutation led to the same aberrant phenotype as in MetAP2 inhibitor-treated embryos [237]. However, it must be noted this work was performed on recombinant Rab37, since the authors were not able to detect endogenous levels of the protein [237]. Independent work studying the role of MetAP2 in modulating definitive hematopoiesis in zebrafish embryos and human umbilical cord blood (CB) models also showed that MetAP2 knockdown or inhibition with fumagillin inhibited CamKII activity and had an effect on ERK phosphorylation, part of the non-canonical Wnt signalling cascade [238]. In contrast, fumagillin had no effects on the canonical Wnt/β-catenin pathway [238]. In all these cases, MetAP2 inhibition affected the development of hematopoietic stem cells (HSCs), new vasculature formation and embryonic development [238]. A different study related MetAP2 inhibition by fumagillin to PI3K and AKT1 upregulation, which have also been linked to Wnt signalling [239], although according to the authors this was due to direct binding of fumagillin to fibroblast growth factor receptor 1 (FGFR1) [240]. Together, these studies might suggest a possible link of MetAP2 inhibition with non-canonical Wnt signaling, although further studies will be needed to clarify the specific impact of MetAP2 inhibition on this pathway.

1.3.5.6 Weight loss effects

The sustained weight loss effects of MetAP2 inhibitors have been well known for decades, and exploited for the treatment of obesity and other metabolic disorders by several pharma companies (cf. sections 1.3.4.2 and 1.3.4.3 above). However, the molecular mechanism behind these effects is still not clear, although some reports seem to indicate it is mediated through direct action of MetAP2 inhibitors on adipocytes [207], as discussed above (cf. 1.3.4.2).

In conclusion, as thoroughly described in this section, a variety of phenotypic effects have been described for MetAP2 inhibitors beyond the well-established inhibition of endothelial cell proliferation, which are summarised in Figure 1.11. The connections between the reported cellular and systemic phenotypes and the effects on specific signalling pathways are still poorly understood.

1.3.6 MetAP2 inhibitors: concluding remarks

As evidenced in previous sections, MetAP2 has been exploited as a target for different diseases by many major pharmaceutical companies through the years. Interestingly, nearly all MetAP2 inhibitors that have reached clinical trials have been irreversible, fumagillin-like covalent MetAP2 inhibitors [169, 190, 211]. These inhibitors show subnanomolar potencies in

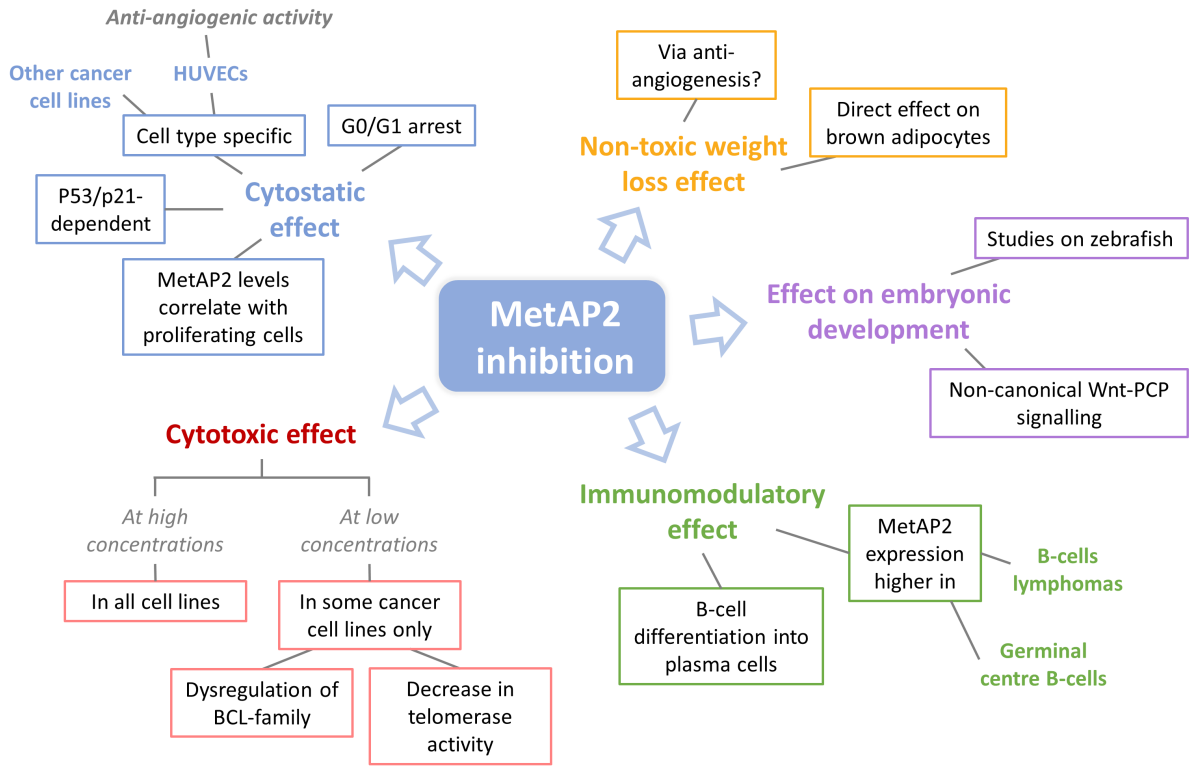


Figure 1.11: MetAP2 inhibition results in distinct cellular effects and phenotypes. These include the cytostatic effect, which is the most studied effect of MetAP2 inhibitors and is responsible for the characteristic anti-angiogenic effect due to inhibition of proliferation of endothelial cells; the non-toxic weight loss and immunomodulatory effects, also exploited for the development of specific inhibitors; and the less studied cytotoxic and developmental activities.

endothelial cells and showed promising results in preclinical models of cancer and obesity. However, fumagillin analogues show one or two reactive epoxide groups that could lead to molecular instability and lack of specificity. While researchers involved in developing these molecules insist on their clean profiles, all of these inhibitors had to be withdrawn from clinical trials due to toxic adverse effects. It is worth noting however, that adverse effects have been diverse: while neurotoxicity was the main issue for TNP-470 [188], the obesity-targeting MetAP2 inhibitors Beloranib and ZGN-1061 showed mainly cardiovascular toxicity [219, 220]. No further reports from the clinical development of PPI-2458 were found beyond the fact that it had entered phase I [190]. Reversible MetAP2 inhibitors arose as cleaner, more 'drug-like' alternatives with better pharmacokinetic properties than fumagillin derivatives. However, until recently reversible inhibitors were not considered potent enough to be of clinical value [211]. An exception was the dual inhibitor bengamide LAF389 (**6**), which did undergo phase I clinical studies. However, targeting both enzymes responsible for N-terminal methionine excision, which has been shown to be an essential process in all organisms (cf. section 1.3.3), was a risky approach and as expected, this strategy was not pursued further. Recently, Merck published the first specific and reversible MetAP2 clinical candidate, which they are now testing in phase I [209]. Only time will tell what comes out of these studies and if they are able

to make this compound sufficiently safe, while maintaining the promising effects of MetAP2 inhibition for cancer treatment.

Moreover, there are still major uncertainties regarding the mode of action of these inhibitors. Different studies have focused on different aspects of the effects of MetAP2 inhibition (i.e. anti-angiogenic and anti-tumour activities, immunomodulatory activity and weight-loss effect) which only seem loosely interconnected, despite good arguments in favour of a MetAP2-mediated mechanism in each case. Whether the weight loss effect is due to the anti-angiogenic activity of MetAP2 on adipose tissues or a direct effect on adipocytes themselves [207] is still not clear and some studies indicate independent mechanisms for each effect [215, 217, 218]. Another explanation for the differential systemic effects shown by MetAP2i's developed for the treatment of obesity or cancer could be distinct tissue distribution. In any case, the link between these different effects will need to be better understood if we are to aim for inhibitors that have a single selective effect. This is important since desired effects for a specific condition (e.g. weight loss in obese patients) might actually become adverse effects in another (e.g. weight loss in patients with advanced solid malignancies, who already suffer from tissue wasting as a consequence of their disease). Some studies seem to indicate a single selective effect might be achievable for MetAP2i's, as it seems MetAP2 inhibitors designed for the treatment of cancer largely avoid the weight-loss effect in mice [164, 204]. Unfortunately pre-clinical studies are not always reflective of their effect in humans. Whether an anti-obesity phenotype can be obtained without any anti-angiogenic effects remains to be seen, since the major adverse effects of Zafgen's compounds seem to be related to deep vein thromboembolism events [219, 220].

The changes induced by MetAP2 inhibition at the cellular level are not clear either. Different studies indicate effects on a wide range of molecular targets and the link between them is vague at best. How the G0/G1 cell cycle arrest relates to the effect seen in the non-canonical Wnt signalling pathway or how this drives cancer-specific targeting in solid tumours is currently unknown. Even less clear are the molecular changes that lead to the systemic weight-loss effect phenotypes. Further, how MetAP2 might drive all these responses and which MetAP2 substrates drive these effects is yet to be discovered. Thus, studies that reveal a comprehensive list of the MetAP2-specifc substrates that are systematically affected upon MetAP2 inhibition will be invaluable to start answering these questions. Chapter 4 has been focused on revealing new MetAP2 substrates that will bridge the gap between MetAP2 inhibition and its phenotypic effects.

1.4 Protein *N*-myristoylation

Protein *N*-myristoylation involves the covalent attachment of a 14-carbon saturated fatty acid, myristic acid, onto the free α-amine of an exposed N-terminal glycine and it is catalysed by myristoil-CoA:protein *N*-myristoyltransferases (NMTs) [241]. *N*-myristoylation happens mostly co-translationally, while the nascent polypeptide is still less than 100 amino acids long and upon exposure of the second amino acid (Gly) by MetAPs [242]. More rarely, it has also been described to occur post-tranlationally upon generation of neo-N-termini by caspase cleavage during apoptosis [243, 244].

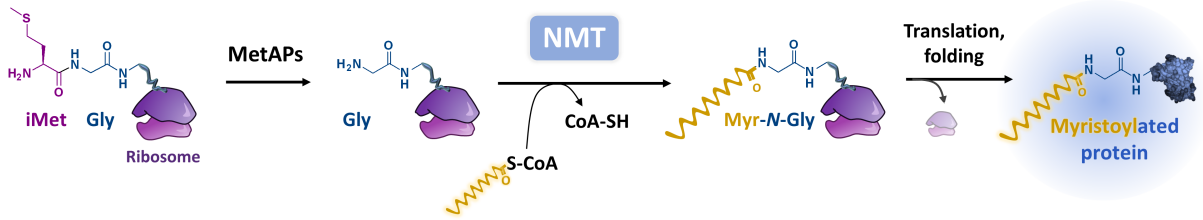


Figure 1.12: NMT catalyses the addition of myristate-CoA onto the N-terminal Gly of protein substrates. This reaction happens co-translationally and upon exposure of the Gly on the P2 position after cleavage of the iMet by MetAPs.

1.4.1 Prevalence of *N*-myristoylation

Estimations suggest 1-4% of all proteins in the proteome are *N*-myristoylated [245]. Around 7% of all proteins present an N-terminal MG- and iMet is generally cleaved when the second amino acid is a Gly residue, so according to those estimates between 15-60% of all N-terminal Gly are myristoylated. The exact prevalence of this modification remains problematic to identify, mainly because it is still challenging to detect the direct presence of the native lipid modification at endogenous levels and because no universal rule on the exact amino acid sequence motif exists to accurately predict its presence. Still, huge advances have been made in the last decade that have allowed for the systematic study of these challenging modifications.

Recent chemical proteomics studies identified >25 myristoylated substrates in *P. falciparum* parasite [246], 30 in *L. donovani* [247] and over 100 co-translationally *N*-myristoylated proteins on human cells as well as >35 proteins that do not have an N-terminal (M)G- but that are myristoylated during apoptosis [244]. Moreover, a recent study using a macroarray of octapeptides that sampled the whole proteome combined with support vector machine (SVM) predictors suggested 1.5% and 1.8% of the proteome (or ca. 20% of the N-terminal Gly (MG-)) is myristoylated in both human and plants, respectively [40].

1.4.2 *N*-myristoyltransferases

1.4.2.1 Types of NMTs

NMTs are ubiquitously expressed in eukaryotes, including parasytic protozoa, yeast, fungi, insects, plants and mammals [9]. While some parasytic prokaryotes and viruses have been shown to benefit and even rely on host NMTs for their replication and survival, they do not possess any *N*-myristoyltransferrase activity of their own [248]. Lower eukaryotes only contain one NMT, whereas in mammals two distinct genes encode for two different NMT isoenzymes: NMT1 and NMT2 [249, 250].

NMT1 is located on chromosome 17 (17q21) and comprises 17 different exons that give rise to three isoforms produced by alternative splicing, whereas NMT2 is located on chromosome 10 (10p13), is composed of 21 exons and gives rise to two transcript variants [55, 251, 252]. Even though each isoenzyme is encoded by a different gene, they share 77% sequence identity, with highly similar catalytic and CoA-binding domains [9, 249]. These subtle structural differences have hindered the development of specific inhibitors to date [253], resulting in limited knowledge on each isoenzyme's specific biological roles and substrate specificities. However, the fact that each NMT is better conserved across different mammalian species (96% identity) than between isoenzymes (77%) of a single species supports the hypothesis of distinct, yet important roles for each enzyme.

Genetic studies further support the hypothesis of separate functions for both isoenzymes. NMT1 was shown to be the principal enzyme in developing mice, with $Nmt1^{-/-}$ not being viable and $Nmt1^{+/-}$ showing severe developmental impairment and early death [254]. NMT2, while expressed showed very little activity and was not able to compensate for NMT1 loss [254]. A more recent study by the same lab showed that NMT1 loss resulted in defective myelopoiesis and macrophage differentiation in the bone marrow [255].

It is still unclear if there exists any specific tissue distribution or subcellular localisations across NMT isoforms [9], although systematic studies such as the Protein Atlas project on the major two isoforms (NMT1 and NMT2) seem to indicate no clear differences between them [55]. Of note, the major differences between the isoforms occur at the N-termini, which are not required for catalytic activity but are believed to be involved in ribosomal binding [9, 256].

1.4.2.2 Mechanism of action and 3D structure

NMT transfers the myristate to protein substrates from myristoyl-CoA (Myr-CoA) following a sequential Bi-Bi mechanism. In this model, Myr-CoA must bind first to induce a conformational change that allows substrate recognition. The reaction is then followed by the nucleophilic attack of the substrate N-terminal Gly amine to the Myr-CoA thioester, resulting in release of the CoA first and the myristoylated substrate second [257].

As NATs, NMTs belong to the GNAT protein superfamily, sharing also their characteristic GNAT fold [32]. NMTs are also monomeric but show two distinct domains (the N-terminal domain and the C-terminal domain), both of which adopt a GNAT fold [258]. Myr-CoA binds to the N-terminal domain while the peptide substrate cavity sits at the interface between the two domains [6].

1.4.2.3 Substrate specificities: Myr-CoA and peptide substrate

The two substrate binding pockets within NMTs show very diverse characteristics. While the Myr-CoA binding pocket is highly conserved across species, the peptide binding pocket is not [9].

The Myr-CoA binding pocket, with only one main substrate to which it binds with very high affinity, has been well characterised. The chain length has been shown to be particularly important for recognition, more than the charge and bulkiness of the fatty acyl chain [259]. Thus, it tolerates addition or reduction of one carbon in the fatty acyl chain, but addition of two extra carbons makes it a very poor substrate for NMT [260]. This makes sense in the cellular context, where 13- or 15-carbon fatty acids are rare but the 16-carbon saturated fatty acid palmitate (palmitoyl-CoA, Pal-CoA) is very abundant. Importantly, Myr-CoA binds NMT in a bent conformation and it has been shown to pre-dispose NMT structure for peptide binding [259].

The peptide binding pocket has been much more challenging to unravel. This is not unexpected, since it needs to be a bit more flexible to accommodate its 100-600 different peptide substrates. However, efforts studying both the structure of NMT1 bound to different substrates as well as the sequence characteristics of the known NMT substrates have shed some light into the peptide recognition mechanism [40, 261–263] (Figure 1.13). Glycine in the second position (exposed after removal of iMet by MetAPs) has been long recognised as an absolute requirement [261]. The amino acid 3 (aa3) pocket is small and acidic, so it disfavours bulky (Arg, Phe, Tyr, Trp) or acidic residues (Glu, Asp) [40]. The aa4 pocket is quite large and allows most residue types, while the aa5 pocket is quite hydrophobic and thus disfavours charged amino acids (Asp, Glu, Lys, Arg) [40]. aa6 is most commonly a Ser residue [261]. The aa7 is position at the interface between the inner peptide cavity and the enzyme surface, but nearby positioning of three aspartic acid residues (Asp183, Asp185 and Asp471) yields in a strong preference for basic amino acids such as Lys or Arg that can stabilize the binding through a salt bridge [40]. Marrying information of this structural studies with peptide sequence information derived in the consensus sequence G [NCAQ] [CASV] [LFC] S [KR] (aa2-aa7) [40]. It must be noted, however, that this consensus sequence is not absolute and even the best machine learning predictors struggle to indicate the presence of this modification with 100% accuracy. Since aa7 is the last amino acid to enter the peptide binding pocket on NMT, amino acids downstream in the peptide sequence are less likely to play a role in

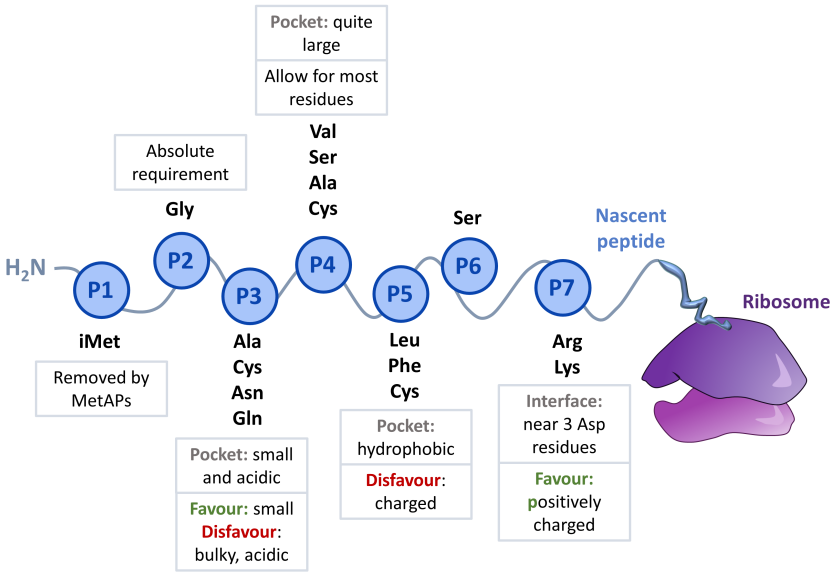


Figure 1.13: Specificity of the peptide substrate pocket of NMT. In addition to the exposed Gly, the NMT peptide substrate pocket shows preference towards peptides displaying small residues on P3 position, small hydrophobic residues on P5, a Ser on P6 and basic residues on P7. Note the prevalence of Cys residues on posituons P3-5.

substrate recognition on an individual basis [40]. However, they might still influence peptide stabilisation through interactions with the surface of the enzyme [261, 262]. Interestingly, peptide sequence analysis revealed a strong bias towards Cys residues in positions 3, 4 or 5 in the sequence [40]. The authors suggested this Cys 'imprinting' might have arisen from co-evolution of myristoylation and palmitoylation, thus favouring a second lipidation site within the N-terminus of a number of myristoylated proteins [40]. It must be noted all these analyses have been performed for human and plant NMT1, as it is considered the most abundant enzyme of the two and the most efficient *in vitro* [6]. Therefore, it is still unclear what are the substrate differences between NMT1 and NMT2.

1.4.2.4 Regulation of NMT activity

In yeast, the single NMT enzyme seems to be constitutively active and cytosolic [264]. According to the Protein Atlas [55], human NMT1 seems to localise to both cytosol and plasma membrane while NMT2 localises to the Golgi apparatus, cytosol and plasma membrane. Since most of these experiments rely on antibody identification, there is currently no information of differential distribution across isoforms.

Lipid heterogeneity in protein lipidation, including protein *N*-myristoylation, has been shown to mirror cellular lipid metabolism [265, 266], implying a higher level of regulation and interconnection with other cellular processes involved in energy uptake, expenditure and storage. Myr-CoA is scarce in the cell [267], so availability to Myr-CoA pools might be a way of regulating NMT activity [9]. Moreover, palmitoyl-CoA, although inherently a poor substrate for

NMT, has been shown to be able to inhibit its activity at high concentrations [268], which might be relevant given its high abundance in the cell and the NMTs presence in membranes [260].

NMTs have been proposed to associate with ribosomes through their N-terminal domain [269], which is not surprising given their shared fold with NATs. Therefore there could exist regulation of NMT activity at the level of ribosome binding. In particular, recent evidence suggest competition with NATs with partial overlap of substrates between NatA and NMTs [39, 40]. This is discussed in more detail in section 1.5.3.

During apoptosis, NMT1 and 2 have been shown to be cleaved by caspase 3 (NMT1/NMT2) and caspase 8 (NMT2) and while this cleavage did not greatly impact activity of NMT enzymes, it did shift their localisation towards a greater cytosolic pool in the case of NMT1 whereas NMT2 became predominantly membrane bound [270]. This localisation shift could influence the known activity of NMTs on myristoylating neo-N-termini formed by caspase cleavage.

1.4.3 Impact of protein *N*-myristoylation on protein function

The importance of protein myristoylation in healthy cell homeostasis has been recognised for decades and NMT has been shown to be essential in various organisms [254, 271–273]. As other lipids, it increases hydrophobicity of the protein it modifies, therefore increasing its affinity for membranes. It has been shown to influence protein function and fate in a variety of ways beyond membrane binding, which include forming part of the protein tertiary structure, creating the binding site for interaction with other proteins or affecting protein stability (Figure 1.14).

1.4.3.1 Reversible membrane binding and two-signal mechanism

Protein lipidation plays a unique role in protein subcellular localisation by increasing membrane association of proteins that would otherwise be fully free in the cytosol [274]. While myristoylation alone is not sufficient to dictate complete membrane targeting [275], it has been shown to generate a dynamic partitioning of myristoylated proteins between the cytosol and specific membranes [276]. The myristoyl moiety has been shown to exist in to states: buried inside a hydrophobic pocket or exposed to promote membrane binding. A model has been proposed in which the transition between those two conformations (called 'myristoyl switch') is promoted by a second signal (Figure 1.15). This way, an otherwise irreversible PTM is allowed to be dynamically regulated by addition/removal of the second signal. Three different types of myristoyl switches exist, depending on the properties of the second signal [9, 277, 278]:

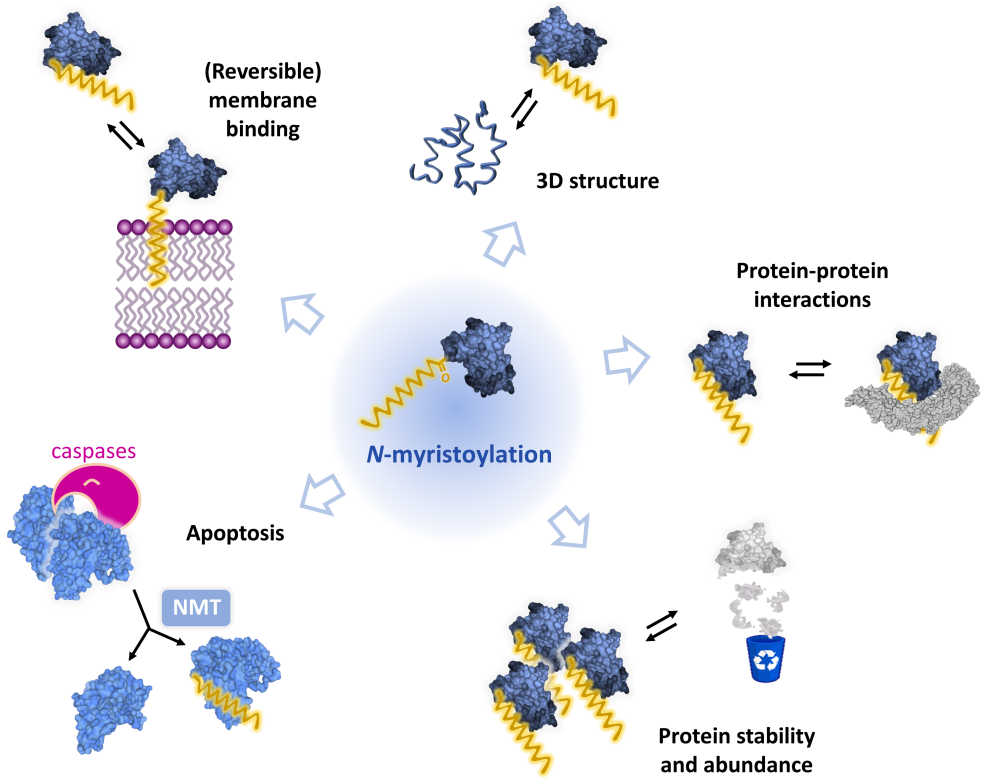


Figure 1.14: Impact of protein *N*-myristoylation on protein function. Addition of myristate can impact protein function in several ways, including increased affinity for membranes, stabilisation of the functional 3D structure, mediate protein-protein interactions and impact on protein stability and abundance (Gly/N-end degron), in addition to the roles of N-myristoylation during apoptosis, upon exposure of neo-N-terminal glycines.

- **Palmitoylation.** Many myristoylated proteins have been shown to also undergo palmitoylation in their first few N-terminal residues. The prevalence of Cys is so exceptionally high in the first few residues of the sequence of myristoylated proteins that some mechanism for genomic imprinting has been proposed, according to which both types of lipidation evolved together, influencing one another [40]. Thus, the proposed mechanism for the combined lipidation suggests myristoylation happens first, while the protein is still at the ribosome. This creates a higher affinity of the newly myristoylated protein for membranes, bringing it into the proximity of membrane-bound DHHC palmitoyltransferases, which then attach the second lipid moiety and force the protein to remain membrane bound. Since *S*-palmitoylation is reversible, this allows for trafficking of proteins between the cytosol and specific membranes when needed [6]. An example of proteins regulated this way are most Src family members and G_α proteins, both of which require a membrane-bound location to exert their function as key elements in signalling transduction pathways [277, 279].

- **Electrostatic switch.** Another type of myristoyl switch is given by the presence of polybasic domains, which provide the necessary Gibbs free energy to increase the affinity and thus anchor the protein to the membrane [278]. These electrostatic switches can

also be reversed, for example by phosphorylation, which introduces a negative charge and promotes dissociation from membranes. Alternatively, proteins not having a polybasic domain might develop permanent membrane binding through interaction with positively charged ions, like calcium [9, 279]. This is the case of recoverin, which can bind to membranes only when Ca^{2+} levels increase, thus serving as a sensor of intracellular calcium levels [280]. In contrast, proteins with a polybasic domain in addition to a myristoyl group include c-Src itself [281] (although most other Src family members are palmitoylated [279]), the avian Rous sarcoma virus v-Src [282], the HIV-1 protein Gag [283] and the myristoylated alanine-rich C kinase substrate MARCKS. [284, 285].

- **Ligand binding.** Finally, membrane anchoring might be driven by interaction of the myristoylated protein with other membrane-bound proteins [278].

Thus, prevention of *N*-myristoylation has been linked to the deficient targeting of NMT substrates to membranes [277].

1.4.3.2 Elemental part of the protein 3D structure

The myristoyl moiety has also been shown to contribute directly to the stabilisation of the 3D fold of proteins. An example of this is cAMP-dependent protein kinase (PKA), in which the myristoyl group is buried within a hydrophobic cleft, stabilising the active conformation of the enzyme [286]. Myristoylation has also been shown to activate c-Src, not only by ensuring its necessary membrane-bound location, but also through direct effect on its catalytic activity [281]. Myristate sits in a hydrophobic pocket at the base of the kinase domain and any mutations to that binding pocket resulted in reduced catalytic activity even if the membrane binding was not affected in most cases [281]. In the c-Abl tyrosine kinase, the myristoyl group not only binds tightly to the base of the kinase domain, but this binding also anchors the SH2 domain against the kinase domain and has an inhibitory effect on c-Abl [287, 288]. Therefore, myristoylation can form an integral part of the 3D structure of a protein, modulating its activity, both positively or negatively.

1.4.3.3 Integral mediator of protein-protein interactions

N-myristoylation has also been shown to mediate calmodulin (CaM) binding of several proteins directly, such as MARCKS [289] and the brain-specific protein kinase NAP22 [290].

1.4.3.4 Determinant of protein stability

N-myristoylation has been shown to mask proteins from the Gly/N-degron pathway [112], and therefore it constitutes a major regulator of protein stability across most of its substrates.

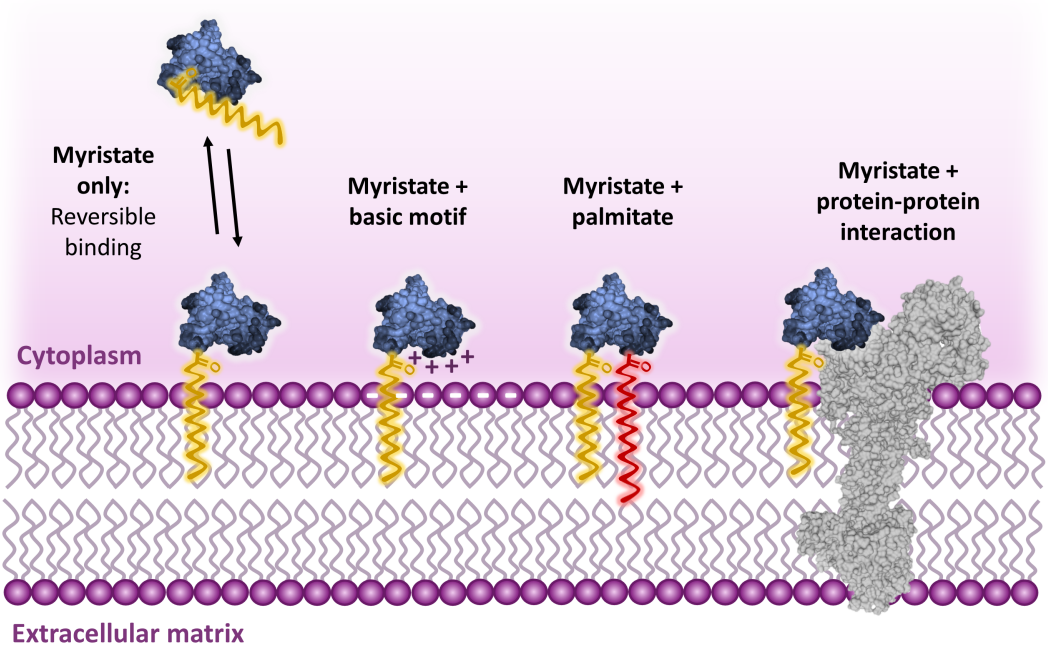


Figure 1.15: Reversible membrane binding and two-signal mechanism. Addition of myristate increases affinity for membranes but is not sufficient to provide permanent membrane anchoring. A second signal is needed to target *N*-myristoylated proteins to the membrane permanently, which can be mediated by a polybasic motif, a second acylation or binding to membrane proteins.

More rarely, it has also been shown to affect stability directly, as is the case of c-Src, in which myristoylation activates c-Src but also promotes its degradation [281], probably as a negative feedback loop to ensure proper regulation of such an important kinase.

1.4.3.5 N-myristoylation during apoptosis

Gly is the most common amino acid after caspase cleavage sites in proteins, making up about one third of the newly formed N-termini [112]. Both NMT1 and NMT2 have been shown to both be cleaved by caspases [270] as well as to take active part in processing of novel N-termini formed upon cleavage of proteins by caspases during the apoptotic cascade [243, 244]. The functional relevance of this is still to be elucidated, but it could play a role in competing against the degradation of specific newly formed protein fragments by the Gly/N-degron pathway (further information on the Gly/N-degron pathway can be found on section 1.2.8.3).

1.4.4 Relevance of *N*-myristoylation in disease

Together, by regulating key aspects of proteins such as localisation, stability, structure and interaction partners, protein *N*-myristoylation affects protein signalling, intracellular trafficking and cell to cell communication [9, 277]. With such pivotal roles, in addition to its direct

implication on apoptosis, protein myristoylation has been linked to a wide range of infectious diseases, as well as to diabetes and cancer [9, 291].

1.4.4.1 NMT in infectious diseases

NMT involvement has been described for almost any type of infectious disease, from parasytic protozoa and fungal infections to viral and bacterial infections, where host NMT is the one involved in myristoylating viral or bacterial proteins.

Diseases caused by eukaryotic organisms such as parasitic protozoa or fungal infections provide a particularly attractive scenario for specific targeting of parasite NMTs, which are significantly different to human NMT. Some parasitic protozoa that have proven evidence on their reliability on NMT include *Plasmodium falciparum*, the parasite causing malaria [292]; *Leishmania donovani* responsible for leishmaniasis [273, 293]; *Trypanosoma brucei*, the parasite causing sleeping sickness [273]; and *Trypanosoma cruzi*, originator of the Chagas disease [294]. NMTs of the opportunistic pathogenic yeast *Candida albicans* and of *Cryptococcus neoformans*, the yeast responsible for fungal meningitis, were both reported to be necessary for their growth [272, 295].

A number of viruses have been shown to produce proteins that are myristoylated by the host NMTs [248]. Among them, the myristoylation of Gag and Nef proteins from HIV-1 and the capsid proteins of picornaviruses have been particularly studied[296, 297], including the rhinovirus responsible for the common cold [298].

Finally, myristoylation of proteins from some plant-affecting bacteria have also been reported, after their secretion into the host cell [248].

1.4.4.2 NMT in cancer

The discovery that some well-known oncogenes are *N*-myristoylated and the importance of this modification in mediating their carcinogenic activities has placed NMT in the spotlight as an anticancer target [9, 299]. Amongst the most prominent examples are c-Src, c-Abl and MARCKS.

- c-Src is a non-receptor tyrosine kinase and a proto-oncogene that is upregulated in a number of solid malignancies [300]. The related v-Src oncogene from the Rous sarcoma virus was reported to lose its transformation capabilities when non-myristoylated and cytosolic [301]. Similarly, non-myristoylated c-Src cannot be targeted to membranes and this relocalisation uncouples it from its activation by a membrane bound phosphatase [302]. Moreover, another study showed that the myristoyl group in c-Src also plays a role in maintaining its active conformation, when most mutations to the myristoyl-binding hydrophobic pocket reduced its activity without affecting its localisation [281].

- Myristoylation of c-Abl, another non-receptor protein tyrosine kinase that is also a known proto-oncogene, has been shown to be essential for the regulation of its function. Mutations in the *ABL1* gene, in particular the translocation into the breakpoint cluster region (*BCR*) gene that forms the constitutively active *BCR-ABL* aberrant kinase form, are associated with chronic myelogenous leukemia (CML) [303]. Myristoylation participates in docking c-Abl in an autoinhibitory conformation [287, 288], so the lack of this modification would render c-Abl constitutively active and devoid of any type of control. In addition, allosteric inhibitors that bind to the myristoyl pocket and lock the protein in an inhibitory conformation have been proposed for the treatment of CML [304].
- MARCKS is critical for cancer development and progression. It is generally membrane bound through its myristoyl moiety and polybasic domain and becomes cytosolic when phosphorylated by PKC, where it functions in actin cytoskeletal remodelling and Ca2+ signalling through CaM binding. By promoting both membrane and CaM binding, myristoylation is an integral part of the regulatory cycle of MARCKS and is therefore likely to play a role in tumourigenesis [305].

However, *N*-myristoylation of a protein previously established to act as a tumour suppressor, FUS1/TUSC2, has also been reported [306]. FUS1/TUSC2 is believed to induce G1 arrest and promote apoptosis in lung cancer cells [306]. Moreover, given the extent of *N*-myristoylation events across the proteome, it would be naive to assume all effects of NMT in cancer are mediated by the myristoylation of three such proteins. However, there is mounting evidence of specific NMT upregulation in a series of carcinomas and brain tumours [307–310] and its expression has been shown to be positively correlated with telomerase activity levels [311]. Moreover, a recent paper from Thinon and colleagues showed that prolonged NMT inhibition can induce apoptosis in cancer cells through cell cycle arrest at G1 phase [312]. Together, these indicate an important role in cancer, likely through a complex network of events, whose intricacies we have barely started to grasp.

It is important to note that while targeting NMT against infectious diseases has shown great promise so far due to the considerable differences between parasitic and human enzymes [313], targeting human NMT in cancerous cells still poses a major challenge [314]. Protein *N*-myristoylation is essential for normal cell homeostasis and it is still unclear whether the dependency of cancer cells on NMT will be sufficient to provide a therapeutic window. Deeper understanding on how *N*-myristoylation is (dys)regulated in cancer on a global scale and how the specific implications of each NMT isoform is needed in order to define the framework in which targeting NMT will provide clinical value.

1.4.5 NMT inhibitors

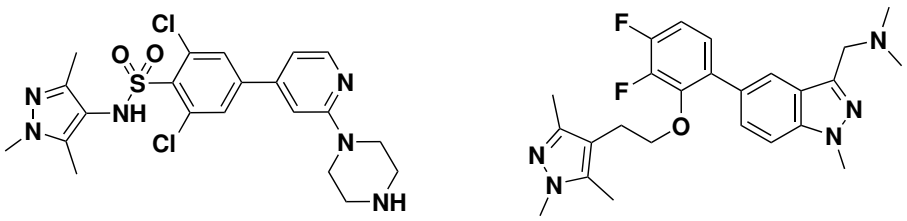
A number of NMT inhibitors have been developed throughout the years. Initial attempts targeted mainly disease-causing fungal and protozoan NMTs, and more recent studies have also shown the potential to target human NMT for both viral infections and cancer.

1.4.5.1 Inhibitors against fungal and protozoan NMTs

The first NMT inhibitors to be reported targeted the opportunistic yeast *Candida albicans* NMT, encouraged by the reliance of this organism on NMT for vegetative growth [295]. Reported inhibitors consisted of both peptidic and non-peptidic inhibitors that bound at the peptide binding pocket of NMT [315, 316]. Similarly, driven by the discovery of NMT essentiality in protozoan parasites NMT was also proposed as a drug target for *Leishmania donovani* and *Trypanosoma brucei*, responsible for leishmaniasis and sleeping sickness diseases, respectively [273]. This lead to the discovery and development of a novel series of inhibitors by the University of Dundee exemplified by the DDD-85646 compound (**1**) [317, 318]. This compound, initially designed to target *T. brucei*, was discovered to have similar potency against the human NMT enzymes and has been thereafter used in multiple studies, proving itself an invaluable chemical tool for the study of NMTs from multiple organisms. New evidence showing NMT was essential for *Plasmodium berghei* in mice [319] suggested a broader antiparasitic potential for NMT and started the search for NMT inhibitors for the two major parasites causing malaria in humans: *Plasmodim falciparum* and *Plasmodium vivax*. These included high-throughput screens that lead to several new series of inhibitors against *L. donovani*, *Plasmodium falciparum* and *Plasmodium vivax* [320, 321]; and a chemistry-driven approach to repurpose and optimize a series of inhibitors originally designed by Roche against *C. albicans* for *P. falciparum* and *P. vivax* [322–324]. NMT inhibition has also been explored for the treatment of Chagas disease, with three compounds of the DDD series showing potent antiproliferative effect [294]. Recently, a novel inhibitor of *P. falciparum* NMT was reported (IMP-1002), which was shown to be 4-fold more potent than the DDD-85646 compound [313].

1.4.5.2 Inhibitors against human NMT

Through N-myristoylation, NMT plays an important role in regulating the function of many oncoproteins and pro-proliferative pathways. Therefore, it has long been proposed as a target for the treatment of cancer. A number of compounds have been reported to inhibit human NMT in cells. In a recent study, however, we showed that many of the compounds claimed to target NMT, including 2-hydroxymyristic acid, D-NMAPPD and tris-DBA palladium, cause cell cytotoxicity by means unrelated to NMT [325]. In contrast, the DDD-85646 compound initially designed against *T. brucei* NMT has been an invaluable chemical probe to understand



1 IMP-366 / DDD-85646
IC_{50} (HsNMT1) = 17 nM
IC_{50} (HsNMT2) = 22 nM

2 IMP-1088
IC_{50} (HsNMT1) < 1 nM
IC_{50} (HsNMT2) < 1 nM

Figure 1.16: Inhibitors of human NMT. DDD-85646 (IMP-366) and IMP-1088 are the two main human NMT inhibitor series. Both show remarkable potency and have been validated by a number of independent studies.

the biological role of human NMTs. It is a dual NMT1/2 inhibitor, with an IC_{50} value of ca. 20 nM for both human NMT1 and NMT2 [244]. It allowed for the identification of most of the known human NMT substrates in combination with chemical proteomics approaches [244]. Moreover, the DDD-85646 compound was shown to induce apoptosis in a panel of cancer cells through ER stress and G1 arrest [312], supporting the potential benefits of targeting cancer cells with NMT inhibitors.

Human NMT has also been exploited for the treatment of viral infections. HIV-1 proteins Gag and Nef have been shown to undergo myristoylation by human host NMT. The polyprotein Gag in particular, is targeted to the membrane through a myristoylation plus basic motif where it is processed into its mature form [326]. Thus, NMT inhibition was shown to inhibit Gag processing at the membrane, significantly impairing viral infectivity [327]. In picornaviruses, the smallest capsid protein VP4 has been shown to be myristoylated by host NMTs. Truncation of the N-terminus of poliovirus VP4 had been shown to have deleterious effects on viral assembly. Driven by this, Corbic-Ramljak and colleagues showed that treating coxsackievirus B3 (another member of the picornaviridae family)-infected cells with the DDD-85646 NMT1/2 resulted in a significant reduction in viral infectivity by inducing severe viral assembly defects [297]. In a more recent study our lab reported a novel, second generation human NMT inhibitor, IMP-1088 (**2**), derived through fragment tethering, which had a potency superior to the DDD-85646 compound [298]. Following on previous studies with DDD-85646, IMP-1088 was shown to be a potent inhibitor of the replicative cycle of several picornaviruses, including the common cold virus, by preventing capsid assembly through deficient myristoylation of the viral capsid protein VP0 (VP4) [298].

1.5 Interplay among N-terminal modifications

As any other process in cells, protein modifications are bound to co-exist and must be tightly regulated to ensure their specificity, scope and spatio-temporal dependence. In the case of protein N-terminal modifications, the need for this exquisite control is even more apparent as proteins only have one N-terminus. Thus, many different components responsible for regulating different types of PTMs must come together to 'make' the decision of which modification (if any) will be placed at this unique site.

1.5.1 Ribosomes, the hubs that coordinate protein synthesis, folding, co-translational modifications and subcellular targeting

The 'triage' for protein N-terminal modifications starts as soon as the first few amino acids have left the exit tunnel of the ribosome. Most enzymes responsible of co-translational N-terminal modifications have been shown to not only interact with the nascent peptide, but also with ribosomes [145, 328]. Until very recently, ribosomes were seen as molecular machines that had little to no role in the regulation of protein synthesis and were regarded as an immutable entity that simply produced new proteins based on the mRNAs it 'encountered' on its way. Increasing evidence suggests this view could not be further from the truth. Interacting with factors involved in a myriad of different cellular processes, ribosomes have started to be conceived as important 'hubs' to control not only protein synthesis, but also co-translational modifications, folding and localisation of the nascent proteins [14]. Following this new concept, the enzymes that come together at ribosomes to decide the new protein fate have been termed ribosome-associated protein biogenesis factors (RPBs). These include enzymes responsible for attachment of co-translational modifications, molecular chaperones and proteins responsible for 'reading' the proteins initial signal sequence and targeting them to the right cellular compartment [14, 123].

1.5.2 Ribosomes, complex and heterogeneous supramolecular structures

Ribosomes are extremely complex structures made of ribosomal RNA (rRNA), core ribosomal proteins (RPs) and ribosome associated factors (Figure 1.17).

- **Ribosomal proteins (RPs)** can be differentiated into RPs of the large ribosomal subunit (RPLs) and RPs of the small ribosomal subunit (RPSs) and, so far, 80 different RPs have been described in humans [329]. These form integral part of the ribosome complex, ensuring assembly of the rRNA into the two ribosomal subunits and providing the necessary stability for the complex. They also serve as docking sites for interaction with ribosome-associated factors.

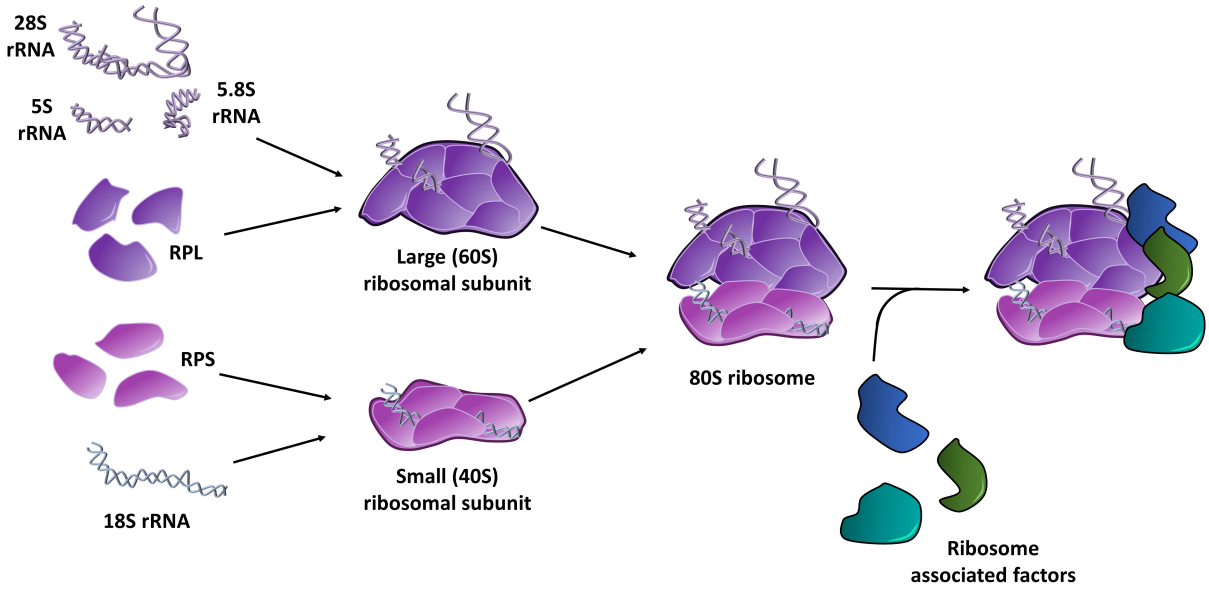


Figure 1.17: Assembly of functional 80S ribosome from ribosomal RNA (rRNA), core ribosomal proteins (RPs) and ribosome associated factors. RPs and rRNAs come together first to form the large (60S) and small (40S) ribosomal subunits. 28S, 5S and 5.8S rRNAs assemble with the large ribosomal subunit RPs (RPLs) and form the 60S ribosomal subunit. 18S rRNA assembles with small ribosomal subunit RPs (RPSs) and form the 40S ribosomal subunit. Upon translation initiation, 60S and 40S ribosomal subunits assemble to form the 80S ribosome and start protein synthesis. Many other proteins, referred to as ribosome-associated factors, interact transiently with both 60S and 40S subunits to regulate protein synthesis. Ribosome-associated factors include translation initiation and elongation factors as well as protein biogenesis factors (RPBs) such as molecular chaperones, organelle-targeting signal reader proteins and enzymes involved in co-translational protein processing.

- **Ribosomal RNAs (rRNAs).** In mammals there are four types of rRNA distributed into two different DNA loci, which are transcribed by two different RNA polymerases [329, 330]. The 18S, 5.8S and 28S rRNAs in one DNA locus, which are transcribed by RNA polymerase I (Pol I), and the 5S in another, transcribed by RNA polymerase III (Pol III). Once transcribed, the 18S rRNA together with RPSs are assembled in the nucleolus into the small ribosomal subunit (also called 40S subunit) and the 5.8S, 28S and 5S rRNAs are assembled together with RPLs into the large ribosomal subunit (or 60S subunit) in the nucleolus and nucleus [330].
  Important to mention are expansion segments, regions of rRNA that are generally located at the surface of the ribosome and that are particularly flexible. They have experienced one of the most dramatic variations throughout evolution, displaying a drastic increase in size from prokaryotes to eukaryotes [330]. They have been shown to serve as interaction points for particular ribosome-associated factors and are thought to play important roles in fine-tuning translational control [145, 330].

- **Ribosome-associated factors**, including RPBs are proteins that do not form part of the core ribosomal proteins, but that transiently interact with the ribosome for determined functions. These can include molecular chaperones, N-terminal localisation signal reader proteins or enzymes involved in N-terminal co-translational modifications [14].

There is increasing evidence of potential ribosomal heterogeneity that could give rise to 'specialised ribosome' pools. While the presence of such 'specialised ribosomes' has not been demonstrated yet, there is accumulating evidence of diversity for each of its different components [330, 331] (Figure 1.18). For RPs, differences in expression (which could lead to different stoichiometries or compositions) and in modification status have been reported [330, 331]. Moreover, within most species paralogues exist for many RPs. Although they arose by gene duplication and share high sequence similarity, genetic alterations of these genes have been shown to give rise to completely distinct phenotypes [330]. In some species, more than one set of different rRNAs have been reported that changed with changing stages of the organism's life cycle [330]. In addition, rRNA modifications such as methylation and pseudouridylation are now known to be ubiquitous, although the functional roles of these modifications are poorly understood [330]. Finally, different RPBs can interact transiently with ribosomes at different times to exert specific functions [332]. It is still to be determined how all these different sources of variation interact with one another to deliver exquisitely regulated ribosomes, which will in turn ensure a finely controlled protein synthesis and processing.

1.5.3 Co-regulation of co-translational N-terminal modifications

The major co-translational N-terminal modifications, iMet removal, N-acetylation and N-myristoylation, are all irreversible modifications. However, evidence for non-myristoylated pools of the known NMT substrate MARCKS was already described 20 years ago [333] and some recent studies have revealed co-occurrence of acetylated and non-acetylated versions of the same proteins [34, 40]. The same appears to be true for MetAPs, for which most proteins appear with either iMet retained or removed, but there is a small fraction of proteins that shows both types of N-termini in native samples [149]. Furthermore, comparative studies on *N*-acetylation and *N*-myristoylation have shown that while the majority of proteins were present in a single state, either free, *N*-acetylated or *N*-myristoylated, there seems to be at least some overlap between different co-translational N-terminal modifications [39, 40]. The fact that some overlap exists and that these modifications do not seem to appear complete might indicate competition between enzymes responsible for those modifications for the same peptide substrates. However, the exact implications of this overlap are still unclear and it seems like there might be a combination of parameters such as affinity, kinetic rates and relative availability that influence the ultimate fate of a protein.

Evidence suggests each of these enzymes (MetAPs, NATs and NMTs) interacts with ribosomes in addition to the nascent peptide substrate [6, 30, 269]. Recent studies have provided novel structural information on the binding of some of these enzymes to the ribosome [145, 328], which has led to a better understanding of how these processes might coexist. Through RNA mutation studies, the two yeast MetAPs (MetAP1 and MetAP2) as well as the catalytic subunit of NatA, Naa10, were shown to interact with the ES27L expansion segment of the ribosome, which is directly on top of the ribosomal exit tunnel [145]. Another study, which

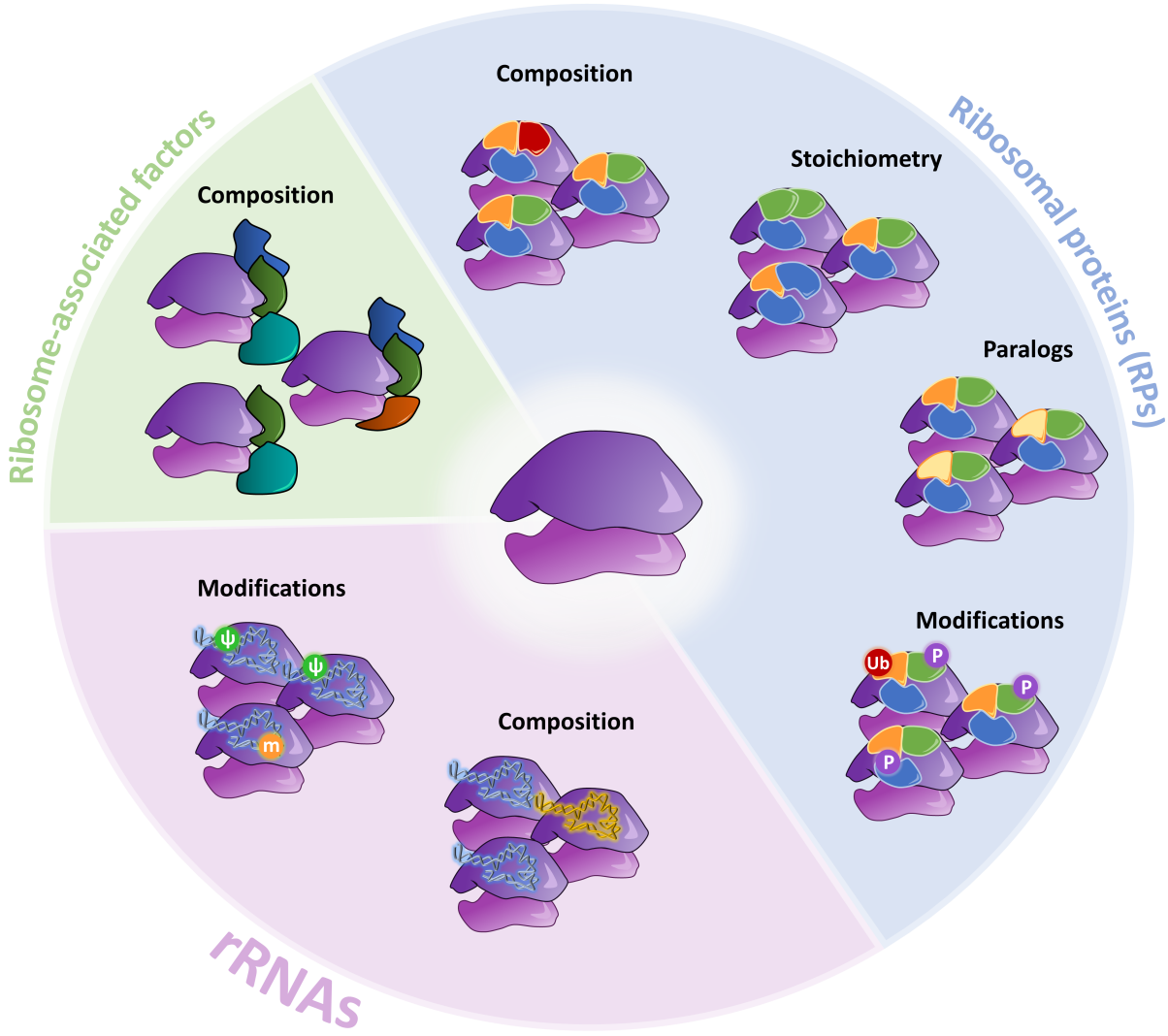


Figure 1.18: Sources of ribosomal heterogeneity. Diversity can arise from changes in composition, stoichiometry and modification status of RPs, changes in composition or modification status of rRNA or in composition of ribosome-associated factors, including RPBs. P = phosphorylation; Ub = ubiquitination; ψ = pseudouridylation; m = methylation.

solved the structure of the ribosome-NatA complex through cryo-electron microscopy (cryo-EM), further showed that the auxiliary subunit of NatA, Naa15, also interacted with the ribosome through both ES27L and ES7L expansion segments [328]. Interestingly, according to their model, MetAPs and NatA seemed to be able to fit simultaneously on the same ribosome [328], thus allowing for sequential processing first by MetAP and then by NatA of the same nascent peptides. No structural information exist for the interaction between NMT and the ribosome and therefore, further studies will be needed before we can fully understand how these enzymes are coordinated at the level of the ribosome. However, since NMTs share the same GNAT fold with NATs, it would be tempting to hypothesise that NMT could bind ribosomes in the same location as NatA, and thus compete for the same binding site at the ribosome [6]. Further, if there was competition between NMT and NatA and only one of them could be loaded onto one ribosome at a time, it could be speculated that this

happens based on their interaction with both the ribosome and the nascent peptide. Thus, differential affinity for nascent peptide sequences could drive the selection of which of the two enzymes stays ribosome-bound. Another possibility could be that, in line with the ribosome heterogeneity theory, NatA and NMTs are 'pre-loaded' onto different subtypes of ribosomes, which are in charge of transcribing different pools of transcripts. Such a case would imply that somehow the N-terminal modification status of each protein is imprinted in their mRNA and that NatA and NMT, by associating with different pools of specialised ribosomes, can only modify pre-defined sets of substrates. This model for the regulation of co-translational N-terminal modifications at the level of the ribosome is represented in Figure 1.19.

Another important fact to bear in mind is that, while abundant in comparison to most of the proteome, MetAPs, NATs and NMTs are expressed in very low quantities when compared to core ribosomal proteins [334], which means that even at maximal occupation it is only a small fraction of ribosomes that is associated to these enzymes at any one time. The model proposed by Fujii and colleagues for PDFs and MetAPs suggests extremely rapid kinetics of these enzymes allowing for fast association, processing and dissociation of each of these enzymes, permitting efficient sequential processing of PDFs and MetAPs at the regular protein synthesis rates [145]. Since N-terminal processing enzymes are only required at the start of every protein synthesis event, it would be reasonable to think that this model of rapid on and off binding kinetics could be extended to the rest of the enzymes involved in co-translational processing. Thus, these enzymes would only interact with the ribosome transiently and dynamically, in an on-and-off manner that allows for efficient processing of multiple nascent proteins that are being synthesised simultaneously. Importantly, such a model could explain the low expression levels of these enzymes relative to the number of ribosomes in a given cell.

1.5.4 Temporal control: co- vs. post-translational N-terminal modifications

Another important level of control one needs to consider when it comes to N-terminal modifications, is the temporal dimension. iMet removal needs to happen first to allow for subsequent co-translational modifications (e.g. *N*-myristoylation), and post-translational modifications will not start to play a role until the protein has left the ribosome. This establishes a staggered process for decision making. Since addition of a PTM at the N-terminus blocks it from further processing, only 'free', exposed N-terminal amines will be available for post-translational processing. This way, post-translational processing events have been proposed as quality control for co-translational events. This is particularly evident with regard to N-degron pathways. N-terminal sequences that target proteins for degradation have been shown to be generally depleted in intact proteins or masked through specific PTMs, folding or through interaction with other proteins [15, 112]. The N-Gly/N-degron pathway, for example, has been shown to be prevented by *N*-myristoylation and has been proposed to have evolved as a quality control mechanism to degrade aberrant proteins that escaped from NMT processing [112]. Interestingly, while N-terminal modification rules seem to be generally conserved across organisms,

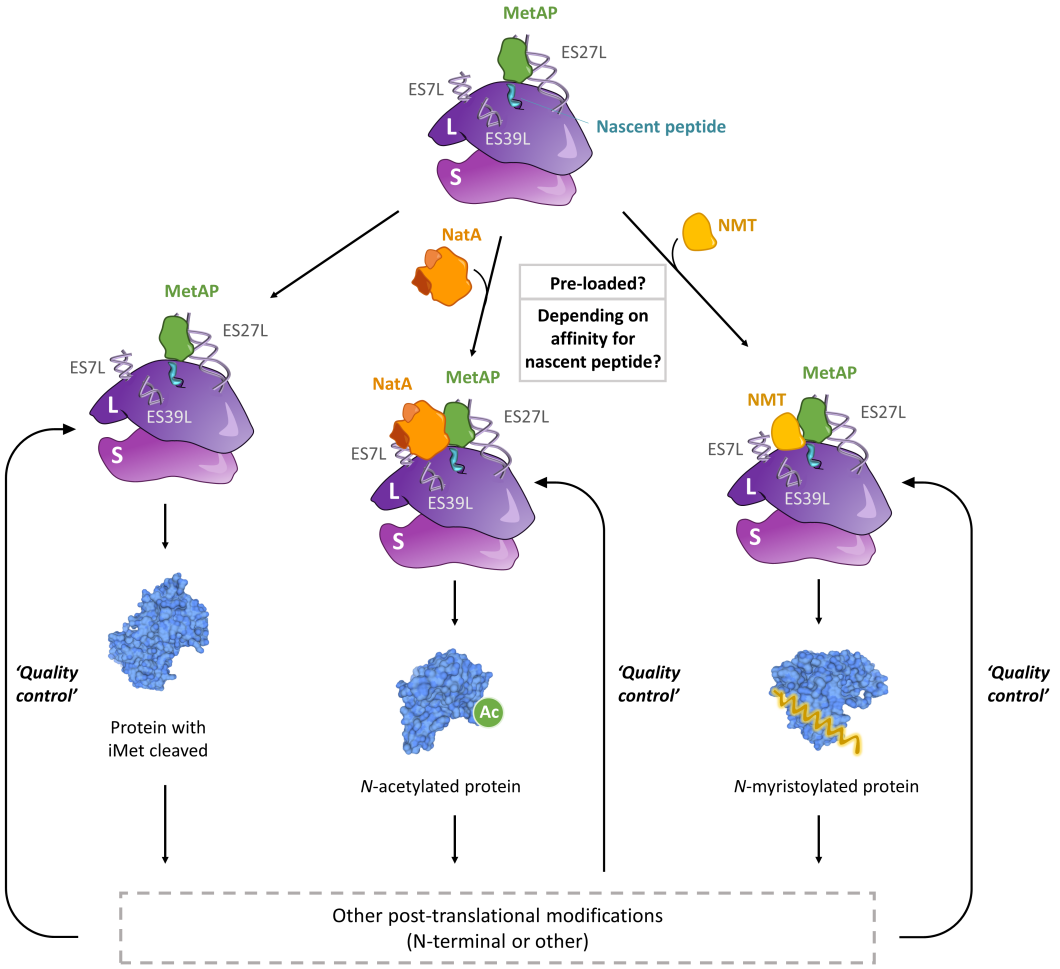


Figure 1.19: Model of regulation of N-terminal modifications at the ribosome level. MetAP and NatA are localised at the ribosome by interacting with ES27L and ES7L ribosome expansion segments. NMT is hypothesised to bind at the same site as NatA and compete for binding. Co-translational modifications happen first and thus, determine whether the nascent protein will have those modifications or will be amenable for further N-terminal processing by post-translationally acting enzymes. Post-translational modifications can therefore act as quality control mechanisms of co-translational events.

N-degron pathways are markedly different between prokaryotes, lower eukaryotes and higher eukaryotes, suggesting that higher organism complexity requires additional levels of control. Evolutionarily, this means that the different N-terminal modifications evolved to fit within the pre-existing pool of modifications. This seems to be the case, for instance, for N-degron pathways. Alternatively, they could have co-evolved together, shaping the prevalence of one another, as it has been proposed for N-myristoylation and Cys prevalence and palmitoylation within the first few residues [40].

1.6.1 Aim 1. To elucidate the substrate differences of *N*-myristoyltransferases 1 and 2

Human cells contain two distinct *N*-myristoyltransferases (NMTs): NMT1 and NMT2 [249, 250]. Their high sequence and structural similarities, particularly at the catalytic site, have hindered the development of isoform-specific inhibitors and there is no NMT inhibitor (NMTi) described to date which can distinguish between NMT1 and NMT2 [253]. Particularly interesting is the contribution of each NMT isoform to the NMT inhibition phenotype, which could reveal possible therapeutic avenues to treat cancer through targeting NMT.

Aim 1 is covered in **Chapter 2** and the main objectives include:

1. To generate NMT1 and NMT2 genetic knockouts (KOs) using CRISPR/Cas9 technology in a cancer cell line of choice (Panc-1).
2. To validate the generated NMT1- and 2-KO clones by Western blot and sequencing.
3. To characterise the generated NMT1- and 2-KO clones for general cell biology differences, including morphology and cell cycle analysis.
4. To assess sensitivity differences of NMT1- and 2-KOs to NMTi's.
5. To apply the previously described chemical proteomics workflow using YnMyr metabolic labelling (REF) to the NMT1- and 2-KO clones to elucidate the specific substrates of NMT1 and NMT2.
6. To study the proteome-wide effects of genetic deletion of NMT1 and NMT2 to shed light on the specific roles of each NMT isoform in cancer.
7. To compare the YnMyr enrichment and whole proteome proteomics results of the generated NMT1- and 2-KO clones to analogous clones generated in another cell line (HeLa) by a postdoc in the lab.

1.6.2 Aim 2. To generate a biomarker assay to profile NMT activity in complex biological samples

Validation of on-target activity of inhibitors targeting a specific enzyme is critical in the advancement of inhibitors of potential clinical value. Current methodologies to assess NMT activity and inhibition upon NMTi treatment rely on the use of the myristate analogue chemical probe YnMyr, which is metabolically incorporated into cells in culture upon addition onto the cell media [335, 336]. While this strategy has proven itself invaluable for the

identification of NMT substrates and the assessment of NMT activity in cells [244, 312], it poses major challenges for its scalability to animal models or human studies. Therefore, new methods are needed that are able to assess NMT activity and inhibition in complex biological samples.

Aim 2 is covered in **Chapter 3** and the main objectives include:

1. To design and develop a new strategy to profile NMT activity that does not rely on metabolic tagging. This new strategy will exploit the substrate overlap of NMT with the bacterial enzyme sortase A (SrtA), which both recognise free N-terminal glycines.
2. To optimise the SrtA labelling conditions to label free N-terminal glycines in whole-cell lysates.
3. To characterise the substrate specificity of SrtA in whole-cell lysates and to validate its labelling capabilities are sufficient to profile a wide range of NMT substrates.
4. Using chemical proteomics, to validate the SrtA labelling strategy is capable of detecting NMT activity changes in a NMTi dose-dependent manner and to compare the SrtA labelling profile to the previously described YnMyr metabolic labelling method [244].
5. To select a few NMT substrates showing a reliable dose-dependent SrtA labelling profile as biomarkers for NMT activity and to extend this strategy to a Wester blot format for ease of application.
6. To extend the SrtA biomarker assay to other gel-based and ELISA formats for increased applicability and throughput.

1.6.3 Aim 3. To develop a novel chemical proteomics approach to elucidate the specific substrates of methionine aminopeptidase 2

Methionine aminopeptidases (MetAPs) remove initiator methionine (iMet) from nascent peptides, a process estimated to occur in >70% of eukaryotic proteins and that precedes the addition of other N-terminal modifications, such as *N*-myristoylation [6, 9]. Type II MetAP (MetAP2) is of particular pharmacological interest since the discovery of natural inhibitors of MetAP2 with potent anti-angiogenic agents with potent anti-tumour activity [164, 176, 180]. However, despite the three decades of extensive research on developing MetAP2 inhibitors (MetAP2i's), little is known about the underlying mechanism of action of these compounds. Given its potential overlap with *N*-myristoylation, collaborations were started with several pharmaceutical companies to deepen the understanding of the mode of action of their preclinical and clinical candidates. A major question regarding the biological role of MetAP2 is the nature of its substrates, which could help bridge the gap between MetAP2

inhibition and its phenotypic effects.

Aim 3 is covered in **Chapter 4** and the main objectives include:

1. To characterise the MetAP2i panel obtained through different collaborations under the same biological and biochemical settings.
   (a) To establish a cellular model for MetAP2 inhibition.
   (b) To develop a phenotypic assay capable of quantifying MetAP2i activity in a robust and consistent way across the MetAP2i panel.
   (c) To develop a biochemical assay to determine binding of MetAP2i to MetAP2 in cells across the MetAP2i panel.
2. To design, develop and optimise a novel chemical proteomics workflow to systematically identify the specific substrates of MetAP2 across the whole proteome.
3. To study the overlap between MetAP2 and NMT and the effects of their inhibitors.

2

Dissecting the individual roles of *N*-Myristoyltransferase 1 and 2 in cancer

2.1 Development of NMT1 and 2 genetic knockouts by CRISPR/Cas9

N-myristoylation involves the attachment of a 14-carbon saturated fatty acid, myristic acid, onto the N-terminal Gly amino group of protein substrates [241]. This lipid modification has been shown to mediate reversible membrane binding, promote interaction with binding partners, and to play a role in protein 3D structure and half life [9, 277–279] (cf. Introduction section 1.4.3). With such important roles, the relevance of this irreversible protein modification in cellular homeostasis and disease has been well established (cf. Introduction section 1.4.4). However, the understanding on how this process is regulated in disease states such as cancer and how pharmacological inhibition mediates specific cytotoxicity of these cells is still poorly understood. Deeper understanding on how NMT is (dys)regulated in cancer and how this affects the global network of *N*-myristoylated proteins is therefore of paramount importance, in particular to advance in the development of NMT inhivitors with the aim of generating new clinical candidates. Something that has remained elusive to date is the contribution of each NMT isoform to the NMT inhibition phenotype, which could reveal possible therapeutic avenues to treat cancer through targeting NMT.

Due to the lack of selective inhibitors capable of distinguishing between NMT1 and NMT2 enzymes [253], genetic manipulation of their genes constitutes the only way of establishing the differences in their substrate specificities and biological roles. Previous studies have used small interfering RNA (siRNA) to specifically knock down (KD) expression of NMT1 or NMT2, which exhibited distinct phenotypes: KD of NMT1 inhibited cell replication, while reduction of NMT2 expression induced apoptosis by altering the expression of BCL family proteins [250]. siRNA KD approaches, however, while allowing for fast induction of the phenotype and potentially resembling chemical knock down by small molecule inhibitors better, usually require a very careful tuning of both the siRNA titer and incubation times and are not always compatible with the set up of further experiments. Moreover, similar to chemical knock down approaches, they need induction of the knock down for every experiment, but in contrast to chemical inhibitors, the dose-response relationship is harder to establish and generally less consistent, potentially leading to cross-experimental variability. This, together with the heterogeneity in responses between cells within a siRNA-treated population, can lead to a gradation in the observed phenotypes that hinders interpretation and prevents direct comparison between different KDs. Therefore, a cell line that stably expresses the genetic modifications of interest was considered a more appropriate model to study the two NMT enzymes. With the advance of the CRISPR/Cas9 technology as a quick and effective way of generating stable genetic knockouts (KOs) in mammalian cells [337], all tools required were either already available or easy to generate. Therefore, we set out to generate *NMT1* and *NMT2* gene KOs in cancer cells. Since *N*-myristoylation is an essential protein modification in all eukaryotes, it was hypothesised double *NMT1* and *NMT2* KOs would not be viable and therefore, these were not attempted to be generated. A previous study which attempted to generate full *Nmt1*-null ($Nmt1^{-/-}$) mice showed NMT1 ablation was embryonically lethal

[254]. This was determined to be due to a specific role of NMT in embryonic development [254], and hence, it was hypothesised cells in culture would not be affected to the same level.

2.1.1 Choice of cell line

Cancer cell lines often show aberrant karyotypes with multiple chromosomal duplications and deletions [338], which can interfere with genetic manipulation techniques [339]. Following on this, the pancreatic cancer cell line Panc-1 was selected, which according to the distributor (ATCC) documentation, only contained two copies of chromosomes 10 and 17, where *NMT2* and *NMT1* are located respectively. In addition, Panc-1 has been widely used as pancreatic cancer cell model both for cell culture and xenograft studies [340]. A postdoc in the lab, Dr. Monica Faronato, had generated similar KOs in HeLa cells, which despite their easier handling, are generally of little biological significance [341]. Thus, the combination of the two cell lines had greater potential to give a meaningful insight to the mode of action of our genes of interest in disease.

2.1.2 CRISPR/Cas9 knockout generation

The procedure for CRISPR/Cas9 mediated genome editing was carried out as shown in Figure 2.1. Six primer pairs were designed and prepared for transfection, three against NMT1 (AG1, AG2 and AG3) and three against NMT2 (AG4, AG5 and AG6), each targeting a different exon of the genes to maximize success rates (gRNA sequences can be found in Methods Table 6.1). Each of these were cloned into a gRNA plasmid for mammalian expression under the U6 promoter (plasmid p41824, Addgene)[342]. After ligation by Gibson assembly and PCR amplification, each of the constructs was sequenced to ensure correct insertion of the gRNA construct into the gRNA plasmid. From all successfully generated constructs, two constructs were randomly selected for transfection per gene (AG1 and AG3 for *NMT1*; AG4 and AG6 for *NMT2*). After bacterial amplification in *E. coli* DH5α, gRNA constructs were co-transfected together with the p44719 Cas9-GFP plasmid (Addgene)[343] into Panc-1 cells. Positive cells were sorted into 96-well plates as single cells per well based on their green fluorescence levels 72 h post-transfection via fluorescence-activated cell sorting (FACS) (FACS gating settings are shown in Appendix Figure A.1). Control cells were transfected with Cas9-GFP plasmid only. *NMT2*-KO candidate clones grew in 9.9% of the wells (controls in 9.5%) in a time comparable with the control cells (ca. two months). In contrast, *NMT1*-KO candidate clones grew in only 4.7% of the wells, and at a significantly slower growth rate, with double *NMT1*-KOs needing six months to grow sufficiently to enable further experiments.

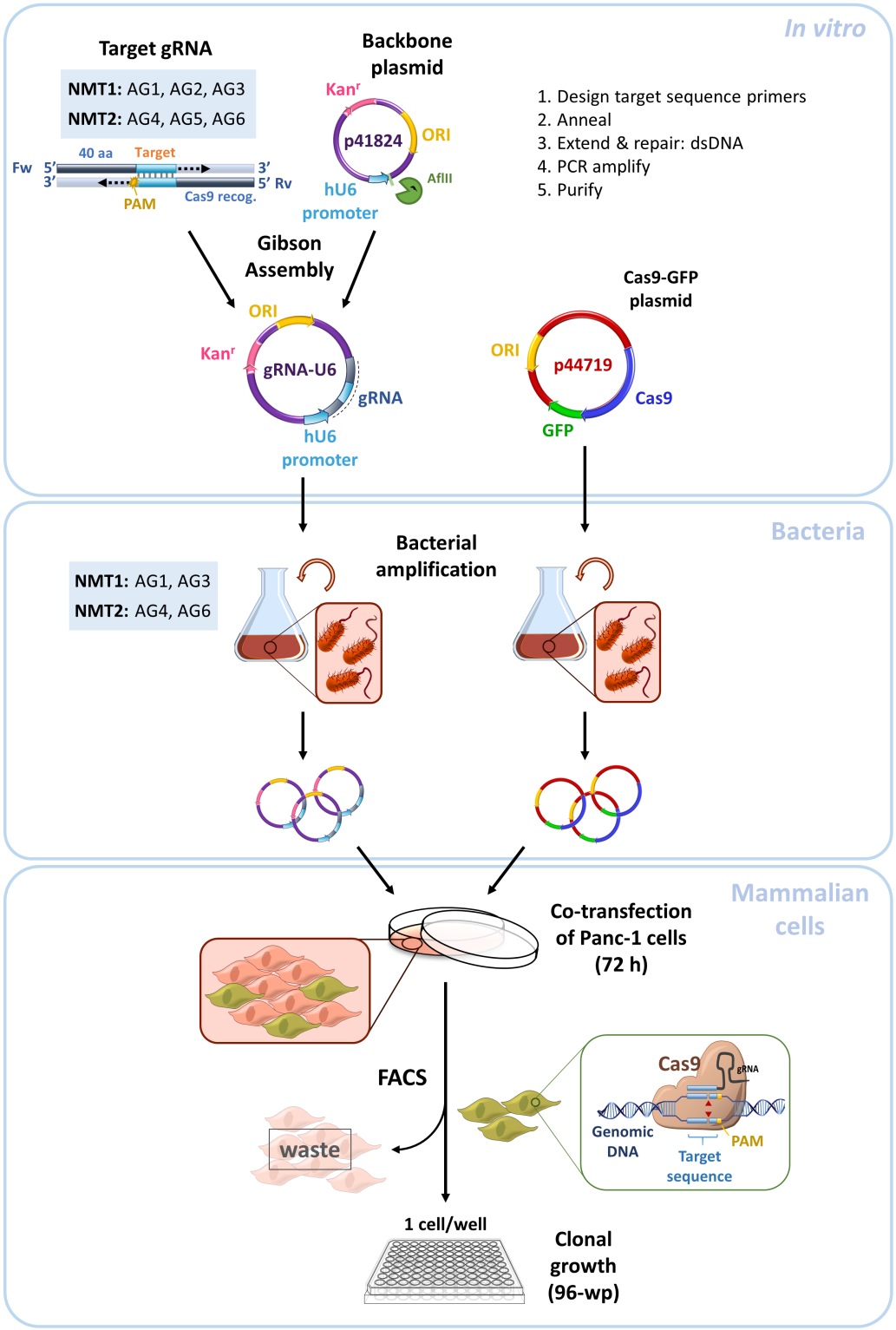


Figure 2.1: CRISPR-Cas9 knockout procedure used to generate NMT1- and 2-KOs in Panc-1 cells. Three different gRNAs were designed to target either *NMT1* (AG1, AG2, AG3) or *NMT2* (AG4, AG5, AG6). Each of those was generated from a single stranded DNA (ssDNA) primer pair that contained a complementary sequence of 20 bp that coded for each target sequence and an overhang region of 40 bp at each side. ssDNA primer pairs were annealed, extended and amplified by PCR before insertion into the p41824 gRNA cloning plasmid (Addgene) by Gibson assembly. Two full gRNA constructs per gene (AG1 and AG3, AG4 and AG6) and the GFP-Cas9 plasmid were then amplified in *E.coli* DH5α, before co-transfection into Panc-1 cells. CTRL cells were transfected with GFP-Cas9 plasmid only. 72 h post transfection, GFP-positive cells were sorted into a 96-well plate as single cells per well to allow for clonal growth.

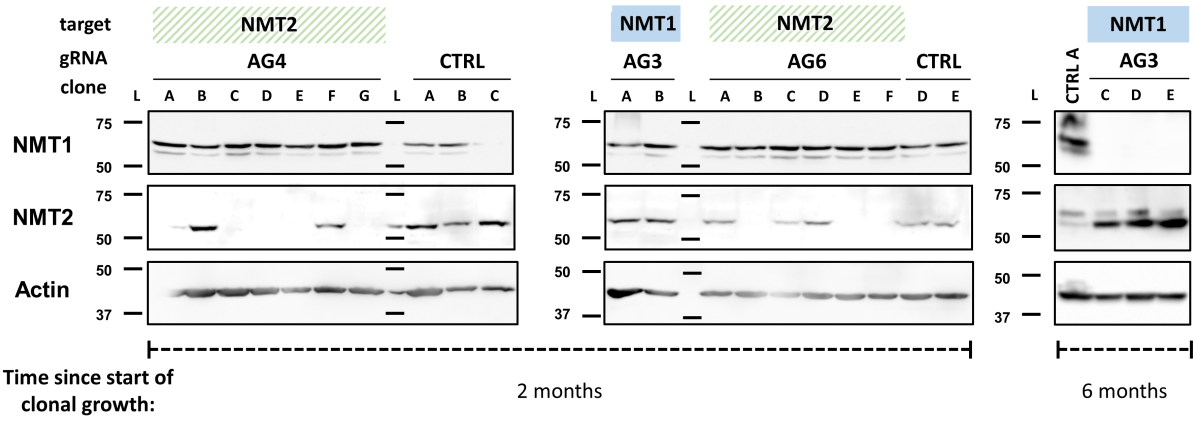


Figure 2.2: Screening for full NMT1- and 2-KOs by western blot. Eight NMT2-KOs (clones A, C, D, E and G of AG4 and clones B, E and F of AG6) showed complete depletion of the NMT2 band indicating full homozygous KO of NMT2 in those clones. All these clones were grown in less than two months, similar growth rate to that shown by CTRL cells. Three complete NMT1 KOs were obtained (clones C, D, and E of AG3) and they required 3 times longer than control cells to grow. No clones of gRNA AG1 survived.

2.1.3 Clone validation and selection

Clones were screened for successful complete NMT1 and 2 ablation by Western blot (Figure 2.2). Eight NMT2 clones showed complete disappearance of the NMT2 protein band, five derived from AG4 gRNA (clones A, C, D, E and G) and three from AG6 gRNA (clones B, E and F). For NMT1, only clones with the AG3 gRNA thrived, and only those which took over six months to grow showed full NMT1 depletion by Western blot. This striking difference in growth rates of NMT1-KO *vs* NMT2-KO clones (2 vs 6 months; CTRL: 2 months) and cell viability (4.7% vs 9.9% initial cell viability) already hinted towards a higher sensitivity of cells to NMT1 ablation than that of its paralog NMT2. Even after six months, NMT1-KO Panc-1 cells showed extremely impaired cell growth. In fact, some clones displayed a complete growth arrest, detected as a small number of cells in some wells of the initial 96-well plate that did not increase in number during the whole clonal growth process. Together, this supported the idea of distinct biological roles for these two enzymes [249, 254], with a significantly more important role for NMT1 in overall cell survival and proliferation [9, 254, 255].

The gRNA region of each of the clones was then sequenced to confirm the presence of insertions and deletions (indels) that could indicate the presence of knockouts by truncation due to frame-shifting. In this CRISPR/Cas9 strategy the regular Cas9 generates a double strand break a few base pares after the protospacer adjacent motif (PAM) sequence [344]. In the absence of a donor DNA template, this is then repaired primarily through non-homologous end joining (NHEJ), which is prone to errors and often results in the generation of random indels [344]. Genomic DNA of each of the positive clones was extracted and subjected to PCR using custom made primers (Sequences of all primers can be found in Table 6.2). After several PCR rounds, the regions around the AG3 target sequence in the *NMT1* gene and around the AG6 target sequence in the *NMT2* gene were successfully amplified (Figure 2.3A). Sequencing

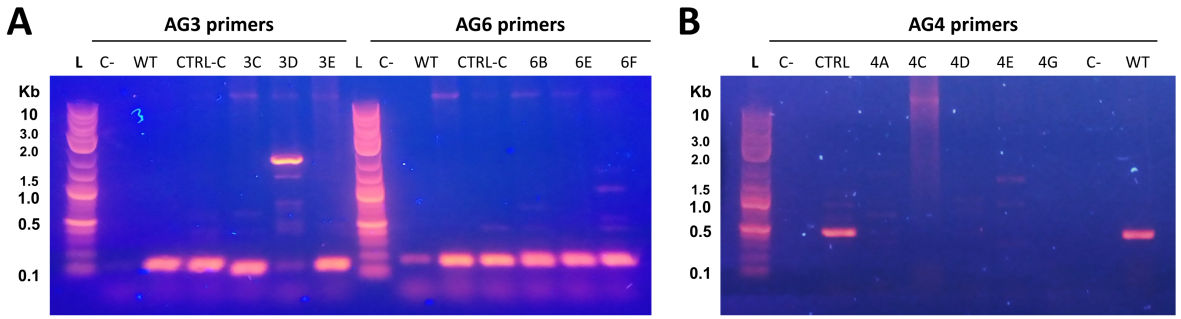


Figure 2.3: PCR amplification of genomic DNA of Panc-1 clones at the gRNA target region. Genomic DNA was extracted and amplified using custom primers for each target gRNA sequence. **(A)** Regions including AG3 and AG6 target sequences were amplified successfully showing mostly small changes in band size, with the exception of AG3D, which showed a big insertion of 1.7 kb. **(B)** The region including AG4 target sequence failed to be amplified in every clone.

analysis of AG3 NMT1-KO clones revealed small deletions of 25 and 3 bp in clones AG3C and AG3E respectively, and a larger insertion in clone AG3D (Appendix Figure A.2). This could not be efficiently determined by the sequencing read but was estimated to be of ca. 1.7 kb according to the apparent electrophoretic band shift (Figure 2.3A). Sequencing analysis of AG6 NMT2-KO clones revealed small indels in all of them: a 3 bp deletion in AG6B, a 1 bp insertion in AG6E and a 3 bp insertion in AG6F (Appendix Figure A.2). In contrast, the AG4 gRNA region (*NMT2* gene) of the clones eluded all efforts of PCR amplification. AG4 gRNA targeted the first exon of the NMT2 gene and the availability to create a forward primer upstream of that sequence was limited. After four different set of primers, only the AG4 region of CTRL and WT cells was amplified but not that of the clones (Figure 2.3B), implying that region might have been deleted altogether during the double-strand break repair process.

The CRISPR/Cas9 procedure has been shown to result in off-target gene editing, even after highly optimised gRNA designs [345]. Since off targets are either non-specific or gRNA-sequence related, two clones were selected per *NMT* gene for further studies, preferentially derived from different gRNAs. By studying the common effects in each clone and discarding the clone-specific effects, it should be possible to differentiate the effect of knocking out either *NMT* from off-target gene-editing effects generated during the CRISPR procedure, as well as from differential adaptational mechanisms adopted by each cell clone to counteract the genetic deficiency. For NMT2, AG4C and AG6B clones were arbitrarily selected from the clones that showed clear complete KO during the Western blot screen (Figure 2.2). For NMT1, unfortunately, only clones derived from AG3 gRNA survived and therefore, two clones generated with the same gRNA had to be selected. Due to the structural differences to the resulting *NMT1* gene (Figure 2.3), AG3D and AG3E were selected for further studies.

2.2 Adaptation to NMT1 ablation comes in different flavours, but all include slower growth rates, higher cellular cytotoxicity and increased cell size

2.2.1 Cell cycle analysis of Panc-1 clones reveals profound effects on NMT1-KO clones

The next step was to characterise the generated NMT1- and 2-KO clones. Encouraged by the differences seen in growth rates between NMT1- and 2-KO clones, they were subjected to cell cycle analysis using flow cytometry. For this purpose, 5-ethynyl-2'-deoxyuridine (EdU) was used, a thymidine analogue bearing an alkyne handle, which gets incorporated into newly synthesised DNA and thus serves as a proxy for cellular proliferation. Moreover, by comparing EdU incorporation rates to DAPI nuclear stain (total DNA content), it is possible to dissect the three main phases of the cell cycle: G1/G0 with low DAPI and low EdU; S phase with low-to-medium DAPI and high EdU; and G2/M with high DAPI and low EdU. An antibody against active caspase-3 was also used, the final effector of the caspase cascade activation leading to apoptosis.

Flow cytometry analysis of Panc-1 clones revealed strong changes in cell cycle patterns of NMT1-KO clones, whereas changes in NMT2-KOs were more modest (Figures 2.4A-D). Surprisingly, NMT1-KOs showed opposite trends in general population distribution across the three major cell cycle phases (Figures 2.4A). AG3D showed a significant decrease in S phase as compared to WT/CTRL with concomitant increase in both G2/M and G1/0 phases, similar to what had been reported for cells under NMT inhibition [312]. In contrast, AG3E showed a higher proportion of cells in S phase than CTRL/WT with reduced G2/M population (Figures 2.4A and B). Importantly, both NMT1-KO clones were considerably bigger in overall cell size (Figure 2.4D), which was readily noticeable under regular phase-contrast microscopy (Figure 2.4E). There are ca. 100 NMT substrates identified to date and >300 estimated [40, 244]. If *N*-myristoylation is significantly affected in these clones, it is likely that there exist several NMT substrates that might be necessary for normal cell cycle progression. Moreover, being an N-terminal modification and with NMTs having close contact with ribosomes, NMT deficiency could lead to general defects in translation regulation and impaired protein synthesis. Larger cells are often associated with slower cell cycles and senescence [346]. Therefore, it was hypothesised that cells try to cope with the lack of NMT1 by slowing down the cell cycle and allowing enough time to generate the necessary components to pass the G1 checkpoint. AG3D cells do not fulfil quality control criteria and partially arrest in G0/1, while AG3E are able of passing the G1 checkpoint but are then unable to complete S phase. Together, these data indicate that NMT2 depletion is well tolerated, whereas NMT1 ablation is detrimental for normal cell growth and survival and hence, cells required to adopt profound changes in the cell cycle in order to overcome the lack of this enzyme.

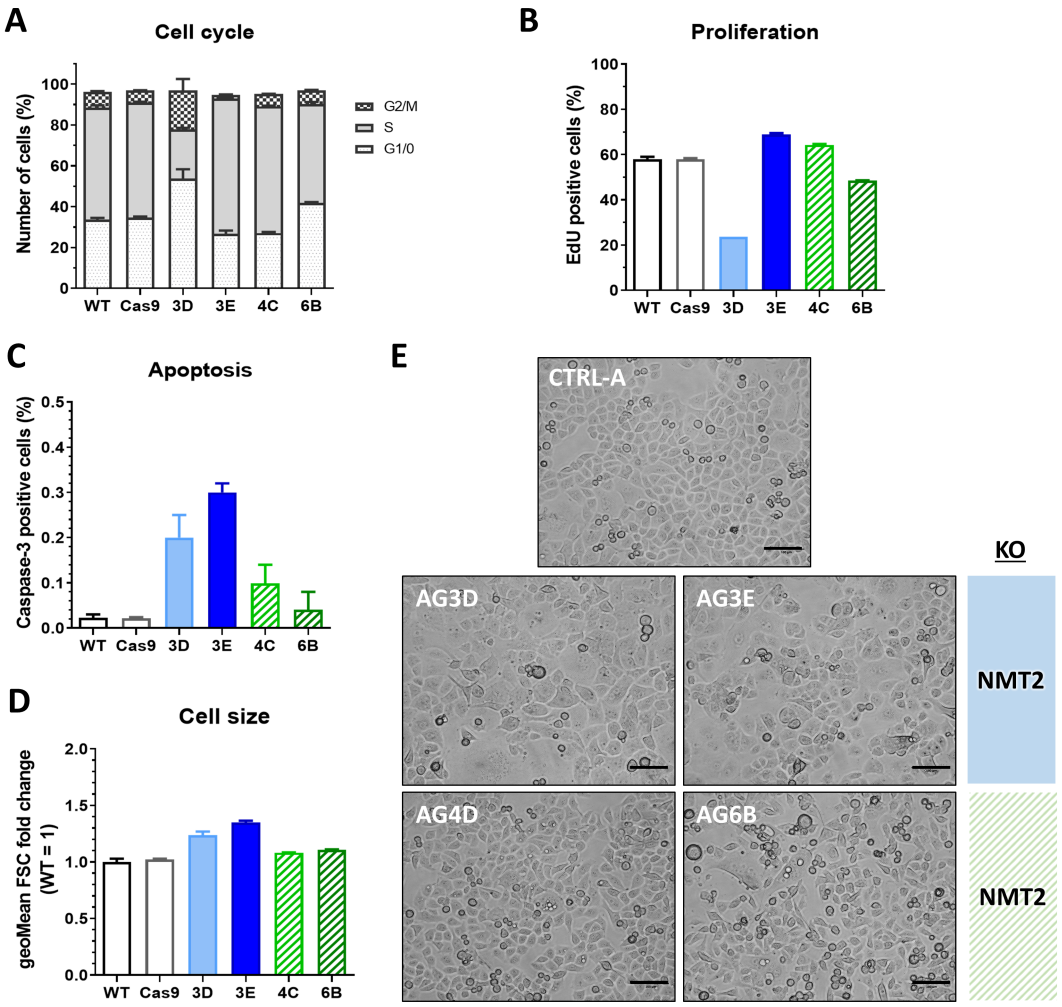


Figure 2.4: Characterisation of Panc-1 NMT1- and 2-KOs: cell cycle distribution. (A) Flow cytometry analysis of the cell population distribution across cell cycle phases (Edu and DAPI), **(B)** proliferation frequencies (EdU), **(C)** apoptosis frequencies (active caspase-3) and **(D)** cell sizes (forward scatter, FSC). Error bars represent SEM in all cases. **(E)** Representative phase-contrast microscopy images showing the sizes of the clones adhered to the culture plates. Scale bars represent 100 µm.

2.2.2 Sensitivity screening of Panc-1 clones to NMT inhibition

Next, the ability of the NMT-KO clones to withstand NMT inhibition by small molecule inhibitors was studied. Panc-1 cells are relatively resistant to NMT inhibitors on short exposures and require longer exposure-times to show the typical patterns of NMT inhibition. Therefore, Panc-1 clones were dosed with two different small molecule inhibitors of NMT1/2 for 6 days before measuring cell viability using an MTS assay (Figure 2.5). A well established tool compound (IMP-366/DDD-85646) [244] and a more potent molecule designed and developed within the lab (IMP-1088) [298] were used, both of which have been extensively characterized and validated as NMT inhibitors by the Tate group and others [244, 298, 312, 317, 318] (Figure 2.5A). In this assay, MTS (negatively charged, and thus non cell permeable) is used in combination with PMS to transfer electrons from NADH in the cytoplasm to MTS in the

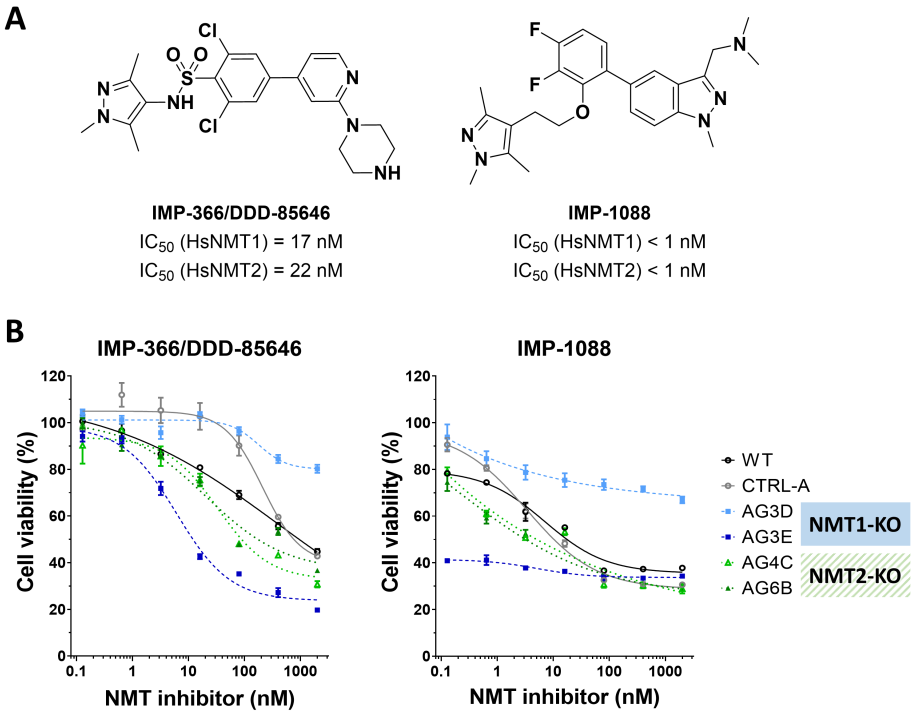


Figure 2.5: Sensitivity screen of Panc-1 NMT1- and 2-KOs to three small molecule inhibitors of NMT. (A) Structures and IC50s of IMP-366/DDD-85464 and IMP-1088 NMT inhibitors. **(B)** Cells were incubated with dual NMT1/2 inhibitors IMP-366/DDD-85464, IMP-1088 and IMP-1320 for 6 days before cell metabolic activity measurement by MTS assay. Error bars represent SEM.

cell culture medium [347], reducing it to generate purple soluble formazan products in the medium. Formation of formazan can be easily monitored by measuring absorbance at 570 nm, and therefore constitutes an easy and direct readout of metabolic activity of cells. For these assays cells are typically seeded in a number that will result in near-full confluence at the end of the experiment, to maximise signal differences across conditions. Since NMT1-KO clones doubled at a much lower rate than NMT2-KO or WT/CTRL cells, double the amount of NMT1-KO cells were seeded at the outset (4,000 cells/well vs 2,000 cells/well in a 96-well plate).

Building on previous results on the cell cycle, which indicated NMT1 ablation was detrimental for normal cell growth, NMT1-KO cells were expected to be particularly sensitive to further NMT inhibition. Surprisingly, this was only true for one of the NMT1-KO clones, AG3E. In contrast, AG3D showed higher resistance to NMT inhibition than CTRL/WT cells (Figure 2.5B). It was considered that the G1/0 arrest displayed by this clone (Figure 2.4A) could interfere with the NMT inhibition phenotype, which requires active protein synthesis. Moreover, the cell cycle arrest could interfere with the MTS assay itself, since cells with lower metabolic activity will result in slower conversion rates of the MTS reagent and thus yield less formazan product. NMT2-KO clones showed a slight left shift of the dose-response curves, indicative of a small increase in sensitivity to NMT inhibition. This further supported the working hypothesis that NMT2 is less important to drive *N*-myristoylation in the cancer cell.

2.3 Proteomics analyses reveal NMT1 as the principal enzyme responsible for *N*-myristoylation in cancer cells

Proteomics is a powerful tool to study biological systems, allowing proteome-wide identification and quantification of proteins [348]. For the study of protein lipidation, chemical proteomics approaches have been some of the most successful to date, particularly for the identification of NMT substrates (cf. references [9, 244, 312, 335, 336] and section 1.4.1). Chemical proteomics approaches rely on metabolic incorporation of lipid analogue probes; in the case of *N*-myristoylation, YnMyr (tetradec-13-ynoic-acid) (Figure 2.6B). YnMyr is readily taken up by cells when added to the culture medium, and is incorporated into the endogenous metabolic pathway of myristic acid through its conversion to YnMyr-coenzyme A (CoA) by cellular fatty acyl CoA synthetases [244, 335]. YnMyr-CoA is then recognised by cellular NMTs and transferred onto its natural protein substrates [335]. Bearing an alkyne (Yn) handle, YnMyr then allows for highly selective and efficient chemical ligation to custom made molecules via copper-catalysed alkyne-azide cycloaddition (CuAAC), often referred to as 'click chemistry' (Figure 2.6B). Through click chemistry, probes bearing an alkyne (e.g. YnMyr) or an azide can be specifically ligated to azide- or alkyne-bearing di- or tri-functional capture reagents containing affinity tags (e.g. biotin) or fluorophores (e.g. TAMRA) upon cell lysis (Figure 2.6A). These then allow for tracking of probe incorporation events in cells, which in the case of YnMyr mirror *N*-myristoylation events. Through ligation to capture reagents containing a fluorophore and subsequent in-gel fluorescence analysis, cellular *N*-yristoylation can be visualised in gels. Alternatively, by ligation to capture reagents bearing an affinity handle, they can be selectively enriched for either western blot detection of specific proteins or sytematic shotgun proteomics identification (Figure 2.6B).

In this section, two complementary proteomics approaches were applied to the panel of Panc-1 clones to extensively characterize the proteome-wide effects of knocking out either NMT isoform. First, the YnMyr metabolic tagging strategy developed within the lab was used to track down protein *N*-myristoylation across the whole proteome [244]. Second, traditional whole proteome (no enrichment) proteomics was used to study the proteome-wide effects of genetically deleting either NMT1 or NMT2 on these cells. Both workflows are summarised in Figure 2.7.

2.3.1 In-gel fluorescence analysis of YnMyr labelling

YnMyr labelling was first analysed by in-gel fluorescence using AzTB (azide-TAMRA-biotin) as capture reagent (structure in Figure 2.6B). After click chemistry, proteins were precipitated before loading them onto 12% SDS-PAGE gels and scanning TAMRA fluorescence. In-gel fluorescence gel images did not show any clear differences among bands between WT/CTRL lanes and KO clones (Figure 2.8A). This was not surprising, however, since gel-based analyses

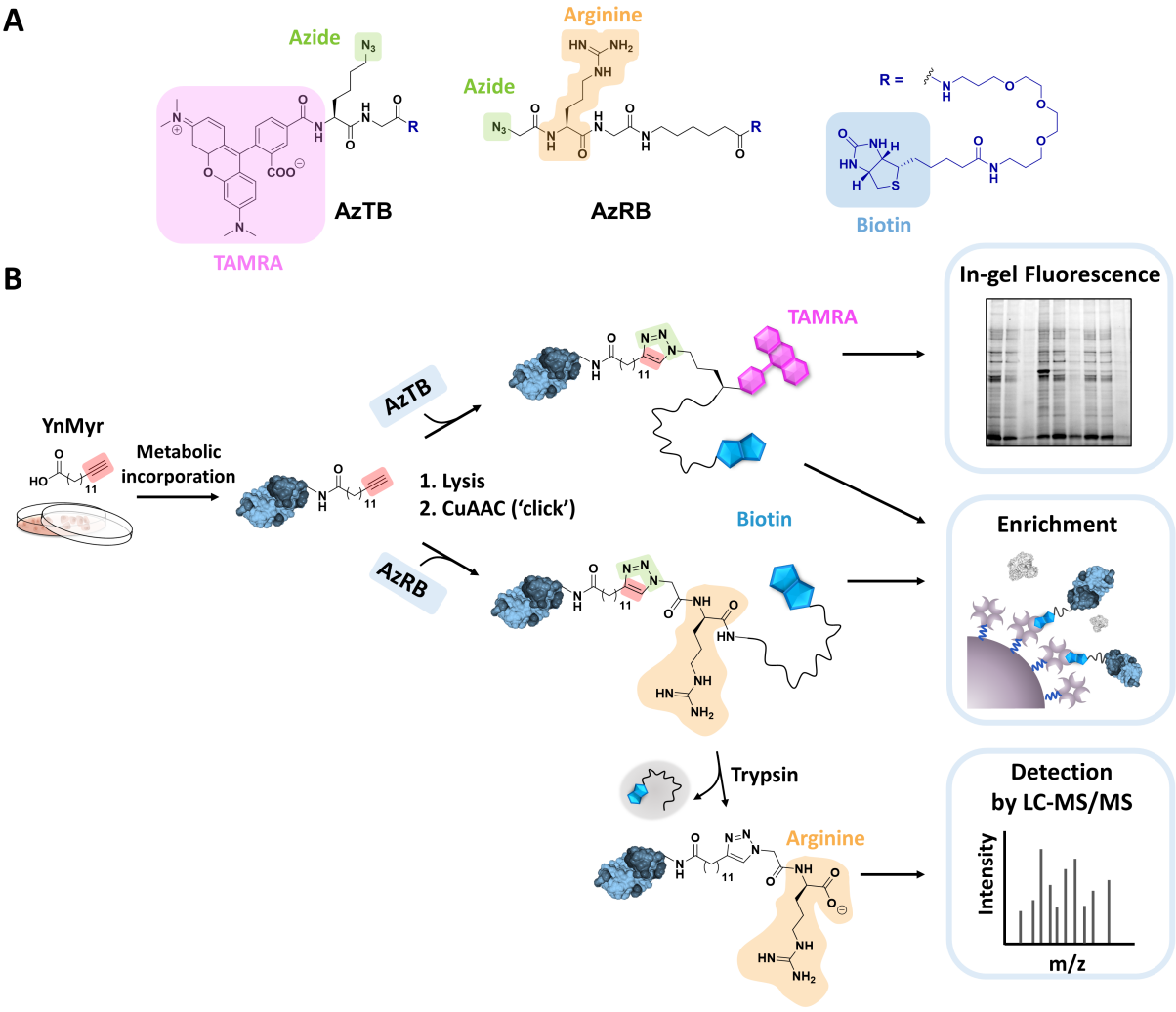


Figure 2.6: Metabolic incorporation of YnMyr followed by click chemistry allows for the systematic study of *N*-myristoylation. (A) Structures of AzTB and AzRB capture reagents. **(B)** YnMyr is supplemented in the cell culture medium, taken up by the cells and transferred onto proteins by cellular NMTs. Upon lysis, protein-bound YnMyr can be specifically ligated to capture reagents such as AzTB and AzRB for analysis of specific *N*-myrostoylated proteins or general *N*-myristoylation patterns across the proteome.

can only reveal bulk changes in complex biological mixtures and might miss specific labelling events. In contrast, addition of NMT inhibitor IMP-1088, systematically decreased labelling of all major bands, thus confirming specificity of the signal for protein *N*-myristoylation events (Figure 2.8B).

2.3.2 YnMyr proteomics analysis of NMT1- and 2-KOs reveals substrate preferences for each NMT isoform

Next, YnMyr-labelled proteomes from each clone were processed for proteomics following the procedure described in Figure 2.7A. For this purpose, a spike-in SILAC approach previously shown to successfully identify a wide-range of NMT substrates [244] was used. In this approach heavy (K8, R10) lysate derived from WT Panc-1 cells was spiked into every sample in a 1:3

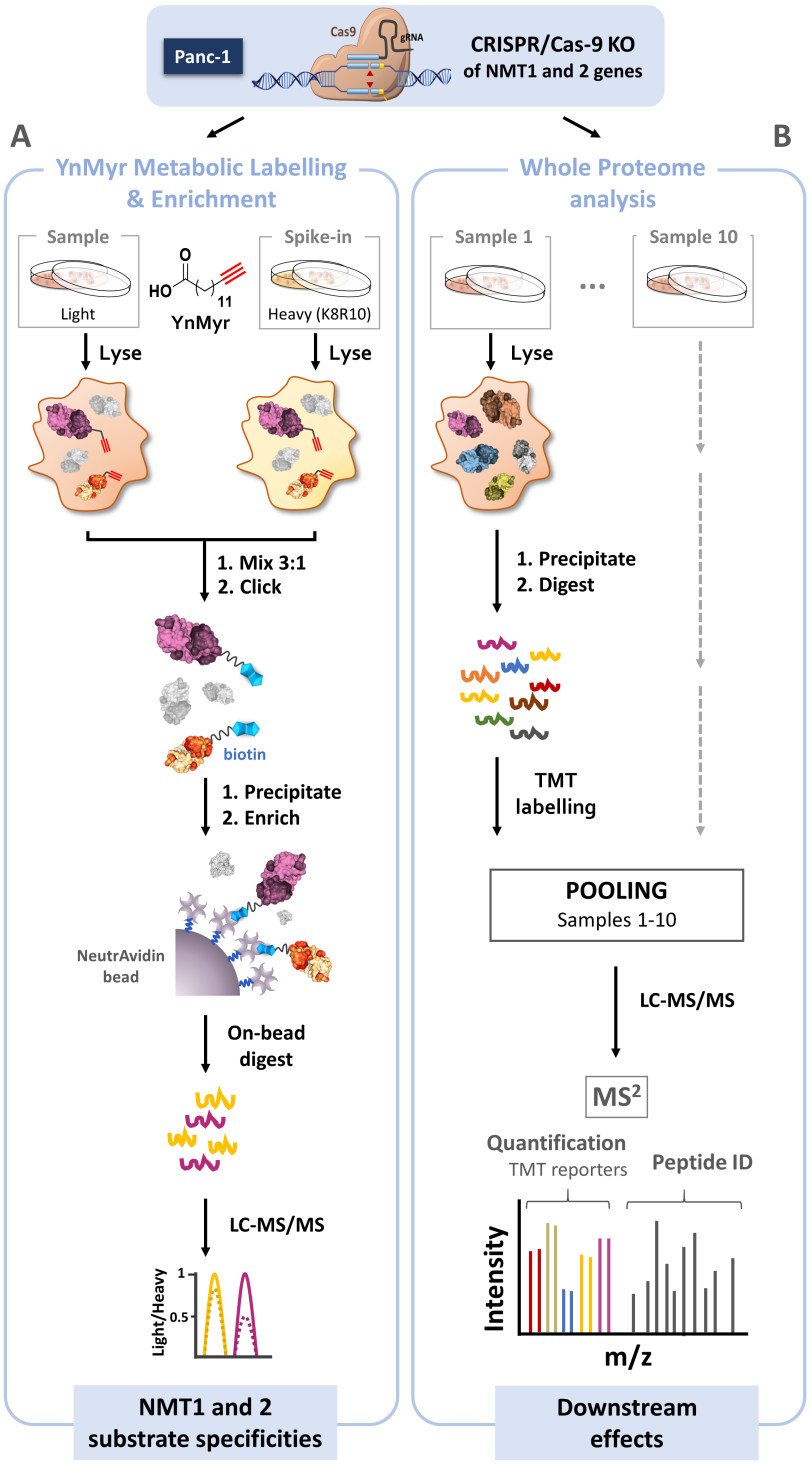


Figure 2.7: Schematic workflow representing the two complementary proteomics approaches applied to NMT1- and 2-KO clones. (A) clones (light) and WT (heavy) were incubated with 30 µM YnMyr for 24 h before lysis. Light protein lysates were mixed with heavy spike-in in a 3:1 ratio before ligation of incorporated YnMyr onto AzRB capture reagent by click chemistry. Proteins were then precipitated, enriched in NeutrAvidin beads and digested on-bead with trypsin, before desalting and LC-MS/MS. Enrihcment levels were calculated as ratios against the WT heavy spike-in. **(B)** For whole proteome analysis, clones were lysed, and protein samples precipitated and digested with trypsin before TMT labelling of peptides. TMT labelled samples were then pooled into 10-plexes and run in the LC-MS/MS system. Relative quantification of protein levels was achieved by measuring TMT reporter ion intensities in the lower m/z region of the MS^2 spectra of each peptide.

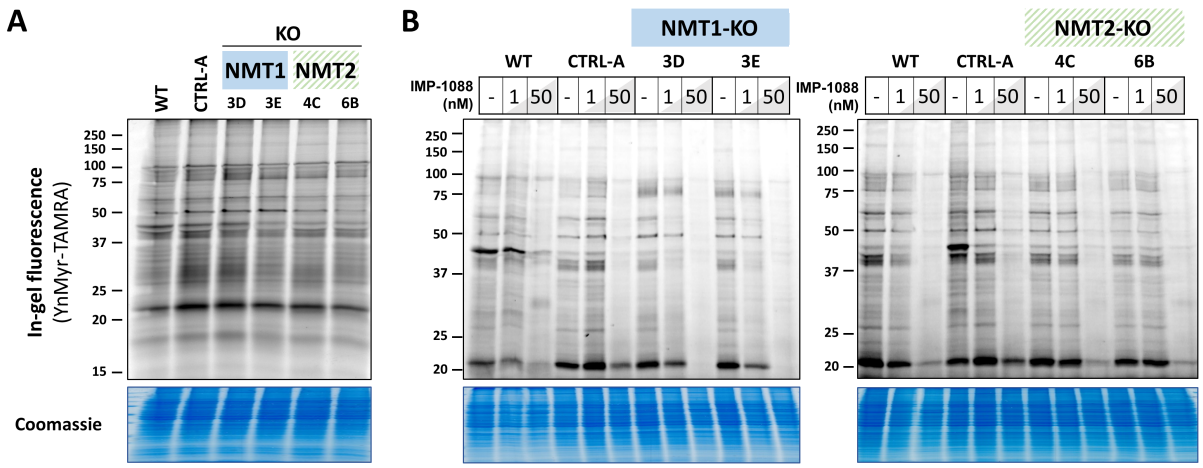


Figure 2.8: In-gel fluorescence analysis of YnMyr metabolic labelling in Panc-1 NMT1- and 2-KOs. **(A)** Clones were cultured in the presence of 30 µM YnMyr for 24 h or **(B)** incubated with with DMSO, 1 or 50 nM IMP-1088 for 72 h before addition of 30 µM YnMyr for additional 24 h, before lysis and chemical ligation to AzTB by CuAAC. Proteins precipitated and loaded into 12% SDS-PAGE gels. Gel images show TAMRA fluorescence measured at 532/575 nm excitation/emission wavelengths.

spike-in-to-lysate ratio at the beginning of the sample preparation procedure. By quantifying against an internal standard for every sample, this then allows for accurate quantifications of the light-to-heavy ratios. After ligation to AzRB (azide-Arginine-biotin) capture reagent (structure in Figure 2.6B) by click chemistry, excess click reagents were discarded by protein precipitation and biotinylated proteins specifically enriched in NeutrAvidin agarose beads. After on-bead digestion and peptide sample clean up, the generated peptide mixture was then analysed by LC-MS/MS (Figure 2.7A).

This experiment identified 74 known *N*-myristoylated proteins across all datasets, annotated based on previous evidence of *N*-myristoylation [244] (Figure 2.9 and Supplementary Table S1). For the statistical analysis of this data, the two clones of each type (AG3D and AG3E as "NMT1-KOs" and AG4C and AG6B as "NMT2-KOs") were treated as biological replicates in order to draw conclusions on the specific roles of NMT1 and NMT2 and minimise artefacts arisen from the CRISPR procedure (cf. section 2.1.3 for further information). Unfortunately, this conservative strategy can also result in lower number of hits due to the increased variability within each group, now formed by biologically distinct clones. Nevertheless, it should still be able to reveal the major effects of knocking out each NMT isoform, mirroring their primary biological targets.

YnMyr pull down proteomics analysis revealed a strong significant decrease in enrichment levels of 21 known NMT substrates in Panc-1 NMT1-KOs (Figure 2.9A), while no significantly differentially enriched *N*-myristoylated proteins were detected for NMT2-KOs, even though a generalised downward trend could be observed (Figure 2.9D). Findings on Panc-1 cells were then compared to those of the clones generated by Dr. Monica Faronato on HeLa cells (Figures 2.9B and E). This showed a clear correlation (Pearson coefficient of 0.711) between

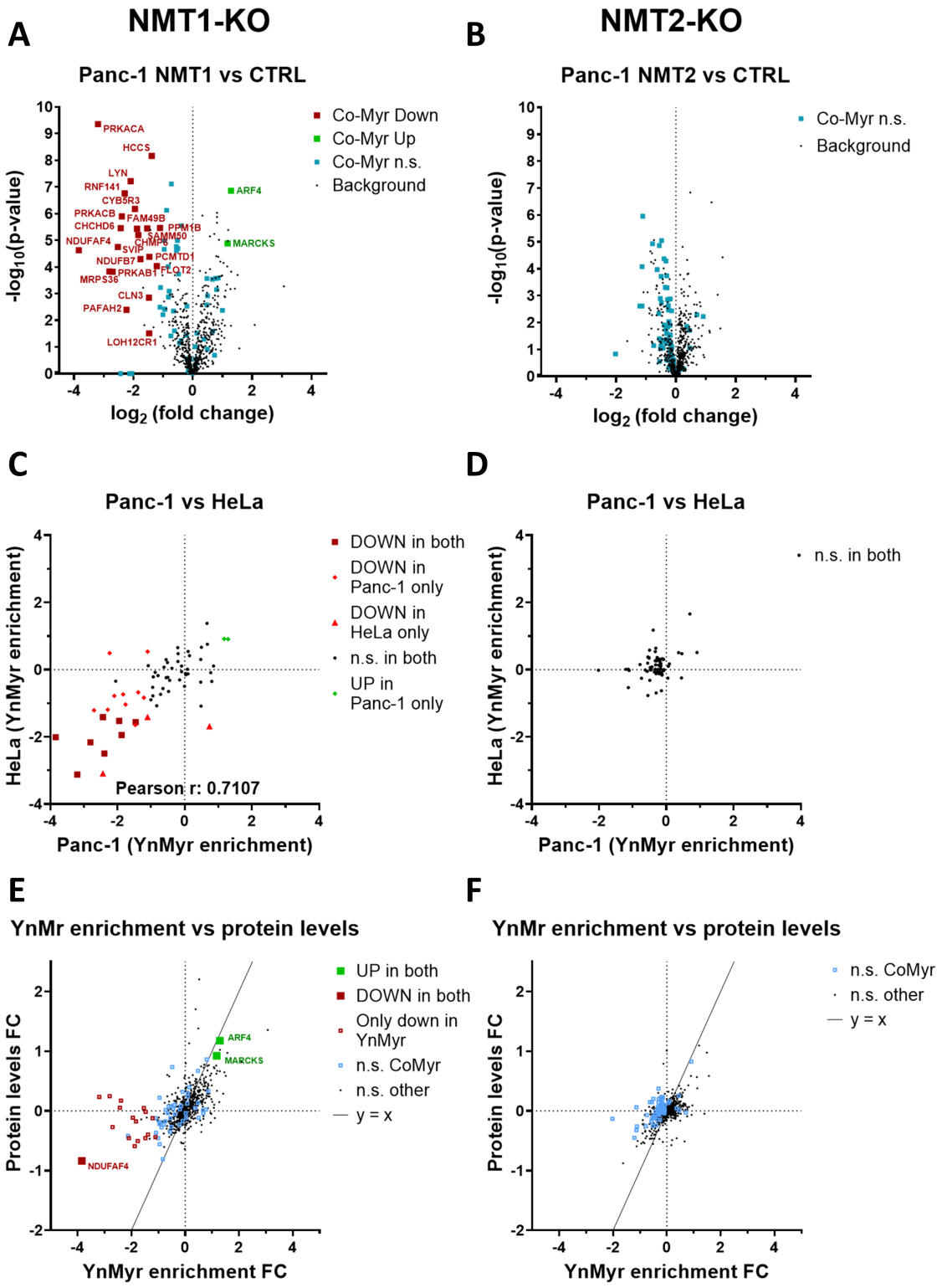


Figure 2.9: YnMyr enrichment proteomics reveals NMT1 is the main driver of co-translational *N*-myristoylation in the cancer cell. (A) Volcano plot showing $\log_2$ fold-change of YnMyr enrichment levels in NMT1-KOs *vs* CTRL against the significance level calculated with a student t-test (FDR=0.01, S_0=1). **(B)** Same as in (A) but for NMT2-KO *vs* CTRL (FDR=0.01, S_0=1). **(C)** Correlation of enrichment levels of known NMT substrates in Panc-1 *vs* HeLa NMT1-KO clones. **(D)** Same as in (C) but for NMT2-KOs vs CTRLs. **(E)** Correlation of YnMyr enrichment levels and total protein levels in Panc-1 NMT1-KOs. **(F)** Same as in (E) but for NMT2-KO clones.

HeLa and Panc-1 NMT1-KOs but no observable correlation in NMT2-KOs (Figures 2.9B and E). Data from the two cell lines indicated a dominant role for NMT1 in maintaining *N*-myristoylation of the cancer cell and showed no evidence for a specific role of NMT2 in cancer cells in culture. Interestingly, this analysis also revealed an increase in enrichment of two NMT substrates (ARF4 and MARCKS) in Panc-1 cells (Figure 2.9A), which was not seen in HeLa cells (Figure 2.9B). However, when compared YnMyr enrichment to total protein levels in the same cells (measured following the procedure in Figure 2.7B), the two co-myristoylated proteins ARF4 and MARCKS were also found to be significantly increased in total protein levels (Figure 2.9C). Since *N*-myristoylation happens co-translationally, the increase in enrichment levels was determined to be due to the upregulation of protein levels, which would then get *N*-myristoylated by NMT2. Comparison of YnMyr enrichment levels to whole proteome levels also revealed that all proteins showing a significant decrease in YnMyr enrichment, with the exception of one (NDUFAF4, an assembly factor for mitochondrial respiratory Complex I), were solely due to the effect on *N*-myristoylation levels upon NMT1 knock down (Figure 2.9C). Taken together, these data suggest NMT1 is the main driver of co-translational *N*-myristoylation of the cancer cell, with NMT2 not being able to compensate for *N*-myristoylation in a number of known NMT substrates when NMT1 is not present. Further, it shows no evidence for any specific role of NMT2 in these cancer cells, since NMT1 alone seems sufficient to fully maintain normal *N*-myristoylation levels. This comes together with the major impairment of cell growth observed for NMT1-KO clones in Figure 2.4 and the absence of any clear diversions of NMT2-KOs away from the WT/CTRL phenotype.

2.3.3 Whole proteome analysis of NMT1- and 2-KO clones

In addition to the study the effect of genetic ablation of either NMT in protein *N*-myristoylation, the consequences of NMT1- and 2-KOs on global cellular protein levels were also studied. Sample preparation was carried out as described in Figure 2.7B, using TMT labelling as the method for quantification.

This experiment identified and quantified 4,958 proteins of the Panc-1 proteome (Figure 2.10 and Supplementary Table S2). After filtering and normalisation (refer to Materials and Methods section *Whole proteome analysis of Panc-1 clones* for further information), two independent two-sample tests (student t-test, FDR $= 0.05$; $S_0 = 1$) were applied to compare the total protein levels of NMT1-KOs and NMT2-KOs to those in CTRL cells. This resulted in 62 proteins showing significant upregulation and 17 showing significant downregulation in NMT1-KOs (Figure 2.10A), whereas only 8 showed significant upregulation and 13 significant downregulation in NMT2-KOs (Figure 2.10D). NMT1 and NMT2 were strongly downregulated in their respective KO clones (Figures 2.10A and D). In TMT (MS^2)-based quantification a baseline noise-level value is expected in every channel where the peptide being quantified is not present. In addition that noise-level value depends on the dynamic range of the measurement, which is much lower than for MS^1 quantification methods. Therefore, the downregulation of

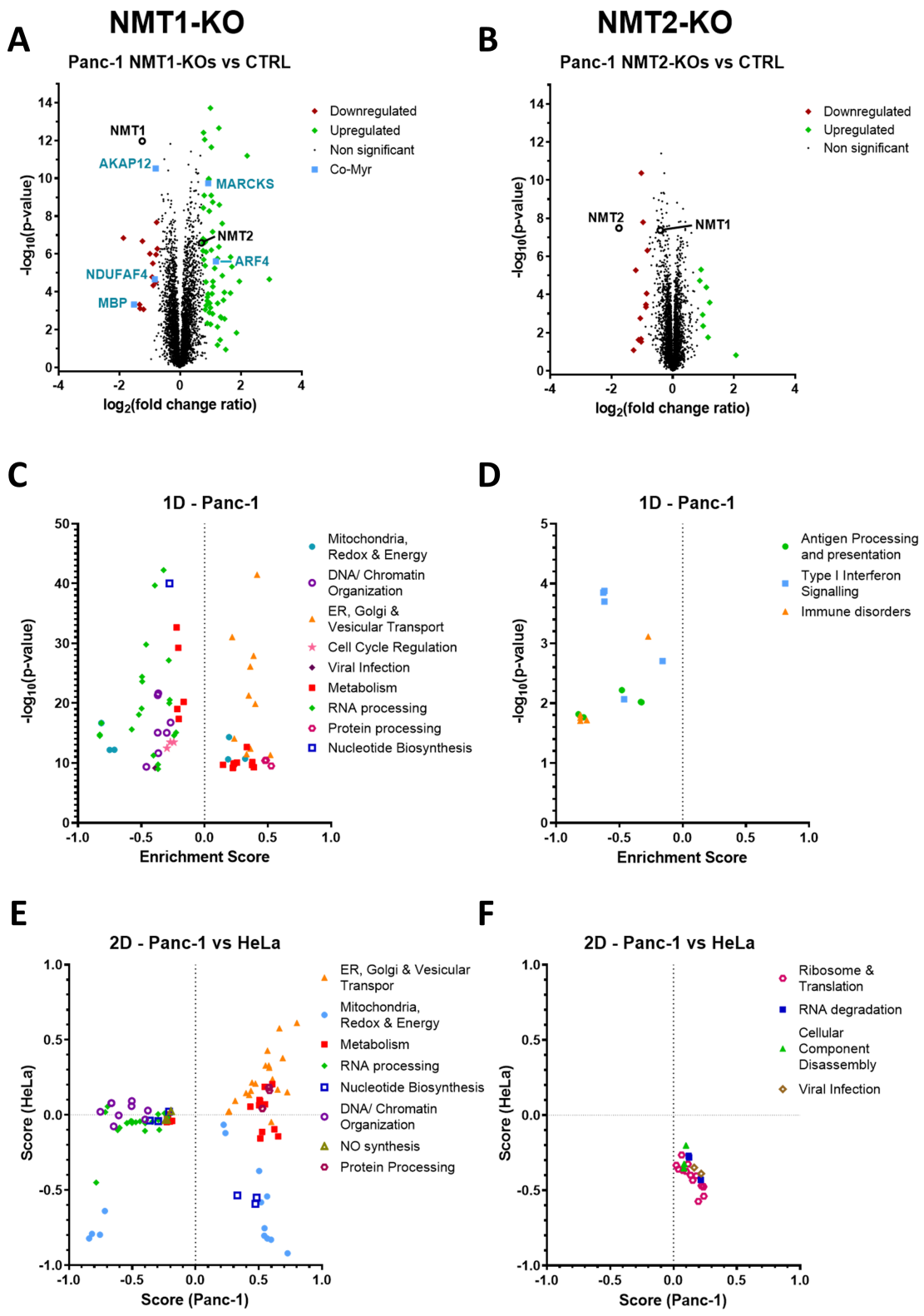


Figure 2.10: Effects of knocking out NMT1 and 2 on the protein levels across the whole proteome. **(A)** Volcano plot of the student t-test on NMT1-KOs *vs* CTRL (FDR=0.05, S_0=1). **(B)** Same as in (A) but for NMT2-KOs *vs* CTRL (FDR=0.05, S_0=1). **(C)** 1D annotation enrichment results (FDR¡0.02) of Panc-1 NMT1-KO clones *vs* CTRL (AG3D and AG3E averaged). Only terms showing a p-value$<10^{-9}$ are shown for ease of visualisation. **(D)** Same as in (C) but for NMT2-KO clones. All significant terms shown. **(E)** 2D enrichment plot of Panc-1 *vs* HeLa NMT1-KOs. **(F)** Same as in (E) but for NMT2-KOs. Annotation enrichment analysis was performed using Perseus (version 1.6.0.2) on annotations derived from KEGG, GOBP, GOMF and GOCC databases.

NMT1 and NMT2 observed in their respective KO clones, close to the maximum fold change ratio, was indicative of full KOs in these clones, in line with previous Western blot results (Figure 2.2). These two-sample tests also revealed significant changes in NMT1-KOs corresponding to some known *N*-myristoylated proteins. Since total protein abundances can also affect YnMyr enrichment, these were contrasted to YnMyr enrichment levels to ensure dissection of both phenomena (Figure 2.9C).

Performing statistical tests in whole proteome experiments only provide a hint of the profoundness of the effects caused by the experimental perturbation and are not able to deliver mechanistic insight in isolation. Annotation enrichment analysis [349] was chosen to extract biologically meaningful information from the statistical results. In this analysis proteins are clustered into different categories/pathways (e.g. GO terms) and a category-level statistic is computed that accounts for how many proteins in each pathway show a coordinated change based on their pre-calculated fold changes. The dataset consisting of $\log_2$ fold changes against CTRL cells was then annotated with KEGG pathway and gene ontology terms corresponding to the biological process (GOBP), the molecular function (GOMF), and the cellular compartment (GOCC), before performing 1D and 2D annotation enrichment. The difference between 1D and 2D annotation enrichments lies in the number of samples the test is performed in: 1D enrichment analyses the extent of representation of each annotation term by proteins showing similar fold change trends in a single sample (*vs* control, since the data is given as fold change values), whereas 2D annotation enrichment performs a 1D enrichment for 2 different samples in the background and then compares the scores in a correlation plot. Consistent with previous results, annotation enrichment of NMT1-KOs resulted in numerous GO and KEGG terms showing significant alterations while very few terms showed significant under- or over-representation in NMT2-KOs (Figure 2.10B and E) (note the difference in significance scales). GO and KEGG terms that were significantly enriched and that were involved in similar biological processes or molecular functions were manually clustered into categories. The manual curation of these categories was performed independently of the enrichment scores, ensuring each category could span the whole enrichment score dimension [-1, 1]. NMT1-KOs showed consistent changes in proteins affecting normal function of ER and Golgi transport, RNA processing and mitochondrial activity. This correlated exquisitely for both Panc-1 NMT1-KOs (Appendix Figure A.4) and showed important similarities between HeLa and Panc-1 NMT1-KO clones, particularly on ER, Golgi and vesicular transport, general biosynthetic metabolism and mitochondrial processes, including redox homeostasis and energy production (Figure 2.10C). NMT2-KOs, in contrast did not show any clear pathway/GO term that was enriched consistently across clones in the same cell line (Appendix Figure A.4) or across clones in different cell lines (Figure 2.10E and F). In general, these data indicate that deletion of NMT2 in these cells has no major effects on general protein levels, while disruption of NMT1 causes deep changes in important cellular processes such as vesicular transport, mitochondrial activity and RNA processing. Deregulation of vesicular trafficking could be mediated by CLN3, which is involved in Golgi network-to-endosome transport and endosomal sorting, or CHMP6, believed

to mediate endosomal sorting [70], two *N*-myristoylated proteins detected as NMT1-selective proteins in the YnMyr enrichment experiment above (Figure 2.9A). Moreover, several other NMT substrates are known be involved in vesicular transport, including ARL1, ARF1, ARF5 and ARF6 ADP-ribosylation factors [70], although these only showed a modest reduction in enrichment in the YnMyr enrichment experiment (Figure 2.9A). In addition, there were several mitochondrial proteins detected among the NMT1-selective subset (Figure 2.9A) that could explain the proteome-wide effects on mitochondrial dysfunction, including ribosomal protein MRPS36, Complex I subunit NDUFB7 and assembly factor NDUFAF4, NADH-cytochrome reductase CYB5R3, cytochrome heme lyase HCCS and two proteins involved in maintaining mitochondrial cristae integrity: CHMP6 and SAMM50 [70]. In contrast, no proteins were found among the NMT1-selective subset that play a direct role in RNA processing, although recent data in the Tate group suggest it is a characteristic of NMT inhibition that is able to predict for cancer cell sensitivity across tumour types (manuscript in preparation).

2.4 Conclusions

In this chapter, full NMT1- or NMT2-KOs were generated in Panc-1 cells in order to distinguish their biological roles in cancer. Genetic ablation of NMT1 strongly affected the cells, with less clones surviving the CRISPR/Cas9 and sorting procedure than in CTRL and NMT2-KO cells. Further, the clones that survived showed a marked reduction in cell growth and increased cell sizes, a characteristic of cells in senescence. Generation of NMT2-KOs, on the other hand, proceeded with no major consequence and cells behaved as unmodified controls. This coincides with previous reports on NMT1-KO mice, in which full ablation was embryonically lethal [254]. Moreover, recent genome-wide CRISPR screens performed by the Broad Institute across a wide range of cancer cell lines (https://depmap.org/) classified NMT1 (but not NMT2) as a commonly essential gene. Interestingly, RNAi interference (RNAi) screens did not detect this dependency, suggesting small amounts of NMT1 might be enough for cells to survive, whereas a complete lack of this enzyme is detrimental for cell viability (Depmap).

Consistent with this, YnMyr enrichment proteomics analysis revealed that NMT1 ablation significantly altered the scope of *N*-myristoylation independently of total protein levels, whereas *N*-myristoylation patterns were generally undisturbed in NMT2-KO clones. Extension of this analysis to KOs generated in another cell line by a colleague confirmed NMT1 as the main responsible enzyme for co-translational *N*-myristoylation in cancer cells. There was still some *N*-myristoylation happening in NMT1-KO cells, indicating NMT2 is able to partially compensate in the absence of NMT1. However, this was not sufficient to sustain normal cell growth and cells resorted to slower cell cycles to withstand the genetic deficiency. Importantly, functional analysis of whole proteome data showed impairment of *N*-myristoylation by NMT1 deletion led to dysregulation of proteins involved in ER/Golgi transport, RNA processing and mitochondrial activity, suggesting a broader implication for *N*-myristoylation in normal cell

proteostasis. These implications need to be taken with care, however, since they could partially have arisen as an adaptation mechanism to the genetic ablation during the clonal growth phase. Moreover, some of the effects might be masked by the senescence phenotype observed in these cells.

Unfortunately, these studies did not reveal any specific substrates of NMT2 or specific effects of its deletion in these cells. It was therefore hypothesised that NMT2 might be redundant in cancer cells, or at least in cancer cells in culture, where cells behave significantly different than in tissues. Alternatively, NMT2 could have a specific role in apoptosis, where both NMTs have been shown to participate actively through the *N*-myristoylation of newly formed neo-N-termini by the action of, for instance, caspases [243, 244]. Interestingly, an earlier report suggested NMT1 and NMT2 gain opposed cellular distributions upon activation of apoptosis [270], which could indicate a more specialised role for NMT2.

Together, these data confirm NMT1 is essential in cancer cells, where it is responsible for *N*-myristoylating proteins that NMT2 cannot process and that are crucial for normal cancer cell growth and viability.

3

Whole proteome profiling of *N*-Myristoyltransferase activity and inhibition using Sortase A

This chapter has been adapted from the published article [350].

3.1 Using Sortase A for the assessment of NMT activity

To understand the process of *N*-myristoylation better and target it efficiently using small molecule inhibitors, methods that allow for a convenient and robust assessment of NMT activity are needed. This is particularly important in drug discovery settings, where validation of on-target activity of the lead compounds in complex biological system is crucial. Due to their generally low abundances in inherently complex samples, identification of modification sites and their stoichiometries at endogenous levels continues to be a major challenge for most PTMs, even with the most sensitive mass spectrometric approaches. Lipid PTMs pose an extra challenge due to their lipophilicity, which makes them laborious to handle as well as difficult to ionise. Further, they display poor antigenicity, impeding the development of good antibodies to detect them by classical molecular biology methods.

As described in Chapter 2, the most successful way of overcoming these difficulties has been the use of chemically tagged myristate analogues such as YnMyr [244, 312, 335, 336] (Structure in Figure 3.1D) that are cell-permeable and readily transferred by endogenous NMT onto protein substrates. As shown in Chapter 2, this allows both for visual assessment of NMT activity upon chemical ligation to fluorophores and enrichment of tagged proteins upon ligation to affinity reagents in vitro. However, while this method has proven to be extremely useful to identify new substrates and assess NMT activity in cells (Chapter 2 and [244, 351]), it has the disadvantage of relying on metabolic tagging. It works remarkably well in cell culture, but is difficult to scale to animal models, let alone human studies. With the lab advancing on the development of NMT inhibitors with potential clinical value, new assays for assessing NMT activity in biological samples beyond cell culture were needed.

Identification of the *N*-myristoylated form of NMT substrates at endogenous levels is challenging. Therefore, a new strategy was sought that could detect the appearance of non-myristoylated NMT substrates upon NMT inhibition (this strategy is represented in Figure 3.1A). For this purpose, it was speculated that the use of *Staphylococcus aureus* Sortase A (SrtA) could be particularly useful, given its selectivity towards N-terminal glycines [352], which overlaps with NMT substrate specificity. In nature, SrtA is a transpeptidase that catalyses the attachment of surface proteins bearing a conserved LPXTG motif to the peptidoglycan cell wall in Gram-positive bacteria [353]. SrtA cleaves between the threonine and glycine residues of the recognition motif and covalently attaches the -LPXT adduct to proteins containing an N-terminal glycine through a regular peptide bond [352]. Due to its substrate promiscuity, SrtA-mediated site-specific ligation has been very widely used in the past years for a plethora of applications that involved ligating entities of very diverse biochemical nature to peptides or proteins containing an unhindered N-terminal glycine. The attached moieties included peptides, proteins, glycopeptides and even nucleic acids or lipidic groups

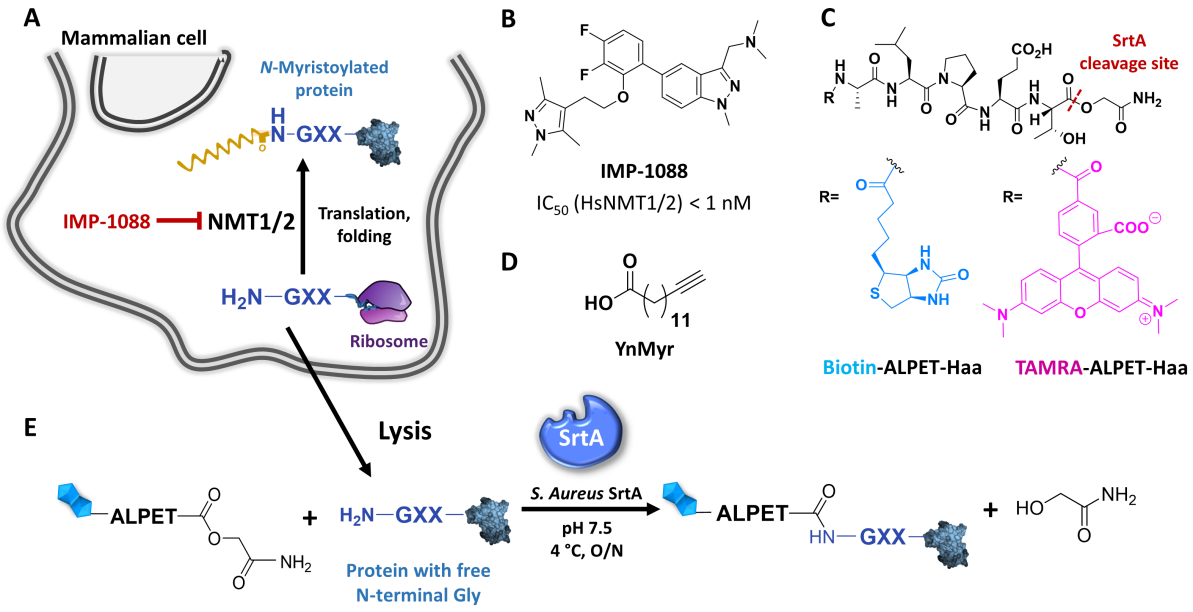


Figure 3.1: Chemical tools for SrtA profiling of inhibition of NMT activity in cells. (A) Depiction of the *N*-myristoylation reaction by NMT in the cell. **(B)** Structure of IMP-1088 and potencies against human NMT1 and NMT2. **(C)** Structures of SrtA ALPET-Haa depsipeptide substrates used in this study (biotin tag in blue and TAMRA fluorophore in pink). **(D)** Structure of *N*-myristoylation probe YnMyr (tetradec-13-ynoic-acid).**(E)** Schematic representation of the SrtA labeling reaction.

[354]. However, this is the first report of using SrtA to label a complex mixture of proteins in a non-engineered system. It is therefore anticipated that this study will open the door to new exciting applications of SrtA for the labelling of proteins containing an N-terminal glycine expressed at endogenous levels.

In this work, SrtA pentamutant (5M) (P94R/D160N/D165A/ K190E/K196T) was combined with a chemically tagged depsipeptide form of the LPXTG peptide (Biotin/TAMRA-ALPET-Haa, where Haa = 2-hydroxyacetamide, Figure 3.1B) to label free N-terminal glycines in whole-cell lysates (Figure 3.1C). Several versions of SrtA have been engineered in the last years through directed evolution and rational design, improving reaction rates and characteristics of SrtA [355]. SrtA 5M has been reported to have 120-fold increase in activity over the natural enzyme [356, 357], while maintaining Ca^{2+} dependency. A newer version of SrtA, called SrtA 7M (additional E105K and E108A mutations), no longer requires Ca^{2+}, making it suitable for *in vivo* applications and other experimental conditions where Ca^{2+} might not be available (e.g. in the presence of Ca^{2+} chelators), but at the expense of lower activity similar to the WT enzyme [358, 359]. Increased activity was hypothesised to be linked to higher promiscuity, beneficial for labelling complex samples. Therefore, since this assay had been designed to be performed *in vitro* and the Ca^{2+} dependency was not an issue, the 5M version was selected as the best candidate. Regarding the LPXTG peptide substrate, the use of a depsipeptide instead of a natural peptide has been shown to increase the efficiency of the reaction by preventing the reverse reaction [360, 361] (Figure 3.1C). The combination of the

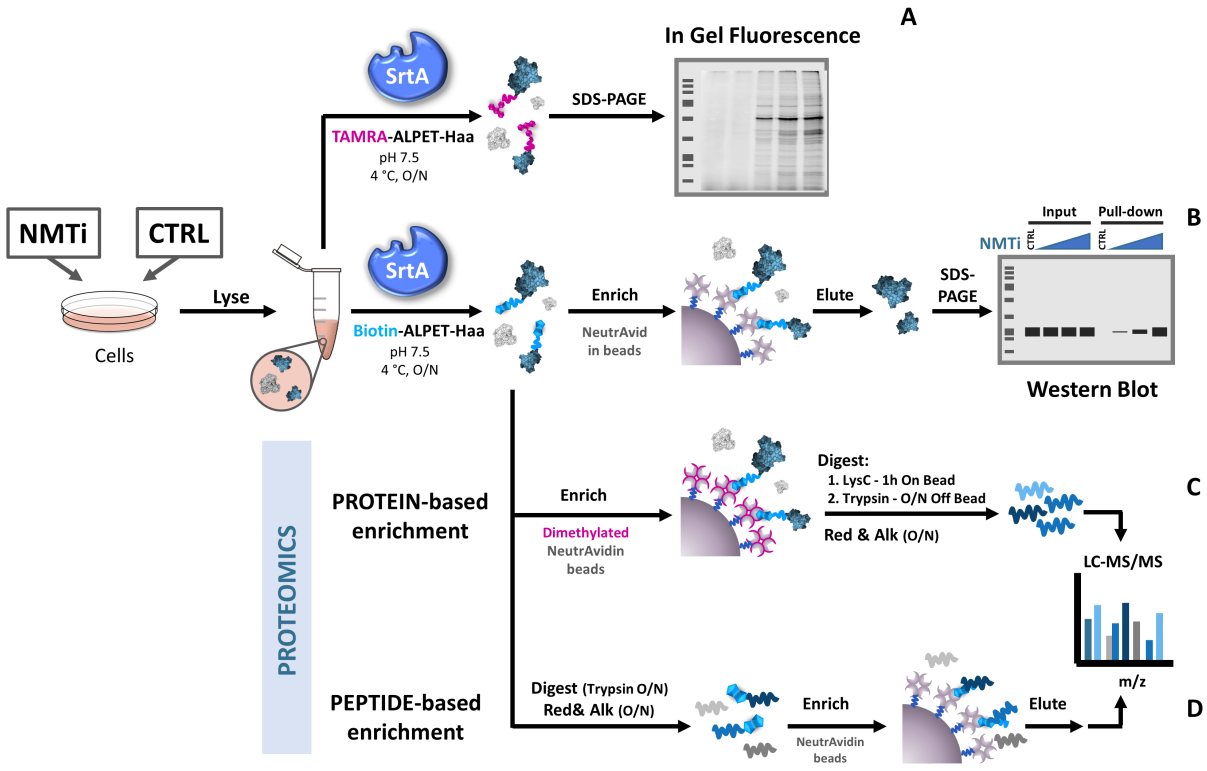


Figure 3.2: SrtA labeling workflows enable multiple detection methods for NMT activity. Protein samples are resuspended in SrtA reaction buffer, here exemplified by cells from tissue culture; however, in principle this approach is equally applicable to tissues exposed to a change in NMT activity. **(A)** Overnight reaction with SrtA and TAMRA modified depsipeptide substrate (TAMRA-ALPET-Haa) enables in-gel fluorescence analysis of the SrtA-labeled protein lysates. Alternatively, samples subjected to SrtA reaction with biotin-modified depsipeptide substrate (Biotin-ALPET-Haa) are enriched and studied by in-gel analysis **(B)** or proteomics approaches (**C** and **D**), following dimethylation of NeutrAvidin-coated beads in the case of protein-based enrichment.

SrtA 5M with the depsipeptide substrate allowed for efficient and selective labelling of free N-terminal glycines in whole cell lysates (Figure 3.1E). By enriching for SrtA-labelled proteins and integrating this workflow with previous knowledge on NMT substrates, NMT activity status can be assessed (Figure 3.2). Most importantly, the SrtA labelling step is performed post-lysis, ensuring a much wider applicability than the metabolic labelling approach.

3.2 Labelling N-terminal glycines in whole-cell lysates with SrtA 5M

SrtA has been extensively used to ligate engineered peptide or protein substrates containing an LPXTG recognition sequence and an N-terminal glycine at a physiological pH [354, 362]. Here, we aimed at extending this ligating capability to label free N-terminal glycine-containing proteins in whole-cell lysates with a tagged (fluorophore/biotin) LPXTG depsipeptide. High activity SrtA pentamutant was a kind gift from Dr. M. Jamshidiha, and had been expressed and purified from Addgene plasmid 86962 [356, 357]. ALPET-Haa depsipeptides (Biotin or TAMRA) (Figure 3.1B) were synthesised by a postdoc in the lab, Dr J. Morales-Sanfrutos as

described in our publication [350].

3.2.1 Optimisation of SrtA labelling conditions

To achieve an efficient labelling with SrtA in whole-cell lysates, the amount of high-activity *S. aureus* SrtA pentamutant and ALPET-Haa depsipeptide needed were optimised using in-gel fluorescence. For this, the TAMRA version of the ALPET-Haa depsipeptide substrate was used together with the in-gel fluorescence workflow described in Figure 3.2A. MDA-MB-231 breast cancer cells were treated with the potent and specific NMT inhibitor IMP-1088 [298] (Figure 3.1B) before mechanical lysis in SrtA reaction buffer. Protein samples in SrtA buffer were then incubated with a concentration range of either TAMRA-ALPET-Haa depsipeptide (Figure 3.3A) or SrtA pentamutant (Figure 3.3B). Based on this results, 100 nM of SrtA and 75 µM of the substrate depsipeptide were selected for subsequent labelling experiments.

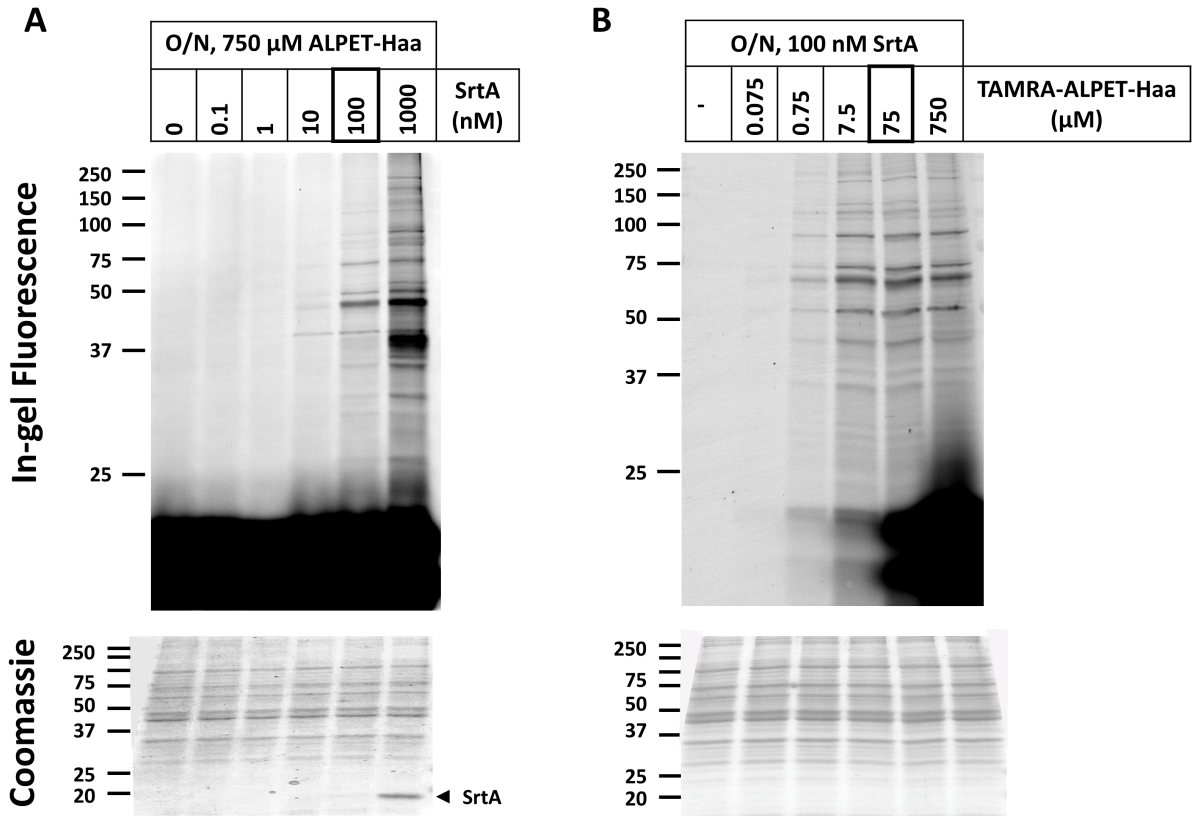


Figure 3.3: Optimization of overnight SrtA labelling reaction of whole-cell lysates. MDA-MB-231 cells were incubated with 100 nM IMP-1088 for 24 h before lysis. Lysates were labeled with SrtA and TAMRA-ALPET[G] overnight at 4 °C and run in 12% SDS-PAGE gels. In-gel fluorescence images were acquired with excitation/emission wavelengths of 532/568 nM. **(A)** Optimization of SrtA concentration with constant depsipeptide substrate. **(B)** Optimization of T-ALPET[G] concentration with constant SrA concentration. The concentrations selected for further experiments are highlighted in thicker squares.

3.2.2 Peptide-based enrichment proteomics to study the substrate specificity of SrtA in whole cell lysates

To characterise the SrtA-labelling reaction in whole-cell lysates and to analyse if there were any biases towards labelling of particular protein substrates, SrtA-labelled peptides were enriched followed by LC-MS/MS analysis (procedure depicted in Figure 3.2D). Cells were treated in triplicate with IMP-1088 for 24 h to ensure nearly complete inhibition of NMT and increase the chances of identifying N-terminally modified peptides of NMT substrates. Samples were labelled with SrtA, precipitated to remove excess depsipeptide and digested with trypsin. SrtA-biotinylated peptides were enriched using NeutrAvidin agarose beads, eluted as reported previously [363], desalted and analysed by LC-MS/MS. The peptide sequences identified were matched to a modified version of the SwissProt human proteome database (UniProt; December 2017; 42,326 entries), in which the ALPET peptide sequence was added to the N-termini of all proteins in the database. Analysis of 120 identified modified peptides (Figures 3.4 and 3.5A, and Supplementary Table S3) revealed good selectivity of *S. aureus* SrtA for glycine in the second amino acid position (after removal of iMet) (63.8%) as compared to the database frequency (7.8%) (Figure 3.4, 2nd Position). Interestingly, direct evidence for low levels of labelling of N-termini other than Gly (e.g. Ala, Met, Pro, Val) was also observed, suggesting that this SrtA pentamutant has somewhat relaxed specificity at the N-terminus under conditions of excess SrtA and depsipeptide substrate. Analysis of amino acid positions 3-11 did not reveal any clear preferences of SrtA for particular amino acids in any of those positions (Figures 3.4 and 3.5A). Surprisingly, no particular preference was observed for a second glycine at the N-terminus (P3 position), which had been reported for *in vitro* studies in multiple occasions as a way to mimic the pentaglycine motifs of their natural substrate, the cell wall peptidoglycan [352, 364]. This had lead to the common practice of using diglycine or triglycine motifs at the N-terminus of recombinant proteins wanted to be efficiently targeted by SrtA [352]. Together, this experiment validated the use of SrtA to label N-terminal glycines in whole-cell lysates in an unbiased manner.

Moreover, the use of IMP-1088 allowed us to identify modified N-terminal peptides corresponding to 30 known NMT substrates (Figure 3.5B), which showed a very similar profile to the combined list of human myristoylated proteins derived from UniProt [70] and previous proteomics experiments using YnMyr [351](Figure 3.5C), as well as the GXXXSK motif described previously [40].

3.3 Neutravidin dimethylation and two-step LysC/trypsin digestion eliminates avidin background in on-bead digests

As shown in Figures 3.4 and 3.5A, SrtA labels a wide range of proteins in the lysate containing an accessible N-terminal glycine, and to a much lesser extent proteins with an N-terminal

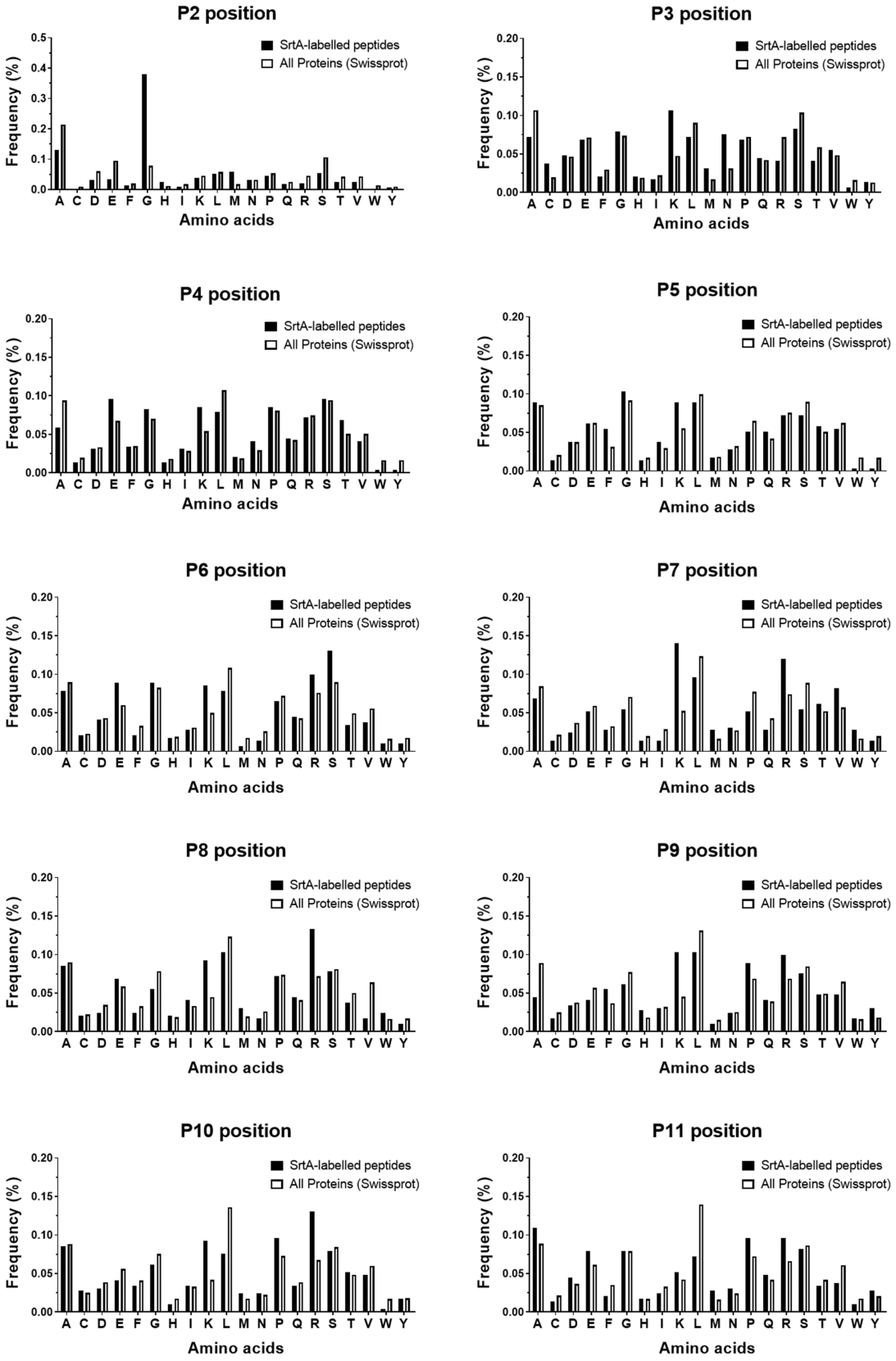


Figure 3.4: Frequencies per amino acid for the first 10 amino acid positions after iMet removal. Comparison between experimentally identified SrtA-labelled peptides and all peptides in the Swissprot database.

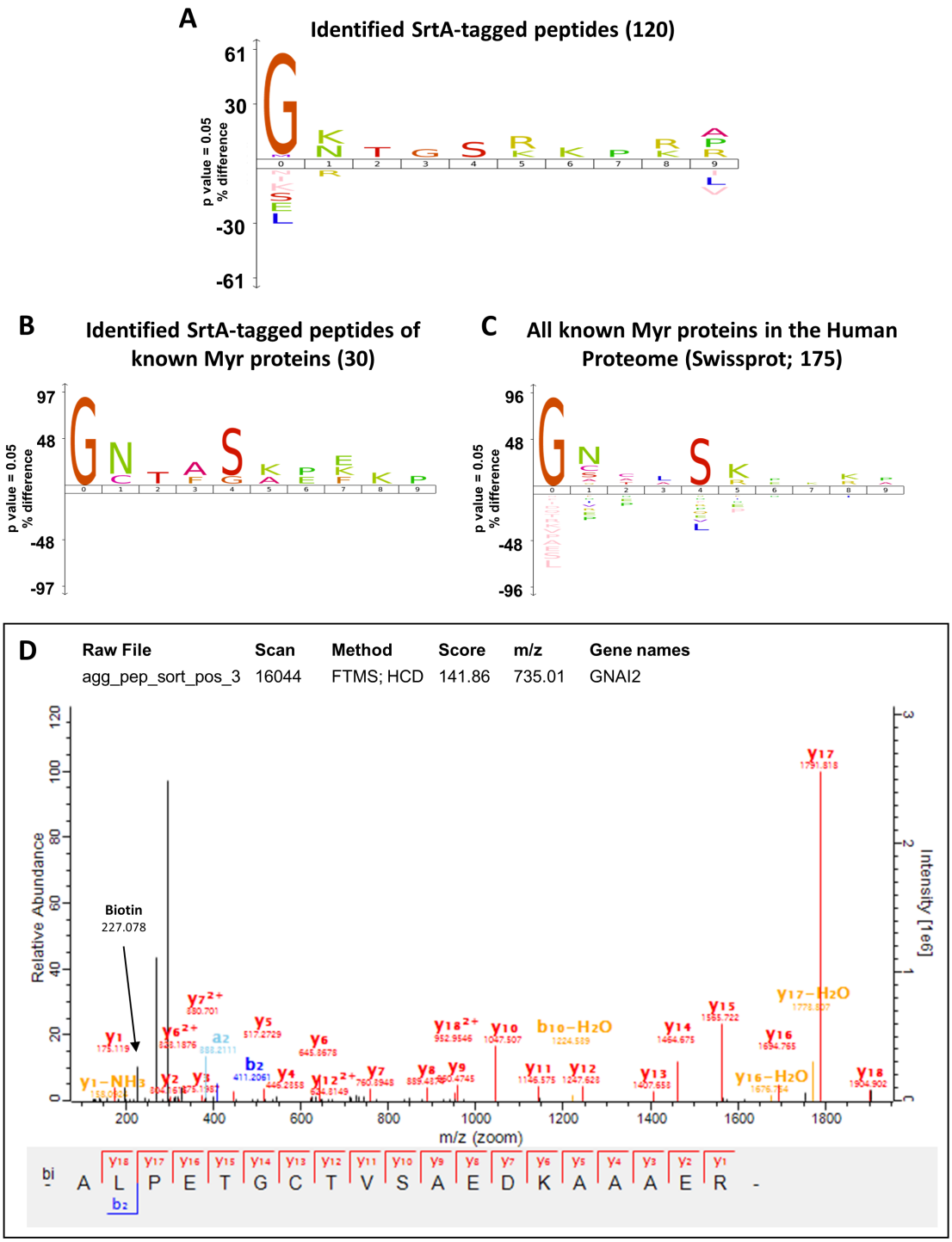


Figure 3.5: SrtA selectively labels N-terminal glycine-containing proteins in complex cell lysates. (A) Sequence analysis using iceLogo [365] of the N-terminal first 10 amino acids of the 120 SrtA-modified peptides identified by nanoLC-MS/MS after peptide-level enrichment of SrtA-labelled peptides. The frequencies of residues at each position were compared to the frequencies at that position in the SwissProt human proteome (pre-compiled in iceLogo, accessed on 17/09/2018). Side-by side sequence analyses of known NMT substrates identified containing the SrtA-labeling modification **(B)**, and of all known human myristoylated proteins **(C)**. Both logos were generated against the Swiss-Prot human proteome reference data set (pre-compiled in iceLogo, accessed on 17/09/18). **(D)** Example MS/MS spectrum of an N-terminal peptide modified with Biotin-ALPET.

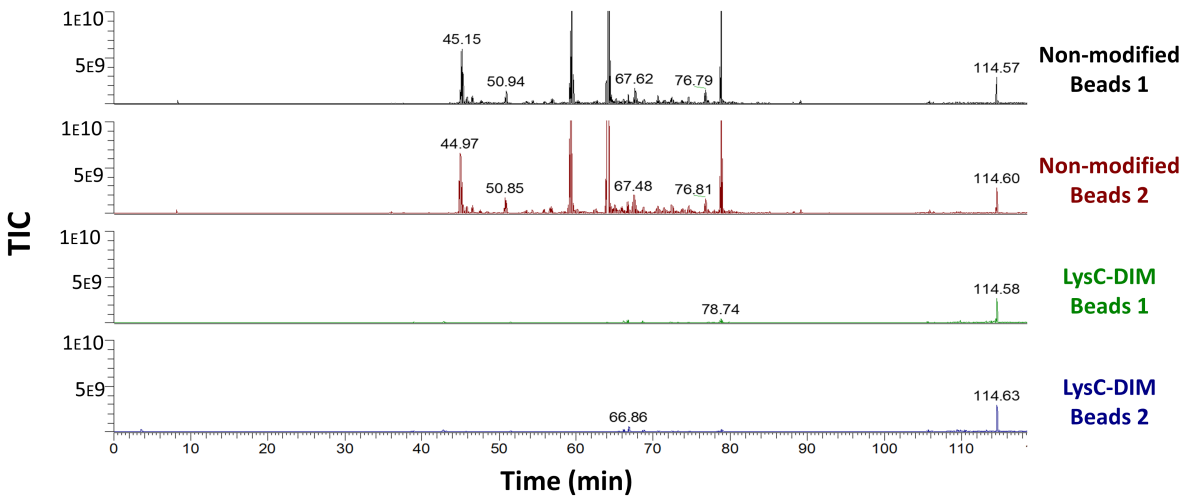


Figure 3.6: Capping NeutrAvidin lysines by dimethylation followed by a two-step on-bead LysC/off-bead trypsin digestion significantly reduces NeutrAvidin-derived high intensity peaks. Samples containing 200 µg of Biotin-ALPET labeled proteins, originated fromMDA-MB-231 cells, were incubated in duplicate with 200 µL of dimethylated NeutrAvidin agarose beads or with 200 µL of non-derivatized NeutrAvidin agarose beads for 2 h at RT. Samples incubated with dimethylated beads were subjected to 1 h LysC digestion followed by digestion off-bead with trypsin overnight. Samples incubated with non-derivatized beads were subjected to an overnight on-bead digestion with trypsin. The plots show total ion current (TIC) chromatograms recorded for each of the four samples.

alanine or other amino acids. Of note, only ca. 25% of these are NMT substrates. This made it challenging to accurately quantify NMT activity at the protein-level of NMT substrates against the high levels of endogenous N-terminal glycine labelling. In a preliminary experiment following the previously reported protein pull down procedure that involved on-bead tryptic digestion with trypsin for SrtA-labelled samples, only 7 NMT substrates could be identified (data not shown). This urged us to seek new strategies to reduce background signals. One prominent source of contamination from enrichment samples comes from the on-bead digestion step, where the protease (generally trypsin) cleaves not only avidin-bound proteins but also avidin itself. We therefore designed an improved protocol that involved derivatisation of commercial avidin resin to make it proteolytically resistant. The derivatisation step consisted in capping Lys residues of the bead-bound NeutrAvidin by dimethylation before the interaction with the protein sample. We then combined this with a two-step on-bead/off-bead digestion with LysC and trypsin. Derivatized NeutrAvidin retained its biotin-binding capacity but was now resistant to LysC cleavage. By eluting resin bound-polypeptides with LysC treatment and then completing the digestion off-bead with trypsin we were able to completely eliminate Neutravidin-derived peptide contamination (Figure 3.6). This resulted in cleaner MS1 spectra, which boosted identification rates significantly. This additional step in the enrichment procedure not only allowed us to obtain the results reported in the next few sections of this Chapter, but also provides a generally applicable enhancement for any other pull-downs on avidin resin.

3.4 Evaluation of SrtA labelling as a method for studying NMT activity by protein-based enrichment proteomics

Having shown that SrtA can be used to label proteins at endogenous levels in complex mixtures such as whole-cell lysates, the next aim was to assess if this could be applied to indirectly measure protein *N*-myristoylation levels and to discriminate between cells possessing different levels of NMT activity. For this purpose, cells were treated with vehicle control, low (1 nM), or high (100 nM) concentration of IMP-1088 NMT inhibitor in triplicate for 24 h, before collection of the proteome lysates and overnight labelling with SrtA and Biotin-ALPET-Haa. To provide direct comparison with the previously described method for the study of NMT activity, another set of cells was cultured under the same conditions in the presence of YnMyr, followed by post-lysis chemical ligation to a biotinylating reagent (AzRB) as reported previously [351]. After protein precipitation, affinity enrichment with NeutrAvidin agarose beads using the optimized procedure described in section 3.3 and the two step on-bead/off-bead digestion, samples were labelled with 9-plex TMT, combined, and analysed by LC-MS/MS (workflow scheme in Figure 3.2C).

In the SrtA-labeled set of samples, 40 known NMT substrates were identified, 24 of which were responsive to NMT inhibition (ANOVA with FDR=0.01 and S_0 =1). On the other hand, the YnMyr experiment, which was designed to specifically enrich for myristoylated proteins, identified a total of 78 NMT substrates, from which 56 showed response to NMT inhibition (ANOVA with FDR=0.01 and S_0 =1) (Figure 3.7A and Supplementary Table S4). The superior identification power of the YnMyr strategy was not unexpected, since the enrichment of YnMyr-tagged proteins has increased selectivity for NMT substrates and this ensures higher chances of identification by LC-MS/MS. However, comparison between the two datasets revealed significant overlap between the two strategies, with 32 NMT substrates identified by both methods. From these, 22 NMT substrates responded significantly to NMT inhibition in both methods, 4 did not show any response to NMT inhibition in any of the two methods, and 6 were significantly less enriched in YnMyr but did not show any changes in the SrtA-labelling pattern (Figure 3.7B). Importantly, no NMT substrates were found that significantly responded to NMT inhibition in the SrtA assay and lacked responsiveness in the YnMyr labelled samples, thus indicating that the SrtA strategy, when combined with a specific NMT inhibitor is as selective as the YnMyr method and is not prone to false positives. As for the cases where SrtA was not able to detect a response to inhibition, this was hypothesised to be due to the requirement for an accessible N-terminal glycine, which can be obstructed by the protein's final 3D fold or other N-terminal modifications, such as *N*-acetylation. While it is a ubiquitous modification that has been shown to affect 80% of proteins in human cells [27], *N*-acetylation has also been shown to have limited overlap with NMT [40] (refer to the Introduction section 1.5.3 for further discussion on this). However, in the absence of functional NMT, it is possible that *N*-acetyltransferases (NATs) or other N-terminal processing enzymes could act co- or

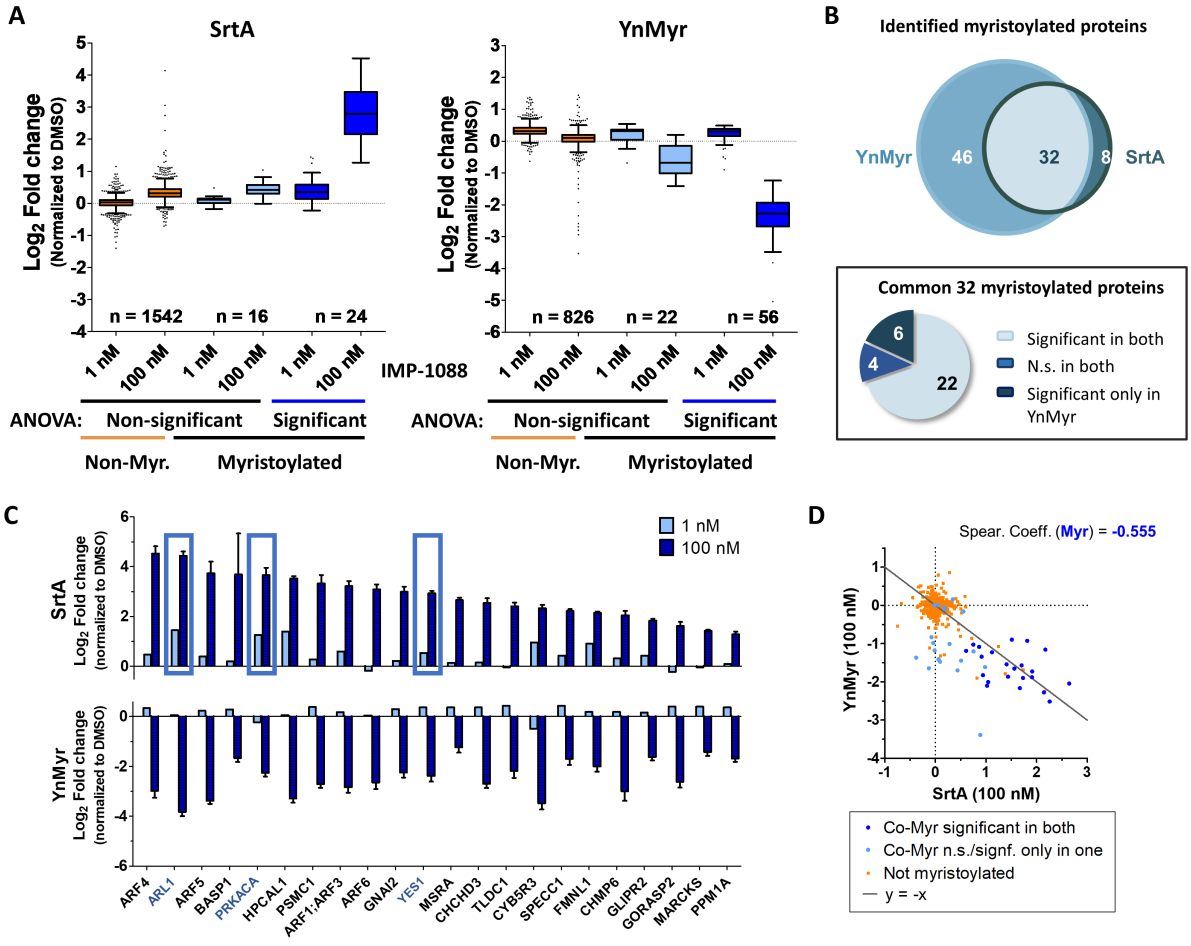


Figure 3.7: SrtA-labeling complements YnMyr metabolic tagging for quantitative profiling of NMT activity by protein-enrichment proteomics analysis. MDA-MB-231 cells were incubated in the presence of control (DMSO), low (1nM) or high (100 nM) concentrations of NMT inhibitor IMP-1088 for 24 h and either biotinylated with SrtA post-lysis or fed with YnMyr for the duration of the NMT inhibitor treatment and biotinylated via CuAAC. Biotin-labeled proteins were enriched on NeutrAvidin beads and analyzed by nanoLC-MS/MS. **(A)** Boxplot representation of fold changes for known myristoylated proteins vs all other identified proteins for both methods, derived from a one-way ANOVA test performed on all identified proteins (FDR = 0.01 and $S_0 = 1$). **(B)** Pie chart and Venn diagram comparing number of identified known myristoylated proteins and their response to NMT inhibition by each method. N.S., non-significant (ANOVA, FDR = 0.01 and $S_0 = 1$) **(C)** Bar plots showing side-by-side comparison of myristoylated proteins identified and significantly changing in both methods. Boxes highlight proteins selected for follow-up studies. **(D)** Correlation of SrtA and YnMyr labeling enrichments in 100 nM IMP-1088 treated samples.

post-translationally to block the N-terminal amino group. Another possible explanation could be NMT substrate destabilization in the absence of *N*-myristoylation, which had been described previously for specific NMT substrates [9, 249, 281] and more recently brought together with the discovery of a glycine-specific part of the N-end degron pathway (Gly/N-end degron pathway) [112] (cf. Introduction section 1.2.8.3 for further details on N-end degron pathways).

More detailed quantitative analysis of proteins that revealed a response to NMT inhibition by both methods showed slight discrepancies in the levels of enrichment (Figure 3.7C). These

probably arose due to the different limitations of each method, which detect opposite sides of the *N*-myristoylation reaction (substrates/products) catalysed by NMT. While YnMyr detects the disappearance of *N*-myristoylation of proteins upon NMT inhibition, SrtA labels new free N-terminal glycines. Thus, each labelling strategy is limited by other factors affecting either side of the *N*-myristoylation reaction, such as the ones described above for SrtA labelling or limited incorporation of the probe for YnMyr metabolic labelling. Most importantly, however, SrtA and YnMyr labelling patterns showed the expected negative correlation for known NMT substrates: SrtA-labelling of NMT substrates increased with NMT inhibition and new exposure of N-terminal glycines, while YnMyr labelling disappeared (Figure 3.7D). This correlation was absent in the rest of the identified proteins, including all containing an N-terminal glycine that are not *N*-myristoylated.

In addition, these experiments revealed six novel proteins not previously known to be myristoylated that responded to NMT inhibition in both methods: BAG5, PSMC2, PSMC5, SPAG1, SPANXB1 and SPANXC (Supplementary Table S4 and Figure 3.8A). One of these, SPANXB1 was identified as a novel NMT substrate as it contains an N-terminal glycine and its N-terminal peptide was found to be modified with YnMyr-AzRB (Figure 3.8B). Because some tightly bound protein complexes often co-elute together in these enrichment procedures, SPAG1 and SPANXC were hypothesised to have co-purified with SPANXB1. Similarly, PSMC2 and PSMC5 are known to be binders of the myristoylated protein PSMC1 [366], and co-elution could explain their presence even if they do not have an N-terminal glycine for them to be *N*-myristoylated.

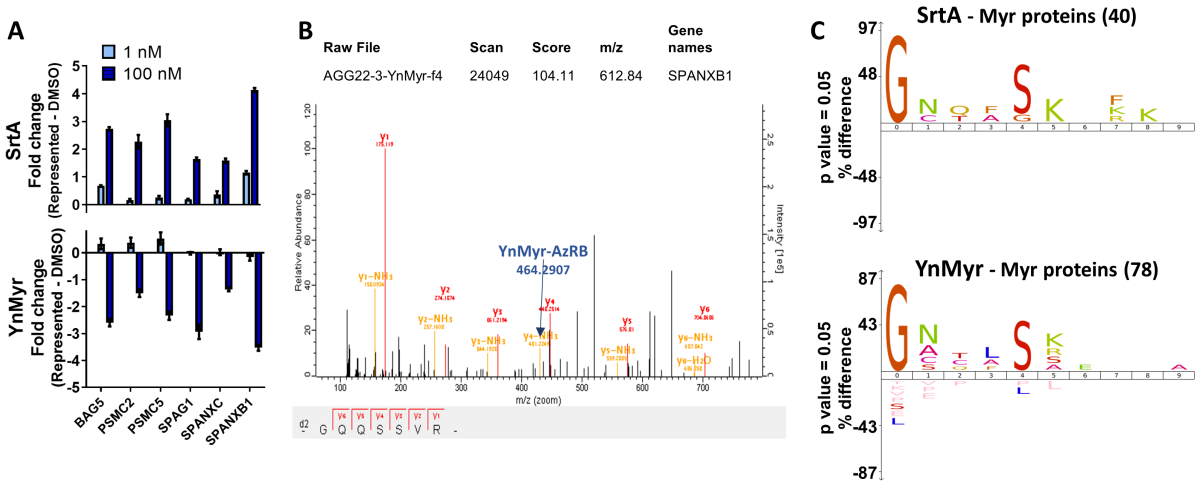


Figure 3.8: SrtA labelling identifies SPANXB1 as a new NMT substrate in MDA-MB-231 cells and fully recapitulates NMT sequence preferences. (A) Enrichment-derived intensities of significant proteins not previously known to be myristoylated. **(B)** Annotated MS/MS spectrum of SPANXB1 N-terminal peptide showing modification with YnMyr-AzRB. **(C)** Sequence logos of all known myristoylated proteins identified in SrtA and YnMyr protein-based enrichment experiments. Logos were generated with iceLogo [365], using the pre-compiled Swissprot human proteome (accessed on 17/09/2018) as the reference set.

Sequence analysis of all *N*-myristoylated proteins identified by SrtA recapitulated that of N-myristoylated proteins identified by YnMyr (Figure 3.8C), and together showed a very similar profile to that of the complete list of known myristoylated proteins in humans shown in Figure 3.5C). This further validates the use of the herein described SrtA labelling method to efficiently sample most abundant *N*-myristoylated proteins irrespective of their protein sequence.

3.5 In-gel analysis and SrtA-Elisa as high-throughput NMT target engagement biomarker assays

Based on the proteomics results, three NMT substrates that showed a dose-dependent signal were chosen as biomarkers for a gel-based assay (Figure 3.7C): ARL1, PRKACA and YES1. These proteins were selected on the basis of their dose-dependent response in the proteomics analysis (Figure 3.7C), as well as accounting for their theoretical molecular weight so that they could be analysed in one single gel together with the selected loading control (GAPDH). Cells were treated with vehicle control or increasing concentration of IMP-1088 and lysed mechanically in SrtA reaction buffer. Proteins in lysates were labeled with SrtA in the presence of Biotin-ALPET-Haa substrate, enriched on NeutrAvidin beads, subjected to SDS-PAGE gels and blotted for the selected biomarkers ARL1, PRKACA and YES1 following the procedure depicted in Figure 3.2B. Increased NMT inhibitor concentration resulted in a higher signal intensity in the pull-down fractions, suggesting an NMT inhibitor-specific dose-dependent increase in labelling by SrtA and confirming the results obtained from the protein-based enrichment proteomics analysis (Figure 3.9). Due to the low molecular weight

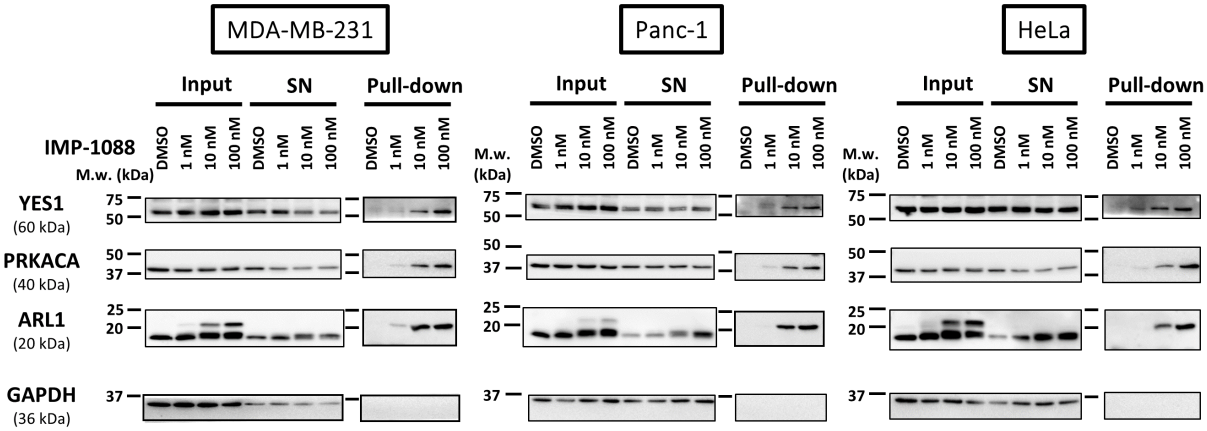


Figure 3.9: Validation of proteomics results by western blot. Pull-down and western blot analysis of SrtA-labeled proteins. MDA-MB-231, Panc-1 and HeLa cells were treated with DMSO or increasing concentration of IMP-1088 NMT inhibitor, lysed and labeled with SrtA overnight; biotinylated proteins were enriched on NeutrAvidin beads, resolved by SDS-PAGE and blotted against YES1 (60 kDa), PRKACA (40 kDa), ARL1 (20 kDa) and GAPDH as loading control (36 kDa). SN: supernatant. NMT substrate proteins show concentration-dependent increase in enrichment with increasing concentration of NMT inhibitor, whilst the low molecular weight of ARL1 also allows for direct identification of the SrtA-labeled fraction in the input sample due to the gel shift induced by ligation to Biotin-ALPET.

of ARL1 (ca. 20 KDa), labelling with Biotin-ALPET-Haa peptide (ca. 800 Da) by SrtA resulted in a visible molecular weight shift. This molecular weight shift of ARL1 was sufficient to indicate NMT inhibition prior to affinity enrichment, thus shortening the assay steps considerably and increasing the throughput.

To extend this phenomenon to the other two chosen biomarkers of greater molecular weight (PRKACA and YES1), a streptavidin-shift assay [367] was used. In this experiment, SrtA-labeled lysates were precipitated to remove excess depsipeptide and incubated with streptavidin before loading the samples onto SDS-PAGE gels. Streptavidin binds to biotinylated proteins, shifting their apparent molecular weight by ca. 30 KDa, corresponding to the molecular weight of the streptavidin dimer. ARL1 and PRKACA gels showed the expected mass shift clearly, whereas this could not be reproduced effectively with YES1 due to the high background of the antibody (Figure 3.10).

Finally, to increase the throughput of this assay one step further, we developed and optimised an ELISA-based assay using streptavidin-coated 96-well plates to retain all SrtA-biotinylated proteins (all N-terminal glycine-containing proteins) and using ARL1 as the detection antibody. The optimisation procedure was carried together with Dr. Markus Ritzefeld, a postdoc in the Tate group. After an extended optimisation process, this ELISA-based assay was able to significantly detect the dose-dependent increase in biotinylated ARL1 levels in MDA-MB-231 cells, thereby being able to reliably distinguish between medium to high NMT inhibition and non-inhibited controls (Figure 3.11C).

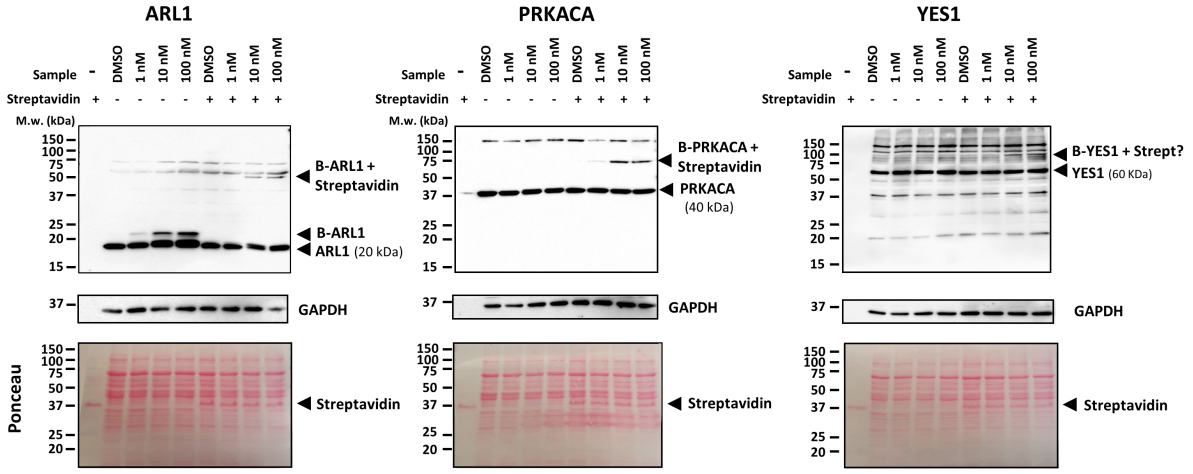


Figure 3.10: Streptavidin shift analysis of SrtA-labeled proteins. SrtA-labelled samples were briefly incubated with streptavidin before SDS-PAGE and blotted against ARL1, PRKACA and YES1. SrtA biotinylated ARL1 (B-ARL1) shows a molecular weight shift induced by the Biotin-ALPET label that gets shifted by 30 kDa on Streptavidin binding. Biotinylated PRKACA (B-PRKACA) shows no apparent shift in the absence of streptavidin, but a clear shift of 30 kDa upon addition of streptavidin. The band shift is less clear in YES1 due to the backgorund labelling of the antibody. These shifts match the apparent molecular weight of streptavidin alone, as shown by Ponceau staining.

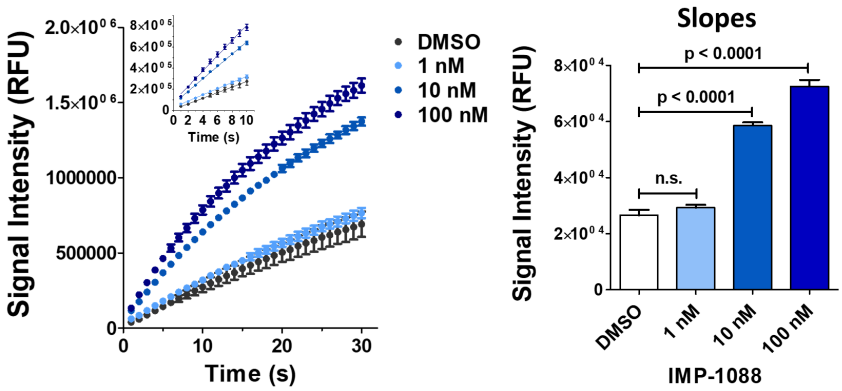


Figure 3.11: SrtA-ELISA analysis of labeled proteins. After protein precipitation to remove excess depsipeptide, 1 µg labeled protein lysate was applied to each well of a streptavidin-coated 96-well plate and incubated for 3 h at RT. After primary (anti-ARL1) and secondary (anti-rabbit HRP) antibody incubations, turnover of QuantaBlu fluorogenic HRP substrate was monitored for 30 min. Slopes were calculated from the first 10 minutes (linear range) after fitting to a straight line ($Y = Slope \times X + Y_{intercept}$) by nonlinear regression (Insert). Data were analyzed by one-way ANOVA followed by Dunnett's multiple comparison test.

3.6 The SrtA-based biomarker assay efficiently shows target engagement of NMT inhibitor IMP-1088 in a tumour xenograft mouse model

Having the SrtA-based biomarker assay in place to assess NMT activity in cells, we tested whether it could also detect changes in NMT activity in more complex samples such as tissue samples from an animal model. All mouse work was performed by Dr. Monica Faronato. We selected a Balb/c Nude mouse model in which cells from the human triple negative breast cancer cell line MDA-MB-231 had been subcutaneously injected. After the tumour volume had reached 50 mm^3, tumour-bearing mice were treated with an NMT inhibitor or PBS control (8 mice per group) through oral gavage twice a day for nine days. The identity of the NMT inhibitor is proprietary and cannot be disclosed. At the end of the nine days, animals were culled and tumours dissected and snap frozen before lysis in SrtA buffer using a cell disruptor. Tumour-derived proteome samples were then subjected to overnight SrtA labelling with Biotin-ALPET-Haa and loaded onto SDS-PAGE gels (Figure 3.12A and B) or ELISA plates (Figure 3.12C) for analysis using the ARL1 antibody. Both formats of the SrtA assay successfully showed target engagement of the NMT inhibitor in the mice tumours, thus confirming the use of the SrtA-based labelling strategy to confirm target engagement of NMT inhibitors in complex biological samples.

3.7 Conclusions

This study extends the use of SrtA to label proteins expressed at endogenous levels for the first time and provides evidence for its usefulness. The substrate specificity of SrtA in whole-cell lysates was examined, verifying its promiscuity beyond the recognised preference for N-terminal

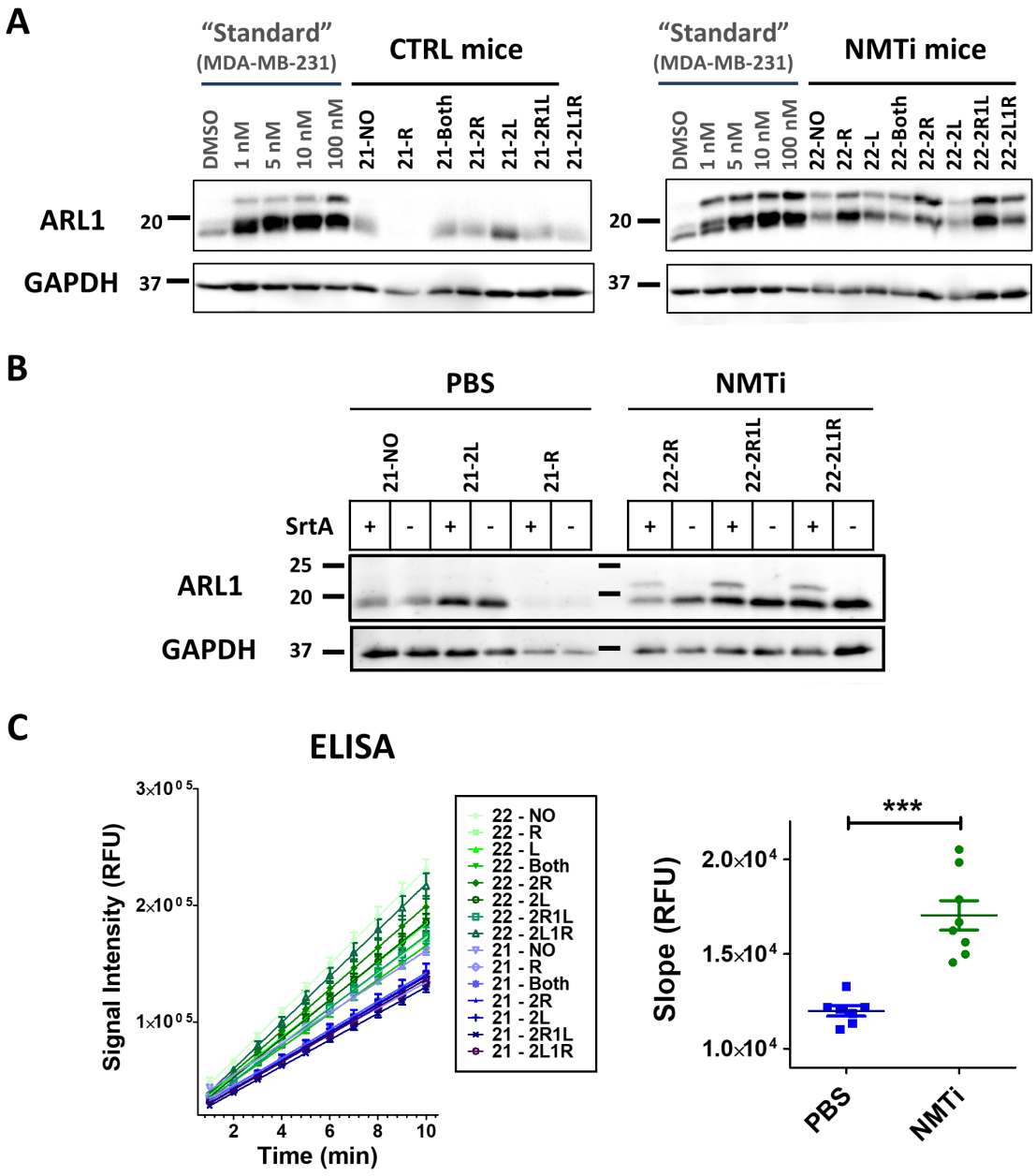


Figure 3.12: Assessing the on-target activity of an NMT inhibitor in a tumour xenograft mouse model using the SrtA-based assay. Tumour samples derived from a mouse model subcutaneously xenografted with the human breast cancer cell line MDA-MB-231 were lysed in SrtA buffer and labelled with SrtA and biotin-ALPET-Haa overnight. **(A)** and **(B)** SrtA-labelled tumour samples were loaded onto SDS-PAGE gels and blotted against ARL-1 side by side with a cell culture-derived sample 'standard'. The presence of a second band indicates NMT inhibition. **(C)** SrtA-labelled tumour samples were analysed using the ARL1 ELISA assay. *** indicates p-value$<$ 0.001 according to a two-tailed non-parametric Mann Whitney test

glycine. In contrast to previous reports in engineered systems [352, 364], our study did not suggest any advantages of having more than one glycine residue at the protein N-terminus or any other amino acid preference in positions 3-11, thus proving SrtA as a versatile method for non-selective labelling of complex mixtures of proteins containing an N-terminal glycine.

To enhance sensitivity of detection of NMT substrates by shotgun proteomics in samples

where the analyte of interest (NMT substrates) is diluted by virtually all proteins that have exposed N-terminal glycines, an improved affinity enrichment procedure was developed in which NeutrAvidin lysine residues are capped by dimethylation and resin-bound proteins are eluted by a short LysC on-bead cleavage step before full digestion with trypsin off-bead. This procedure removed typical avidin-derived high intensity peaks without affecting the yield of enriched protein-derived peptides, thus increasing the chances for successful identification and overall performance of the method. Notably, this derivatisation and two-step digestion protocol could potentially be applicable to any biotin-avidin affinity enrichment procedure to provide improved results independently of the sample source.

SrtA labelling of whole-cell lysates was successfully combined with a potent and selective NMT inhibitor to develop an assay capable of providing a readout of cellular NMT activity. This new SrtA-based labelling method was shown to have good overlap with YnMyr metabolic tagging strategy [244] and similar capability to measure cellular NMT activity. While YnMyr reflects the presence of *N*-myristoylation upon NMT inhibition, SrtA mirrors the absence of *N*-myristoylation on newly synthesized proteins exposing an N-terminal glycine. Therefore, the SrtA-based NMT assay described here is proposed as a complimentary method to the YnMyr strategy to assess NMT activity in cells. Moreover, it was shown here that this robust SrtA-based NMT activity assay can be coupled to a variety of detection techniques, ranging from MS-based proteomics for in-depth analysis, through gel-based analyses, to ELISA for increased throughput. Western blot of ARL1 seems particularly appealing as it can provide a quick and visual readout of NMT activity status by analysing the appearance of a second band of higher molecular weight when NMT activity is inhibited, without the need for affinity enrichment or any additional steps after SrtA-labelling.

This work constitutes the first report of applying SrtA not only as an enzymatic ligation mechanism, but as a sensor that detects the status of the cellular *N*-myristoylated proteome, thereby broadening the scope of the applicability of SrtA as a chemical biology tool. Most importantly, since the SrtA labelling reaction is performed post-lysis, this labelling strategy provides a readout of the cell status in the absence of any external perturbation, giving clear advantages over other in-cell labelling strategies and being applicable to any type of biological sample. To test the potential of this new assay, it was applied to tumour tissue samples derived from a mouse xenograft model and showed that it efficiently detected on-target activity of NMT inhibitors in those complex biological samples. We anticipate that this method will enable the exploration of new avenues in the usage of NMT inhibitors for therapeutic purposes, an area where the previous strategy involving metabolic tagging posed serious limitations.

4

Understanding the biological role of Methionine Aminopeptidase 2 and the effect of its inhibitors

4.1 Introduction

As discussed in the Introduction section 1.3, initiator methionine (iMet) excision plays an essential role in protein homeostasis in all kingdoms of life. In eukaryotes, all proteins are synthesised with an initiator methionine, which in >70% of the cases is removed by methionine aminopeptidases while the protein is still ribosome-bound [6]. The specific roles of iMet removal are not completely clear and are probably protein-specific, but it has been shown to influence important aspects of protein regulation such as protein stability and half life, protein interaction with binding partners and even L-Met metabolism [15, 160, 163] (cf. Introduction section 1.3.3). Moreover, this process affects all other protein N-terminal modifications occurring downstream of iMet removal, such as *N*-acetylation and *N*-myristoylation, amongst others (cf. ref [6, 9] and Introduction section 1.5).

The interest in MetAP2 started with the discovery of this protein as the specific target of fumagillin and ovalicin [176], two fungal-derived compounds that showed potent antiangiogenic and antitumoral activities [164, 180]. Both compounds were shown to bind MetAP2 irreversibly by covalently modifying a His residue at the catalytic site [133, 368]. A number of different inhibitor series have since been reported against MetAP2, both irreversible (mainly fumagillin-like compounds) and reversible (cf. Introduction section 1.3.4 for a full review on MetAP2 inhibitors). For three decades, MetAP2 inhibitors (MetAP2i) have been evaluated clinically for the treatment of both cancer and obesity, with little success so far despite the extensive efforts of multiple pharmaceutical companies. However, continued promising preclinical and clinical efficacy results for MetAP2i's over the years have helped to retain the interest of the drug development community.

There are still substantial uncertainties regarding the mode of action of these inhibitors and the underlying biological role of MetAP2. While the most reported therapeutic hypothesis for MetAP2 had been focused on reduction of endothelial cell proliferation and the subsequent anti-angiogenic effect [164, 168, 369], a myriad of other cellular activities have also been reported through the years. These include immunomodulatory activity [179], with some possible direct effects on B-cell differentiation [232], and sustained non-toxic weight-loss effects [207, 218]. Moreover, numerous studies have reported non-cytotoxic anti-proliferative effects in a range of other cancer cell lines beyond primary endothelial cells [149, 200, 370]. Whether the anti-cancer or anti-obesity activity of these inhibitors relates solely to the anti-angiogenic capabilities or it derives from a synergistic effect together with direct targeting of tumour cells and adipocytes is still not clear. Furthermore, the changes induced by MetAP2i at the molecular level are also inconclusive. MetAP2 inhibition has been implicated with a number of cellular signalling events, ranging from cell cycle regulation, non-canonical Wnt signalling and PI3K signalling [234, 237, 240], to some apoptosis-related pathways in a cell specific manner [171, 173]; but the basis on which MetAP2 contributes to those effects is still unknown.

A major question regarding the biological role of MetAP2 is the nature of its substrates.

Both MetAP1 and MetAP2 are multi-substrate enzymes. A broad substrate profile for both is known [131, 146, 147, 149]; however, the specific differences between the two are unclear. For decades, pharmaceutical companies have invested titanic efforts in developing more selective and more potent inhibitors with better phamacokinetic properties, but have paid little attention to the molecular effects underlying their desired systemic phenotypes. Studies concerning the molecular basis and consequence of perturbing native MetAP2 activity have thus far not been able to build a clear picture of how these phenotypes result from MetAP2 inhibition. To address this challenge, this chapter describes a novel method to characterise MetAP2 activity and shed light on the molecular mechanism of action of its inhibitors. This was mediated by collaboration with three pharmaceutical companies to obtain a diverse library of compounds, reversible and irreversible, which were characterised extensively using cellular and biochemical assays. Moreover, a new chemical proteomics workflow was developed in combination with these specific MetAP2i's, which enabled the identification of a comprehensive list of MetAP2 substrates in cells for the first time. Finally, with iMet removal being the process directly preceding *N*-myristoylation, the potential overlap of MetAP2i with NMT inhibitors was also studied. This work paves the way for future studies that will ultimately bridge the gap between molecular inhibition of MetAP2 and its phenotypic effects, unlocking the full potential of MetAP2i's in the clinic.

4.2 Preliminary studies: Establishing new collaborations to build a panel of MetAP2 inhibitors and validating a cellular model to study their phenotypic effects

Due to our research group's growing interest in understanding the therapeutic scope of *N*-myristolylation inhibition (cf. Results Chapters 2 and 3, Introduction section 1.5 and [244, 298, 312]), we decided it was pertinent to understand whether the MetAP2 inhibition effects could be also be mediated by downstream *N*-myristolylation blockade. In order to determine the overlap between MetAP2 and NMT, further understanding of the MetAP2 substrate profile was required, which was envisaged through a combination of specific pharmacological MetAP2 inhibition and mass-spectrometry based proteomics. Collaborations were established with three pharmaceutical companies, which provided inhibitors as well as related information regarding their biochemical and in cell potencies. Unfortunately, since the provided data had been generated through different assays and distinct cellular/animal models, the data were difficult to reconcile between companies. Therefore, assays that could systematically characterise all inhibitors in the panel head-to-head were necessary before any substrate-profiling experiments could be performed. However, since method development of the challenging mass spectrometry experiments is time consuming, it was decided to carry out these two lines of research in parallel. On the one hand, a novel proteomics workflow that would enable the systematic discovery of new MetAP2 substrates was designed, tested and optimised; and on the other hand, cellular target engagement assays were sought that could validate the on-target

effect of MetAP2i and identify the optimal concentration range for comparative studies. In this chapter, efforts for determining the most appropriate cell model for MetAP2 inhibition and dose-response ranges for all MetAP2i are described first, and the proteomics optimisation process is explained second, although both lines of research were carried out in parallel.

The MetAPi panel was built over the course of the project through collaborations with three pharmaceutical companies. GSK provided two compounds: a tool molecule from Abbott (Abbott, **1**) and a compound they developed which had similar potency but improved physicochemical properties (GSK, **2**). Takeda also sent two of their lead preclinical MetAP2i candidates: Tak1 (**3**) and Tak2 (**4**). Finally, Merck KGaA (referred to as Merck hereafter) provided three compounds to complete the MetAP2i panel: one reversible inhibitor from Merck (Merck+, **5**) [204], its inactive enantiomer (Merck-, **6**) [204] and a fumagillin-like irreversible compound from Zafgen, ZGN-1061 (ZGN, **7**) [221].

Since GSK had focused their research on the effect of MetAP2 inhibitors on B-cell differentiation (unpublished work, based on previous research [231, 232]), BL41 Burkitt lymphoma cell line was first selected to test the first two compounds provided by GSK, Abbott (**1**) and GSK (**2**). While the first proteomics experiment on this cell line showed very promising results (Figure 4.9 and accompanying text), no phenotypic effect was detected in this suspension cell line in cell viability assays or flow cytometry studies (data not shown). Aware that most of the studies in the literature had been performed on endothelial cells, an immortalised somatic endothelial cell line, EA.hy926, was selected next. However, despite numerous efforts in trying to detect a compound-induced phenotypic response in this new cell line, the struggle to observe any clear and consistent effect persisted (Figure 4.2). While a subtle increase in G1/0 population could be observed upon exposure at very high concentrations of the GSK compound (Figure 4.2B), this was at a concentration 100 times higher than its reported EC_{50} ([198] and personal communication).

After consultation with MetAP2 experts at Merck, it was suggested to switch to using three different cancer cell lines: the fibrosarcoma cell line HT1080, the lung carcinoma cell line A549 and the glioblastoma cell line U-87 MG. Using a real-time proliferation assay carried out on an Incucyte live cell imager as described in the next section (cf. section 4.3), a clear dose-response effect was finally detected in all three cell lines (Appendix Figures A.5 and A.6). The fibrosarcoma cell line HT1080 was selected for the majority of further studies.

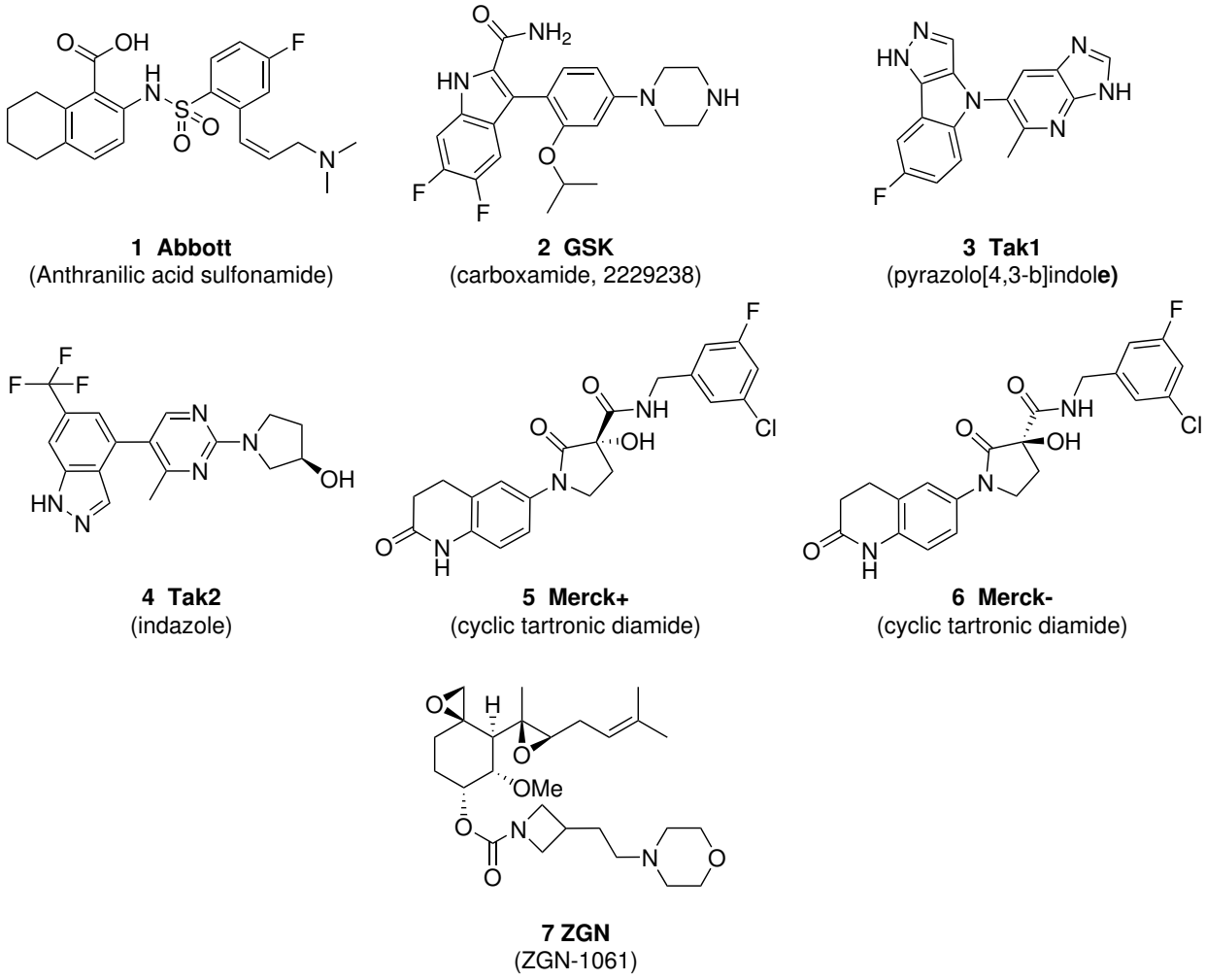
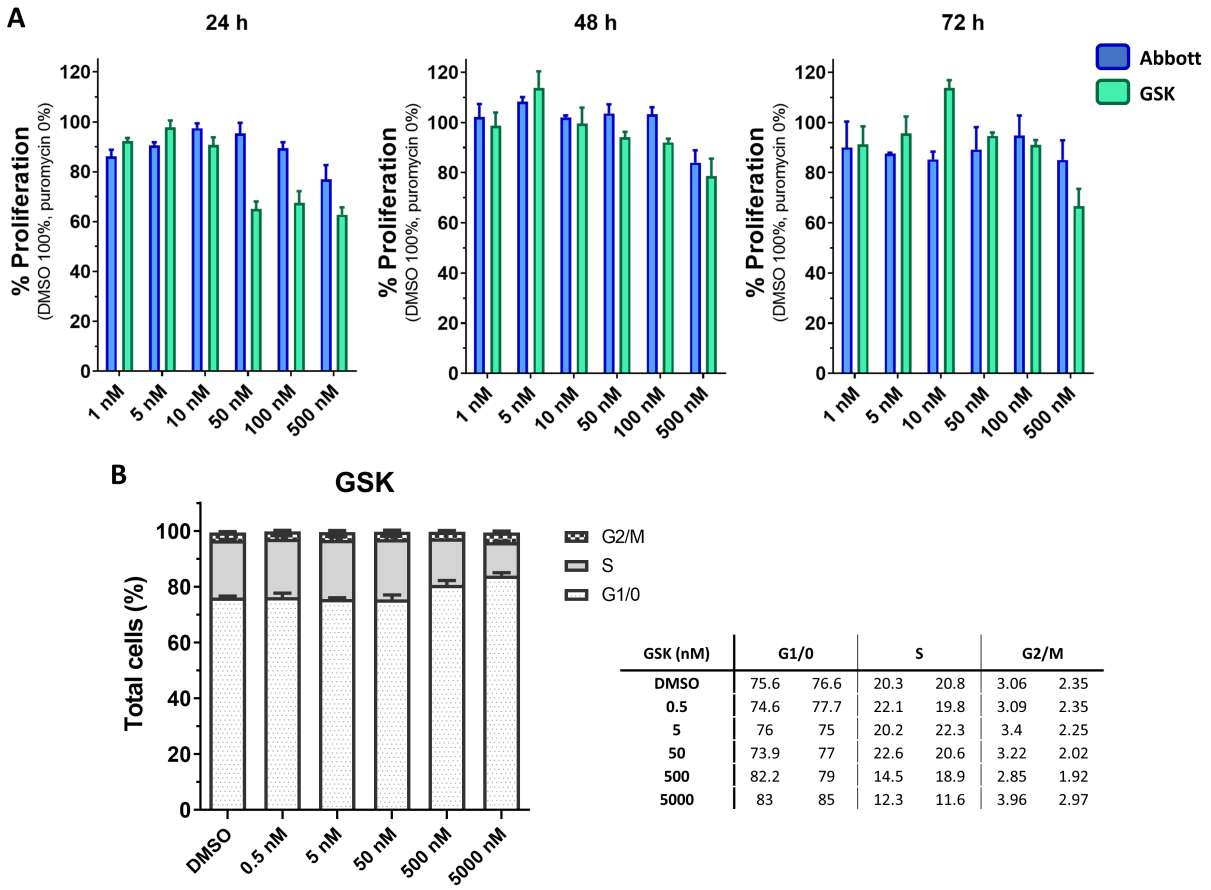


Figure 4.1: Our panel of MetAP2 inhibitors. Abbott (**1**), an anthranilic acid sulfonamide (24n in [198]); **GSK** (**2**), a carboxamide (also compound 9 in Introduction Figure 1.9) [204] ; **Tak1** (**3**), an indole (compound 10 in [206] and also compound 11 in Introduction Figure 1.9); **Tak2** (**4**), an indazole (compound 38 in [205] and also compound 10 in Introduction Figure 1.9); **Merck**+ (**5**), a cyclic tartronic diamide (also compound 14 in Introduction Figure 1.9) and its inactive enantiomer, **Merck**- (**6**) [204]; and **ZGN** (**7**), an irreversible inhibitor developed for the treatment of obesity (ZGN-1061, also compound 17 in Introduction Figure 1.10) [221].

4.3 Real-time proliferation monitoring provides a quantitative measurement of MetAP2 inhibitor activity

Literature suggested a profound cytostatic activity of MetAP2 inhibitors in certain cell lines. After unproductive initial studies using colorimetric assays such as Sulforhodamine B (SRB) assay (Figure 4.2) or MTS (not shown), attention turned to using a recently aquired Incucyte live cell imager. The Incucyte is an automated robotic microscope platform placed within a regular cell culture incubator that allows for real-time monitoring of cell cultures. Importantly, it comes with a built-in image recognition software that converts the acquired image data into quantitative data for subsequent analysis. Being able to measure the time variable in a highly frequent and continuous manner was extremely useful for initial testing, when the



GSK (nM)	G1/0		S		G2/M	
DMSO	75.6	76.6	20.3	20.8	3.06	2.35
0.5	74.6	77.7	22.1	19.8	3.09	2.35
5	76	75	20.2	22.3	3.4	2.25
50	73.9	77	22.6	20.6	3.22	2.02
500	82.2	79	14.5	18.9	2.85	1.92
5000	83	85	12.3	11.6	3.96	2.97

Figure 4.2: Cell viability assay and cell cycle profile of the EA.hy926 cell line upon MetAP2 inhibition. **(A)** SRB colorimetric cell viability assay of EA.hy926 cells upon exposure to a series of concentrations of Abbott (**1**) and GSK (**2**) MetAP2i. **(B)** Flow cytometry analysis of cell cycle profiles of EA.hy926 cells upon exposure to GSK (**2**) MetAP2i.

time-frame of the expected cellular effect is unknown in addition to compound concentration. Moreover, this system allows for the systematic study of cells in real time in their growth environment, avoiding any handling errors or sudden changes in the culturing environment. The main advantage of the Incucyte compared to more traditional viability assays that are often based on metabolic activity or DNA or protein content, however, is that it provides phenotypic insight. By analysing the imaging data in addition to the quantitative data, it is possible to detect alterations in cell morphology in a time-resolved manner. Additionally, having integrated phase contrast and fluorescence microscopes, it can be coupled to specific live-cell dyes to study other cell parameters in addition to proliferation status and confluence. In this case, SYTOX Green (Invitrogen) was used, a commercially available, cell impermeable cytotoxicity dye originally designed for flow cytometry analyses. SYTOX Green is a nucleic acid stain that binds to DNA upon disruption of the cell plasma membrane integrity, acting as a marker for cell death. Together, this combined assay provided a platform to test one of the main characteristics of MetAP2 inhibitors reported in the literature: their cytostatic activity with no apparent cytotoxicity.

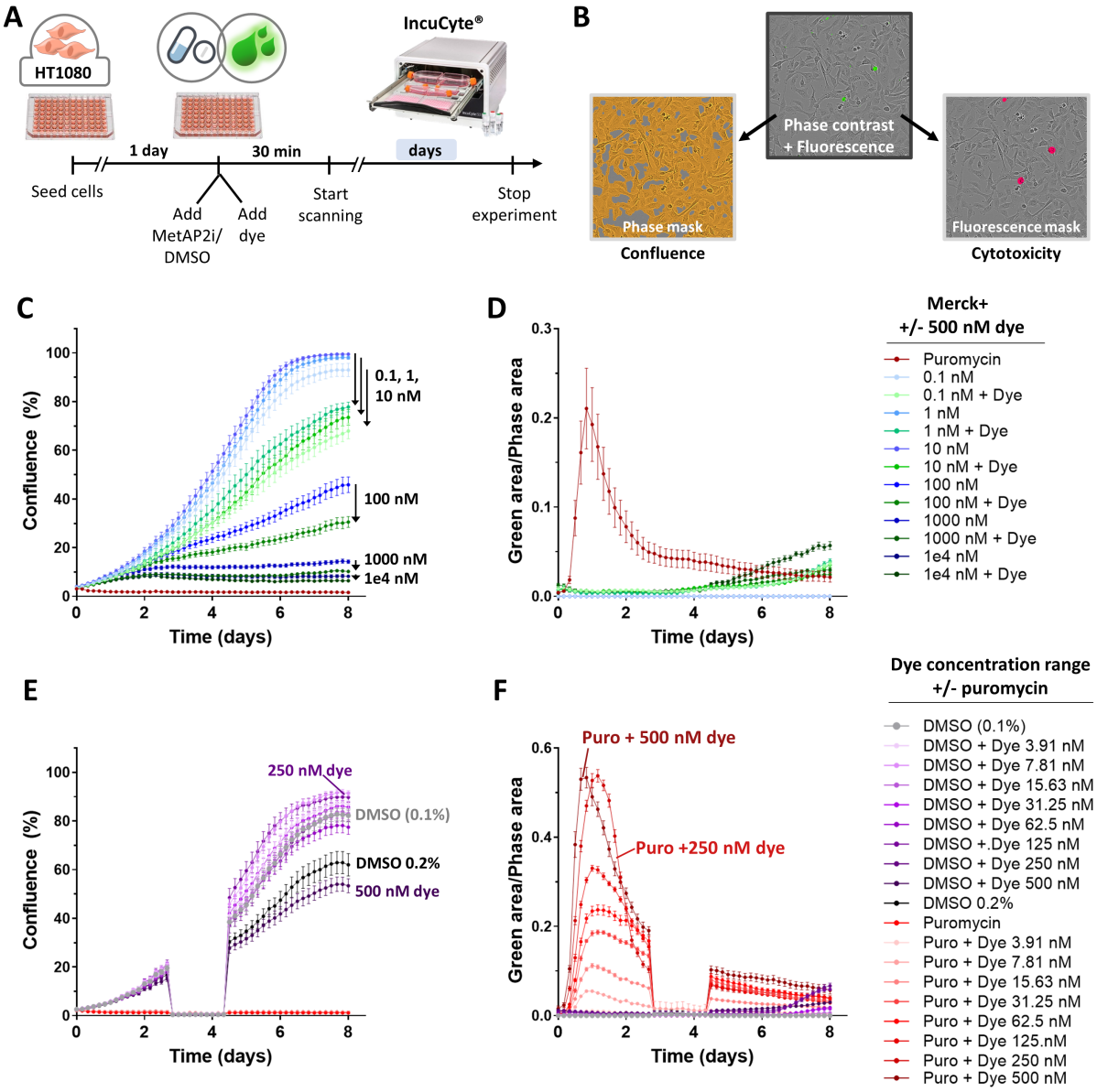


Figure 4.3: Synergy of SYTOX Green dye with MetAP2 inhibitors and dye concentration optimisation. **(A)** Schematic representation of the experimental procedures for this experiment using an Incucye. **(B)** Representative image taken by the Incucyte, showing the overlayed image taken by the phase-contrast and fluorescence microscopes. These are then analysed with the integrated image-recognition software to build experiment-specific 'masks' that allow for conversion of image data into quantitative data. **(C)** and **(E)** show the proliferation curves over time, measured as the percentage of total imaged area that was covered by cells at each time point (phase-contrast) while **(D)** and **(F)** show the displayed cellular cytotoxicity over the same period of time. Cytotoxicity was measured as the ratio of green fluorescence-positive area to total cell area. Cell area was calculated as in (C) and (E) from phase-contrast images. The fluorescence threshold was manually selected for each experiment based on puromycin signal (positive control for cytotoxicity) and commonly applied to all data sets of the same experiment. **(C-D)** Addition of SYTOX Green dye showed an added effect on cell proliferation when dosed alone or in combination with MetAP2i Merck+ (**5**), distorting the real effect of the inhibitor. **(E-F)** Titration of SYTOX Green dye concentration confirmed the results observed in (C-D) and showed that 250 nM final dye concentration suppressed this synergistic effect (E) while maintaining same signal/noise (F). Time-points with no signal correspond to scans where the auto-focus feature of the microscopes failed.

First tests with MetAP2 inhibitors on this platform showed an added effect of the SYTOX Green dye on cellular viability that negatively impacted the activities of MetAP2 inhibitors (Figure 4.3A-B). This was hypothesised to be a combined effect between addition of high concentrations of dye, with increased final concentration of DMSO, as this value was already 0.1% due to addition of inhibitors. With the dye initially diluted down from a 1000X stock in DMSO, the final concentration of DMSO was 0.2%, which resulted nearly as detrimental as the highest concentration of dye tested (Figure 4.3E). Titration of the final concentration of dye showed that the next 1:2 dilution down solved the cellular cytotoxicity issue, while maintaining the cytotoxicity fluorescence levels at a level comparable to the highest concentration (Figure 4.3C-D). Therefore, a final dye concentration of 250 nM was selected for all subsequent experiments, together with an increase in the concentration of the parent DMSO dye stock used to reduce the final DMSO concentration to a minimum (final DMSO concentration 0.1025

With all these parameters optimised, the next step was to test the MetAP2i's using this platform. Concentration ranges were selected for each inhibitor in our panel based on previous preliminary Incucyte experiments (not shown) to ensure full coverage of the dose-response curve. The proliferation index was calculated as the slope of the straight line, fitted around the common linear range of the curves (2.5-4 days). This was then used to calculate the EC_{50}, by plotting it against the $\log_{10}$(MetAP2i concentration) and fitting it to a four-parameter curve (variable slope) (Figure 4.4C). The dye release is characterised by an initial rapid release with the start of the cytotoxicity, followed by a period of diffusion that stretches long after the cells have died. Calculating the area under the curve (AUC) constituted an appropriate way to profile this, as it allowed comparison across conditions independently of the underlying mechanism, and allowed for the use of an independent positive control (puromycin) (Figure 4.4D). This analysis constituted a robust workflow to obtain high quality dose-response curves for all inhibitors and calculate their potencies. Moreover it confirmed the absence of any apparent cytotoxicity for all of the inhibitors even at the highest concentration tested (Figure 4.4E-F and Appendix Figures A.7 and A.8), consistent with previous reports in the literature [168]. Importantly, this assay provided a rationale for selecting the phenotypically relevant inhibitor concentrations for proteomics experiments. The ZGN (**7**) compound showed exquisite sub-nanomolar potency (51 pM), similar to that reported for other fumagillin-like compounds [164, 168, 221]. This was at least 3 orders of magnitude lower than the next most potent reversible inhibitor in our panel. Abbott (**1**), GSK (**2**) and Merck+ (**5**) compounds showed similar potency and EC_{50} values (10 nM, 35 nM and 64 nM, respectively) in line with those reported by each pharma company (personal communication and [198, 204]). Merck- (**6**) compound was not completely inactive, consistent with the data given by Merck (personal communication), but provided a sufficient activity window for comparative studies with its active enantiomer Merck+ (**5**), making this inhibitor pair particularly valuable for proteomics studies. Of note, while the dose-response curve shape showed the same pattern of inhibition for the reversible compounds Abbott (**1**), GSK (**2**), Merck+ (**5**) and Merck- (**6**) and the

irreversible MetAP2i

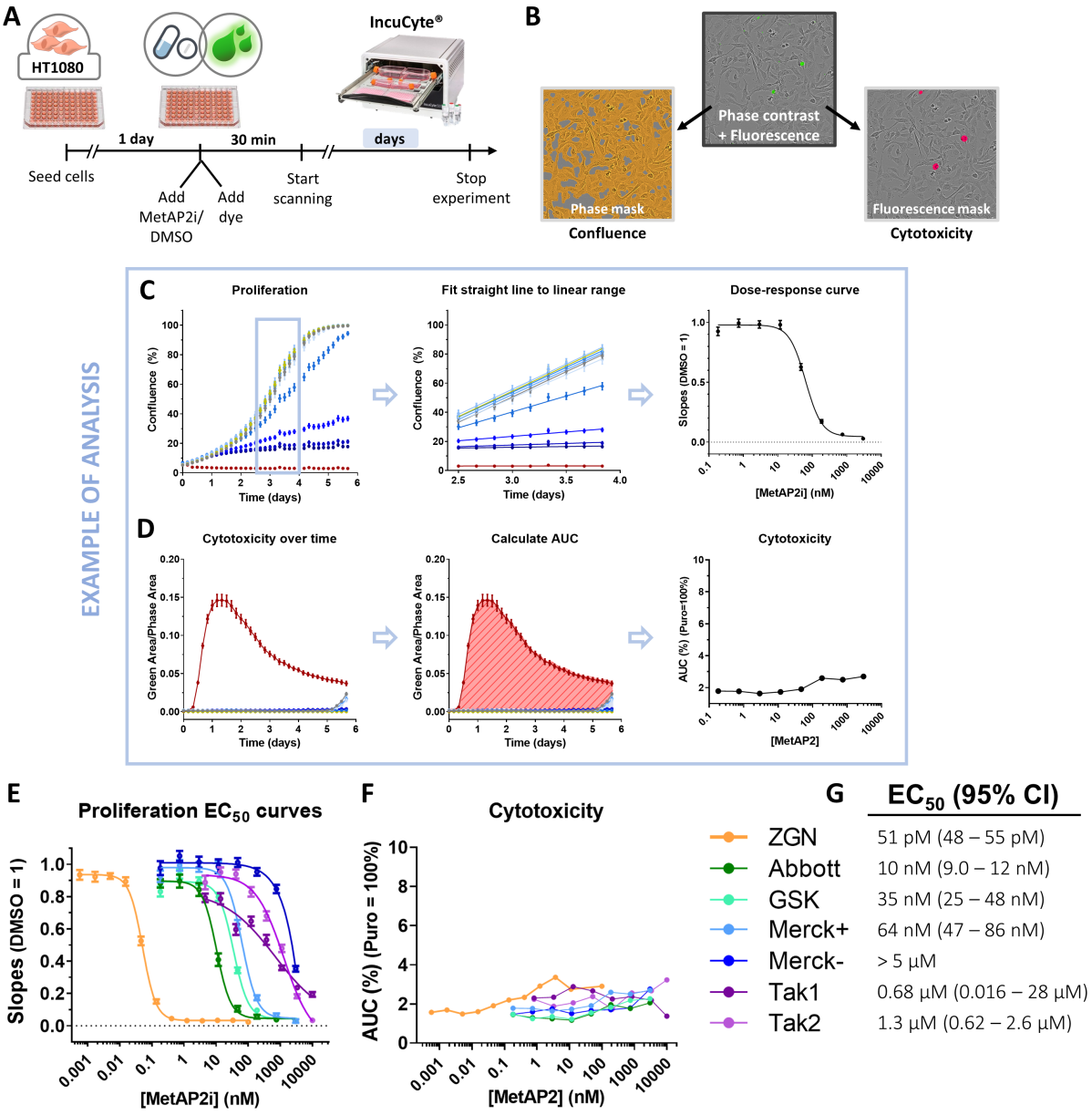


Figure 4.4: Real-time cell proliferation and cytotoxicity assay confirms cytostatic effect of all MetAP2 inhibitors without any apparent cytotoxicity. **(A)** Schematic representation of the experimental procedures for this real-time proliferation and cytotoxicity assay using the Incucyte. **(B)** Representative image of a time point of the experiment, showing the overlayed image taken by the phase-contrast and fluorescence microscopes. These are then analysed with the integrated image-recognition software to build experiment-specific 'masks' that allow for conversion of image data into quantitative data. **(C)** Representation of the analysis performed on the proliferation data for each inhibitor. A common linear range for all inhibitors was manually selected (days 2.5-4) and fitted to a straight line. The slopes were then plotted against the $\log_{10}$(MetAP2i concentration) for each inhibitor and a 4-parameter curve fitted to determine the EC_{50} value. **(D)** Representation of the analysis performed on the cytotoxicity data for each inhibitor. For each compound, the area under the curve (AUC) for days 0-5 was calculated, normalised to the positive control (puromycin, representing 100%) and plotted against the $\log_{10}$(MetAP2i concentration). Data points after day 5 were excluded to avoid altering the values with cytotoxicity resulting from overconfluence, observed for DMSO control and lowest MetAP2i concentrations. **(E)** Dose-response curves for all inhibitors as calculated based on their inhibition of proliferation. **(F)** None of the compounds showed any apparent cytotoxicity even at the highest concentrations tested. **(G)** EC50 values and 95% confidence intervals (CI) for each MetAP2i calculated from proliferation dose-response curves in (E).

ZGN (**7**), this was not shared by the two Takeda compounds (Tak1 (**3**) and Tak2 (**4**)). This observation hinted that these compounds may possess different mechanism of action. Moreover, the EC_{50} values obtained in this proliferation assay deviated significantly from the reported potencies (680 nM and 1.3 µM vs. 12 nM and 15 nM for Tak1 and Tak2, respectively) [205, 206] . Interestingly, the reported EC_{50} values of these compounds had been obtained using an iMet-14-3-3γ antibody assay that has been shown to also account for MetAP1 activity [193, 371]. Potencies of these compounds against MetAP1 have not been reported.

4.4 Cellular Thermal Shift Assay (CETSA) confirms cellular target engagement of MetAP2 inhibitors

With a validated assay to measure the anti-proliferative effects of our MetAP2i panel now in place, the next step was to confirm inhibitor engagement of MetAP2 in cells. Evidence from a preliminary proteomics experiment showed clear iMet retention differences upon inhibition (Figure 4.9), suggesting MetAP2 target engagement indirectly. However, an assay that could indicate direct and quantitative target engagement was considered useful. Due to the fact that there are robust antibodies for MetAP2 detection, a cellular thermal shift assay (CETSA) followed by a western blot readout (CETSA-WB) was selected for this purpose. CETSA is based on the same principle as another well-known biochemical assay, isothermal titration calorimetry (ITC), routinely used to characterise ligand binding *in vitro*. In essence, the intrinsic thermal stability of a protein (normally determined by the melting temperature (T_m), the temperature by which half of the protein is denatured in a specific buffer condition) increases upon binding to a stabilising partner, which can be a substrate or a ligand; in this case, a small molecule inhibitor. This creates a detectable shift in the melting temperature (ΔT_m) (Figure 4.5B). This principle has been shown to be applicable in the cellular environment, where the heat-treatment can be performed on intact cells [372]. After centrifugation of the protein lysate at high speeds, it is possible to separate the unfolded proteome from the more thermally stable, folded proteome (schematic representation of a typical CETSA procedure is depicted in Figure 4.5A). This allows target engagement of compounds to be determined and has the potential to identify on and off-targets of novel compounds by coupling CETSA with shotgun proteomics [373, 374]. However, for the purpose of this investigation, CETSA-WB was considered more appropriate to test the fate of MetAP2 upon addition of MetAP2i followed by thermal treatment. Screening a two dimensional matrix of conditions by western blot can be time consuming and difficult to interpret. Instead, a strategy similar to the one used in ITC experiments was applied: the apparent T_m of the target (MetAP2) was determined first, to then apply an isothermal dose-response with a concentration range of MetAP2i at the determined T_m.

The apparent T_m of MetAP2 in HT1080 cells was determined to be 72 °C, markedly higher

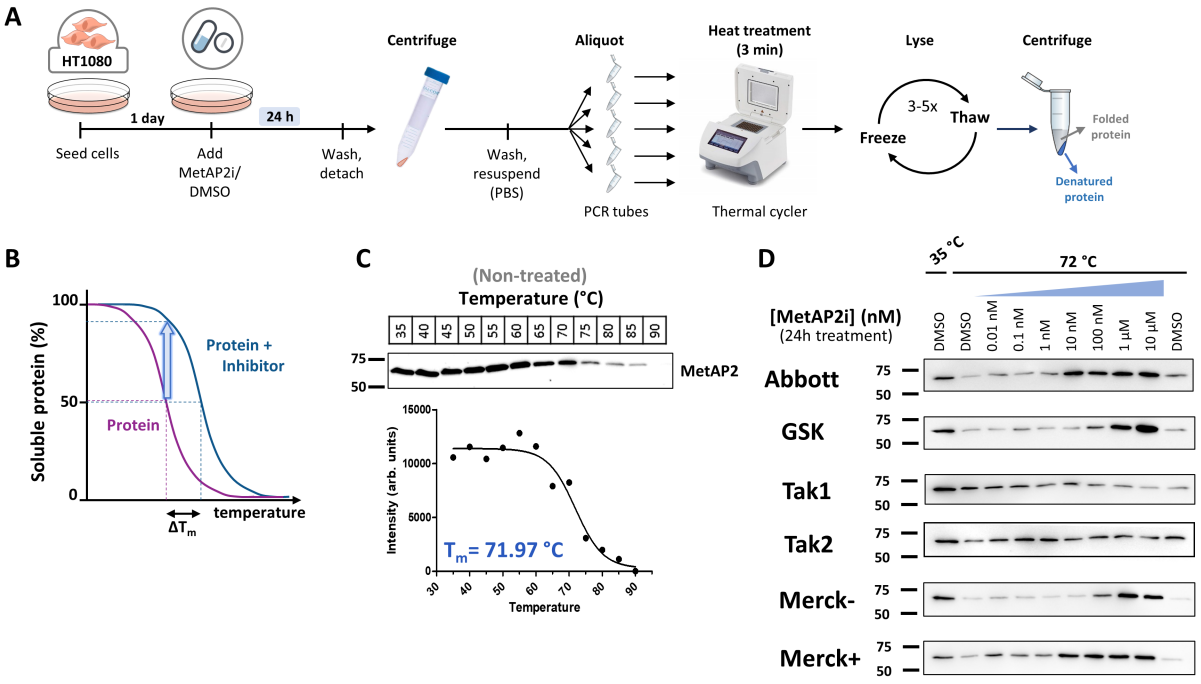


Figure 4.5: CETSA-WB validates on-target activity of MetAP2 inhibitors. (A-B) Schematic representation of the experimental procedures involved in this experiment. One day after seeding, cells were dosed with MetAP2i or DMSO for 24 h before harvesting them with trypsin. Cells were centrifuged and resuspended in PBS before aliquoting into PCR tubes. Live cells were subjected to a 3 min heat-shock treatment before mechanical lysis by freeze-thaw cycles. Lysates were then centrifuged at maximum speed to ensure separation of the denatured (pellet) and folded protein pools in the sample. Cleared lysates (folded protein) were taken forward for SDS-PAGE analysis and western blot. **(B)** Schematic representation of the CETSA principle. A protein will generally gain thermal stability upon binding to an stabilising ligand, this case a small molecule inhibitor. This will produce a shift in the melting temperature (ΔT_m), but also a detectable shift in the amount of soluble protein at one given temperature (blue arrow). This effect will be maximal around the T_m of the protein. **(C)** Non-treated control samples were subjected to a temperature-gradient CETSA experiment to determine the apparent T_m of MetAP2 (compound-free) in this cell system (HT1080). T_m detected was of 72 °C. **(D)** Isothermal dose-response (ITDR) CETSA experiment. Cells were dosed with different concentrations of MetAP2i before heat-shock at a single temperature of 72 °C (T_m) or 35 °C (non-denatured control).

than the average protein melting temperature (ca. 60 °C) [374] (Figure 4.5C). Cells were then dosed with a series of MetAP2i concentrations and subjected to either a 3 min heat-shock at 72 °C (T_m) or at a 35 °C control. This revealed clear stabilisation and thus target engagement of MetAP2 for all compounds (except for the Takeda MetAP2i's) that correlated with the same inhibitor concentration window of anti-proliferative activity observed in the Incucyte experiments (Figure 4.5C and Appendix Figure A.9A). The ZGN compound was added later to these analyses, together with a more refined selection of the concentration ranges of the two Merck compounds to determine the differential window more clearly in an independent experiment carried out by another PhD student in the lab, Cassandra Kennedy (Appendix Figure A.9B). After a few trials, the target engagement of the two Takeda compounds (Tak1 (**3**) and Tak2 (**4**)) could not be confirmed, leading to the removal of these two compounds from the MetAP2i panel for future experiments.

4.5 A novel metabolic tagging proteomics approach based on azidohomoalanine enrichment allows for systematic MetAP2 substrate elucidation

4.5.1 Previous efforts to identify MetAP2 substrates

Few studies have been performed to profile MetAP2-specific substrates. Most, undertaken by pharmaceutical companies in the search for novel biomarkers, detected only a handful of MetAP2 substrates. GAPDH and CypA were the first MetAP2-specific substrates to be revealed by 2D SDS-PAGE followed by MALDI-TOF identification of in-gel tryptic digests [181]. This list was slightly expanded with the identification of TSH, SH3BGRL and eEF2 by combining MALDI-/SELDI-MS profiling and either FT-ICR-MS identification of small intact proteins or in-gel tryptic digestion and regular bottom-up proteomics [370]. The MetAP substrate 14-3-3γ was initially discovered using the dual MetAP1/MetAP2 inhibitor LAF389 (compound **6** in Introduction Figure 1.9) [193] but has later been reported as a MetAP2-specific substrate in numerous occasions [205–207, 210]. However, this protein has also been employed as a biomarker for MetAP1-specific inhibitors [371], and it is still unclear how this protein seems to necessitate both active MetAP1 and MetAP2 to be fully processed and neither of them seems to be able to compensate in the absence of the other. An independent study on the effect of MetAP2 inhibitors in non-canonical Wnt-PCP signalling indirectly revealed Rab37 as a novel MetAP2 substrate, although the authors were unable to detect it at endogenous levels [237].

A few other systematic studies focused on determining the substrate specificity differences between MetAP1 and MetAP2 have also been reported. Three of these studies, comparing the two yeast enzymes [131], the bacterial vs archaeal enzymes [148], or the two human enzymes [147], revealed little differences between the substrate specificities of both enzymes (discussed in more detail in Introduction section 1.3.2.4). Moreover, performed either based on recombinant protein expression (with altered N-terminal sequences) [131] or *in vitro* synthetic peptide array analysis [147, 148], they did not identify any new endogenous substrates of either MetAP.

One discovery proteomics experiment has been reported previously that analysed general patterns of all N-terminal peptides arising from MetAP2 inhibition with fumagillin [149]. They used the SILProNAQ approach for N-terminal peptide enrichment, which consists on enrichment of N-terminal peptides through heavy SCX fractionation after chemical heavy *N*-acetylation with *N*-acetoxy-d_3-succinimide. However, their analysis was only based on the presence or absence of each type of peptide (iMet retained or iMet cleaved) rather than on a quantitative measure, and they did not provide a list of confidently determined MetAP2-specific substrates. Unfortunately, they did not make their proteomics data publicly available, not allowing us to perform independent searches on their raw data.

Therefore, the field would benefit from a new systematic study of the substrate profile of MetAP2 that could provide a mechanistic insight for the phenotypic effect of MetAP2 inhibition. Moreover, a more complete list of MetAP2 substrates could also provide new and improved biomarkers to show target engagement of MetAP2 inhibitors in complex biological samples, which is essential for drug candidate clinical development, as discussed in Results Chaper 3. Currently reported biomarker assays are either cumbersome, such as the IEF electrophoresis assay to detect GAPDH band shift upon iMet retention [200], or based on the dubious dual MetAP1/MetAP2 substrate 14-3-3γ [210, 371].

4.5.2 Previously reported proteomics methods for the study of N-terminal peptides: positive and negative enrichment approaches

Several proteomics methods have been described for the systematic study of protein N-terminal peptides, both for the identification of N-terminal modification sites as well as proteolytic cleavage events. They can be generally classified into two main categories: positive enrichment methods and negative enrichment methods. In the former, N-terminal peptides are specifically tagged and enriched, while the latter consist of labelling internal and C-terminal peptides to allow for their depletion. Generally, negative enrichment methods have been more widely used, while positive enrichment methods have not been so successful [375]. The main drawback of positive enrichment methods, which rely on chemical [376] or enzymatic [377] labelling of free N-termini for their subsequent affinity enrichment, has been their inability to identify endogenously modification sites as these peptides are depleted during the enrichment procedure. Therefore, these methods have been limited to the identification of new proteolytic sites.

Negative enrichment methods rely on initial capping of the free amines in N-termini and lysines at the protein level (generally by chemical dimethylation, acetylation or isobaric labelling), followed by protease digestion, which generates newly exposed free amines in internal and C-terminal peptides. They then apply different strategies in order to separate capped amine-bearing N-terminal peptides from exposed amine-containing internal and C-terminal peptides, which can be grouped into the following categories:

- *Negative enrichment of N-terminal peptides based on chromatographic separation.* These include COmbined FRActional DIagonal Chromatography (COFRADIC) [378–380], Charged-based FRActional DIagonal Chromatography (ChaFRADIC) [381] and Stable Isotope-Labelled Protein N-terminal Acetylation Quantification (SILProNAQ) [382, 383] (Figure 4.6). COFRADIC is based on extensive fractionation by reverse-phase liquid chromatography (RP-HPLC) [378]. After trypsin digestion, a pre-enrichment step of N-terminal peptides is performed using SCX chromatography at low pH (pH 3) [379]. The peptide mixture is further separated in two RP-HPLC fractionation rounds [378–380] (Figure 4.6A). In the first, hydrophobic N-terminal peptides (capped amines) are

separated from amino-exposed internal peptides. Next, all obtained fractions are subjected to chemical modification with 2,4,6-trinitrobenzenosulfonic acid (TNBS), which specifically modifies free amines in internal peptides, dramatically increasing their hydrophobicity. A second round of deep RP-HPLC fractionation further separates the now more hydrophobic internal peptides from N-terminal peptides. The initial protocol of this method yielded over 100 fractions [378, 379], but these were further pooled into 36 fractions [380], which considerably reduced the machine run time, often a limiting factor due to the high costs involved. ChaFRADIC was introduced as a simpler version of COFRADIC, this time based on two rounds of SCX fractionation, separated by the acetylation of internal peptides by NHS d_3-acetate, which changes the charge state of internal peptides to ensure full separation from N-terminal peptides (Figure 4.6B). Because in ChaFRADIC each SCX round is only separated into 5 fractions (according to charge states $+1$, $+2$, $+3$, $+4$ and >4), this method requires significantly less starting material (50 µg) and needs markedly less machine time [381]. SILProNAQ was introduced for the specific analysis of endogenous *N*-acetylation and relies on acetylation with isotopically labelled heavy *N*-acetoxy-d_3-succinimide before trypsin digestion followed by SCX fractionation [382, 383] (Figure 4.6C).

- *Negative enrichment of N-terminal peptides through direct scavenging of exposed amines of internal peptides.* These methods are based on aldehyde-functionalised platforms, which are covalently reacted with free primary amines in internal peptides through reductive amination. They include Dimethyl Isotope-coded Affinity selection (DICAS), which uses a commercially available aldehyde-functionalised immobilised matrix (POROS-AL) [384]; and Terminal Amine Isotopic Labelling of Substrates (TAILS) (Figure 4.7A), which instead uses an in-house generated, high molecular weight, aldehyde-derivatised polyglycerol polymer (HPG-ALDII) that can be then separated from non-reacted N-terminal peptides by ultrafiltration [385, 386].

- *Negative enrichment of N-terminal peptides by labelling exposed amines on internal peptides.* Such approaches include the enzymatic phosphorylation of internal peptides by glyceraldehyde-3-phosphate (G3P) before phospho-tag (PTAG) enrichment on TiO_2 affinity columns [387], biotinylation of internal peptides with a biotin NHS ester followed by avidin affinity enrichment [388] or modification with saturated fatty aldehydes such as hexadecanal (HYdrophobic Tagging-Assisted N-termini Enrichment, HYTANE) [389] or undecanal (High efficiency Undecanal-based N-Terminal EnRichment, HUNTER) [**Weng2019b**] (Figure 4.7B).

In all these methods, initial labelling of free amines with isotope-labelled dimethylation or acetylation reagents or isobaric-labels (TMT, iTRAQ) then allow for a more accurate quantitative analysis of the negatively enriched N-terminal peptides. The use of isotope-coded dimethylation reagents or isobaric labelling allows for sample multiplexing, which considerably

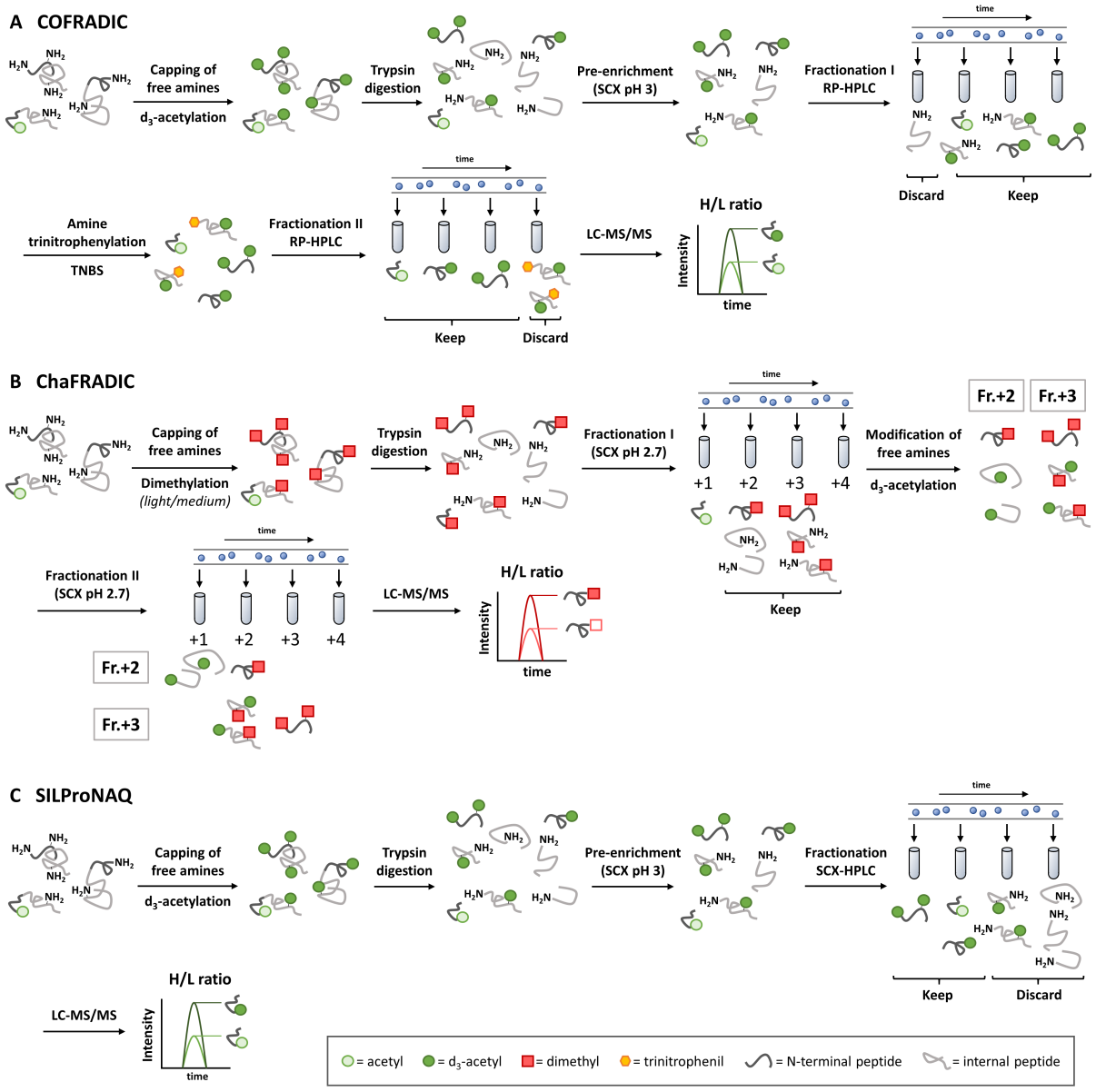


Figure 4.6: Main methods for N-terminal peptide enrichment through chromatographic separation. All primary amines on N-termini and Lys are capped before trypsin digestion. (A) In COFRADIC, N-terminal peptides are then pre-enriched by SCX chromatography before a first round of RP-HPLC fractionation. Free primary amines on internal peptides are then modified by treatment with TNBS before a second round of RP-HPLC, which fully separates N-terminal from internal peptides before LC-MS/MS analysis. **(B)** In ChaFRADIC N-terminal peptides are separated based on their charge state by a first round of SCX chromatography. N-terminal peptides are mainly eluted in fractions corresponding to charge states $+2$ and $+3$. Each of those fractions is then d_3-acetylated to reduce the charge state of internal peptides, which are then separated from N-terminal peptides in a second round of SCX chromatography. **(C)** SILProNAQ is a simpler version of COFRADIC in which only partial enrichment of N-terminal peptides is achieved by a single round of SCX chromatography.

reduce both LC-MS/MS run times as well as initial sample requirements. However, they do not allow for the quantification of endogenously acetylated protein N-termini that do not contain a neighbouring lysine. Taking into account over 80% of human proteins are estimated to be N-terminally acetylated [27], the number of peptides in this pool that would be systematically

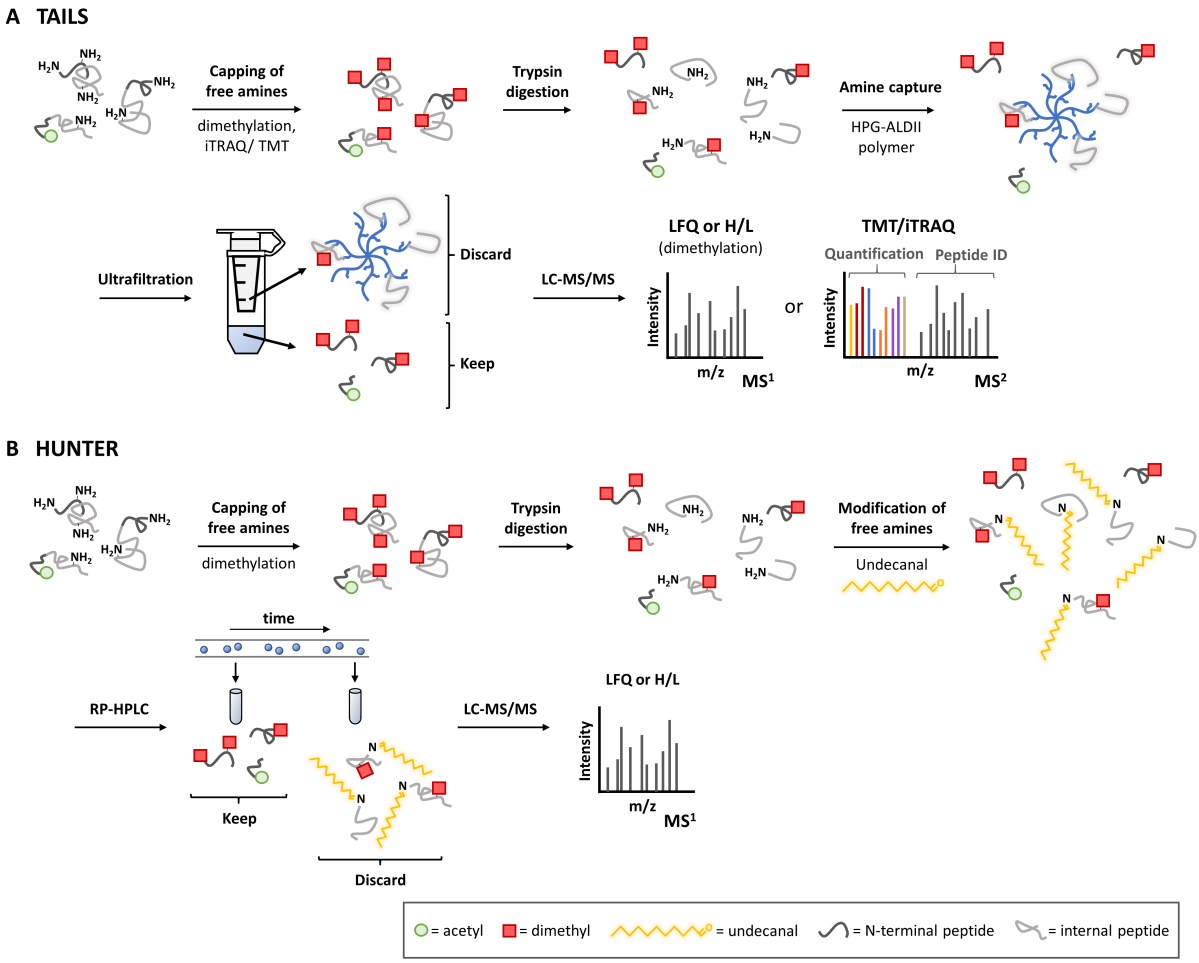


Figure 4.7: Main methods for N-terminal peptide enrichment through scavenging (TAILS) or labelling (HUNTER) free amines of internal peptides. (A) In TAILS free amines in N-termini and Lys are capped by either dimethylation or isobaric labelling (iTRAQ/TMT) before trypsin digestion. Exposed amines on internal peptides are then reacted with a high molecular weight aldehyde-derivatised polymer (HPG-ALDII) and separated out by ultrafiltration before LC-MS/MS analysis. **(B)** In HUNTER free amines are labelled by dimethylation before trypsin digestion. Newly formed primary amines of internal peptides are then reacted with a saturated fatty aldehyde, undecanal, which then allows for separation through RP-HPLC.

missed in this analysis could be significant. Alternatively, for the quantification of the N-acetylome, acetylation with isotopically labelled *N*-acetoxy-d_3-succinimide allows for accurate quantification of the ratios of endogenous to chemical acetylation as shown by the SILProNAQ approach [149, 382, 383].

Two main things were considered when deciding which method to implement for the study of N-terminal iMet retention: first, accessibility to necessary equipment or specific reagents; and second, overall cost of the experiment. Given that no similar experiments had been performed in the group, optimisation of the procedure was inevitable. In addition, the cost of reagents and machine time would increase exponentially dependent on the level of optimisation required. From all the methods listed, the two most successful ones have been COFRADIC and TAILS, with both yielding samples highly enriched in N-terminal peptides (90% for COFRADIC when performed with the additional SCX step [380] and >95% for TAILS [375]). Due to the

required depth of fractionation and consequent LC-MS/MS run time, COFRADIC was not deemed appropriate. TAILS, on the other hand, required high amounts of sample (recommended minimum amount of sample: 1 mg per sample) unless willing to multiplex using TMT (recommended minimum amount of sample: 200 µg per sample). In this sense, TMT labelling was considered an attractive approach due to the fact that reagent usage can be optimised to reduce reagent-related costs. In TAILS, however, the labelling step is performed at the protein level, where it is significantly less efficient. According to the Overall lab protocol [386], each 200 µg sample required a whole 0.8 mg vial, which increased the price per experiment significantly. On the other hand, if one employed dimethylation for the initial amine-capping step, the polymer costs became a problem ($500 per 20 mg batch + $200 in delivery costs, (http://www.flintbox.com/public/project/1948/), since it needs to be added in a two-fold excess (w/w) as compared to the protein sample (1 mg recommended per sample in the case of dimethylation) [386].

Reducing sample complexity is crucial for the identification of new modifications or proteolysis sites due to the data-analysis difficulties posed by the poorly defined search space. This, however, is not such an issue for the quantification of iMet retention, where both possibilities (iMet retained or iMet cleaved) are well defined. Moreover, all methods described above systematically discard an important fraction of *N*-acetylated peptides (except for those that used heavy acetylation as the quantification method). Taken together, this suggests that there is still room for improvement for methods to measure N-terminal modifications. Therefore new strategies specifically tailored to detect iMet excision where sought.

4.5.3 Set-up of the proteomics workflow

Azidohomoalanine (AHA) (Figure 4.8A) is a methionine analogue that has been widely used in the study of proteome dynamics [363, 390–392]. It is recognised by methionyl-tRNA synthetase and so it is incorporated into $tRNA^{Met}$ as if it was methionine, albeit at lower efficiency [390]. Thus, by addition of AHA to the cell culture medium in place of Met, AHA is taken up by the cells and incorporated into newly synthesised proteins as a surrogate for Met. This is often referred to as 'metabolic incorporation' or 'metabolic tagging'. Importantly, AHA is a non-canonical amino acid bearing an azido group (R-N_3) that is amenable to click chemistry through copper-catalysed alkyne-azide cycloaddition (CuAAC). AHA therefore allows for specific enrichment of newly synthesised methionine-containing proteins or peptides upon ligation to alkyne-bearing capture reagents after cell lysis. It was hypothesised that AHA could be incorporated into newly synthesised proteins both in place of internal methionines as well as in the place of iMet and be cleaved by cellular MetAPs as if it was the endogenous substrate (Figure 4.8B). This way, by comparing MetAP2-inhibited to control samples it should be possible to monitor the appearance of new iMet retention events in response to MetAP2 inhibition by detecting an increase in the levels of enrichment of iMet-retained N-terminal peptides (Figure 4.8C). Since methionine is an abundant amino acid in proteins, this new

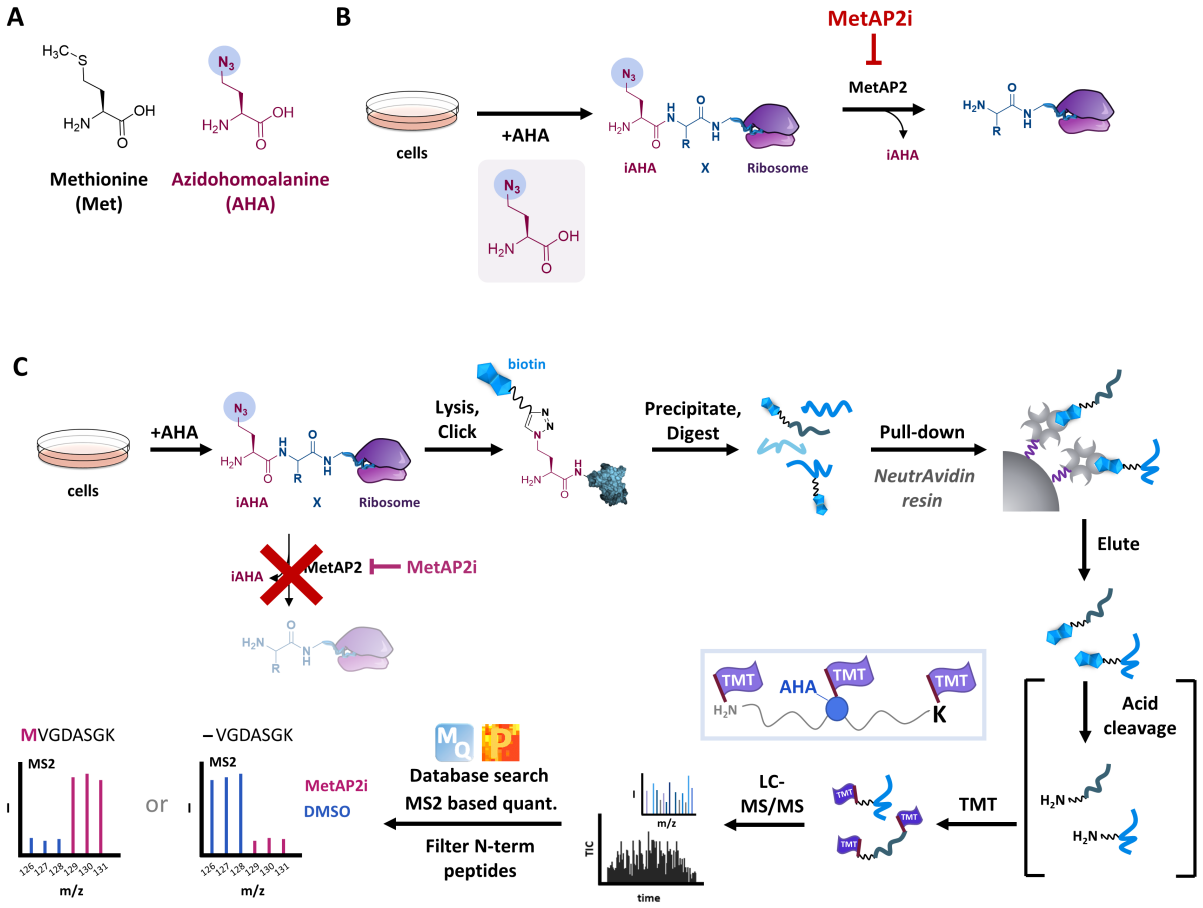


Figure 4.8: New proteomics workflow for the study of iMet retention based on metabolic tagging with AHA. (A) Structures of methionine (left) and the methionine surrogate non-canonical amino acid azidohomoalanine (AHA)(right). **(B)** Metabolic incorporation of AHA can serve as a marker for iMet cleavage events by MetAP2. **(C)** Schematic representation of the final proteomics workflow after the optimisation steps. Cells are metabolically tagged with AHA in the absence of methionine and AHA is incorporated into newly synthesised proteins in the place of Met, including the iMet position. Upon lysis, AHA is chemically ligated to a biotinylated alkyne capture reagent by click chemistry (CuAAC). Biotinylated proteins are then precipitated and digested with trypsin before specific enrichment of AHA containing peptides in NeutrAvidin beads. Bound peptides are eluted from the beads in an acidic, high acetonitrile solution that breaks the interaction between biotin and NeutrAvidin before acid cleavage of the capture reagent to reveal a novel TMT-labelling site. The acid-cleavage step appears in brackets as it was not part of the original protocol but was incorporated during the optimisation steps. Peptides are then labelled with TMT, fractionated and analysed by LC-MS/MS.

workflow is based on peptide-level pull down, which allows for specific enrichment of iMet-retained N-terminal peptides, together with other internal peptides containing Met that can serve as internal controls of general protein levels. As the analysis is performed at the peptide-level, the use of isobaric labels such as TMT presented an important advantage, allowing for the detection of the same peptides across the different conditions tested, such as compound type and concentration.

The first experiment using this workflow used a commercially available capture reagent (Yn-PEG4-biotin, Sigma Aldrich) (Figure 4.9B). Since previous reports showed proteome-wide effects of longer AHA labelling pulses on bacterial cells [390], a short 2 h AHA pulse was

selected following an arbitrarily chosen pre-incubation with MetAP2i of 4 h (Figure 4.9A). This preliminary experiment showed promising initial results despite not yet being optimised. The two structurally diverse reversible MetAP2 inhibitors used, the Abbott (**1**) and GSK (**2**) compounds, revealed an almost identical pattern across 23 ANOVA-significant N-terminal peptides, all of which retained iMet (Figure 4.9C-D). Moreover, comparison of the N-terminal peptide enrichment fold change to the internal peptide enrichment fold change (where another peptide was identified for the same protein) revealed N-terminal specific enrichment changes (Figure 4.9D), indicating that overall protein levels were not affected and that the apparent enrichment was due to changes in iMet retention. However, the N-terminome coverage was not optimal, with only 197 N-terminal peptides identified across all samples. Importantly, similar to other methods described above (cf. section 4.5.2), this quantification method does not quantify naturally acetylated N-terminal peptides that end in Arg after trypsin digestion, due to those peptides lacking free amines for TMT labelling.

4.5.4 Optimisation of the proteomics workflow

4.5.4.1 Repurposing a previously reported trifunctional capture reagent

A switch to a LysC endopeptidase was first considered to address the issue of missing half the acetylated N-terminal peptides for quantification. Using LysC instead of trypsin should ensure all N-terminal peptides end with a Lys residue, making them all amenable to TMT tagging. However, higher price and lower efficiency make digestion with LysC nearly 50 times more expensive than with trypsin. Thus, it was not considered appropriate for a workflow that employed peptide-level enrichment, where considerable amounts of material (500-1000 µg) need to be digested before enrichment. Therefore, new ways of artificially introducing a TMT labelling site in *N*-acetylated N-terminal peptides ending in Arg were sought. A previous study had reported the use of a trifunctional capture reagent that beared an alkyne tag for click chemistry, a biotin handle for affinity enrichment and an acid-cleavable (AC) moiety within the linker between the two [393] (Figure 4.10C). Of note, once cleaved with acid, the newly formed adduct revealed an exposed primary amine, amenable for TMT tagging (Figure 4.10D). With the help of Dr. Christelle Soudy, who synthesised this compound, the new acid-cleavable capture reagent (Yn-AC-Biotin) was compared to the commercially available Yn-PEG4-biotin capture reagent (Figure 4.10E-G). For this purpose, proteome samples of cells that had been treated with MetAP2i and metabolically tagged with AHA were used, which were then ligated to either Yn-PEG4-biotin or the acid cleavable Yn-AC-biotin before digestion with trypsin. After enrichment of biotinylated peptides with NeutrAvidin beads, the Yn-AC-biotin capture reagent was cleaved using concentrated TFA to expose a new amine for every AHA incorporation event. After TMT-labelling, the peptide samples were analysed by LC-MS/MS. TMT labelling efficiency analysis of samples ligated to Yn-AC-biotin revealed a marked preference for TMT reagents to react with the newly exposed amine in the acid

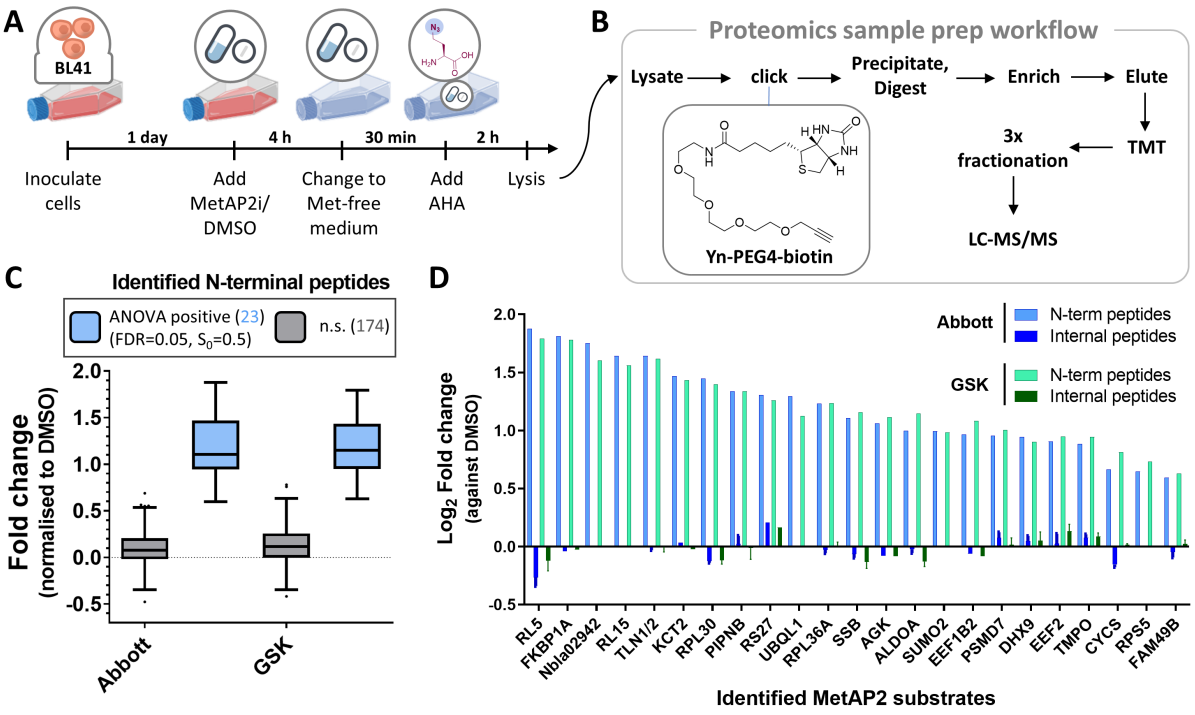
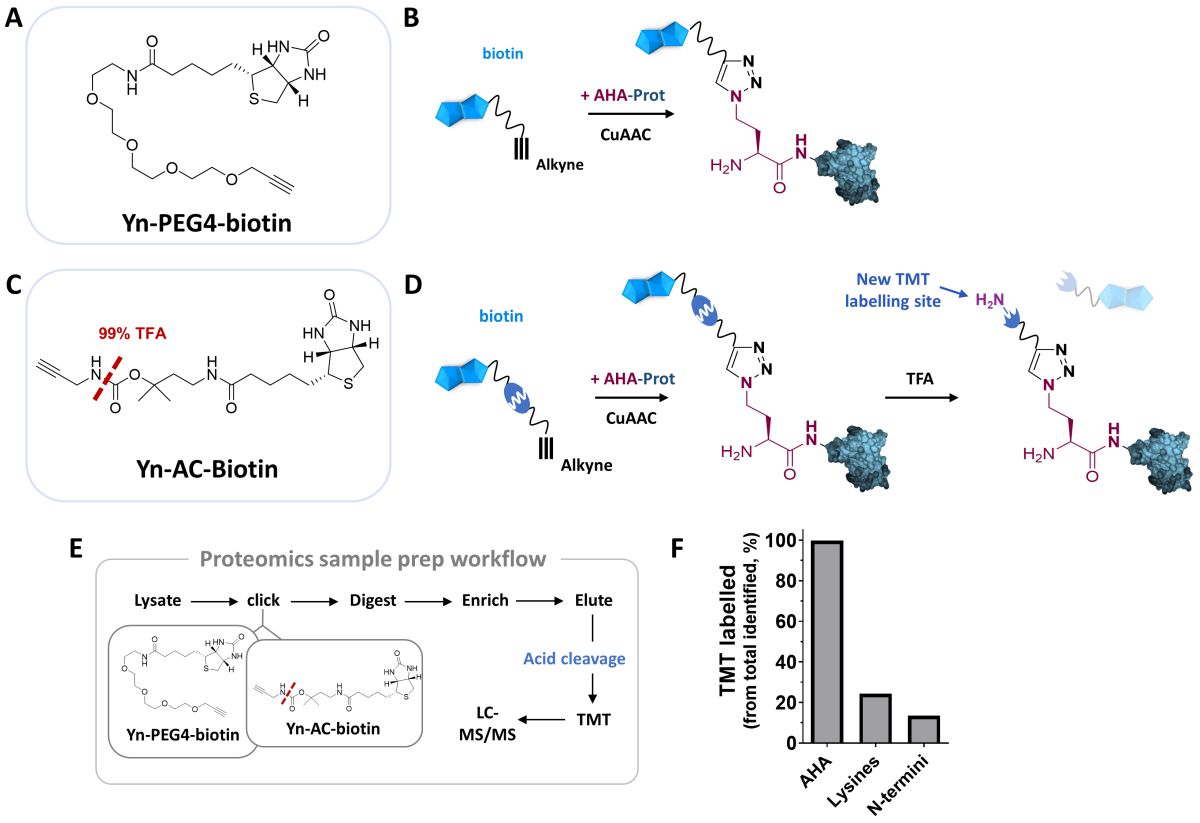


Figure 4.9: Preliminary experiment using AHA metabolic tagging shows promising results and highlights routes for improvement. (A) BL41 cells were allowed to grow for one day before pre-incubation with MetAP2i or DMSO control for 4 h. The cell medium was then replaced by Met-free medium and allowed the cells to equilibrate before 2 mM AHA addition. After 2 h AHA pulse, cells were lysed and proteomes extracted. **(B)** Step by step proteomics workflow. AHA-containing proteins were clicked to Yn-PEG4-biotin, precipitated to remove excess click reagents and digested overnight with trypsin. AHA-containing peptides were then enriched in NeutrAvidin agarose beads, eluted and labelled with TMT, before 3-way fractionation and proteomics analysis. **(C)** Statistical analysis (ANOVA, FDR=0.05 and S_0=0.5) revealed 23 iMet-retaining N-terminal peptides that were significantly more enriched in both GSK and Abbott MetAP2i compared to control. **(D)** Barplot of the 23 N-terminal peptides that showed significant enrichment, ordered by detected fold change and compared to other Met-containing peptides in the same protein (if detected).

cleavable capture reagent than with other amines in proteins such as in Lysines or N-termini (Figure 4.10F), leaving Lys and N-termini only partially labelled. This was an issue for the searches as it required Lys and N-terminal TMT labels to be selected as variable modifications, increasing the search times considerably and potentially influencing the quantification results. To avoid this, it was decided to increase the TMT-to-peptide ratio in future experiments 3-fold (from 1/10 to 1/3 of a 0.8 mg TMT reagent vial per sample). One positive of using this new capture reagent is that it boosted the number of identifications in comparison to the commercially available Yn-PEG4-Biotin (Figure 4.10G).

4.5.4.2 Assessing the sensitivity of the assay

With the workflow in place, the next step was to study if this assay was sensitive enough to detect the dose-dependent response in iMet retention upon MetAP2 inhibition. Having observed in Incucyte experiments that the MetAP2i effect did not induce a full phenotype



G

Reagent	Quantification (MS^2)	Fixed modifications	Variable modifications	Total IDs	N-term peptides
Yn-PEG4-biotin	TMT	Met > AHA-Yn-PEG4-Biotin Carbamidomethyl (Cys)	Acetylation (N-term)	1,767	360
Yn-Cleav-Biotin	AHA-Cleav-TMT	Carbamidomethyl (Cys)	Acetylation (N-term) TMT10plex (Lys) TMT10plex (N-term)	4,045	489

Figure 4.10: An acid-cleavable capture reagent allows for the identification and quantification of all the main types of N-terminal peptides. (A) Structure of Yn-PEG4-Biotin. **(B)** Click ligation reaction of Yn-PEG4-biotin with a protein containing AHA. **(C)** Structure of the acid cleavable (AC) capture reagent, Yn-AC-Biotin, **(D)** Click ligation reaction of Yn-AC-biotin with a protein containing AHA and subsequent acid cleavage step to reveal an exposed primary amine, amenable for TMT labelling. **(E)** Schematic representation of the proteomics sample preparation workflow with the additional acid cleavage step (highlighted in blue). **(F)** Testing Yn-AC-Biotin in of samples treated with MetAP2i and metabolically tagged with AHA revealed preferential reaction of TMT reagents with the newly exposed amine at the capture reagent and incomplete labelling of Lys and N-termini. **(G)** Summary of the search performed for the comparison of the commercially available Yn-PEG4-Biotin capture reagent to the acid cleavable Yn-AC-Biotin capture reagent.

until 48 h, the MetAP2i incubation times were increased to 24 h to ensure maximal effect of MetAP2i's on iMet retention but before the effects on proliferation were evident. It was also hypothesised that increasing AHA pulse time could be beneficial as it could provide longer times for accumulation of iMet-retained N-terminal peptides, thereby increasing the fold-change difference between MetAP2i-treated and control samples. A previous report using AHA had reported detrimental proteome-wide effects of AHA labelling in the absence of Met for periods longer than 2 h in *E. coli* [390]. However, this issue was hypothesised to be due to Met-deprivation rather than AHA toxicity due to the nearly 400-fold decrease in $tRNA^{Met}$ loading

of AHA as compared to Met. Importantly, the issue was solved by re-introduction of small quantities of Met [390]. After checking that a pulsed 50:1 ratio of AHA:Met did not affect cell viability compared with AHA alone (data not shown), a proteomics experiment was set up to test the new workflow against a concentration gradient of both Tak1(**3**) and Tak2 (**4**) (Figure 4.11A). It should be noted that at this time the results from the Tak1 and Tak2 Incucyte proliferation assay (Figure 4.4) and CETSA target engagement assay (Figure 4.5) were not available. Following the pipeline developed above for the acid cleavable capture reagent and increasing the TMT-to-peptide ratio 3-fold as compared to the previous experiment, this experiment identified ca. 600 N-terminal peptides (acetylated and non-acetylated) with efficient TMT labelling in all sites (Lys, N-termini and Yn-AC-Biotin adduct). While the total number of identified and quantified N-terminal peptides had increased 3-fold, only half the number of significantly changing proteins were detected (ANOVA, FDR=0.05 and S_0=0.5) (Figure 4.11D). The perceived lack of sensitivity was postulated to be due to the fact that EA.hy926 cells did not display any phenotypic response (Figure 4.2 and accompanying text) or the fact that the Takeda compounds showed both reduced potency (Figure 4.4), as well as no target engagement (Figure 4.5) in responding cells (HT1080). Thus, the procedure was repeated with HT1080 cells and well-characterised compounds Abbott (**1**) and Merck+ (**5**) (Figure 4.11B). Surprisingly, this yielded even lower sensitivity, with only 7 significantly changing proteins from the ca. 870 N-terminal peptides identified and much smaller fold-changes that distinguished MetAP2i-treated from DMSO controls (Figure 4.11E).

4.5.4.3 Optimising the labelling conditions

Reluctant to abandon these efforts yet, the experimental conditions used in all previous experiments were systematically reviewed to understand how the signal-to-noise ratio had become so poor. Both the AHA and MetAP2 incubation times were found to have been consistently increased from one experiment to the next. Knowing from the literature and the Incucyte cell proliferation studies that MetAP2i halted cell proliferation on these cells, it was hypothesised this could be affecting general protein synthesis. Cell proliferation inhibition became apparent after 24h of MetAPi incubation, with the following 24h allowing quantitative resolution of anti-proliferative properties at different inhibitor concentrations (Figure 4.4E). Since AHA incorporation relies on new protein synthesis taking place, and most protein turnover rates are less than 48 h, this could in fact be reducing the accumulation of newly synthesised proteins with iMet retained. Proteomics experiments are labour intensive and there are long waiting times for machine usage. Hence, in-gel fluorescence was selected to test this hypothesis (Figure 4.12). AHA was used to monitor global protein synthesis, which was then clicked to a fluorphore containing capture reagent, Yn-TAMRA (YnT) for in-gel fluorescence (IGF) analysis. A 24 h incubation with a panel of MetAP2i only resulted in a subtle effect on overall protein synthesis (Figures 4.12C-D and F-G). However, at longer exposure times (48 h) treatment with GSK MetAP2i did show a clear effect on protein synthesis 4.12E and H).

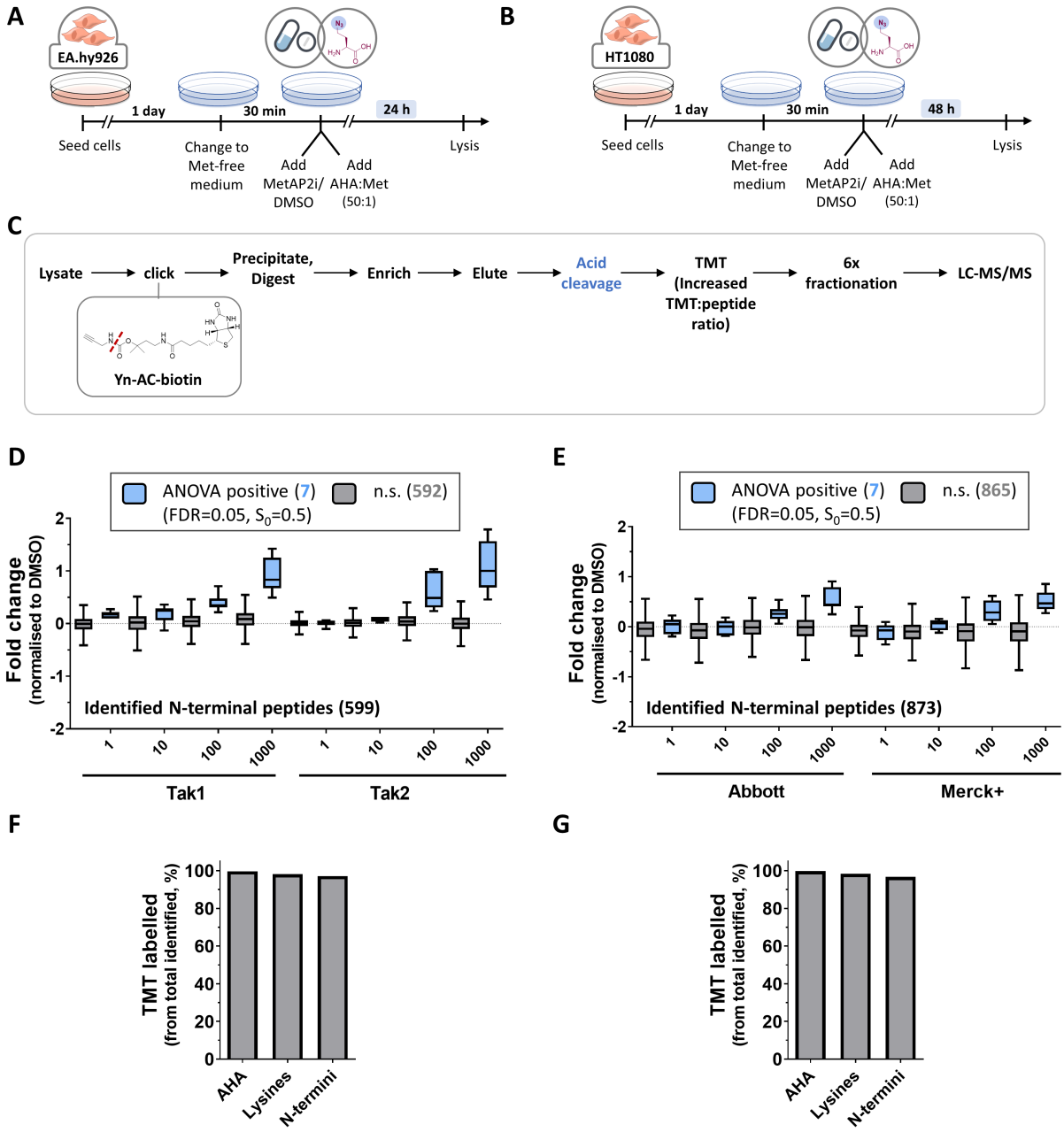


Figure 4.11: Concentration range MetAP2i proteomics experiments. (A) EA.hy926 cells were allowed to grow for one day before changing the medium to Met-free medium. After a 30 min equilibration period, a concentration series of MetAP2i or DMSO control was added, followed by AHA:Met (50:1). Cells were incubated in the presence of MetAP2i and AHA for 24 h before lysis. **(B)** Similar experimental procedure as in (A) but in HT1080 cells and with an incubation of MetAP2i and AHA of 48 h. **(C)** Step by step sample preparation procedure for experiments described in (A-B). **(D)** Statistical analysis of experiment described in (A) revealed 7 candidate MetAP2 substrates of the ca. 600 N-terminal peptides identified. **(E)** Statistical analysis of experiment described in (B) revealed 7 candidate MetAP2 substrates of the ca. 870 N-terminal peptides identified. **(F)** TMT labelling efficiency of AHA (Yn-AC adduct), Lys and N-termini of experiments shown in (A) and (D). **(G)** TMT labelling efficiency of AHA (Yn-AC adduct), Lys and N-termini of experiments shown in (B) and (E).

MetAP2 has been shown to have a non-enzymatic role in maintaining general protein synthesis through its protection of eIF2α phosphorylation (POEP) activity [137, 142]. However, this is

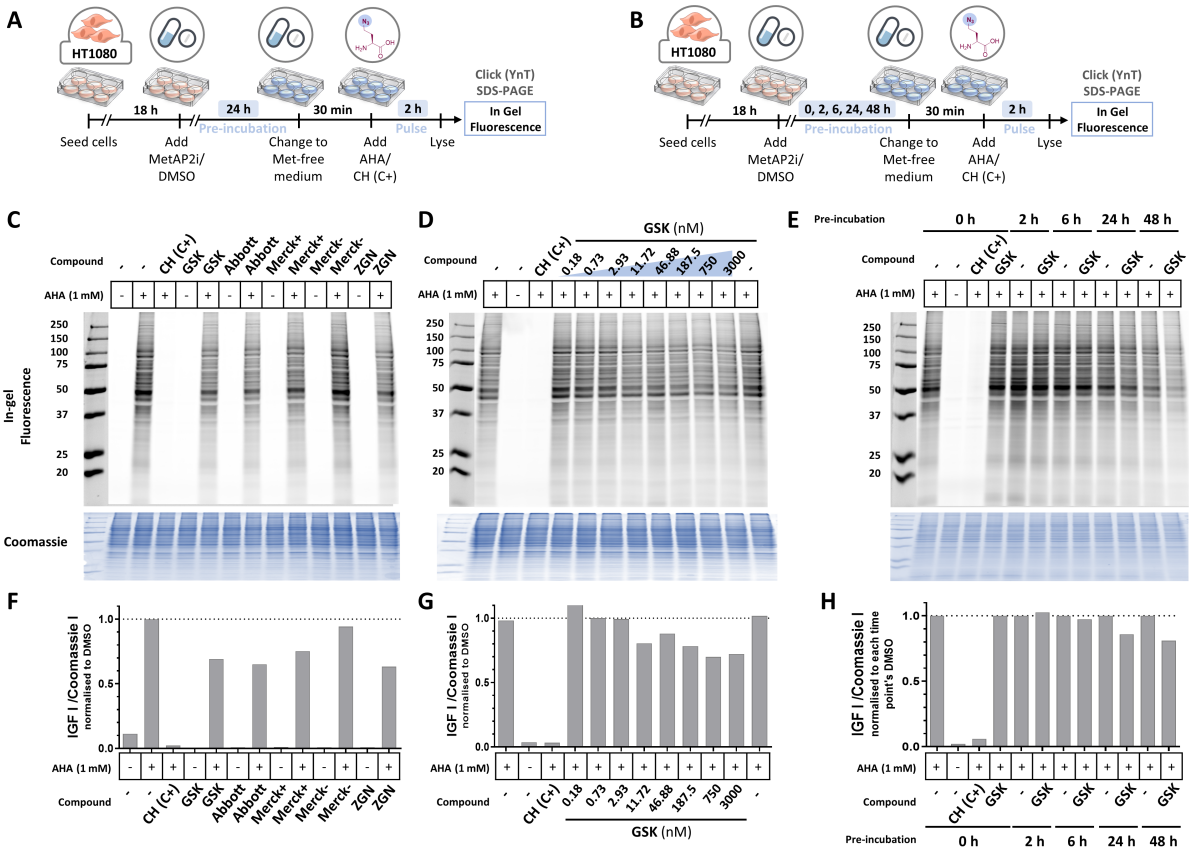


Figure 4.12: In-gel fluorescence analysis of the effect of MetAP2 inhibition on general protein synthesis. (A) Experimental procedures of gels shown in (C-D) and barplots (F-G). Cells were seeded in 6-well plates and allowed to grow overnight before addition of MetAP2i or DMSO control. After 24 h under MetAP2i treatment, cells were equilibrated in Met-free medium, and AHA added. Cycloheximide (CH) was used as the positive control (C+). Due to the high potency of CH in completely blocking protein synthesis, which is cytotoxic upon long exposure, it was added at the same time as AHA and only incubated for 2 h before lysis. All MetAP2i were added to a final 250 nM concentration except for ZGN (1 nM final concentration). **(B)** Experimental procedures of gel shown in (E) and barplot (H). Cells were seeded in 6-well plates and allowed to grow overnight before addition of MetAP2i or DMSO control. Cells were incubated with MetAP2i for a series of exposures of up to 48 h, before equilibration in Met-free medium and AHA addition. CH, cycloheximide, positive control (C+). **(C-E)** In-gel fluorescence (IGF) analyses of AHA labelled samples ligated to Yn-TAMRA (YnT) capture reagent, with a panel of MetAP2i in a single concentration (all 250 nM, except ZGN (1nM))(C), a concentration range of GSK MetAP2i (D) or GSK MetAP2i incubation time series (E). **(F-H)** Barplots showing the IGF-to-coomassie signal for each condition, normalised to DMSO control. In (H), the IGF/Coomassie signal of each lane was normalised to the DMSO at each time point.

mediated by the N-terminal domain of MetAP2, which is not involved in its catalytic activity, and had been shown to be unaffected by binding of MetAP2i to the MetAP2 active site [176]. On the other hand, the reduction of protein synthesis could be mediated solely by the inhibition of proliferation (Figure 4.4E) and reported G0/1 arrest [168]. Driven by these considerations and the promising results obtained in the first proteomics experiment (Figure 4.9 and accompanying text), the MetAP2 incubation time for these experiments was from there on kept to a minimum and always $<$24 h.

4.5.5 Combining AHA labelling and a range of diverse MetAP2 inhibitors reveals >70 novel MetAP2 substrates

An experiment was then designed to test if reverting back to low MetAP2i and AHA incubation times would solve the low signal-to-noise issues. Three MetAP2i's were selected, Abbott (**1**), GSK (**2**) and Merck+ (**5**), and its less active enantiomer Merck- (**6**), which together with the DMSO control could be multiplexed into a single TMT10plex in biological duplicates. Based on previous in-gel fluorescence results indicating that shorter MetAP2i incubation periods showed increased sensitivity (Figure 4.12), HT1080 cells were subjected to a short 2 h pulse of both MetAP2i and AHA, with no pre-incubation period (Figure 4.13A).

This experiment identified 693 N-terminal peptides, from which 43 (corresponding to 41 unique proteins) significantly changed in enrichment levels upon MetAP2i addition (ANOVA, FDR=0.05, S_0=0.5) (Figure fig:C4fig11 and Supplementary Table S5). From these, 38 showed a significant increase in enrichment upon MetAP2i and their sequence showed their iMet retained; whereas the other 5 N-terminal peptides showed a statistical significant decrease and had their iMet cleaved (Figure 4.13C-D). Importantly, all peptides that showed any significant changes shared the same trend across all MetAP2i tested and decreased to a modest fold-change enrichment change for the less active Merck- compound (Figure 4.13E), consistent with our previous EC_{50} calculations (Figure 4.4E). From all MetAP2i tested, Abbott exerted the highest response, followed by Merck+ and GSK, a trend consistently maintained across both iMet retained and iMet cleaved peptides (Figure 4.13E). While the disappearance of iMet excised peptides might seem counter-intuitive, some proteins may undergo secondary modifications or become unstable upon forced iMet retention, thus preventing the identification of their retained iMet by shotgun proteomics.

With appropriate incubation times and sample preparation procedures finally established, attention turned to optimising coverage of MetAP2 substrates. For this two cell lines were selected, HT1080 and A549, both of which had showed good response to MetAP2 inhibition in our proliferation assay (Figure 4.4 and Appendix Figure A.5). In order to compare proteome-wide iMet retention patterns across the two major categories of MetAP2i, reversible and irreversible MetAP2i (for an overview on this please refer to Introduction section 1.3.4), three compounds were chosen: the active/inactive reversible MetAP2i pair, Merck+ (**5**) and Merck- (**6**), and the irreversible MetAP2i ZGN (**7**). An extra protease was also added into the workflow, chymotrypsin, to ensure increased coverage of N-terminal peptides. Chymotrypsin was selected based on its orthogonal digestion specificity to trypsin and the fact that several previously reported MetAP2 substrates showed N-terminal sequences compatible with chymotrypsin, which were not detectable with trypsin: (first underlined residue corresponds to tryspin specificity, second to chymotrypsin) GAPDH [181, 200, 370], N-term sequence: MG<u>K</u>VKVGVNG<u>F</u>; TXN [370], N-term sequence: MV<u>K</u>QIESKTA<u>F</u>; SH3BGRL [370], N-term sequence: MVI<u>R</u>V<u>Y</u>; 14-3-3γ/YWHAG [183, 193, 210], N-terminal sequence: MVD<u>R</u>EQ<u>L</u>.

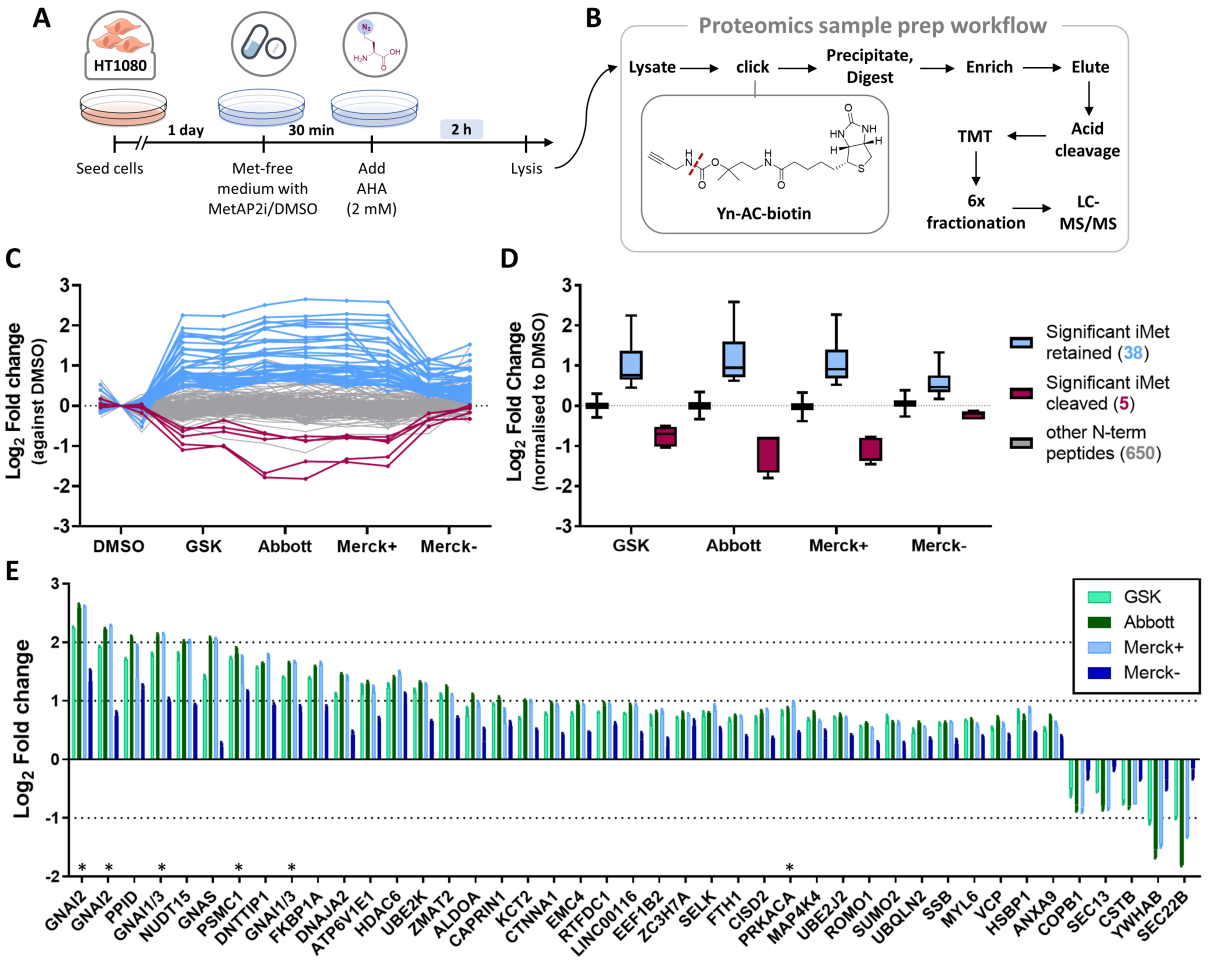


Figure 4.13: Shorter MetAP2i and AHA pulses increase the signal-to-noise ratio and allow for identification of 43 MetAP2 substrates. (A) Cells were allowed to grow for one day before equilibrating them to Met-free medium. MetAP2i/DMSO were then added, followed by 2 mM AHA and pulsed for 2 h in the absence of methionine. **(B)** Step by step proteomics sample preparation workflow. **(C-D)** Profile and box plots summarising the results of the statistical analysis of the experiment (ANOVA, FDR=0.05, S_0=0.5). Each line in the profile plot (C) represents an N-terminal peptide. Peptides showing a significant increase in enrichment are highlighted in blue (all show iMet-retained) and peptides showing a significant decrease in enrichment in magenta (all show iMet excised). **(D)** Barplot representing the profiles of each of the 43 significantly changing N-terminal peptides corresponding to 41 unique proteins. Known NMT substrates are annotated with an asterisk (*).

Together, this experiment identified >1000 unique N-termini in each cell line (1197 in HT1080 and 1061 in A549) (Figure 4.14C and Supplementary Tables S6 and S7). Samples digested with trypsin consistently resulted in 3-to-4 fold more identifications than samples digested with chymotrypsin. Statistical analysis of each dataset revealed 38 hits for HT1080-trypsin, 16 for HT1080-chymotrypsin, 32 for A549-trypsin and 10 for A549-chymotrypsin (Figure 4.14 and Appendix Figure A.11C-D). Together, this resulted in the discovery of 63 MetAP2 substrate candidates: 50 significantly more enriched in MetAP2i-treated samples and showing iMet retained, and 13 with iMet cleaved and showing a reduction in AHA-enrichment (Figure 4.14E). Of note, overlap was much higher across cell lines than across proteases (Figure 4.14E and Appendix Figure A.10), thus supporting the use of

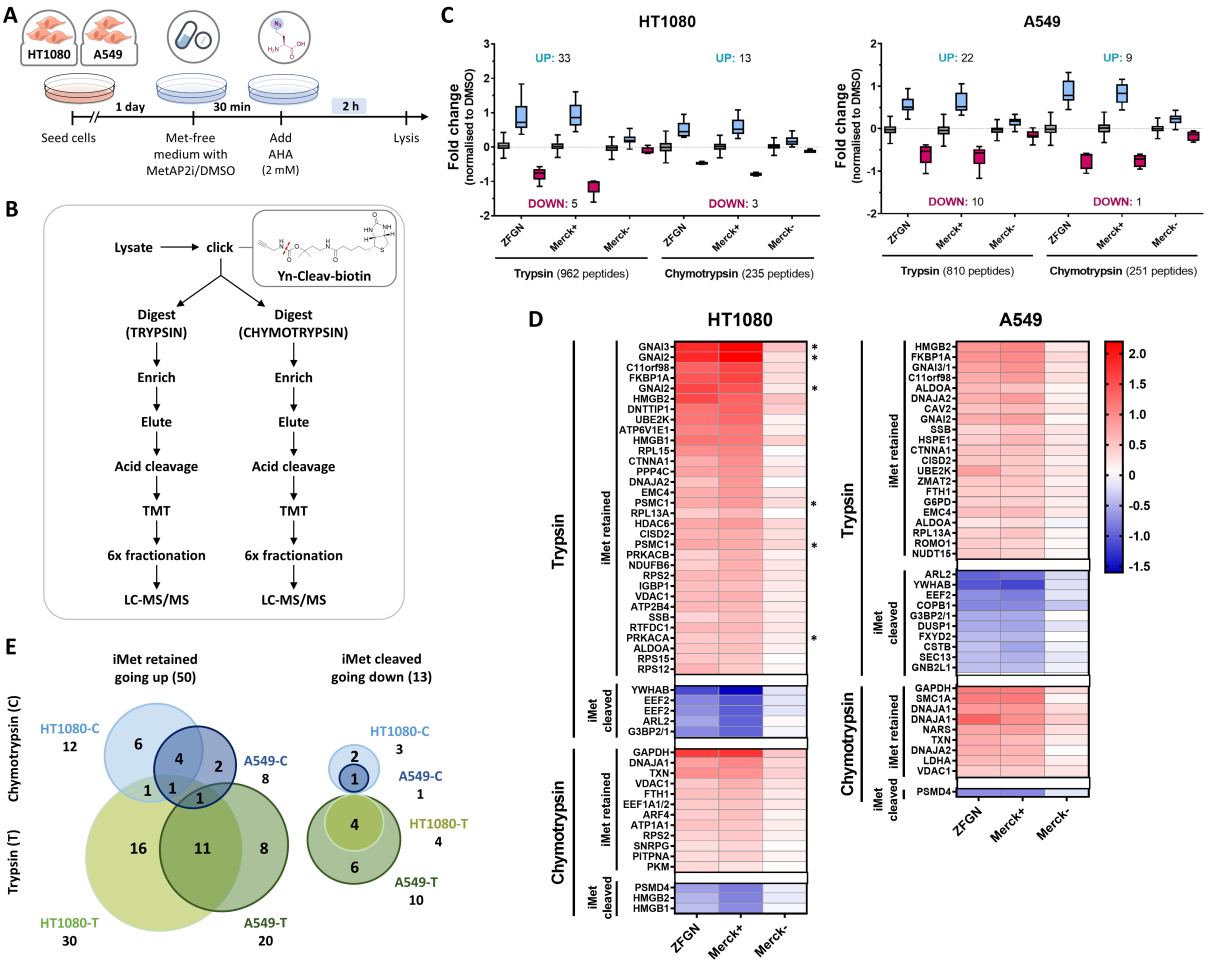


Figure 4.14: Increasing the coverage to two cell lines and two proteases identifies 63 MetAP2 substrates, expanding the list of candidates to 78, including 4 previously reported susbtrates. (A) Cells were allowed to grow for a day before being equilibrated in Met-free medium and pulsed with MetAP2i/DMSO and 2 mM AHA for 2 h. **(B)** Schematic representation of the proteomics sample preparation workflow for each of the samples. **(C)** Boxplots summarising the results of the statistical analysis (ANOVA, FDR=0.05, S_0=0.5), performed for each dataset separately (4 datasets: HT1080-Trypsin, HT1080-Chymotrypsin, A549-Trypsin, A549-Chymotrypsin). Blue boxes represent iMet-retaining N-terminal peptides showing increase in enrichment upon exposure to MetAP2i, whereas boxes in magenta represent iMet-cleaved N-terminal peptides with reduced levels of enrichment upon MetAP2 inhibition. **(D)** Heatmaps representing the fold change against DMSO of all significantly changing peptides across the three compounds. Duplicate N-terminal peptides arose from protease missed cleavages. Scale bar: double-gradient colour map, blue represent decreased enrichment as compared to control while red represents increased enrichment. Known NMT substrates are annotated with an asterisk (*). **(E)** Venn diagrams representing the overlap across hits from each of the four datasets for either iMet-retained (increased enrichment upon MetAP2 inhibition) or iMet-cleaved (reduced enrichment upon MetAP2 inhibition). Numbers represent protein MetAP2 candidates, where non unique N-terminal peptides arising from missed cleavages have been combined into a single protein hit. C = chymotrypsin; T = trypsin.

a second protease, despite the considerable addition of sample preparation and machine run times involved. Moreover, this validated several substrates previously reported in the literature, including GAPDH [181, 200, 370], TXN (also known as TRX1) [370], 14-3-3β/YWHAB [193] and EEF2 [370] (Figure 4.14D and Appendix Figure A.12). The underlying mechanism of MetAP2 inhibition, reversible (exemplified by Merck+ (**5**)) or covalent (exemplified

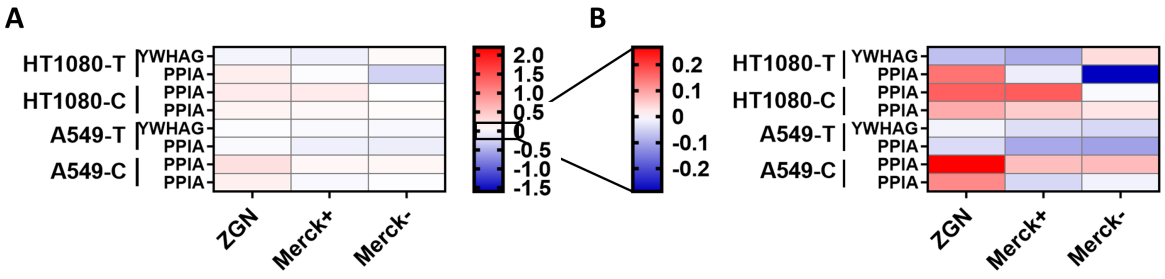


Figure 4.15: 14-3-3γ/YWHAG and CYPA/PPIA N-terminal peptides do not show significant enrichment upon addition of MetAP2i. (A) Heatmap showing the N-terminal peptide levels of 14-3-3γ/YWHAG and CYPA/PPIA across studied conditions using the same scale bar settings as in Figure 4.14. **(B)** Same data as in (A), with reduced scale to distinguish differences across conditions.

by ZGN (**7**)), did not make any notable difference across all the N-termini detected and overlap in N-terminal peptides responding to MetAP2 inhibition between the two compounds was 100%. The few differences in enrichment fold changes may have been due to target engagement differences at the given dosages (250 nM for Merck+ and 1 nM for ZGN), and were in all cases minimal (Figure 4.14D and Appendix Figures A.12). Further, the inactive enantiomer Merck- showed strongly reduced enrichment differences in all cases (Figure 4.14D and Appendix Figure A.12), indicating the robustness of the assay.

Unfortunately, from the MetAP2 substrates already listed in the literature, there were two substrates that were not detected: SH3BGRL [370] and Rab37, although the latter is expressed at extremely low levels and even the authors of the paper that identify it could not detect it at endogenous levels [237]. Moreover, there were two instances were the N-terminal peptide of proteins previously reported as being MetAP2-specific substrates, although identified, did not show any significant changes in enrichment levels. This was the case of 14-3-3γ/YWHAG and CYPA/PPIA (Figure4.15). Amongst our hits, however, 14-3-3β/YWHAB and PPID were identified instead (Figure 4.14D and Apependix Table A.1), which share 27.8% and 74.8% identity with the aforementioned proteins, respectively (calculated with EMBOSS Needle global pairwise alignment online tool).

Combination of the hits from this experiment with those from Figure 4.13 expanded the total list of identified MetAP2 substrates in cells to 78, of which 74 are reported here for the first time (List in Appendix Table A.1).

4.5.6 Identified MetAP2 substrates display canonical MetAP substrate preference and are involved in a variety of important cellular processes

The combined list of 78 MetAP2 substrate candidates identified from the previous two sets of experiments (Appendix Table A.1) was then used to study MetAP2 substrate specificity. Analysing the amino acid in P2 position revealed a preference the for small, uncharged amino

acids such as Ala, Gly, Val, Thr, Ser, Pro, Met and Cys (Figure 4.16A), fully overlaping with the canonical consensus sequence for MetAP cleavage (cf. Introduction section 1.3.2.4 and Figure 1.6)[131, 146, 147]. MetAP2 showed highest processing of proteins having an Ala or Gly at the P2 position, followed by Val and Thr, which together constituted 80% of the substrates identified (Figure 4.16A). This coincided with previous reports on MetAP2 substrate specificity, where MetAP2 had been proposed to be the main enzyme responsible for iMet removal of proteins containing Thr, Val or Gly in the P2 position [147, 149], although this study revealed a higher proportion of substrates bearing a Gly in the P2 position. To study if there was any effect of the following amino acid residues on MetAP2 substrate recognition, a sequence logo was built for positions P2-11 (first 10 amino acids after iMet cleavage) using IceLogo [365] (Figure 4.16B). This was constructed by comparing the sequences of the 78 MetAP2 substrate candidates to the same amino acid positions of the curated Swissprot human proteome (UniprotKB, accessed May 2019, 42425 entries). As shown in Chapter 3 Ala is the most common amino acid in the proteome at the P2 position (Figure 3.4), which reduced its enrichment from the logo, leaving Gly, Val and Thr as the preferred P2 positions by MetAP2 (Figure 4.16B). Positions P3-11 revealed preference for Lys residues in positions 3, 5 and 7-8, while it disfavoured it at positions 4 and 6 (Figure 4.16B). Moreover, the analysis showed a significant preference against Arg in positions P3-6, indicating this slightly bulkier amino acid residue might not fit in the pockets for P4 and P6, where Lys was accepted.

Using PANTHER [394], MetAP2 substrates were annotated for Gene Ontology (GO) Molecular Function (GOMF) terms and classified according to their protein class (Figure 4.16C and Appendix Table A.2). This analysis revealed proteins belonging to more than 14 different protein classes, with strong prevalence of proteins involved in translation and protein folding. These included 5 ribosomal proteins (RPS2, RPS12, RPS15, RPL13A and RPL15), 3 translation elongation factors (EEF2, EEF1A1 and EEF1B2) and 6 proteins with chaperone activity (DNAJA1, DNAJA2, FKBP1A, HSPE1, IGBP1 and PPID). Moreover, several of the MetAP2 substrates were found to be involved in ubiquitin or ubiquitin-like protein tagging (UBE2J2, UBE2K, UBQLN2) and proteasomal degradation (PSMC1, PSMD4) (classified under the 'Protein-modifying enzyme' class) (Figure 4.16C). In addition, 13 of the identified proteins were involved in signalling pathways, including kinases/phosphatases (PRKACA, PRKACB, MAP4K4, PPP4C and DUSP1) and other signalling molecules (GNAI1, GNAI2, GNAS, ARF4, ARL2, CISD2, CSTB and EMC4), as well as 5 proteins involved in metabolism (ALDOA, GAPDH, G6PD, PKM and LDHA). Finally some proteins involved in redox homeostasis were also found (ROMO1, NDUFB6 and TXN), which could support the proposed role of MetAP2 in controlling cellular glutathione redox status [149]. gPROFILER [395] over-representation analysis was applied to pinpoint the main biological processes MetAP2 substrates might be involved in (performed over GO Biological Process (GOBP) annotations). This revealed a cluster of proteins involved in regulation of protein localisation and co-translational protein targeting, mainly mediated by the ribosomal proteins (Figure 4.16D). In addition, MetAP2 substrates were found to be involved in cellular responses to stress and oxidative damage,

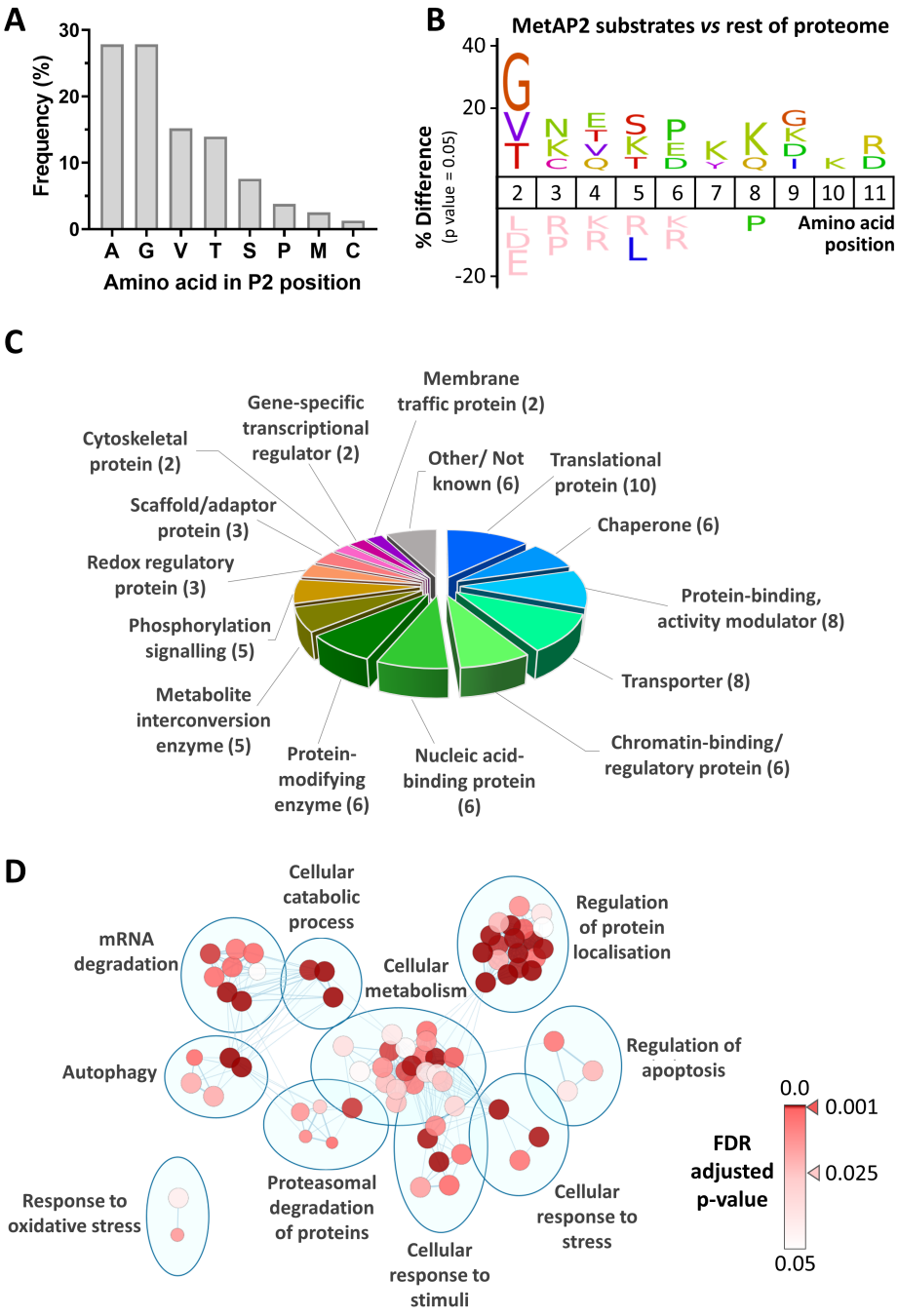


Figure 4.16: Sequence and functional analyses of identified MetAP2 substrates. (A) Frequencies of the amino acid residues at the P2 position of the identified 78 MetAP2 substrates. **(B)** Sequence logo of the first 10 amino acid positions of identified MetAP2 substrates as compared to the general distribution of amino acids for the same positions in the human proteome (Swissprot with isoforms, accessed May 2019, 42425 entries). Logo was generated using IceLogo [365]. **(C)** Classification of identified MetAP2 substrates according to their protein class. The initial classification was performed using PANTHER [394]. Unannotated proteins were then manually annotated based on the function listed in Uniprot [70] and their most prevalent GOMF annotations in QuickGO [396]. The full list of protein class annotations for the 78 MetAP2 substrates can be found in Appendix Table A.2. **(D)** Major biological roles of MetAP2 substrates, as analysed by functional over-representation analysis using GOBP annotations in gPROFILER [395]. Results were visualised using the EnrichmentMap App in Cytoscape [397]. Each node represents a statistically significant GO term and is colour-coded based on its significance level.

protein degradation by the proteasome, mRNA degradation, autophagy and apoptosis as well as in central cellular metabolism (Figure 4.16D).

4.6 Studying the effect of MetAP2 inhibitors on protein myristoylation: overlap between MetAP2 and NMT activities

Removal of methionine by methionine aminopeptidases precedes the addition of other N-terminal modifications, such as *N*-myristoylation of glycine-presenting protein substrates by *N*-myristoyltransferases (NMTs) [9]. *N*-myristoylation has been well established to play a crucial role in protein function and maintaining general protein homeostasis by regulating membrane-binding affinity, trafficking, binding partners and protein stability [9, 277–279] (cf. Introduction section 1.4.3). Therefore, it was hypothesised that if an overlap existed between MetAP2 and NMT, *N*-myristoylation (or rather its absence) could provide an explanation for the phenotypic effects seen upon MetAP2i treatment. The last two MetAP2 substrate discovery experiments identified several hits that coincide with known NMT substrates (see gene names highlighted with an asterisk (*) in Figures 4.13E and 4.14D). However, measuring the levels of iMet retention only, the AHA strategy does not provide a full picture of the iMet-retained to iMet-cleaved (and therefore, myristoylated) ratios. To understand if this shift to increased iMet retention upon MetAP2 inhibition caused measurable shifts in *N*-myristoylation patterns of these substrates, the previously described YnMyr metabolic tagging experiment was applied (cf [244, 312, 335, 336] and Results Chapters 2 and 3). Briefly, YnMyr gets metabolically converted into the coenzyme A (CoA) myristate analogue and transferred onto proteins by cellular NMT as if it was the endogenous myristate [335]. Bearing and alkyne (Yn) handle, it can then be chemically ligated via click chemistry (CuAAC) onto azido capture reagents either containing a fluorophore or affinity handles for subsequent visualisation or enrichment, respectively.

To determine if there were any visible changes in overall myristoylation patterns upon MetAP2 inhibition, in-gel fluorescence analysis was performed first. For this purpose, a TAMRA containing capture reagent, azido-TAMRA-biotin (AzTB) [351] (Figure 2.6), was used. In-gel fluorescence analysis did not show any detectable differences on YnMyr labelling patterns for any MetAP2i's (Figure 4.17A-B), consistent with what had been reported in a similar, albeit much less sensitive, assay using [^{3}H]-labelled myristic acid [181]. However, due to the low resolving power of 1D gels, in-gel fluorescence analyses are not always capable of detecting small changes in labelling across conditions, especially with less abundant proteins. Proteomics analysis was therefore applied to study the effect of MetAP2i on *N*-myristoylation in a more rigorous manner. A short pulse strategy was followed in order to keep time-frames comparable with AHA studies, while ensuring sufficient metabolic incorporation for the experiment (4 h pulse). This approach identified 82 *N*-myristoylated proteins in HT1080 cells. From these, the majority showed a significant reduction on YnMyr incorporation upon NMT inhibition

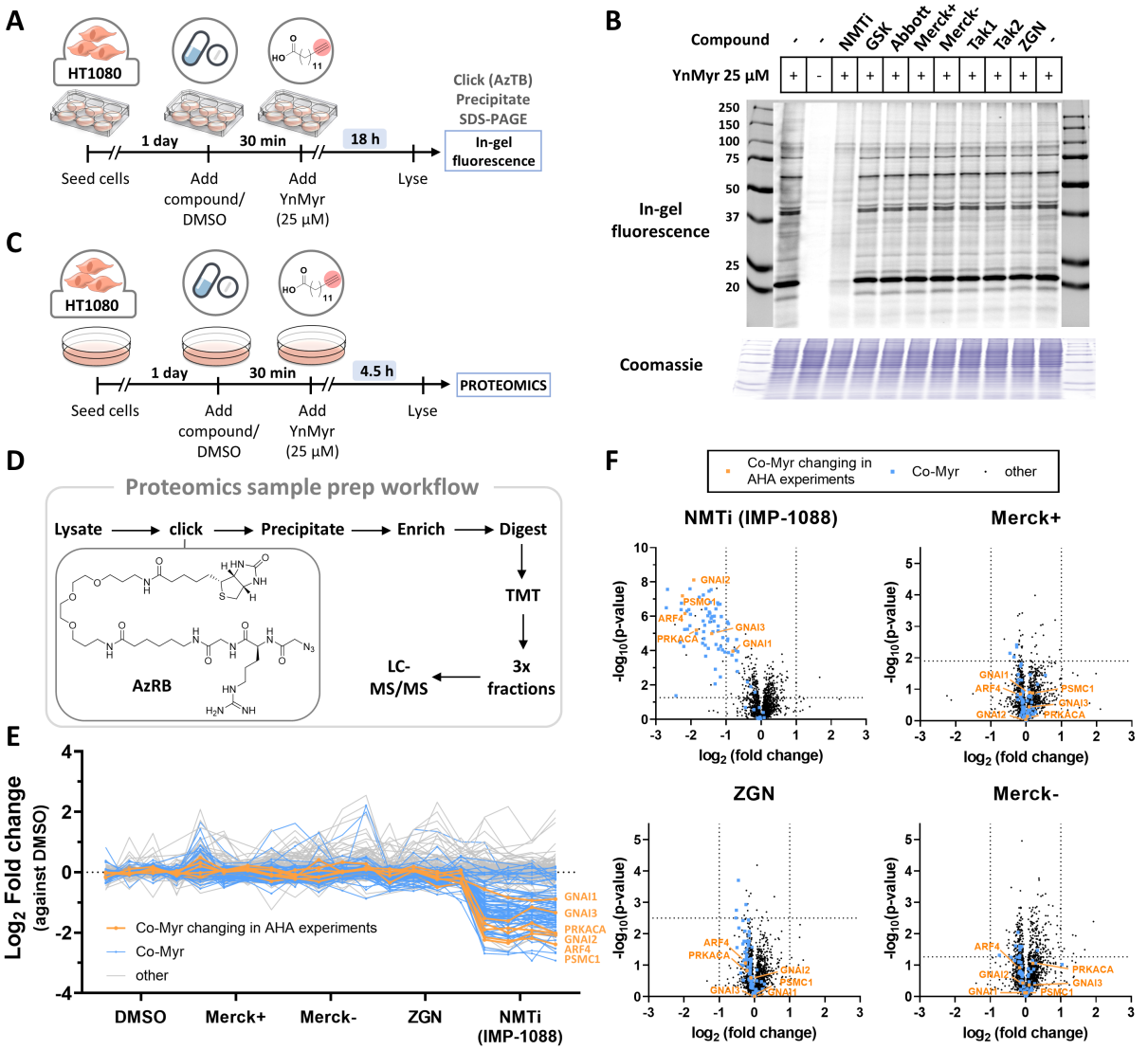


Figure 4.17: N-myristoylation profiling reveals no apparent overlap between MetAP2 and NMT. **(A)** For in-gel fluorescence analysis, cells were seeded in 6-well plates and allowed to grow for one day before the addition of MetAP2i, NMTi or DMSO. After 30 min pre-incubation, YnMyr (25 µM) was added and metabolically incorporated for further 18 h. After lysis, labelled proteins were ligated to azido-TAMRA-biotin (AzTB) capture reagent, precipitated and loaded into SDS-PAGE gels for in-gel fluorescence analysis. **(B)** Results of experiment described in (A). IMP-1088 NMTi (100 nM) was added as positive control. All MetAP2 inhibitors were dosed at 250 nM, except for ZGN, which was added at 1 nM. DMSO/YnMyr+ (negative control) was loaded on both sides for ease of visual inspection. **(C)** For proteomics analysis, cells were seeded in 10-cm dishes in four biological replicates. One day after seeding, cells were pre-incubated with MetAP2i, NMTi or DMSO control for 30 min before addition of 25 µM YnMyr, which was pulsed for 4 h before cell lysis. **(D)** Schematic representation of the proteomics sample preparation workflow for each of the samples. Azido-Arginine-Biotin (AzRB) was used as the capture reagent for the click chemistry, allowing for tag cleavage upon trypsin digestion. **(E)** Profile plot summarising the YnMyr enrichment fold changes against DMSO control of all proteins identified in this experiment. Each column represents a sample; there are four biological replicates per condition. Each line represents an identified protein. All known NMT substrates are highlighted in blue, with the exception of the six NMT substrates that were identified as possible MetAP2 substrates in experiments in Figures 4.13 and 4.14, which are highlighted in orange. **(F)** Volcano plots showing the average YnMyr enrichment fold changes against control for each of the experiments. Horizontal threshold lines represents FDR = 0.05 for each comparison, calculated based on random permutations of the data. Vertical threshold lines are at $\log_2$FC = 1 and -1, representing a 2-fold increase or 1/2-fold reduction of quantified enrichment levels.

with NMT inhibitor (NMTi) IMP-1088 [298] (Figures 4.17E-F). However, none of the MetAP2 inhibitors significantly decreased the YnMyr incorporation levels in any of the NMT substrates detected, not even of those which had been previously identified as MetAP2 substrate candidates in AHA studies (Figures 4.17E-F). A closer inspection revealed a small leftward shift in general patterns of YnMyr enrichment in NMT substrates of MetAP2i-treated samples, especially in the case of ZGN (Figure 4.17F). Yet, in general, it appears that iMet retention events caused by MetAP2 inhibition were not profound enough to significantly reduce YnMyr incorporation levels in the conditions tested.

Detection of YnMyr and AHA depend on their metabolic incorporation rates, which include metabolic activation of the probes (i.e. conversion of YnMyr to the CoA-bound form or recognition of AHA by the Met tRNA) and incorporation into newly synthesised proteins. These two processes are less efficient than for the endogenous myristate or methionine, and limit the proportion of overall *N*-myristoylation and iMet cleavage detected. Therefore, if the proportion of labelled proteins is not sufficient, the labelling levels detected might not be representative of the bulk impact on the overall *N*-myristoylated and retained iMet population, which could explain why iMet retention differences do not cause any detectable changes in YnMyr incoporation.

Alternatively, these results might offer some insight into the stoichiometry of YnMyr incorporation and MetAP2 inhibition. If iMet retention was not complete for a certain *N*-myristoylated substrate even in the absence of MetAP2i, a significant increase in iMet retention levels upon MetAP2i might only show negligible changes in *N*-myristoylation. For example, imagine an scenario where 1000 proteins were synthesised during the pulsed 4 h. In the absence of MetAP2i 400 showed iMet retained and 600 iMet cleaved, and from these only 200 were *N*-myristoylated. Upon addition of MetAP2i, iMet retention increased to 800 proteins and left only 200 with iMet cleaved, from which still 200 showed *N*-myristoylation. In this case, N-myristoylation would be the rate-limiting step, and an increase in iMet retention would not affect YnMyr incorporation levels unless it crossed a certain threshold (in this case, leaving less than 200 proteins available for *N*-myristoylation). In another scenario, an incomplete MetAP reaction followed by efficient *N*-myristoylation could yield a similar result. Imagine that in this second case from the 1000 proteins synthesised during the 4 h pulse, only 40 showed iMet retained and 960 iMet cleaved, with the 960 *N*-myristoylated. Upon MetAP2 inhibition, there was a significant two-fold increase in iMet retention to 80 proteins, but the remaining 920 were all again efficiently *N*-myristoylated. This slight decrease in YnMyr incorporation levels is only of ca. 4% and therefore, no significant alterations were detected in YnMyr enrichment levels. The completeness of N-terminal modifications is still poorly understood, and while there is evidence of incomplete iMet retention, *N*-acetylation or *N*-myristoylation, how this is regulated in the absence of eraser enzymes is still not clear [6, 39, 40]. Further, the relative rates of these PTMs compared with other N-terminal modifications are unknown. Therefore, universal methods that allow for the accurate quantification of the stoichiometries

of multiple N-terminal modifications are in great need to shed light on how these processes are inter-regulated.

4.7 Conclusions: Looking for the missing link in the mechanism of action of MetAP2 inhibitors

Despite the initial difficulties in finding the right cell model, this chapter described two different cellular assays that were successfully developed to provide quantitative MetAP2 target engagement for most of the MetAP2 compounds tested. The fact that structurally and mechanistically (covalent vs reversible) distinct compounds displayed a similar readout in both assays suggested an experimental set up powered to detect target engagement for various chemotypes. From the MetAP2i's tested, the two Takeda compounds did not show target engagement in the CETSA-WB assay and their cellular proliferation inhibition profile was unclear and distinct from all other compounds, suggesting that these compounds may exert their phenotypic effect via some other pathway. Previous evidence indicated it is possible to dissect anti-angiogenic activity from weight-loss effects [164, 204]. Takeda had developed these compounds for the treatment of metabolic disorders [205, 206] and therefore, it was considered they might still be engaging MetAP2 but have lost some of their anti-proliferative effects. This, however, was not the case for ZGN-1061, which was also developed for the treatment of metabolic disorders [221]. While a lack of thermal stabilisation does not necessarily imply that the Takeda compounds are not binding MetAP2, the cellular potencies (EC_{50}) reported for these inhibitors were performed using a 14-3-3γ assay that does not discriminate between MetAP1 and MetAP2 inhibition [193, 371], so it could be that the main target for these compounds is actually MetAP1. Nonetheless, the unclear profile for the Takeda inhibitors suggested it was best to remove them from the MetAP2 inhibitor panel.

With all MetAP2i's thoroughly characterised, this knowledge was applied for the discovery of novel MetAP2 substrates, the main aim of this Chapter. For this purpose, a novel proteomics approach was developed and optimised. Based on the metabolic incorporation of the methionine probe AHA, this method allowed for the systematic detection of iMet retention changes upon exposure to MetAP2 inhibitors. During the optimisation process, a previously reported trifunctional capture reagent was repurposed to provide a novel site for TMT labelling, which ensured quantification of all N-terminal peptides of interest. This revealed a list of ca. 80 specific substrates of MetAP2, of which 95% corresponded to novel MetAP2 substrate candidates. This opens new avenues for understanding the phenotypic effects of inhibiting MetAP2.

To delve a bit further into the phenotypic effect exerted by MetAP2 inhibitors, we have recently established an academic collaboration with the group of Dr. Alexis Barr (MRC London Institute of Medical Sciences and Imperial College London), an expert on cell cycle control. Using high content single-cell imaging, Dr. Barr characterised the G0 cell cycle

arrest inhibitory phenotype in HT1080 and RPE1 cells (after confirming their response in our proliferation assay, see Appendix Figure A.13). Importantly, these data allowed us to link the specific MetAP2 inhibitory phenotype to an array of specific cell cycle markers. Future plans include the use of the identified specific cell cycle markers to perform a targeted siRNA screen against all the candidate MetAP2 substrates identified in proteomics experiments in order to obtain mechanistic insight into how MetAP2 inhibitors exert their cell arrest phenotype. Moreover, through the described cell imaging experiments, Dr. Barr was able to confirm p53 dependency for the growth inhibitory phenotype, in accordance with previous reports [230]. This p53 dependency provides an invaluable positive control for the proposed phenotypic readout for the screen (work in progress).

By revealing a comprehensive list of MetAP2 substrates for the first time, these studies provide the foundation to unravel the missing link between MetAP2 inhibition and its phenotypic effects. Moreover, they also present a list of candidates to serve as biomarkers for improved, more robust target engagement studies, which was difficult to achieve to date due to the limited number of known MetAP2 substrates. MetAP2 inhibitors have shown great promise for the treatment of cancer in both preclinical and clinical studies but failed in late clinical stages due to toxicities that are not yet understood [188, 194, 220]. It is anticipated that future studies to understand the biology underlying the effects of MetAP2 inhibitors, enabled by the discoveries made here, will shape the fate of MetAP2 inhibitors in the clinic, providing the necessary framework to make them efficacious and safe for cancer patients.

5

General conclusions and Future Perspectives

5.1 General Conclusions

Proteomics is a powerful technique that allows for identification and quantification of thousands of proteins in complex biological samples. However, the main limitation to date is the coverage depth, which implies missing on lower abundance features. Unfortunately, in biological systems most important proteins are usually required at trace levels, allowing for dynamic fine-tuning of biological processes. Protein modifications (often referred to as protein post-translational modifications, PTMs), with their diversity and ubiquity, have revealed an unprecedented complexity to human proteomes and provide a much richer cellular milieu, precisely distributed both temporally and spatially. Due to their complexity, relatively low abundances, and their dynamic nature, protein modifications remain largely unexplored. Chemical probes arise as a means to dissecting this complexity, providing the necessary tools to hijack complex biological systems to provide specific tracking of less represented PTMs. The use of chemical tools combined with proteomics, allows for selective pre-enrichment of the tagged proteins, focusing the coverage power of proteomics experiments into the analytes of interest. This combination, coined 'chemical proteomics', provides a robust and effective way to shed light into the functional implications of protein modifications like no other technique is able to offer.

In this work, the potential of chemical proteomics was exploited to study two types of N-terminal modifications, chemically very different to one another: initiator methionine (iMet) excision and *N*-myristoylation. The studies described in this Thesis show how the combination of these techniques can be used not only as a resource to identify these PTM events, but also to characterise small molecule inhibitors targeting these pathways. By allowing for the assessment of on-target activity and providing an overview of appropriate candidates for biomarker assays, chemical proteomics is not only important for understanding of basic biology but also an invaluable tool with impact on drug discovery and translation. The SrtA-based NMT activity assay developed during this PhD project is now being used by Myricx, a company heavily focussed on developing NMT inhibitors with clinical application. Moreover, brought by our advances in MetAP2 substrate identification, a new collaboration with Merck KGaA has been established to share information and provide new functional applications to our findings. These examples illustrate the benefits of collaborations between pharma and academia, which offer academic researchers the possibility of making a greater impact on translational research while providing pharma companies with relevant basic knowledge of the underlying biology.

Finally, this work also highlights the importance of choosing the right chemical tools (i.e. probes, inhibitors, etc.) for their application in biology. It is of crucial importance that all compounds used in biological systems are validated by independent techniques before any biologically meaningful conclusions are drawn. This orthogonal validation can be achieved in multiple ways, such as by testing for their on-target activity, by setting up the right controls, by combining diverse tools to test their overlap (e.g. combination of chemical probes with

specific inhibitors), or by using a panel of compounds of diverse chemical structure.

5.2 Future perspectives

The understanding of N-terminal modifications is still at its infancy, and there is yet much to be explored in this area. Efforts in determining the prevalence of these protein modifications across the proteome, together with studies on how these modification are spatio-temporally regulated are bound to determine the true biological relevance of these PTMs.

Relative stoichiometries of different N-terminal modifications are still difficult to determine. With only one N-terminal peptide per protein, accurate quantification is problematic and can only be achieved by isotopic (e.g. SILProNAQ [382]) or isobaric (e.g. TMT) labelling approaches. Unfortunately, these approaches limit the scope of N-terminal peptides that are quantified. The chemical acetylation with isotopically labelled *N*-acetoxy-d_3-succinimide of SILProNAQ is useful to determine ratios of acetylated-to-non acetylated N-terminal peptides but does not provide information about the prevalence of any other N-terminal modification. TMT approaches miss N-terminally modified peptides that do not have a free amine, such as *N*-acetylated N-terminal peptides ending in Arg (cf. Results Chapter 4 section 4.5.2). This was overcome here by using a capture reagent that chemically introduced an additional TMT labelling site. However, chemical proteomics approaches such as the ones used in this Thesis are also not capable of providing insight into the full picture. By relying on selective enrichment of specifically labelled peptides or proteins, they only mirror the abundances of one side of the reaction. Targeted proteomics approaches (e.g. parallel reaction monitoring, PRM), in contrast, could provide an overall perspective on the relative stoichiometries of these modifications if they are specifically queried. However, this technique is limited to selected substrates only. Alternatively, data-independent acquisition (DIA) experiments could provide the necessary framework to answer these questions. However, the accuracy of the label-free quantification strategy for peptide-level analysis might become an issue in such experiments and it should be carefully tested before any important conclusions are drawn. A universal technique that will allow for the accurate quantification of differentially modified N-terminal peptides will strongly impact our understanding on how different N-terminal modifications are regulated in relation to one another.

On a different note, recent evidence suggests ribosomes might act as the central hub for integrated control of N-terminal modifications by direct regulation of the enzymes responsible for N-terminal processing; and in turn, these enzymes might also help shape the ribosomal network. New technologies have arisen to map ribosomal networks, such as ribosome profiling, which uses deep sequencing to uncover the translating proteome, or other ribosome characterisation techniques that allow for determination of ribosome composition. We anticipate that the application of these techniques in the context of N-terminal modifications will have a strong impact on our view of these central cellular machineries, which might turn to be much

more dynamic than initially expected. Such an analysis might help solve many questions regarding the decision on which N-terminal modification a nascent protein will have. Incredible efforts have been made to characterise the sequence motifs of many N-terminal modifications and yet, most of them seem to be incomplete. If the protein sequence alone does not fully determine the presence or absence of a modification, what is then driving this decision? Is it imprinted on the mRNA sequence? Or is it dictated by protein-protein and protein-RNA interactions between enzymes processing protein N-termini and other core or transient ribosome components? Research on these new avenues will provide a new perspective on translation, a process generally assumed to be fully defined. Moreover, it will help understand how these processes are co-regulated in health and disease, providing new means for targeting them for therapeutic purposes.

6

Materials and Methods

6.1 General Procedures (Except LC-MS/MS-based Proteomics)

Cells. MDA-MB-231, Panc-1 and HeLa cells were purchased from the American Type Culture Collection (ATCC; HTB-26, CRL-1469 and CCL-2 respectively). Cells were cultured in low-glucose Dulbecco's Modified Medium (DMEM) (Merck, D6046) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco, 10270). MDA-MB-231 cells were cultured at 5% CO2 concentration and 37 °C and Panc-1 and HeLa were cultured at 10% CO2 and 37 °C.

Western Blot. Samples were normalised to equal protein concentrations and separated in 10 - 12%, 15-well polyacrylamide gels (10 – 20 µg loaded per lane) (SDS-PAGE running buffer in *List of Buffers*). Proteins were blotted to a nitrocellulose membrane (GE Healthcare Life Sciences) by wet transfer (transfer buffer in *List of Buffers*) and membranes blocked for 1 h in TBST (see *List of Buffers*) containing 5% semi skimmed milk (Blocking Buffer A in *List of Buffers*) or 3% bovine serum albumin (BSA) (Blocking Buffer B in *List of Buffers*). Membranes were incubated with the corresponding primary antibody (list of all antibodies and the dilutions used in 6.10.2) in blocking solution overnight at 4 °C. On the following day, membranes were washed with TBST three times for 10 min each and incubated with the secondary HRP-conjugated antibody solution in blocking solution for 1 h at room temperature (RT). Membranes were washed with TBST again three times for 10 min before being developed with the Luminata Crescendo Western HRP substrate (Millipore) and imaged using a ImageQuant LAS 4000 scanner (GE Healthcare Life Sciences).

Methanol-chloroform precipitation of proteins. Clean up of protein samples by methanol-chloroform precipitation was used before in-gel fluorescence analysis and in all proteomics sample preparation experiments. It allows for removal of small molecule reagents, buffer exchange and for the purification of cellular proteins from other cellular components. Methanol and chloroform are added to protein samples or lysates in this order and to a 4:1 ratio of methanol to chloroform. Water was sometimes added to aid in the precipitation of smaller volumes of sample in a ratio of 1:2 water to methanol. For example, for a typical 50µL (1 mg/mL) protein sample, 4 volumes of methanol, 1 volumes of chloroform and 2 volume of water were added. For a sample of 100 - 300 µL, 2 volumes of methanol, 0.5 volumes of chloroform and 1 volume of water was added. For a sample of >500 µL, however, 1 sample volume of methanol and 0.25 volumes of chloroform were sufficient to ensure proper separation of the phases. After addition of the solvents and mixing, the mixture is separated into two phases by centrifugation at 4,000 $\times g$ for 5 min, with the proteins generating a disc at the interphase. The top aqueous phase is discarded, and the rest washed with 300 - 500 µL methanol three times. Each methanol wash consists of adding methanol, mixing and sonicating until the pellet breaks completely and re-formation of the pellet by centrifugation at 7,500 $\times g$ for 5 min at 4 °C. The supernatant is discarded and the, before discarding the supernatant. After the last wash, the protein pellet was air-dried for no more than 5 min to

allow the methanol to evaporate and resuspended into the desired buffer. Full resuspension is achieved by addition of small volumes (typically 1/10 volume of starting sample volume) of buffer containing high concentration of detergents such as SDS and water bath sonication. Although they were not used here, strong denaturants such as urea or guanidine hydrochloride can also be used instead of SDS. The concentration of detergent is then typically brought down to a concentration compatible with the next steps of the sample preparation procedure.

Click reaction (CuAAC). Metabolically tagged cells were lysed directly on plates using Lysis Buffer A (see *List of Buffers*). Protein concentration was measured and samples diluted to 1 mg/mL with Lysis Buffer A. A minimum material of 50 µg protein per sample was used as starting point to allow for efficient precipitation after CuAAC. The click mixture was prepared as described previously [244] 1 µL 10mM capture reagent (AzTB, AzT, YnT), 2 µL 50 mM CuSO4, 2 µL 50 mM TCEP and 1 µL 10 mM tris(benzyltriazolulmethyl)amine (TBTA). 6 µL of the click mix were then added for every 100 µL protein and the samples incubated for 1 h at RT with gentle shaking. 5 mM EDTA were added to quench copper and stop the reaction. Proteins were precipitated as described previously (see *Methanol-chloroform precipitation of proteins*). Samples were dissolved in 5 µL 2% SDS and then diluted down with PBS to 0.2% SDS. Samples were run in SDS-PAGE gels and imaged as described in *In-gel fluorescence*.

In-gel fluorescence. After *Click reaction (CuAAC)* or *SrtA reaction*, 15 µg labelled protein (AzTB/AzT/YnT or TAMRA-ALPET-Haa) was loaded onto a 12% SDS-PAGE gels, run at 180 V for 1 h and imaged using a Typhoon FLA 9500 imager (GE Healthcare Life Sciences). For TAMRA, 532/575 nm excitation/emission wavelengths were used. For the ladder, Cy5 fluorescence settings were used, with 635/670 nm excitation/emission wavelengths.

6.2 Generation of CRISPR/Cas9 knock outs of NMT1 and NMT2

Design of the CRISPR/Cas9 gRNAs. Three primer pairs were designed for each of the two genes being targeted (NMT1/NMT2). Primers were designed with a complementary region of 20 bp consisting of the gRNA target sequence, which was selected using the CHOPCHOP webtool (`https://chopchop.cbu.uib.no/`), and a 40 bp overhang at opposite sides coding for the Cas9 recognition sequence and to serve as amplification handle. Full sequences of all designed gRNAs are in Table 6.1.

gRNA preparation. Each gRNA primer pair was cloned into a gRNA cloning vector (Addgene, 41824) as described previously [342]. Specifically, each gRNA primer pair was annealed by mixing 8 µL 5× Phusion HF buffer (Thermo Fisher) with 2 µL of each primer

Table 6.1: gRNA sequences designed to target NMT1 and NMT2 by CRISPR/Cas9.

Target Gene	Name	Direction	gRNA target sequence	Exon targeted
NMT1	AG1	Fw	(Cas9)...GGCATCGGTCCCTAAACCCA	8
		Rv	(Cas9)...TGGGTTTAGGGACCGATGCC	
	AG2	Fw	(Cas9)...GATCCGAGAGATCACCAGGC	7
		Rv	(Cas9)...GCCTGGTGATCTCTCGGATC	
	AG3	Fw	(Cas9)...TTCCGATTTGATTATTCCC	5
		Rv	(Cas9)...GGGAATAATCAAATCGGAAC	
NMT2	AG4	Fw	(Cas9)...GAGGAGACGGAGCACGCCAA	1
		Rv	(Cas9)...TTGGCGTGCTCCGTCTCCTC	
	AG5	Fw	(Cas9)...GAACCGTATTCTTTGCCACA	5
		Rv	(Cas9)...TGTGGCAAAGAATACGGTTC	
	AG6	Fw	(Cas9)...GTGCCATCCCAGCAAACATT	7
		Rv	(Cas9)...AATGTTTGCTGGGATGGCAC	
Cas9 recognition sequence[1]		Fw	TTTCTTGGCTTTATATATCTTGTG GAAAGGACGAAACACC...(target seq.)	-
		Rv	GACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAAC...(target seq.)	

[1] The full gRNA would be composed of the fusion between the Cas9 recognition sequence and the target sequence. As an example, full AG1 gRNA would be: (Fw) TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGCATCGGTCCCTAAACCCA, (Rv) GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGGGTTTAGGGACCGATGCC. Both shown in direction 5'→3'. This generates a 40 bp overhang in the Fw primerfollowed by the ca. 20 bp target sequence (complementary in both Fw and Rv primers)and another 40 bp overhang in the Rv primer. Once annealed and extended each ofthese yield a 100 bp dsDNA full gRNA sequence.

(forward/reverse) at 100 μM, and with nuclease-free H_2O to 40 μL in 1.5 mL centrifuge tubes. Lids were sealed with parafilm and tubes floated over 1 L boiling water overnight. 20 μL of the annealed product were then mixed with 8 μL 5× Phusion HF buffer, 0.8 μL 10 mM dNTP mix, 0.4 μL Phusion DNA polymerase (all Thermo Fisher) and 10.8 μL nuclease free H_2O and incubated at 72 °C for 10 min in a thermal cycler to extend the overhangs and create a double-stranded (ds) gRNA sequence of 100 bp. 1 μL dsDNA product were then amplified by PCR. The PCR reaction was set by mixing 5 μL 5× Phusion HF buffer, 1 μL 10 mM dNTP mix, 0.5 μL of each primer (CRISPRfw/CRISPRrv on Table 6.2), 0.5 μL Phusion DNA polymerase and nuclease free H_2O to 50 μL. The PCR reaction was carried out with a initial denaturation step of 30 s at 98 °C followed by 25 cycles of denaturation (15 s at 98 °C), annealing (15 s at 50 °C) and extension (15 s at 72 °C) and a final extension period of 2 min at 72 °). The PCR product was resolved in 1.5% agarose gel and spin column purified using a QIAquick PCR purification kit (QIAGEN) and following manufacturer's instructions. Concentration of the purified PCR products was measured using a NanoDrop spectophotometer (Thermo Scientific). 5 μg p41824 gRNA cloning vector were linearised using 2 μL AFIII restriction enzyme and 10 μL 10× CutSmart buffer (all New England BioLabs) in a final reaction volume of 100 μL and incubating the reaction mixture overnight at 37 °C. Linearisation was confirmed by resolving 2 μL of the product in a 1% agarose gel side by side with 100 ng of non-digested supercoiled plasmid. The remainder of the linearised plasmid was then incubated at 65 °C for 10 min to heat-inactivate the restriction enzyme before being

Table 6.2: Primer sequences for PCR amplification of gRNAs, with their respective annealing temperatures and resulting amplicon size.

Primer	Sequence	Phusion Annealing Temperature	Amplicon size
CRISPRfw CRISPRrv	TTTCTTGGCTTTATATATCT GACTAGCCTTATTTTAACTT	50 °C	100 bp
AG3 Fw AG3 Rv	AGAGCCTGGAACTGGACGAC TCTCAAGAACCTCAAGAAAGCC	67 °C	162 bp
AG4 Fw AG4 Rv	GGATAGACGGGGACAATGAG CACAAAGGGAAGGGCGTAG	63 °C	418 bp
AG6 Fw AG6 Rv	ACTGGTGCCTGTTTTCTTCTTGTCT ATATAAAAACCGCATGGACGAG	67 °C	187 bp

purified using a QIAquick PCR purification kit. Concentration of the product was measured with a NanoDrop spectrophotometer. dsDNA gRNA insert was ligated to the AFIII-linearised p41824 gRNA cloning vector using a Gibson Assembly cloning kit (New England BioLabs). 50 ng of linearised plasmid were mixed with 6 ng 100 bp dsDNA gRNA insert, 10 µL of Gibson Assembly mix and nuclease free H_2O to 20 µL and incubated for 1 h at 50 °C. 2 µL of reaction product were then used to transform one vial of the provided chemically competent NEB 5-alpha *E.coli* cells. Bacteria were incubated for 30 min on ice, heat shocked for 30 s at 42 °C and transferred back to ice for 2 min. After addition of 950 µL of SOC medium, bacteria were allow to recover for 60 min at 37 °C. 100 µL of the transformed bacteria was then spread into agar plates containing kanamycin as selection antibiotic. Plates were incubated overnight at 37 °C. On the following day, a single colony per plate was picked and transferred to a 50 mL Falcon tube containing 5 mL Luria Broth (LB) media with 5 µg/mL kanamycin. Bacterial suspensions were incubated overnight at 37 °C with substantial shaking. Next day, 20% of the suspension was kept for subsequent scaling up and the remaining 80% was used for DNA extraction, which was performed using the QIAprep Spin Miniprep kit (QIAGEN) and according to the manufacturer's instructions. The DNA product was checked for purity using a NanodropTM 2000 spectrophotometer (Thermo Fisher) and sent for sequencing to Beckman Coulter Genomics in order to confirm insertion of the designed primers into the vector. Selected positive bacterial clones were then scaled up to 150 mL LB media with appropriate selection antibiotic, incubated overnight at 37 °C and the DNA extracted using the QIAGEN plasmid Maxi kit (QIAGEN).

Transfection of Panc-1 cells. 1×10^6 Panc-1 cells were seeded in a 10-cm-dish one day prior to the transfection. Cells were transfected with 4 µg DNA (2 µg primer-plasmid + 2 µg GFP-Cas9 plasmid (Addgene, 44719) or 4 µg GFP-Cas9 plasmid for the control) using 35 µL Lipofectamine 2000 transfection reagent (Thermo Fisher) and 1 mL OptiMEM medium (Gibco). 5 h after transfection, the medium was replaced by normal culturing medium and the cells were allow to recover for 72 h.

Fluorescence-activated cell sorting (FACS). Cells were harvested by centrifugation

at 1000 $\times g$ for 3 min at 10 °C, washed with PBS and centrifuged again. Cells were resuspended in 1 mL PBS, 0.1% FCS, 5 mM EDTA and transferred to a FACS compatible tube with filter. The FACS sorting was performed by Jane Srivastava at the Imperial College Flow Cytometry Facility. The output of the sorting was one 96-well plate per CRISPR/Cas-9 primer (two primers against NMT1 and two primers against NMT2) with one cell per well in 20% FBS medium, in addition to one 96-well plate with control cells (cells transfected only with GFP-Cas9 plasmid).

Clonal growth. Clones were grown in normal culturing medium. 5 days after sorting, the 96-well plates were screened for living cells. Wells that contained cells were monitored daily and their medium changed every other day. Clones were initially grown in 96-well plates, subsequently scaled up to 48-well plates when confluent and so on until there were enough cells to be frozen for long-term storage.

Amplification of the gRNA target region by PCR. DNA was extracted following the DNeasy Blood and Tissue kit (QIAGEN) and DNA concentration measured by NanoDrop (Thermo Fisher Scientific). The PCR was set up using 200 µg template genomic DNA and Phusion High Fidelity DNA Polymerase (NEB) in a 25 µL PCR reaction following the manufacturer's instructions (35 cycles). Annealing temperatures were calculated using the NEB Tm calculator web tool (https://tmcalculator.neb.com/) and are listed together with the primers used in Table 6.2.

6.3 Characterisation of NMT1 and NMT2 knock outs

Western Blot of NMT1 and NMT2 in Panc-1 knock outs. Cells were washed with PBS once and lysed in Lysis Buffer A (see *List of Buffers*) directly on the plate. Lysates were collected, centrifuged at maximum speed for 10 min and the supernatant transferred into a new tube. Lysates were stored at -80 °C until use. Protein concentration was determined using the DC Protein Assay Kit II (BioRad) and absorbance was measured at 750 nm. Samples were normalised to equal protein concentrations and separated in 12%, 15-well SDS-PAGE gels (15 µg loaded per lane). The western blot procedure was carried out as described in *Western Blot*, using 5% milk as blocking agent.

Flow cytometry analysis of Panc-1 clones. Cells were seeded one day before the start of the experiment in a 12-well plate in a density of 40,000 cells per well. Cells were incubated with 10 µM EdU 2 h at 37 °C before harvesting cells with trypsin. The cell suspension was then transferred to a U bottom 96-well plate, washed with PBS and stained with 0.5 µL Zombie NIR (Biolegend) per 100 µL cell suspension for 5 min at RT to discard dead cells for analysis. After one wash with PBS cells were fixed with 4% formaldehyde in the dark for 20 min, washed with FACS buffer (see *List of Buffers*) once

and permeabilized with 100 µL Cytofix/Cytoperm buffer (BD biosciences) for 20 min at RT. All subsequent washes were performed using the Wash/Perm buffer (BD biosciences). Cells were washed once before EdU ligation to 6-FAM azide 488 fluorophore by copper-catalyzed alkyne-azide cycloaddition (CuAAC or click reaction). The click mix was prepared by mixing the following reagents per 1 mL of reaction mix in the order given and vortexing between additions: 970 µL PBS, 4 µL $CuSO_4$ 500 mM 10 µL buffer additive (Component G, Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit) and 1 µL 6-FAM azide 488 (AAT Bioquest) 20 mM. 100 µL of the click reaction mix was added per well and plates incubated for 1 h at RT in the dark. After two washes, apoptotic cells were stained for active caspase 3 (PE Active Caspase-3 Apoptosis Kit, BD Biosciences) for 1h at RT, washed with Wash/Perm buffer once, washed with FACS buffer once and nuclear content stained with DAPI (in FACS buffer) for 30 min at RT. Samples were kindly measured by Gregor Lueg on a MACSQuant VYB (Miltenyi Biotec) and analysed using FlowJo software (v10.4, Tree star).

MTS cell viability assay of Panc-1 clones. Cells were seeded in 96-well plates in 50 µL/well one day before the start of the experiment. WT, CTRL-A, 4C and 6B cells were seeded at a density of 2,000 cell/well, while 3D and 3E clones were seeded at 4,000 cells/well. Cells were incubated with 1/5 dilutions of three NMT inhibitors (IMP-1088, IMP-1320 and DDD85646) starting at 2 µM and ranging down to 0.1 nM for 144 h (6 days). Inhibitor dilutions were added at 1.5× concentration the final concentration in 100 µL to make up a final 150 µL in each well. Media was replaced and treatment repeated on day 3 to ensure availability of nutrients for proliferating cells. 2 µg/mL puromycin was used as positive control. After the 6 day incubation period, the detection working solution was prepared fresh by mixing MTS (1.6 mg/mL in PBS) and PMS (0.92 mg/mL in H_2O) (Sigma-Aldrich) reagents at 20:1 ratio. 20 µL of the MTS/PMS working solution were added to each well and plates incubated for 2.5 h (WT, CTRL-A, 4C, 6B) or 3.5 h (3D, 3E) at 37 °C. Absorbance was read at 492 nM.

6.4 Labelling N-terminal glycines in whole-cell lysates with SrtA 5M

Preparation of cell culture samples for SrtA reaction. MDA-MB-231, Panc-1 or HeLa cells were seeded in 100 mm dishes one day before the start of the experiment and treated next day with DMSO, 1 nM, 10 nM or 100 nM IMP-1088 NMT inhibitor. 24 h later, cells were washed with PBS and scraped in SrtA buffer (see *List of Buffers*). Cell suspensions were transferred to microcentrifuge tubes, lysed by passing through a 21-gauge needle and protein concentration measured using the DCTM Protein Assay Kit II (BioRad).

Preparation of tissue samples for SrtA reaction. Tumour dissections were weighted before transferring them to vials containing Lysing Matrix D ceramic beads (MP Biomedicals). A volume of SrtA Buffer (see *List of Buffers*) (measured in µL) equivalent to four-times

the tissue weight (measured in mg) was added to each vial. Tissues were lysed in a FastPrep24-5G cell disruptor (MP Biomedicals) using 2 cycles of 20 s each at 5.0 m/s. Samples were cooled on ice between cycles. These steps were repeated until the solid tissues disappeared. Samples were then centrifuged for 3 min at 5,400 $\times g$ before being transferred to eppendorf tubes for further processing or storage. Protein concentration was measured using the DCTM Protein Assay Kit II (BioRad).

SrtA reaction. Samples (20 – 300 µg protein content) were adjusted to 1 mg/mL protein concentration in SrtA buffer (see *List of Buffers*). The SrtA reaction was carried out by adding 75 µM ALPET-Haa substrate and 0.1 µM SrtA 5M (a kind gift from Dr. M. Jamshidiha, Imperial College London, expressed and purified from Addgene plasmid 86962 [356]) to each sample and incubating with mild shaking overnight (16 h) at 4 °C. The reaction was stopped by addition of 5 mM EDTA. Proteins where then precipitated by addition of 4:1:2 methanol to chloroform to lysate ratio as described previously (please refer to *Methanol-chloroform precipitation of proteins*). Samples were re-suspended in PBS containing 2% SDS, 10 mM DTT. Once dissolved, the samples were diluted to a final concentration of 1 mg/mL of proteins and 0.2% SDS by addition of PBS.

Affinity enrichment and western blot analysis. Samples were labelled by SrtA overnight and precipitated as described above (see *SrtA reaction* and *Methanol-chloroform precipitation of proteins*). For the affinity enrichment, 50 µL of Pierce NeutrAvidin agarose bead slurry (Thermo Fisher), previously equilibrated with 0.2% SDS in PBS, were used for 50 µg of SrtA-labelled protein sample. Samples were incubated with the beads for 2 h at RT. The non-bound solution was collected to be used as the supernatant fraction. The beads where then washed three times with 0.2% SDS/PBS before elution in 2X Protein Loading Buffer (see *List of Buffers*) for 15 min at 95 °C. Typically, proteins enriched from 50 µg of protein were eluted in 20 µL 2X Protein Loading Buffer. Equal volumes of every sample were loaded into 12% polyacrylamide gels, followed by western blotting against ARL1, PRKACA, YES1 or loading control (GAPDH), performed as described previously in section *Western Blot*.

Streptavidin shift assay. Samples were subjected to *SrtA reaction*, precipitated and resuspended as described in *Methanol-chloroform precipitation of proteins*. Protein Loading Buffer (see *List of Buffers*) was added, and protein samples boiled for 5 min at 95 °C. Samples were cooled down to room temperature before addition of 1 µL of 100 µM Streptavidin stock for every 5 µg of SrtA-labeled protein lysate. Streptavidin was allowed to bind to biotinylated proteins for 5 min at RT before loading the protein samples to 12% polyacrylamide gels. Gels were run and blotted as described in *Western Blot*, using ARL1, PRKACA, YES1 or loading control (GAPDH) antibodies.

ELISA. MDA-MB-231 cell or mouse tumour tissue samples were lysed in SrtA Buffer (see *List of Buffers*), subjected to *SrtA reaction* overnight, precipitated (see *Methanol-chloroform precipitation of proteins*) and resuspended to a final concentration of 0.02 mg/mL in PBS containing 0.2% SDS. Pierce streptavidin coated black ELISA 96-well plates (ThermoScientific) were washed four times for 3 min with ELISA Wash Buffer (see *List of Buffers*). All subsequent washing steps were performed in the same way. 50 µL of the SrtA-labelled protein sample containing 1 µg protein (0.02 mg/mL) were added and incubated for 3 h at RT. The plate was then washed and incubated with 100 µL ARL1 primary antibody at 1:500 dilution for 1 h at RT. After washing, the HRP-conjugated anti-rabbit secondary antibody was added in 100 µL and 1:8000 dilution, and incubated for 1 h at RT. The plate was washed, and the kinetics of substrate conversion measured on an EnVision Xcite 2104 (Perkin Elmer; excitation filter: 320 nm, emission filter: 460 nm) every minute for 30 min after addition of the QuantaBluTM fluorogenic peroxidase substrate (ThermoScientific).

6.5 Characterisation of MetAP2 inhibitors

Cellular thermal shift assay (CETSA). Determination of MetAP2 T_m in non-treated cells: HT1080 cells in two t75 flasks at 90 confluence were detached with trypsin, trypsin quenched by addition of medium and cells transferred to 15 mL falcon tubes. After centrifugation (5 min at 1500 $\times$ *g*), cells were washed with PBS once and centrifuged again (5 min at 1500 $\times$ *g*) before re-suspending them in PBS containing protease inhibitors. Cells in PBS were then aliquoted into PCR tubes in 100 µL aliquots. Using a thermal cycler, cells were heat-shocked for 3 min at a 12 temperature gradient spanning 35-90 °C. Because the thermal cycler only allowed gradients of up to 30 °C temperature difference between top and bottom temperatures, two gradients were set up, one of 35-60 °C and a second of 65-90°C and heat treatment performed in two batches. After the 3 min heat treatment, cells were allowed to cool for 3 min before lysis by repeated freeze-thaw cycles. 5 freeze-thaw cycles were performed in total, by transferring the samples from liquid nitrogen to a water bath at RT. Lysates were then transferred into 1.5 mL microcentrifuge tubes and centrifuged in a benchtop centrifuge at maximum speed for 40 min at 4 °C. The supernatant containing the soluble protein was transferred into a new vial and stored at -80 °C or used straight away. For MetAP2 T_m determination by western blot, 15 µL from each sample (without equalising protein concentration) were mixed with 5 µL 4X Protein Loading Buffer (see *List of Buffers*) and 19 µL of the mixture loaded onto a 12% SDS-PAGE gel. Western blotting for MetAP2 was performed as described on *Western Blot*. CETSA ITDR: Cells were treated with a concentration series of MetAP2i or DMSO control for 24 h in t25 flasks (5 mL per flask). Cells were detached with trypsin, washed once with PBS and re-suspended in PBS containing protease inhibitors. Cells in PBS were then aliquoted into PCR tubes (100 µL per aliquot) and heat shocked for 3 min at 72 °C (the calculated T_m for MetAP2). Samples

were then lysed by freeze-thaw cycles (5 cycles), transferred into 1.5 mL microcentrifuge tubes and centrifuged at maximum speed for 40 min at 4 °C. The supernatant containing the soluble protein was transferred into a new vial and protein concentration measured. Samples were then stored at -80 °C or loaded onto 12% SDS-PAGE gels (equal protein content per lane) for western blot analysis (performed as described in the *Western Blot* section).

Real time proliferation monitoring using the Incucyte. Cell treatment: 1,000 HT1080 cells per well were plated in 100 μL medium onto 96-well plates (Grenier, 666180) and allowed to attach overnight. Cells were laid out leaving the outside wells empty, which were filled with 200 μL medium only to prevent evaporation. The following day, 2X concentrated solution of cell culture medium containing SYTOX green dye (final 250 nM, 2X = 500 nM) and MetAP2i (from a 1000X DMSO stock), DMSO (final 0.1%, 2X = 0.2%) or puromycin (final 2 μg/mL, 2X = 4 μg/mL) were prepared on a dilution plate. Using a multichannel, 100 μL of the 2X dye/compound concentrated media were added on top of each corresponding well to make up a final 200 μL per well. After ensuring no bubbles were left in any wells, plates were placed on the Incucyte tray adaptor for multi-well plates. Imaging set-up: 4 images per well were taken using a 10X objective every 4 h (6 scans a day). Phase-contrast and fluorescence acquisition settings were left as default, with auto-focusing enabled. Data processing: Translation of imaging data into quantitative data was carried out using the integrated software (version IncuCyte 2019B). For the phase-contrast settings, segmentation was kept at 0.9 and the area filter was set at min = 100 μM to prevent debris from being accounted as cell area. For the green fluorescence settings, the Top-hat segmentation method was selected, with a maximum radius of 100 μM and a threshold of 25 GCU. All the settings were manually selected for these cells and visually inspected for validity in all scanned plates before starting the analysis. Data analysis: On the Incucyte software, 'Vessel Information' tab, a plate map was created for each scanned plate. This allowed the software to integrate replicate data, which was exported into a .txt file including the average per replicate and standard error (calculated per image) using the 'Graph Metrics' tab. The exported data was then imported into GraphPad Prism (version 5.0). For proliferation plots, confluence area (%) was plotted against the elapsed time (days) to show the full proliferation curve. The linear range was then manually selected as the period spanning 2.5 to 4 days (same for all treatments) and fitted to a straight line. Slopes were then plotted against the $\log_{10}$ MetAP2i concentration (nM) and fitted to a 4-parameter curve to determine the EC_{50} of each compound. For cytotoxicity plots, the ratio of green fluorescence area to phase-contrast area (confluence) were plotted against the elapsed time (days). For each compound concentration, the total area under the curve (AUC) was calculated for the period spanning 0 to 5 days. This ensured cytotoxicity due to over-confluence did not affect the results, which was observed for non-treated control and lower MetAP2i concentrations. Each calculated AUC was then plotted against $\log_{10}$ MetAP2i concentration (nM) to show the cytotoxcity caused by each concetration of compound.

YnMyr pull down spike-in SILAC of Panc-1 clones. This procedure corresponds to the experiment displayed in Figure 2.10. We compared the YnMyr incorporation profiles of NMT1- and NMT2-KO Panc-1 clones without any further experimental perturbation in order to investigate the contribution of each enzyme isoform using a spike-in SILAC quantification approach. Sample harvesting: Heavy cells (WT) were obtained by continuously culturing Panc-1 cells in heavy media (arginine- and lysine-free DMEM supplemented with heavy arginine (K8) and heavy lysine (R10) and 8% dialysed FCS) for 8 passages before expanding the culture to needed volume. To prepare the heavy spike-in sample, cells were seeded in 10-cm dishes, allowed to grow to 75% confluence and incubated with 30 μM YnMyr for 24 h before lysis in Lysis Buffer A (see *List of Buffers*) and pooling the lysate. Similarly, light cells (CTRL, 3D, 3E, 4C and 6B clones) were seeded in 10-cm dishes and allowed to grow to 75% confluence, before addition of 30 μM YnMyr and lysis after a 24 h incubation period. 100 μg of pooled heavy spike-in protein sample were added to 300 μg of each (light) sample. Click reaction: Samples were clicked to azido-Arg-Biotin (AzRB) capture reagent [351] following the procedure described in *Click reaction (CuAAC)*. Precipitation: Quenched samples were precipitated following the standard procedure (see *Methanol-chloroform precipitation of proteins*). Dried pellets were resolubilised in 40 μL 2% SDS in HEPES 50 mM pH 8.0 and then diluted down to 0.2% SDS with HEPES. Protein-level enrichment with on-bead reduction, alkylation and digestion: Samples were applied to 40 μL pre-equilibrated Pierce NeutrAvidin agarose bead slurry (Thermo Fisher) and allowed to bind for 2 h at RT. Beads were washed twice with 0.2% SDS in HEPES and four times with HEPES. All bead washes were performed by centrifugation on a benchtop centrifuge at 4 × *g* for 3 min. Each wash step was followed by a centrifugation step to allow discarding of the supernatant. Beads were then resuspended with 50 μL HEPES containing 5 mM TCEP and 10 mM chloroacetamide. 0.5 μg trypsin (Promega, V5111) was added to each sample and on-bead digestion allowed to carry on overnight at 37 °C in the presence of reducing and alkylating agents. Supernatants were recovered by centrifugation, beads washed with extra 100 μL of H_2O and supernatant combined with previous ones. Samples were acidified with 0.5% trifluoroacetic acid (TFA) to quench trypsin and ensure consistent charging of peptides for stage tipping. Desalting on stage tips: Samples were desalted on stage tips following the stop and go procedure reported previously [398]. Briefly, three stacked disks of SDB-XC membranes (Supelco) were inserted in no-filter 200 μL tips to the same distance to the tip. For all subsequent steps, stage tips were spin down at 1,100 ×*g* to push the liquid through the membrane. Membranes were activated with 150 μL methanol and equilibrated with 150 μL LC-MS grade H_2O before addition of the charged peptide samples. Samples were desalted by washing with 150 μL LC-MS grade H_2O twice before elution with 60% acetonitrile. Desalted peptide samples were then dried on a Savant SPD1010 SpeedVac Concentrator (Thermo Fisher) at 45 °C and stored at -80 °C until submission.

Measurement of the peptide sample: Samples were redissolved in 30 µL Resuspension Buffer (see *List of Buffers*) and 1.5 µL injected into the nLC-MS/MS system.

Whole proteome analysis of Panc-1 clones. This procedure corresponds to the experiment displayed in Figure 2.6. In this experiment, we compared the whole-proteome profiles of NMT1- and NMT2-KO Panc-1 clones in homeostatic conditions, without any further experimental perturbation. Each clone type was seeded and harvested in three biological replicates. Sample harvesting: Cells were seeded in 6-cm dishes and grown until 95% confluent. Lysates were generated with Lysis Buffer B (see *List of Buffers*) and cleared by centrifugation for 5 min at maximum speed. In-solution reduction and alkylation: 100 µg starting protein material were reduced with 5 mM TCEP for 1h at 55 °C, cooled to room temperature and then alkylated with 15 mM chloroacetamide for 1 h at RT. Precipitation: Proteins were precipitated as described in *Methanol-chloroform precipitation of proteins*) and resuspended in 10 µL 1% SDS in HEPES 50 mM pH 8.0. Digestion: Once fully redissolved, samples were diluted down to 0.1% SDS with HEPES before addition of 1 µg of trypsin and overnight incubation at 37 °C. Trypsin was quenched with 1 mM PMSF before TMT labelling. TMT labelling: The equivalent of 10 µg of initial protein (1/10 of each sample) were labelled with 0.08 mg TMT10-plex reagent (1/10 of a 0.8 mg TMT reagent vial). Each 0.08 mg TMT reagent dried aliquot was equilibrated to room temperature, dissolved in acetonitrile and added to the corresponding peptide sample in a 1:1 volume ratio to achieve 50% acetonitrile in the final reaction mixture. The TMT labelling reaction was incubated for 2 h at RT. The reaction was quenched by addition of 1 µL 5% hydroxylamine. Samples in each 10-plex were then combined at equal amounts and dried on the SpeedVac overnight with no heating. Six-layer fractionation: Dried TMT-labelled peptide samples were redissolved in 150 µL 1% TFA, their pH checked to be pH$<$3 and loaded into SCX stage tips containing 3 stacks of polystyrene-divinylbenzene copolymer membranes modified with sulfonic acid (Supelco). The 6-layer fractionation was carried out as described previously [399] with slight modification in the composition of the fraction solutions (Table 6.3). Briefly, membranes were activated with 150 µL acetonitrile and equilibrated with 150 µL LC-MS grade H_2O. All solutions were passed through the stage tip by centrifugation at 1,100 $\times$ *g* for 3 min at RT. Samples were applied to the membranes, desalted by washing twice with 150 µL 0.2% TFA containing LC-MS grade H_2O and eluted with buffers SCXx1-SCXx5 and Buffer X (Table 6.3). Fractionated peptide samples were then dried on the SpeedVac at 45 °C and stored at -80 °C until submission. Measurement of the peptide sample: For injection, samples were redissolved in 16 µL Resuspension Buffer (see *List of Buffers*) and 2 µL injected for fractions 1 and 2, 1.5 µL injected for fractions 3 and 4, and 1 µL injected for fractions 5 and 6.

SrtA peptide-based enrichment proteomics. This procedure corresponds to the experiment displayed in Figures 3.4 and 3.5. To study the substrate specificity of SrtA in whole-cell lysates we compared enrichment of labelled peptides of MDA-MB-231 cell lysates to peptides from control (SrtA-untreated) samples. Three biological replicates

Table 6.3: Six-layer fractionation elution buffers.

Buffer	Composition
SCXx1	75 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx2	125 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx3	200 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx4	300 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx5	400 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
Buffer X	5% (v/v) Ammonium hydroxide, 80% (v/v) ACN

* ACN, acetonitrile. FA, formic acid.

were used, which were separated into two equal aliquots, resulting in two identical sets of triplicates, one of which was subjected to a normal overnight SrtA reaction while the second set constituted a negative control in which incubation was carried out in the absence of enzyme. Cell treatment and sample harvesting: Cells were incubated in the presence of 100 nM IMP-1088 NMT inhibitor for 24 h before lysis in SrtA buffer (see *List of Buffers*). SrtA reaction: Each biological replicate was separated into two 500 µg aliquots, which were subjected to SrtA labelling overnight in the presence or absence of the enzyme. All reactions were quenched by addition of 5 mM EDTA. Precipitation: Proteins were then cleaned up by precipitation (see *Methanol-chloroform precipitation of proteins*) and resuspended in 500 µL of 50 mM HEPES pH 8.0. Trypsin digestion: Final concentration of 5 mM TCEP and 10 mM chloroacetamide were added (from 10X stocks, prepared fresh in HEPES) and incubated at RT for 15 min before addition of 4 µg trypsin per sample and overnight digestion, reduction and alkylation at 37 °C. Trypsin was quenched with the addition of 1X protease inhibitor cocktail. Peptide-based enrichment: 500 µL NeutrAvidin agarose bead slurry were used for each sample containing 500 µg of protein sample. Beads were equilibrated with 50 mM HEPES pH 8.0, before incubation with the biotinylated peptide sample for 2 h at RT. The unbound fraction was discarded, and the beads washed twice with 50 mM HEPES pH 8.0, twice with H_2O, and three times with 10% acetonitrile in H_2O. All bead washes were performed by centrifugation on a benchtop centrifuge at 4 $\times$ *g* for 3 min. Bead-bound peptides were eluted in two 15-minute-incubations with 150 µL of 80% acetonitrile,0.1% TFA and 0.2% formic acid as described previously [363] and pooled. Enriched samples were dried in a centrifugal evaporator at 45 °C. Stage tip desalting was carried out as described above for protein-based enrichment sample preparation (see *Whole proteome analysis of Panc-1 clones*). Measurement of the peptide samples: Samples were resuspended in Resuspension Buffer (see *List of Buffers*) and 30% of the peptide samples were injected into the LC-MS/MS machine.

SrtA protein-based enrichment proteomics - Part I: SrtA labelling and two step on-bead/off-bead digestion. This procedure corresponds to the experiment displayed in Figure 3.7. We compared protein-level enrichment of MD-MB-231 triple-negative breast cancer cell proteomes labeled with SrtA post-lysis to metabolically tagged cells with

YnMyr after 24 h exposure to DMSO, 1 nM or 100 nM IMP-1088 NMT inhibitor. Three biological replicates were used per condition, making up a total of 9 samples labelled with SrtA. Cell treatment and sample harvesting: Cells were harvested and lysed in SrtA buffer (see *List of Buffers*). 200 µg protein per sample were subjected to SrtA reaction and precipitated as described above. Protein samples were resuspended in 0.2% SDS, 1 mM DTT, 50 mM Tris pH 8.0 to 1 mg/mL. Bead derivatisation: NeutrAvidin agarose beads (1 µL per µg of SrtA-labeled protein) were washed five times with 5-bead volumes of 100 mM TEAB pH 8.0 to remove all traces of possible free-amines, and lysines on NeutrAvidin di-methylated with 5-bead volumes of 0.2% formaldehyde and 25 mM sodium cyanoborohydride in 100 mM TEAB pH 8.0 for 1 h at RT. The reaction was quenched with 100 mM TEAB pH 8.0 containing 1% ethanolamine. Excess ethanolamine was washed away with 50 mM Tris pH 8.0 and beads equilibrated to 0.2% SDS/50 mM Tris pH 8.0. Protein-based enrichment and two-step on-bead/off-bead digestion: Samples were added to the derivatised beads and incubated for 2 h at RT. The supernatant was discarded, and the beads washed twice with 0.2% SDS/ 50 mM Tris pH 8.0 and three times with 50 mM Tris pH 8.0 to wash out the SDS. All bead washes were performed by centrifugation on a benchtop centrifuge at 4 $\times$ *g* for 3 min. Bound proteins were partially digested by incubation with 0.4 µg LysC (Promega, V1671) in 50 mM Tris pH 8.0 for 1 h at 37 °C. The supernatant containing the cleaved proteins was then transferred to a new tube for full digestion of the proteins with 0.5 µg trypsin overnight at 37 °C in the presence of 5 mM TCEP and 10 mM chloroacetamide. Buffer exchange/desalting on stage tips: Samples were acidified with 0.5% (v/v) trifluoroacetic acid (TFA) and loaded onto stage tips containing three SDB-XC poly(styrenedivinyl-benzene) copolymer discs (Merck). The stage tipping procedure was carried out as described previously(17). Peptide samples were eluted in 60% acetonitrile in water and the solvent removed by incubation on the SpeedVac at 45 °C. 9-plex TMT labeling: The TMT reaction was performed by adding 0.08 mg TMT reagent dissolved in 15 µL acetonitrile to each of the samples resuspended in 15 µL 50 mM HEPES pH 8.0 and incubating the mixture for 2 h at RT. Each reaction was quenched with 1 µL 5% hydroxylamine before combining all nine conditions into one tube. The sample was dried by centrifugal evaporation at 45 °C. In both cases, one of the 129 channels was excluded, as it has previously been shown that that is one of the pairs that suffers greatest reporter ion coalescence issues [400]. Desalting and six-layer fractionation: The peptide sample was first desalted on a stage tip and dried as described previously (see *YnMyr pull down spike-in SILAC of Panc-1 clones*). Samples were resuspended in 150 µL 1% (v/v) TFA/H_2O and loaded into a second stage tip containing SCX stage tips and separated into six fractions as described above (see *Whole proteome analysis of Panc-1 clones* and Table Tab:C6Tab3). Samples were evaporated to dryness on the SpeedVac at 45 °C. Measurement of the peptide sample: Peptide fractions were dissolved in Resuspension Buffer (see *List of Buffers*). 1/6 of fractions 1 and 2, 1/9 of fractions 3 and 4, and 1/12 of fractions 5 and 6 were injected into the LC-MS/MS system.

SrtA protein-based enrichment proteomics - Part II: YnMyr labelling. This procedure corresponds to the experiment displayed in Figure 3.7. This experiment was designed together with the SrtA-labelling protein-level enrichment above in order to compare the two labelling strategies. MD-MB-231 triple-negative breast cancer cells were metabolically tagged with YnMyr after 24 h exposure to DMSO, 1 nM or 100 nM IMP-1088 NMT inhibitor. Three biological replicates were used per condition, making up a total of 9 samples labelled with YnMyr. Cell treatment and sample harvesting: Cells were lysed in Lysis Buffer A (see *List of Buffers*) and protein concentration adjusted to 1 mg/mL. Click reaction: 400 µg of protein were ligated to the AzRB capture reagent by copper-catalyzed alkyne-azide cycloaddition (CuAAC) as described in *Click reaction (CuAAC)*. Protein precipitation: After quenching the click reaction with 5 mM EDTA, samples were precipitated as described previously (see *Methanol-chloroform precipitation of proteins*). Protein samples were resuspended in in 50 mM HEPES pH 8.0 containing 0.2% SDS and 1 mM DTT. Protein-based enrichment and on-bead digestion: 8 µL NeutrAvidin agarose beads and 32 µL agarose blank beads (Pierce Control Agarose Resin, ThermoScientific) were used for every sample. The beads were equilibrated with 0.2% SDS in 50 mM HEPES pH 8.0 before incubation with the samples (2 h, RT). The supernatant was discarded, and the beads washed twice with 0.2% SDS in 50 mM HEPES pH 8.0, and three times with 50 mM HEPES. Samples were then resuspended in 50 µL HEPES buffer containing 5 mM TCEP and 10 mM chloroacetamide and 0.5 µg trypsin added for overnight on-bead digestion at 37°C. 9-plex TMT labeling, desalting and six-layer fractionation and measurement of the samples was performed as described for SrtA samples.

Peptide-level enrichment of proteomes metabolically tagged with AHA for MetAP2 substrate identification (final optimised procedure). This workflow was designed to systematically detect MetAP2-specific substrates by combining metabolic incorporation of the methionine-analogue azidohomoalanine (AHA) and MetAP2 specific inhibitors. The protocol was optimised through several rounds of iteration as thoroughly described in Results Chapter 4. The procedure described in detail here corresponds to the final optimised protocol of the workflow and is the one used in the experiments described in Figures 4.13 and 4.14. Cell treatment and harvesting: 2×10^6 A549 or HT1080 cells were seeded in 10-cm dishes in 2 (experiment in Figure 4.13) or 8 replicates (experiment in Figure 4.14) per condition one day before the start of the experiment. Cells were then washed once with PBS and changed to methionine-free DMEM (Gibco 21013-024; supplemented with 0.2 mM L-Cys (Sigma-Aldrich, C6852), 4 mM L-Gln (Thermo-Fisher Scientific, 25030-024), 1 mM sodium pyruvate (Sigma-Aldrich, S8636) and 8% FBS (all final concentrations) in order to mimic the composition of the original medium cells were in (Gibco 41966-029)) in the presence of MetAP2i or DMSO control. Cells were allowed to equilibrate for 30 min before the addition of 2 mM AHA and pulsed for 2 h before lysis in Lysis Buffer C (see *List of Buffers*). For the experiment in Figure 4.14, two replicate samples of the same condition were combined into a single vial to obtain protein lysate equivalent to two 10-cm dishes (at least 2 mg per

sample) in 4 biological replicates. Click reaction: 1 mg of protein per sample were ligated to Yn-AC-biotin capture reagent (or Yn-PEG4-biotin for preliminary experiments) as described in *Click reaction (CuAAC)*, with a higher final concentration of capture reagent to account for the fact that AHA is incorporated into newly synthesised proteins to a much higher extent than YnMyr. Briefly, a click master mix was prepared by combining 1 volume 50mM capture reagent (Yn-PEG4-biotin or Yn-AC-biotin), 2 volumes 50 mM CuSO4, 2 volumes 50 mM TCEP and 1 volume 10 mM tris(benzyltriazolulmethyl)amine (TBTA). 6 μL of the click mix were then added for every 100 μL protein and samples incubated for 1 h at RT with gentle shaking. 5 mM EDTA were added to quench copper and stop the reaction. For the experiment described in Figure 4.14, biotinylated protein samples were divided in two for digestion with different proteases (trypsin/chymotrypsin). Protein precipitation was carried out as described previously (see *Methanol-chloroform precipitation of proteins*) and protein samples re-dissolved first in 20 μL 50 mM HEPES pH 8.0 containing 2% SDS and then diluted down to 0.1% SDS with 50 mM HEPES pH 8.0 for digestion. In solution reduction, alkylation and digestion: final 5 mM TCEP and 10 mM chloroacetamide (both from 10X stocks prepared fresh in 50 mM HEPES pH 8.0) were added and samples incubated for 15 min at RT with gentle shaking. 1 μg trypsin or 1.7 μg chymotrypsin (Promega, V1061) were added per starting 100 μg protein (1:100 ratio for trypsin, 1:60 for chymotrypsin) and digestion allowed to carry on overnight at 37 °C (trypsin) or 25 °C (chymotrypsin). Proteases were quenched by addition of 1 mM PMSF (from a 200 mM ethanol stock, prepared fresh). Peptide-level pull down: 100 μL Pierce NeutrAvidin agarose bead slurry per condition were transferred into 1.5 mL vials and washed with 1 mL 50 mM HEPES pH 8.0 once. All bead washes were performed by centrifugation on a benchtop centrifuge at 4 × *g* for 3 min. Peptide samples were centrifuged at maximum speed for 5 min at 4 °C before transferring the samples into the corresponding bead-containing tube. Biotinylated peptides were allowed to bind for 2 h at RT with gentle shaking before extensive washing to remove non-specific binders: two washes with 700 μL 50 mM HEPES pH 8.0, two washes with H_2O and three washes with with 10% acetonitrile in H_2O. Peptides were eluted from the beads with a 15 min incubation in 150 μL of 80% acetonitrile,0.1% TFA and 0.2% formic acid as described previously [363]. Beads were pelleted by centrifugation and the supernatant containing the eluted peptides transferred to a new 1.5 mL microcentrifuge tube containing a glass-insert. This step was repeated once more and both supernatants combined. Enriched samples were dried on the SpeedVac overnight with no temperature. Cleavage of Yn-AC-biotin capture reagent: 30 μL TFA were added to each dried sample (in glass inserts) and incubated for 2 h at RT with gentle shaking. TFA was dried under N_2 gas stream (ca. 2 min per sample) followed by further 30 min in the centrifugal evaporator at 45 °C to ensure complete removal. TMT labelling: samples were redissolved in 15 μL 50 mM HEPES pH 8.0 and pH of every sample checked by addition of 1 μL into a pH strip. In the rare event of a sample showing acidic pH, this was reverted by addition of 1 μL of 1 M HEPES pH 8.0. The equivalent of 1/3 of a 0.8 mg TMT reagent vial was resuspended in 15 μL acetonitrile and the samples and TMT reagents mixed 1:1 (final 50%

acetonitrile). The TMT reaction was allowed to carry on for 3 h at RT with gentle mixing before quenching with 3 µL 5% hydroxylamine (diluted in 50 mM HEPES pH 8.0). Where any TMT channels had to be excluded, one of the 130 and 129 channels were selected following previous evidence that these channels are more prone to coalesce [400]. Samples were then combined into their corresponding TMT multiplex and dried on the SpeedVac overnight with no temperature. Six-layer fractionation: Dried TMT-labelled peptide samples were redissolved in 150 µL 1% TFA, their pH checked to be pH<3 and loaded into SCX (Supelco) stage tips for 6-layer fractionation. 6-layer fractionation was carried out as described in *Whole proteome analysis of Panc-1 clones*, with slight modification of the elution buffers (Table 6.4). Fractionated peptide samples were then dried on the SpeedVac at 45 °C. Desalting: Due to the high ammonium acetate used for fractionation, samples were desalted prior to injection as described in *YnMyr pull down spike-in SILAC of Panc-1 clones* and stored at -80 °C until submission. Measurement of the peptide sample: For injection, samples were redissolved in 10 µL Resuspension Buffer (see *List of Buffers*) and 2 µL injected for all fractions.

Table 6.4: Six-layer fractionation elution buffers (AHA).

Buffer	Composition
SCXx1	200 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx2	350 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx3	450 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx4	500 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
SCXx5	550 mM Ammonium acetate, 20% (v/v) ACN, 0.5% (v/v) FA
Buffer X	5% (v/v) Ammonium hydroxide, 80% (v/v) ACN

* ACN, acetonitrile. FA, formic acid.

Protein-level enrichment of YnMyr metabolic tagging in MetAP2i-treated samples. This procedure corresponds to the experiment displayed in Figure 4.17 and was designed to study the effect of MetAP2 inhibition on YnMyr incorporation to determine the overlap between NMT and MetAP2. An NMT inhibitor, IMP-1088 [298], was included as positive control. Cell treatment and sample harvesting: 1.2×10^6 HT1080 cells were seeded in 10-cm dishes in 4 replicates per condition and allowed to grow overnight. Cells were then pre-incubated with NMTi, MetAP2i or DMSO for 30 min, before addition of 25 µM YnMyr. Cells were allowed to incorporate YnMyr in the presence of NMTi/MetAP2i for 4.5 h before lysis in Lysis Buffer C (see *List of Buffers*). Click reaction: 900 µg of protein per sample were ligated to the AzRB capture reagent as described in *Click reaction (CuAAC)*. Protein precipitation: After quenching the click reaction with 5 mM EDTA, samples were precipitated as described previously (see *Methanol-chloroform precipitation of proteins*). Bead derivatisation: 30 µ NeutrAvidin agarose bead slurry per condition were dimethylated as described above (see *SrtA protein-based enrichment proteomics*) one day before use and kept at 4 °C overnight. Protein-based enrichment and two-step on-bead/off-bead digestion: Derivatised beads were washed twice with 1 mL 50 mM HEPES pH 8.0 containing 0.2% SDS before addition of the samples and 2 h incubation at RT with gentle shaking. The supernatant was discarded,

and the beads washed once with 1% SDS/50 mM HEPES pH 8.0, twice with 0.2% SDS/50 mM HEPES pH 8.0 and three times with 50 mM HEPES to wash out the SDS. All bead washes were performed by centrifugation on a benchtop centrifuge at 4 $\times$ *g* for 3 min. Bound proteins were resuspended in 20 µL 50 mM HEPES pH 8.0 and partially digested by incubation with 0.4 µg LysC (Promega, V1671) for 2 h at 37 °C. The supernatant containing the cleaved proteins was then transferred to a new tube for full digestion with 0.5 µg trypsin overnight at 37 °C in the presence of 5 mM TCEP and 10 mM chloroacetamide. Trypsin was quenched by acidification of the sample with addition of 1 % TFA. On-stage tip TMT labelling: Stage tips were prepared by inserting three stacked discs of C18 membranes into 200 µL tips. Membranes were activated with 50 µL methanol, washed once with 50 µL 50% acetonitrile containing 0.1% formic acid and equilibrated with 75 µL H_2O containing 0.1% formic acid. All solutions were passed through the stage tip by centrifugation at 1,100 $\times$ *g* for 3 min at RT. Samples were centrifuged on a benchtop centrifuge at maximum speed for 5 min before addition of the supernatant to their corresponding stage tip, and washed twice with H_2O containing 0.1% formic acid. 1/10 of each 0.8 mg TMT reagent vial were dissolved in 2 µL acetonitrile and this added into 200 µL 50 mM HEPES pH 8.0. The TMT-reagent solution was then passed through the membrane for at least 15 min at 100 $\times$ *g*. Stage tips were then washed with H_2O containing 0.1% formic acid three times before elution with 50 µL of 60% acetonitrile containing 0.1% formic acid. TMT channels were combined and dried on the SpeedVac at 45 °C. Three-layer fractionation: Stage tips were prepared containing 3 stacks of polystyrene-divinylbenzene copolymer partially modified with sulfonic acid SDB-RPS discs (Supelco). Membranes were equilibrated with 150 µL H_2O. Dry TMT-labelled samples were redissolved in 150 µL H_2O containing 1% TFA and centrifuged for 5 min at maximum speed on a benchtop centrifuge. Supernatants were then added to each corresponding stage tip and passed through the membrane by centrifuging at 1,100 $\times$ *g*. Stage tips were then washed three times with 60 µL 0.2% TFA before elution with SDB-RPSx1, SDB-RPSx2 and Buffer X elution buffers (see Table 6.5). Fractionated samples were dried on the SpeedVac at 45 °C. Measurement of the peptide sample: Samples were redissolved in 15 µL Resuspension Buffer (see *List of Buffers*) and 2 µL injected for all fractions.

Table 6.5: Six-layer fractionation elution buffers (AHA).

Buffer	Composition
SCXx1	100 mM Ammonium formate, 40% (v/v) ACN, 0.5% (v/v) FA
SCXx2	150 mM Ammonium formate, 60% (v/v) ACN, 0.5% (v/v) FA
Buffer X	5% (v/v) Ammonium hydroxide, 80% (v/v) ACN

* ACN, acetonitrile. FA, formic acid.

Proteomics data acquisition of spike-in SILAC samples. These were the settings used for the YnMyr pull down experiment with the NMT1- and NMT2-KO Panc-1 clones. Peptides were separated on an Acclaim PepMap RSLC column 50 cm $\times$ 75 µm inner diameter (Thermo Fisher Scientific) using a 120 min acetonitrile gradient in 0.1% aqueous formic acid at a flow rate of 250 nl/min. Easy nLC-1000 was coupled to a QExactive mass spectrometer via an easy-spray source (all Thermo Fisher Scientific). The QExactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000 at m/z 200. Scans were acquired from 350 to 1650 m/z. The maximum ion injection time for the survey scan was set to 20 ms and the ion target to 1×10^6. Up to 10 of the most abundant isotope patterns from the survey scan with charge $+2$ or higher and an intensity of 1×10^2 or higher were selected with an isolation window of 2.0 m/z and fragmented by HCD with normalised collision energy of 25. The MS/MS scans were acquired with a resolution of 17,500 at m/z 200, a maximum ion injection time of 120 ms and the ion target value set to 1×10^5.

Proteomics data acquisition of LFQ samples. These were the settings used for the SrtA-labelled peptide-enrichment experiment of MDA-MB-231 proteomes. – Peptides were separated using an Acclaim PepMap RSLC column 50 cm $\times$ 75 µm inner diameter (Thermo Fisher Scientific) using a 2 h acetonitrile gradient in 0.1% aqueous formic acid at a flow rate of 250 nl/min. Easy nLC-1000 was coupled to a QExactive mass spectrometer via an easy-spray source (all Thermo Fisher Scientific). The QExactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000 at m/z 200. Scans were acquired from 350 to 1650 m/z. Up to 10 of the most abundant isotope patterns with charge $+2$ or higher from the survey scan were selected with an isolation window of 2.0 m/z and fragmented by HCD with normalized collision energy of 25. The maximum ion injection times for the survey scan and the MS/MS scans (acquired with a resolution of 17,500 at m/z 200) were 20 and 120 ms, respectively. The ion target value for MS was set to 10^6 and for MS/MS to 10^5, and the intensity threshold was set to 8.3×10^2.

Proteomics data acquisition of TMT-labelled samples. These settings were applied to all the rest of the samples measured that had been labelled with TMT isobaric reagents. Peptides were separated on an Acclaim PepMap RSLC column 50 cm $\times$ 75 µm inner diameter (Thermo Fisher Scientific) using a 180 min acetonitrile gradient in 0.1% aqueous formic acid at a flow rate of 250 nl/min. Easy nLC-1000 was coupled to a QExactive mass spectrometer via an easy-spray source (all Thermo Fisher Scientific). The QExactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000 at m/z 200. Scans were acquired from 350 to 1800 m/z. The maximum ion injection time for the survey scan was set to 20 ms and the ion target to 1×10^6. Up to 10 of the most abundant isotope patterns from the survey scan with charge $+2$ or higher and an intensity

of 1.7 $\times$ 10^3 or higher were selected with an isolation window of 1.6 m/z and fragmented by HCD with normalised collision energy of 31. The MS/MS scans were acquired with a resolution of 35,000 at m/z 200, a maximum ion injection time of 120 ms and the ion target value set to 2 $\times$ 10^5.

6.8 Database searches of proteomics experiments

The general database search parameters have been combined here for each quantification method. Specific changes to these parameters are listed in Tables 6.6 and 6.7 below, together with the specific versions of MaxQuant and Perseus as well as fasta files accession dates.

Protein database search of spike-in SILAC samples using MaxQuant. Raw files were uploaded into MaxQuant [401] and searched against the curated Swiss-Prot human proteome (with isoforms) [251] using the built-in Andromeda search engine. Quantification type was set to "Standard" and the multiplicity set to two, with the "Light" label as default and Arg10 and Lys8 selected for the "Heavy" label. The false discovery rate was set to 0.01 for peptides, proteins, and sites. Other parameters were used as preset in the software (maximal mass error = 4.5 ppm and 20 ppm for precursor and product ions, respectively, minimum peptide length = 7, minimum razor + unique peptides = 2, minimum scores for unmodified and modified peptides = 0 and 40, respectively). "Re-quantify" was turned on, which gives a background value to the heavy/light counterpart when it fell under the detection limit, allowing to still get a H/L ratio value, even if this will be under/over-estimated. Second peptides were allowed, to maximize identification of peptides within the same MS1 spectrum. "Match between runs" option (time window 0.7 min) was allowed to allow for feature ID transfer between MS runs. and "Unique and razor peptides" mode was selected to allow identification and quantification of proteins in groups (razor peptides are uniquely assigned to protein groups and not to individual proteins).

Database search LFQ samples using MaxQuant. The search was performed in MaxQuant [401] using the built-in Andromeda search engine against the curated Swissport human proteome (with isoforms) [251]. The false discovery rate was set to 0.01 for peptides, proteins, and sites. The minimum scores for unmodified and modified peptides were set to 0 and 25, respectively. Other parameters were used as preset in the software (maximal mass error = 4.5 ppm and 20 ppm for precursor and product ions, respectively, minimum peptide length = 7, minimum razor + unique peptides = 1). Identification of second peptides was allowed and "Unique and razor peptides" mode was selected to allow identification and quantification of proteins in groups. Special case: *SrtA peptide-based enrichment proteomics*. In this case the search was performed against an adapted version of the Swissprot human proteome [251], which was modified using R [402] so that the first two "MX" residues of every protein sequence were replaced by "MALPETX". This allows for the automated assignment

of ions resulted from HCD-induced fragmentation of the peptidic element of the N-terminal modification. We further allowed a biotin modification at the N-terminus of any protein (see 6.8 for details on the formula or exact mass) that we set as a variable modification to act as an internal control that would reinforce the validity of the identification by combination of the two independently found elements of Biotin-ALPET.

General database search of TMT-labelled protein-centric samples using MaxQuant. This consists of the 'default' settings at which MaxQuant was operated for all TMT-labelling proteomics experiments not listed in this section otherwise. The list of MaxQuant versions and fasta file download dates as well as all variable and fixed modifications and any further changes to these settings for each particular experiment are listed in Tables 6.6 and 6.7. Raw files were uploaded into MaxQuant [401] and searched against the curated Swiss-Prot human proteome (with isoforms) [251] using the built-in Andromeda search engine. Quantification method was set to "Reporter ion MS2" and "10plex TMT" selected for every N-termini and lysine side-chains. TMT correction factors were specified for every TMT reagent set used. Cysteine carbamidomethylation was selected as fixed modification and methionine oxidation and acetylation of protein N-termini as variable modifications. Trypsin/P was selected as the in silico digest protease, which allows for cleavage after arginine or lysine followed by any amino acid. Up to two missed cleavages were allowed. The false discovery rate was set to 0.01 for peptides, proteins, and sites. Other parameters were used as preset in the software (maximal mass error = 4.5 ppm and 20 ppm for precursor and product ions, respectively, minimum peptide length = 7, minimum razor + unique peptides = 2, minimum scores for unmodified and modified peptides = 0 and 40, respectively). "Match between runs" option (time window 0.7 min) was allowed and "Unique and razor peptides" mode was selected. For TMT quantification (MS2 mode) the minimal ratio count = 2 was selected. Second peptides were not allowed.

TMT labelling efficiency search. For every TMT experiment performed, a second search was carried out to check for TMT labelling efficiency and the quantification of the experiment only assumed valid if the TMT efficiency was >97% across TMT labelling sites. For each search, the same fasta file as in the quantificatio ('quant') search was used. In group-specific parameters, quantification type 'Standard' was selected. Cysteine carbamidomethylation and any other fixed modifications selected for that experiment 'quant' search were set as fixed modifications. Any modification set as variable in the 'quant' search was also selected here as variable modification, together with TMT10plex-Lys and TMT10plex-Nterm. These custom modifications were generated as duplicates from one of the corresponding TMTplex channels already built in within MaxQuant, converting tthem to type 'Standard' and removing the diagnostic peak corresponding to that reporter ion to make it general for all TMT channels. For TMT10plex-Lys, position was set to 'Anywhere' and specificities, 'K'. For TMT10plex-Nterm, position was set to 'Any N-term' and specificities,

'-'. These modifications are also specified in Table 6.8. For AHA peptide-level enrichment experiments, where TMT was also reacted with the cleaved adduct of the Yn-AC-biotin capture reagent, these were also added as variable modifications (Table 6.8). Digestion was set up as for the corresponding 'quant' search. Label-free quantification was set as 'None' to increase the speed of the search. Second peptides were disabled and match between runs was allowed. The rest of the parameters were set as default. The calculation on TMT labelling efficiency were based on the number of identifications with TMT label present/absent (*evidence.txt* output file).

Table 6.6: MaxQuant, Perseus and fasta file versions as well as *in silico* digestion protease used in each search.

Chapter	Experiment	MaxQuant version	Perseus	Uniprot accession date	Number of entries	Protease
2 NMT1- and NMT2-KOs	YnMyr enrichment	1.6.1.0	1.6.0.2	Dec-17	42,326	Trypsin/P (2 missed)
	Whole Proteome	1.6.1.0	1.6.0.2	Dec-17	42,326	Trypsin/P (2 missed)
3 SrtA-based NMT activity biomarker assay	SrtA peptide-level enrichment	1.6.1.0	1.6.0.2	Dec-17	42,326	Trypsin/P (2 missed)
	SrtA protein-level enrichment	1.6.1.0	1.6.0.2	Dec-17	42,326	Trypsin/P (2 missed)
	YnMyr protein-level enrichment	1.6.1.0	1.6.0.2	Dec-17	42,326	Trypsin/P (2 missed)
4 MetAP2 substrate identification	AHA peptide-level enrichment (Trypsin)	1.6.10.43	1.6.0.2	May-19	42,425	Trypsin/P (2 missed)
	AHA peptide-level enrichment (Chymotrypsin)	1.6.10.43	1.6.0.2	May-19	42,425	Chymotrypsin (4 missed)
	YnMyr protein-level enrichment	1.6.10.43	1.6.0.2	May-19	42,425	Trypsin/P (2 missed)

Table 6.7: Main parameters selected for each search including the quantification method, fixed and variable modifications as well as requantify and second peptides.

Chapter	Experiment	Quantification Method	Fixed modifications	Variable modifications	Re-quantify	Second peptides
2 NMT1- and NMT2-KOs	YnMyr enrichment	SILAC	Cys carbamido-methylation	Met oxidation YnMyr-AzRB	Yes	Yes
	Whole Proteome	TMT10plex (Lys and N-termini)	Cys carbamido-methylation	Met oxidation Acetyl (Prot N-termini) Myristic acid	No	No
3 SrtA-based NMT activity biomarker assay	SrtA peptide-level enrichment	LFQ	Cys carbamido-methylation	Met oxidation Acetyl (prot N-term) Biotin (Prot N-term)	No	Yes
	SrtA protein-level enrichment	TMT9plex (Lys and N-termini)	Cys carbamido-methylation	Met oxidation Acetyl (prot N-term) Biotin (Prot N-term Gly)	No	No
	YnMyr protein-level enrichment	TMT9plex (Lys and N-termini)	Cys carbamido-methylation	Met oxidation Acetyl (prot N-term) YnMyr-AzRB	No	No
4 MetAP2 substrate identification	AHA peptide-level enrichment (Trypsin)	TMT8plex (AHA+Yn-AC adduct, Lys and N-termini)	Cys carbamido-methylation	Acetyl (prot N-term)	No	No
	AHA peptide-level enrichment (Chymotrypsin)	TMT8plex (AHA+Yn-AC adduct, Lys and N-termini)	Cys carbamido-methylation	Acetyl (prot N-term)	No	No
	YnMyr protein-level enrichment	TMT10plex (Lys and N-termini)	Cys carbamido-methylation	Met oxidation Acetyl (prot N-term) YnMyr-AzRB	No	No

Table 6.8: List of custom modifications used in proteomics searches, with their chemical formula and exact mass.

Search type	Modification	Position, Specificities	Type	Formula	Exact mass
Quant	YnMyr-AzRB	Prot N-term, "G"	Standard	H(37) O(4) C(22) N(7)	463.2907027125
Quant	Biotin (before ALPET-)	Any N-term, "N"	Standard	H(14) O(2) C(10) N(2) S	226.077598394
Quant	Met AHA +Yn-AC+TMT *	Anywhere, "M"	Isobaric label	H(22) O(2) C(10) Cx(4) N(5) Nx S(-1)	279.218807536
TMT efficiency	TMT10plex-Lys	Anywhere, "K"	Standard	H(20) C(8) Cx(4) N O(2) Nx	229.162932141
TMT efficiency	TMT10plex-Nterm	Any N-term, "-"	Standard	H(20) C(8) Cx(4) N O(2) Nx	229.162932141
TMT efficiency	Met AHA +Yn-AC+TMT *	Anywhere, "M"	Standard	H(22) O(2) C(10) Cx(4) N(5) Nx S(-1)	279.218807536

* Generated one modification per channel, adding diagnostic peaks on corresponding reporter ion.

YnMyr pull down spike-in SILAC of Panc-1 clones. Data analysis was performed using Perseus (version 1.6.0.2) [403]. Data processing: MaxQuant *proteinGroups.txt* output files (normalised H/L ratio) were filtered against contaminants, reverse and proteins identified only by site. Only proteins (rows) containing 8 valid values or more were kept for further analysis. Ratios were inverted (transform "1/X") to reflect the value of Light(sample)/Heavy(spike-in). Because the light(sample) to spike in ratio had been 1:3, L/H ratios were then divided by 3 (transform "X/3"). Base 2 logarithm was applied to all values before column-wise median subtraction. Proteins were annotated with previous evidence of myristoylation [70, 244, 336]. Statistical analysis: Both NMT1-KO (AG3D and AG3E) and NMT2-KO (AG4C and AG6B) clones were combined and treated as two single biological sample types in order to extract the changes to the proteome made by genetic ablation of either NMT1 or NMT2 and to minimise off-target or adaptational effects. NMT1- or NMT2-KOs were compared to CTRL cells using a student t-test (Permutation-based FDR = 0.05; $S_0 = 1$) and the results plotted as volcano plots (x axis: mean fold change, y-axis: statistical significance (p-value)). Panc-1 fold changes were compared to HeLa fold changes (calculated following the exact same procedure) in a correlation plot. Correlation of YnMyr enrichment fold-changes (this experiment) with whole proteome protein levels (Whole proteome analysis of Panc-1 clones experiment) was studied.

Whole proteome analysis of Panc-1 clones. Data analysis was performed using Perseus (version 1.6.0.2) [403]. Data processing: MaxQuant *proteinGroups.txt* output files were filtered against contaminants, reverse and proteins identified only by site. Base 2 logarithm (log2(x)) was applied to all measured values. Each sample (column) was categorically annotated for condition (clone type or CTRL) and TMTplex run ("1" or "2"). For every peptide (row), the median across all conditions within each TMTplex was subtracted to every condition in that TMTplex to remove multiplicational bias arisen from separate runs and bring conditions across TMTplexes to a common comparable baseline. This was followed by column-wise median subtraction to all peptides in that column, which renders the distribution of peptides aligned (centred around "0") for every condition. This normalisation step is performed to take into account differences in overall peptide levels coming from differences in digestion or overall peptide loses throughout the sample preparation procedure. Statistical analysis: Both NMT1-KO (AG3D and AG3E) and NMT2-KO (AG4C and AG6B) clones were combined and treated as two single biological sample types in order to extract the changes to the proteome made by genetic ablation of either NMT1 or NMT2 and to minimise off-target or adaptational effects. NMT1- or NMT2-KOs were compared to CTRL cells using a student t-test (Permutation-based FDR = 0.05; $S_0 = 1$) and the results plotted as volcano plots (x axis: mean fold change, y-axis: statistical significance (p-value)). Proteins were annotated for their corresponding gene ontology (molecular function (GOMF),

biological process (GOBF) and cellular compartment (GOCC)) and KEGG pathway terms and subjected to 1D or 2D annotation enrichment analysis based on the fold changes calculated on the previous step (FDR=0.02, calculated based on the Benjamini Hochberg correction algorithm). Annotation enrichment analyses test whether there is a systematic bias for each annotation term (i.e. GO or KEGG terms) to be larger or smaller compared to the overall distribution across the whole dataset. Significantly changing terms were grouped manually into larger biological categories before visualisation. The difference between 1D or 2D is that in the 1D enrichment analysis, only fold-changes of one sample type are used for the analysis (e.g. NMT1-KO - CTRL fold change) whereas 2D analyses are performed on the fold-changes of two samples (e.g. NMT1-KO FC and NMT2-KO FC, each calculated against control), being the output the correlation between the two enrichment scores.

SrtA peptide-based enrichment proteomics. Data analysis was performed using Perseus (version 1.6.0.2) [403], into which MaxQuant *peptides.txt* output file was loaded. Data processing: Data were filtered against contaminants and reverse identifications. Only peptides with at least 2 out of 3 valid values and containing the N-terminal biotinylation followed by the ALPET- sequence were kept. Sequence analysis: To generate sequence logos all input sequences must have the same length. Therefore, each identified peptide was annotated with the first 10 amino acid sequence of the protein isoform it had been mapped to. To ensure quality of the sequence analysis, every single peptide spectrum was inspected manually, leaving only the spectra that showed at least three b- and y-ions supporting peptide backbone fragmentation in the sequence region following the added "ALPET". Peptides unique to a particular protein group were kept and in cases were peptides matched to several proteins within a protein group, the first 10 amino acid sequences of all matched proteins were inspected manually, only to keep the IDs that contained a single, unambiguous N-terminal amino-acid sequence across matching proteins within the group. N-terminal 10 amino acid sequence duplicates were removed before creating N-terminal sequence logos using iceLogo [365].

SrtA protein enrichment datasets (SrtA-labelling and YnMyr tagging experiments). Data analysis was performed using Perseus (version 1.6.0.2) [403]. Data processing: MaxQuant *proteinGroups.txt* output files were filtered against contaminants, reverse and proteins identified only by site. Base 2 logarithm was applied to all measurements and the median values within each column (TMT channel) subtracted. Protein groups with at least three valid values were kept. Protein groups were annotated with previous evidence of myristoylation [70, 244, 336] and nature of the 2nd amino acid position (Gly or other) according to their isoform protein ID.Statistical analysis: An ANOVA test (Permutation-based FDR = 0.01; $S_0 = 1$) was applied to all proteins in the dataset and results analysed according to their statistical significance and myristoylation evidence.

AHA peptide-level enrichment proteomics experiments for identification of MetAP2 substrates. Data analysis was performed using Perseus (version 1.6.0.2) [403]. Data processing: MaxQuant *peptides.txt* output files were filtered against contaminants and reverse, before $\log_2$ transformation (log2(x)). In the cases were there was more than one TMT multiplexed sample, each sample (column) was categorically annotated for condition (MetAP2i type or DMSO) and TMTplex run ("1" or "2"). Only peptides that contained at least one Met (AHA) and that had at least 8 valid values were kept. If there was more than one TMT multiplexed sample in that experiment, for every peptide (row), the median across all conditions within each TMTplex was subtracted to every condition in that TMTplex. Column-wise median subtraction was then applied to all peptides in that column (see *Whole proteome analysis of Panc-1 clones* for explanations on this normalisation procedure). Once the dataset was normalised and all conditions were comparable, all peptides that were not N-terminal were filtered out by either filtering in rows with "Start position" of "1" (iMet retained) or "2" (iMet cleaved). Statistical analysis: N-terminal peptides were then subjected to an ANOVA test (Permutation-based FDR = 0.01 and $S_0 = 0.5$) and results analysed according to their statistical significance and iMet status.

Sequence Logo: Significant peptide hits from different experiments were combined into a single list of MetAP2 substrates and duplicates removed. Frequencies of the first 10 amino acid residues after iMet removal against the general frequencies for the same positions across the proteome were compared using iceLogo [365]. Classification of MetAP2 substrates according to their protein class: The combined list of MetAP2 substrates (78 substrates) were classified according to their protein class (derived from Gene Ontology (GO) Molecular Function (GOMF) terms) using PANTHER gene list analysis [394]. The gene names of the 78 substrates were used as query input and 'Functional classification viewed in graphic charts' was selected. Data was exported and unnanotated terms manually manually annotated based on the function listed in Uniprot [70] and their most prevalent GOMF annotations in QuickGO [396]. Functional over-representation analysis: To get functional insight of the major biological processes the list of MetAP2 substrates might be involved in, gPROFILER functional over-representation analysis [395] was used. The list of gene names was used as input query and 'g:SCS' was selected as the multiple testing correction method, with threshold set to 0.05. The output of gPROFILER over-representation test for GO Biological Process (GOBP) terms was then exported into a GEM file, which was then uploaded into the Cytoscape [397] EnrichmentMap App for visualisation.

YnMyr protein-level enrichment of samples treated with MetAP2i. Data analysis was performed using Perseus (version 1.6.0.2) [403]. Data processing: MaxQuant *proteinGroups.txt* output files were filtered against contaminants, reverse and proteins identified only by site. Base 2 logarithm was applied to all measurements. Each sample (column) was categorically annotated for condition (NMTi, MetAP2i type or DMSO) and TMTplex run ("1" or "2"). For every peptide (row), the median across all conditions within each TMTplex was

subtracted to every condition in that TMTplex, followed by column-wise median subtraction to all peptides in that column (see *Whole proteome analysis of Panc-1 clones* for explanations on this normalisation procedure). Only peptides that contained at least 10 valid values were kept. Protein groups were then annotated with previous evidence of myristoylation [70, 244, 336] and their presence in our putative MetAP2 substrate list according to their isoform protein ID. Statistical analysis: An ANOVA test (Permutation-based FDR = 0.05; $S_0 = 1$) was applied to all proteins in the dataset and results analysed according to their statistical significance and myristoylation evidence.

6.10 Miscellaneous

6.10.1 List of Buffers

- *Blocking buffer A:* 5% (w/v) semi skimmed milk in TBST
- *Blocking buffer B:* 3% (w/v) BSA in TBST
- *ELISA wash buffer:* 0.1% (w/v) BSA, 0.05% (v/v) Tween20, 50 mM NaCl, 25 mM Tris-HCl pH 7.2
- *FACS buffer:* 0.9% FCS in PBS
- *Lysis buffer A:* 1% (v/v) Triton X-100, 0.1% (w/v) SDS, PBS containg freshly added EDTA-free protease inhibitors (Roche)
- *Lysis buffer B:* 1% triton-X, 0.1% SDS, 50 mM HEPES pH 8.0 containg freshly added EDTA-free protease inhibitors (Roche)
- *Lysis buffer C:* 1% triton-X, 0.25% SDS, PBS containg 2 U benzonase and freshly added EDTA-free protease inhibitors (Roche)
- *Protein loading buffer (SDS-PAGE)(4X stock):* A 5X stock needs to be prepared first, which contains 0.05% (w/v) bromophenol blue, 30% (v/v) glycerol and 10% (w/v) SDS in 250 mM Tris-HCl pH 6.8. This stock needs to be freshly diluted with β-mercaptoethanol (reducing agent) in a 4:1 (loading buffer-to-β-mercaptoethanol) ratio to provide the working 4X Protein loading buffer.
- *Resuspension Buffer (for injection into the LC-MS/MS):* LC-MS grade H_2O containing 2% (v/v) acetonitrile and 0.5% (v/v) TFA
- *SDS-PAGE running buffer (10X):* 1.92 M glycine, 1% (w/v) SDS, 250 mM Tris (base) (pH around 8.3, no adjustment needed)
- *SrtA Buffer or SrtA reaction buffer:* 10 mM $CaCl_2$, 150 mM NaCl, 50 mM Tris-HCl pH 7.5 (EDTA-free protease inhibitors added when used as lysis buffer for mechanical lysis)

- *SrtA storage buffer:* 10% (v/v) glycerol, 150 mM NaCl, 20 mM Tris-HCl pH 8.0
- *Mild stripping buffer:* 0.1% (w/v) SDS, 1% Tween20, 1.5% (w/v) glycine (pH 2.2)
- *Tris buffered saline (TBS):* 150 mM NaCl, 50 mM Tris-HCl pH 7.4
- *TBST:* Tris buffered saline containing 0.1% (v/v) Tween20
- *Transfer buffer:* 20% (v/v) methanol, 192 mM glycine, 25 mM Tris (base) (no pH adjustment needed)

6.10.2 List of Antibodies

Table 6.9: List of antibodies used with their source codes and dilutions used.

Target	Mw (kDa)	Source	Code	Host	Type	Dilution
β-actin	42	Abcam	ab6276	Rabbit	Polyclonal	1:5000
ARL1	20	Proteintech	16012-1-AP	Rabbit	Polyclonal	1:500
GAPDH	36	Abcam	Ab9485	Rabbit	Polyclonal	1:5000
MetAP2	52 (detect: 67)	Abcam	ab124953	Rabbit	Monoclonal	1:1000
NMT1	57	Prestige (Atlas)	HPA022963	Rabbitt	Polyclonal	1:1000
NMT2	57	BD Biosciences	611310	Mouse	Monoclonal	1:1000
PRKACA	40	Cell Signalling	5842S/D38C6	Rabbit	Monoclonal	1:1000
YES1	60	Abcam	ab133314	Rabbit	Monoclonal	1:1000
Mouse IgG	-	Advansta	R-05071-500	Goat	HRP-conjugate	1:10,000
Rabbitt IgG	-	Advansta	R-05072-500	Goat	HRP-conjugate	1:10,000

References

1. Pray, L. A. Eukaryotic Genome Complexity. *Nat. Educ.* **1,** 1 (2008).
2. Reyes, A. & Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **46,** 582–592 (2018).
3. Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am. J. Hum. Genet.* **102,** 11–26 (2018).
4. Laurencikiene, J., Källman, A. M., Fong, N., Bentley, D. L. & Öhman, M. RNA editing and alternative splicing: the importance of co-transcriptional coordination. *EMBO Rep.* **7,** 303 (2006).
5. Bazykin, G. A. & Kochetov, A. V. Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res.* **39,** 567–77 (2011).
6. Giglione, C., Fieulaine, S. & Meinnel, T. N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie* **114,** 134–146 (2015).
7. Minguez, P., Parca, L., Diella, F., Mende, D. R., Kumar, R., Helmer-Citterich, M., Gavin, A., van Noort, V. & Bork, P. Deciphering a global network of functionally associated post-translational modifications. *Mol. Syst. Biol.* **8,** 599 (2012).
8. McCartney, R. R., Garnar-Wortzel, L., Chandrashekarappa, D. G. & Schmidt, M. C. Activation and inhibition of Snf1 kinase activity by phosphorylation within the activation loop. *Biochim. Biophys. Acta - Proteins Proteomics* **1864,** 1518–1528 (2016).
9. Wright, M. H., Heal, W. P., Mann, D. J. & Tate, E. W. Protein myristoylation in health and disease. *J. Chem. Biol.* **3,** 19–35 (2010).
10. Ponomarenko, E. A., Poverennaya, E. V., Ilgisonis, E. V., Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., Lisitsa, A. V. & Archakov, A. I. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016,** 7436849 (2016).
11. Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **10,** 186–187 (2013).
12. Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., Kiessling, L. L., Krogan, N. J., Larsen, M. R., Loo, J. A., Ogorzalek Loo, R. R., Lundberg, E., MacCoss, M. J., Mallick, P., Mootha, V. K.,

Mrksich, M., Muir, T. W., Patrie, S. M., Pesavento, J. J., Pitteri, S. J., Rodriguez, H., Saghatelian, A., Sandoval, W., Schlüter, H., Sechi, S., Slavoff, S. A., Smith, L. M., Snyder, M. P., Thomas, P. M., Uhlén, M., Van Eyk, J. E., Vidal, M., Walt, D. R., White, F. M., Williams, E. R., Wohlschlager, T., Wysocki, V. H., Yates, N. A., Young, N. L. & Zhang, B. How many human proteoforms are there? *Nat. Chem. Biol.* **14,** 206–214 (2018).

13. Raue, U., Oellerer, S. & Rospert, S. Association of Protein Biogenesis Factors at the Yeast Ribosomal Tunnel Exit Is Affected by the Translational Status and Nascent Polypeptide Sequence. *J. Biol. Chem.* **282,** 7809–7816 (2007).

14. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16,** 589–597 (2009).

15. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci.* **116,** 358–366 (2019).

16. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61,** 1455–1474 (2004).

17. Polevoda, B. & Sherman, F. N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins. *J. Mol. Biol.* **325,** 595–622 (2003).

18. Lee, S. H., Kyung, H., Yokota, R., Goto, T. & Oe, T. N-Terminal α-Ketoamide Peptides: Formation and Transamination. *Chem. Res. Toxicol.* **27,** 637–648 (2014).

19. Lee, S. H., Kyung, H., Yokota, R., Goto, T. & Oe, T. Hydroxyl Radical-Mediated Novel Modification of Peptides: N-Terminal Cyclization through the Formation of α-Ketoamide. *Chem. Res. Toxicol.* **28,** 59–70 (2015).

20. Kajita, R., Goto, T., Lee, S. H. & Oe, T. Aldehyde Stress-Mediated Novel Modification of Proteins: Epimerization of the N-Terminal Amino Acid. *Chem. Res. Toxicol.* **26,** 1926–1936 (2013).

21. Masson, N., Keeley, T. P., Giuntoli, B., White, M. D., Puerta, M. L., Perata, P., Hopkinson, R. J., Flashman, E., Licausi, F. & Ratcliffe, P. J. Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants. *Science (80-.).* **365,** 65–69 (2019).

22. Buglino, J. A. & Resh, M. D. in *Vitam. Horm. Chapter 10 Hedgehog Signal.* 1st ed., 229–252 (Elsevier Inc., 2012).

23. Huang, R. Chemical Biology of Protein N-Terminal Methyltransferases. *Chembiochem* **20,** 976–984 (2019).

24. Akimov, V., Barrio-Hernandez, I., Hansen, S. V. F., Hallenborg, P., Pedersen, A.-K., Bekker-Jensen, D. B., Puglia, M., Christensen, S. D. K., Vanselow, J. T., Nielsen, M. M., Kratchmarova, I., Kelstrup, C. D., Olsen, J. V. & Blagoev, B. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25,** 631–640 (2018).

25. Foyn, H., Van Damme, P., Støve, S. I., Glomnes, N., Evjenth, R., Gevaert, K. & Arnesen, T. Protein N-terminal acetyltransferases act as N-terminal propionyltransferases in vitro and in vivo. *Mol. Cell. Proteomics* **12,** 42–54 (2013).

26. Kumar, A. & Bachhawat, A. K. Pyroglutamic acid: Throwing light on a lightly studied metabolite. *Curr. Sci.* **102,** 288–297 (2012).

27. Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M., Varhaug, J. E. & Lillehaug, J. R. Identification and characterization of the human ARD1–NATH protein acetyltransferase complex. *Biochem. J.* **386,** 433–443 (2005).

28. Falb, M., Aivaliotis, M., Garcia-Rizo, C., Bisle, B., Tebbe, A., Klein, C., Konstantinidis, K., Siedler, F., Pfeiffer, F. & Oesterhelt, D. Archaeal N-terminal Protein Maturation Commonly Involves N-terminal Acetylation: A Large-scale Proteomics Survey. *J. Mol. Biol.* **362,** 915–924 (2006).

29. Strous, G., Berns, A. & Bloemendal, H. N-terminal acetylation of the nascent chains of α-crystallin. *Biochem. Biophys. Res. Commun.* **58,** 876–884 (1974).

30. Gautschi, M., Just, S., Mun, A., Ross, S., Rücknagel, P., Dubaquié, Y., Ehrenhofer-Murray, A. & Rospert, S. The Yeast Nα-Acetyltransferase NatA Is Quantitatively Anchored to the Ribosome and Interacts with Nascent Polypeptides. *Mol. Cell. Biol.* **23,** 7403 (2003).

31. Polevoda, B., Brown, S., Cardillo, T. S., Rigby, S. & Sherman, F. Yeast Nα-terminal acetyltransferases are associated with ribosomes. *J. Cell. Biochem.* **103,** 492–508 (2008).

32. Vetting, M. W., S. de Carvalho, L. P., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L. & Blanchard, J. S. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433,** 212–226 (2005).

33. Polevoda, B, Norbeck, J, Takakura, H, Blomberg, A & Sherman, F. Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae. *EMBO J.* **18,** 6155 (1999).

34. Arnesen, T., Damme, P. V., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J. E., Vandekerckhove, J., Lillehaug, J. R., Sherman, F. & Gevaert, K. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8157 (2009).

35. Hole, K., Van Damme, P., Dalva, M., Aksnes, H., Glomnes, N., Varhaug, J. E., Lillehaug, J. R., Gevaert, K. & Arnesen, T. The Human N-Alpha-Acetyltransferase 40 (hNaa40p/hNatD) Is Conserved from Yeast and N-Terminally Acetylates Histones H2A and H4. *PLoS One* **6** (ed Imhof, A.) e24713 (2011).

36. Evjenth, R, Hole, K, Karlsen, O., Ziegler, M, Arnesen, T & Lillehaug, J. Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity. *J. Biol. Chem.* **284,** 31122–31129 (2009).

37. Van Damme, P., Hole, K., Pimenta-Marques, A., Helsens, K., Vandekerckhove, J., Martinho, R. G., Gevaert, K. & Arnesen, T. NatF Contributes to an Evolutionary Shift in Protein N-Terminal Acetylation and Is Important for Normal Chromosome Segregation. *PLoS Genet.* **7** (ed Snyder, M.) e1002169 (2011).

38. Aksnes, H., Van Damme, P., Goris, M., Starheim, K., Marie, M., Støve, S., Hoel, C., Kalvik, T., Hole, K., Glomnes, N., Furnes, C., Ljostveit, S., Ziegler, M., Niere, M., Gevaert, K. & Arnesen, T. An Organellar Nα-Acetyltransferase, Naa60, Acetylates Cytosolic N Termini of Transmembrane Proteins and Maintains Golgi Integrity. *Cell Rep.* **10,** 1362–1374 (2015).

39. Utsumi, T, Sato, M, Nakano, K, Takemura, D, Iwata, H & Ishisaka, R. Amino acid residue penultimate to the amino-terminal gly residue strongly affects two cotranslational protein modifications, N-myristoylation and N-acetylation. *J. Biol. Chem.* **276,** 10505–13 (2001).

40. Castrec, B., Dian, C., Ciccone, S., Ebert, C. L., Bienvenut, W. V., Le Caer, J.-P., Steyaert, J.-M., Giglione, C. & Meinnel, T. Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. *Nat. Chem. Biol.* **14,** 671–679 (2018).

41. Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15,** 2385–2401 (2015).

42. Scott, D. C., Monda, J. K., Bennett, E. J., Harper, J. W. & Schulman, B. A. N-Terminal Acetylation Acts as an Avidity Enhancer Within an Interconnected Multiprotein Complex. *Science (80-.).* **334,** 674–678 (2011).

43. Hofmann, I. & Munro, S. An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility. *J. Cell Sci.* **119,** 1494–1503 (2006).

44. Dikiy, I. & Eliezer, D. N-terminal Acetylation Stabilizes N-terminal Helicity in Lipid- and Micelle-bound α-Synuclein and Increases Its Affinity for Physiological Membranes. *J. Biol. Chem.* **289,** 3652–3665 (2014).

45. Forte, G. M. A., Pool, M. R. & Stirling, C. J. N-Terminal Acetylation Inhibits Protein Targeting to the Endoplasmic Reticulum. *PLoS Biol.* **9** (ed Walter, P.) e1001073 (2011).

46. Hwang, C.-S., Shemorry, A. & Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science (80-.).* **327,** 973–7 (2010).

47. Shemorry, A., Hwang, C.-S. & Varshavsky, A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Mol. Cell* **50,** 540–551 (2013).

48. Holmes, W. M., Mannakee, B. K., Gutenkunst, R. N. & Serio, T. R. Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nat. Commun.* **5,** 4383 (2014).

49. Dormeyer, W., Mohammed, S., van Breukelen, B., Krijgsveld, J. & Heck, A. J. R. Targeted Analysis of Protein Termini. *J. Proteome Res.* **6,** 4634–4645 (2007).

50. Zhang, X., Ye, J. & Højrup, P. A proteomics approach to study in vivo protein Nα-modifications. *J. Proteomics* **73,** 240–251 (2009).

51. Todd King, M., Reiss, P. D. & Cornell, N. W. Determination of Short-Chain Coenzyme A Compounds by Reversed-Phase High-Performance Liquid Chromatography. *Methods Enzymol.* **166,** 70–79 (1988).

52. Tooley, J. G. & Schaner Tooley, C. E. New roles for old modifications: Emerging roles of N-terminal post-translational modifications in development and disease. *Protein Sci.* **23,** 1641–1649 (2014).

53. Petkowski, J. J., Schaner Tooley, C. E., Anderson, L. C., Shumilin, I. A., Balsbaugh, J. L., Shabanowitz, J., Hunt, D. F., Minor, W. & Macara, I. G. Substrate Specificity of Mammalian N-Terminal α-Amino Methyltransferase NRMT. *Biochemistry* **51,** 5942–5950 (2012).

54. Petkowski, J. J., Bonsignore, L. A., Tooley, J. G., Wilkey, D. W., Merchant, M. L., Macara, I. G. & Tooley, C. E. S. NRMT2 is an N-terminal monomethylase that primes for its homolog NRMT1. *Biochem. J.* **456,** 453 (2013).

55. Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A.-K., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Ödling, J., Oksvold, P., Olsson, I., Öster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, Å., Sköllermo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S. & Pontén, F. A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932 (2005).

56. Stock, A., Clarke, S., Clarke, C. & Stock, J. N-terminal methylation of proteins: Structure, function and specificity. *FEBS Lett.* **220,** 8–14 (1987).

57. Pettigrew, G. W. & Smith, G. M. Novel N-terminal protein blocking group identified as dimethylproline. *Nature* **265,** 661–662 (1977).

58. Henry, G. D., Winstanley, M. A., Dalgarno, D. C., M. Scott, G., Levine, B. A. & Trayer, I. P. Characterization of the actin-binding site on the alkali light chain of myosin. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **830,** 233–243 (1985).

59. Desrosiers, R. & Tanguay, R. M. Methylation of Drosophila histones at proline, lysine, and arginine residues during heat shock. *J. Biol. Chem.* **263,** 4686–4692 (1988).

60. Chen, T., Muratore, T. L., Schaner-Tooley, C. E., Shabanowitz, J., Hunt, D. F. & Macara, I. G. N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **9,** 596–603 (2007).

61. Hao, Y. & Macara, I. G. Regulation of chromatin binding by a conformational switch in the tail of the Ran exchange factor RCC1. *J. Cell Biol.* **182,** 827–36 (2008).

62. Cai, Q., Fu, L., Wang, Z., Gan, N., Dai, X. & Wang, Y. α-N-Methylation of Damaged DNA-binding Protein 2 (DDB2) and Its Function in Nucleotide Excision Repair. *J. Biol. Chem.* **289,** 16046–16056 (2014).

63. Bailey, A. O., Panchenko, T., Sathyan, K. M., Petkowski, J. J., Pai, P.-J., Bai, D. L., Russell, D. H., Macara, I. G., Shabanowitz, J., Hunt, D. F., Black, B. E. & Foltz, D. R. Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11827–32 (2013).

64. Dai, X., Otake, K., You, C., Cai, Q., Wang, Z., Masumoto, H. & Wang, Y. Identification of Novel α-N-Methylation of CENP-B That Regulates Its Binding to the Centromeric DNA. *J. Proteome Res.* **12,** 4167–4175 (2013).

65. Kimura, Y., Kurata, Y., Ishikawa, A., Okayama, A., Kamita, M. & Hirano, H. N-Terminal methylation of proteasome subunit Rpt1 in yeast. *Proteomics* **13,** 3167–3174 (2013).

66. Jakobsson, M. E., Małecki, J. M., Halabelian, L., Nilges, B. S., Pinto, R., Kudithipudi, S., Munk, S., Davydova, E., Zuhairi, F. R., Arrowsmith, C. H., Jeltsch, A., Leidel, S. A., Olsen, J. V. & Falnes, P. Ø. The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat. Commun.* **9,** 3411 (2018).

67. Chamoun, Z., Mann, R. K., Nellen, D, von Kessler, D. P., Bellotto, M, Beachy, P. A. & Basler, K. Skinny Hedgehog, an Acyltransferase Required for Palmitoylation and Activity of the Hedgehog Signal. *Science (80-.).* **293,** 2080–2084 (2001).

68. Pepinsky, R. B., Zeng, C., Went, D., Rayhorn, P., Baker, D. P., Williams, K. P., Bixler, S. A., Ambrose, C. M., Garber, E. A., Miatkowski, K., Taylor, F. R., Wang, E. A. & Galdes, A. Identification of a palmitic acid-modified form of human Sonic hedgehog. *J. Biol. Chem.* **273,** 14037–14045 (1998).

69. Miura, G. I., Buglino, J., Alvarado, D., Lemmon, M. A., Resh, M. D. & Treisman, J. E. Palmitoylation of the EGFR Ligand Spitz by Rasp Increases Spitz Activity by Restricting Its Diffusion. *Dev. Cell* **10,** 167–176 (2006).

70. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47,** D506–D515 (2019).

71. Buglino, J. A. & Resh, M. D. Hhat Is a Palmitoylacyltransferase with Specificity for N-Palmitoylation of Sonic Hedgehog. *J. Biol. Chem.* **283,** 22076–22088 (2008).

72. Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15,** 3059–87 (2001).

73. Rubin, L. L. & de Sauvage, F. J. Targeting the Hedgehog pathway in cancer. *Nat. Rev. Drug Discov.* **5,** 1026–1033 (2006).

74. Rodgers, U. R., Lanyon-Hogg, T., Masumoto, N., Ritzefeld, M., Burke, R., Blagg, J., Magee, A. I. & Tate, E. W. Characterization of Hedgehog Acyltransferase Inhibitors Identifies a Small Molecule Probe for Hedgehog Signaling by Cancer Cells. *ACS Chem. Biol.* **11,** 3256–3262 (2016).

75. Lim, S., Lim, S. M., Kim, M.-J., Park, S. Y. & Kim, J.-H. Sonic Hedgehog Pathway as the Prognostic Marker in Patients with Extensive Stage Small Cell Lung Cancer. *Yonsei Med. J.* **60,** 898 (2019).

76. Lee, J. D., Kraus, P., Gaiano, N., Nery, S., Kohtz, J., Fishell, G., Loomis, C. A. & Treisman, J. E. An Acylatable Residue of Hedgehog Is Differentially Required in Drosophila and Mouse Limb Development. *Dev. Biol.* **233,** 122–136 (2001).

77. Chen, M.-H., Li, Y.-J., Kawakami, T., Xu, S.-M. & Chuang, P.-T. Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. *Genes Dev.* **18,** 641–59 (2004).

78. Lanyon-Hogg, T., Masumoto, N., Bodakh, G., Konitsiotis, A. D., Thinon, E., Rodgers, U. R., Owens, R. J., Magee, A. I. & Tate, E. W. Synthesis and characterisation of 5-acyl-6,7-dihydrothieno[3,2-c]pyridine inhibitors of Hedgehog acyltransferase. *Data Br.* **7,** 257–281 (2016).

79. Kleuss, C. & Krause, E. Galpha(s) is palmitoylated at the N-terminal glycine. *EMBO J.* **22,** 826–32 (2003).

80. Lai, Z. W., Petrera, A. & Schilling, O. Protein amino-terminal modifications and proteomic approaches for N-terminal profiling. *Curr. Opin. Chem. Biol.* **24,** 71–79 (2015).

81. Kehlen, A., Haegele, M., Böhme, L., Cynis, H., Hoffmann, T. & Demuth, H. U. N-terminal pyroglutamate formation in CX3CL1 is essential for its full biologic activity. *Biosci. Rep.* **37,** 1–14 (2017).

82. Van Coillie, E., Proost, P., Van Aelst, I., Struyf, S., Polfliet, M., De Meester, I., Harvey, D. J., Van Damme, J. & Opdenakker, G. Functional Comparison of Two Human Monocyte Chemotactic Protein-2 Isoforms, Role of the Amino-Terminal Pyroglutamic Acid and Processing by CD26/Dipeptidyl Peptidase IV. *Biochemistry* **37,** 12672–12680 (1998).

83. Cynis, H., Hoffmann, T., Friedrich, D., Kehlen, A., Gans, K., Kleinschmidt, M., Rahfeld, J.-U., Wolf, R., Wermann, M., Stephan, A., Haegele, M., Sedlmeier, R., Graubner, S., Jagla, W., Müller, A., Eichentopf, R., Heiser, U., Seifert, F., Quax, P. H. A., de Vries, M. R., Hesse, I., Trautwein, D., Wollert, U., Berg, S., Freyse, E.-J., Schilling, S. & Demuth, H.-U. The isoenzyme of glutaminyl cyclase is an important regulator of monocyte infiltration under inflammatory conditions. *EMBO Mol. Med.* **3,** 545–58 (2011).

84. Wirths, O., Bethge, T., Marcello, A., Harmeier, A., Jawhar, S., Lucassen, P. J., Multhaup, G., Brody, D. L., Esparza, T., Ingelsson, M., Kalimo, H., Lannfelt, L. & Bayer, T. A. Pyroglutamate Abeta pathology in APP/PS1KI mice, sporadic and familial Alzheimer's disease cases. *J. Neural Transm.* **117,** 85–96 (2010).

85. Schlenzig, D., Manhart, S., Cinar, Y., Kleinschmidt, M., Hause, G., Willbold, D., Funke, S. A., Schilling, S. & Demuth, H.-U. Pyroglutamate Formation Influences Solubility and Amyloidogenicity of Amyloid Peptides. *Biochemistry* **48,** 7072–7078 (2009).

86. Tomidokoro, Y., Tamaoka, A., Holton, J. L., Lashley, T., Frangione, B., Revesz, T., Rostagno, A. & Ghiso, J. Pyroglutamate formation at the n-termini of abri molecules in familial British dementia is not restricted to the central nervous system. *Hirosaki Med. J.* **61,** S262–S269 (2010).

87. Song, H., Chang, Y. J., Moon, M., Park, S. K., Tran, P.-T., Hoang, V.-H., Lee, J. & Mook-Jung, I. Inhibition of Glutaminyl Cyclase Ameliorates Amyloid Pathology in an Animal Model of Alzheimer's Disease via the Modulation of γ-Secretase Activity. *J. Alzheimer's Dis.* **43,** 797–807 (2014).

88. Cynis, H., Kehlen, A., Haegele, M., Hoffmann, T., Heiser, U., Fujii, M., Shibazaki, Y., Yoneyama, H., Schilling, S. & Demuth, H.-U. Inhibition of Glutaminyl Cyclases alleviates CCL2-mediated inflammation of non-alcoholic fatty liver disease in mice. *Int. J. Exp. Pathol.,* n/a–n/a (2013).

89. Kehlen, A., Haegele, M., Menge, K., Gans, K., Immel, U.-D., Hoang-Vu, C., Klonisch, T. & Demuth, H.-U. Role of glutaminyl cyclases in thyroid carcinomas. *Endocr. Relat. Cancer* **20,** 79–90 (2013).

90. Weits, D. A., Giuntoli, B., Kosmacz, M., Parlanti, S., Hubberten, H.-M., Riegler, H., Hoefgen, R., Perata, P., van Dongen, J. T. & Licausi, F. Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nat. Commun.* **5,** 3425 (2014).

91. White, M. D., Klecker, M., Hopkinson, R. J., Weits, D. A., Mueller, C., Naumann, C., O'Neill, R., Wickens, J., Yang, J., Brooks-Bartlett, J. C., Garman, E. F., Grossmann, T. N., Dissmeyer, N. & Flashman, E. Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nat. Commun.* **8,** 14690 (2017).

92. Breitschopf, K., Bengal, E., Ziv, T., Admon, A. & Ciechanover, A. A novel site for ubiquitination: the N-terminal residue, and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. *EMBO J.* **17,** 5964–5973 (1998).

93. Reinstein, E., Scheffner, M., Oren, M., Ciechanover, A. & Schwartz, A. Degradation of the E7 human papillomavirus oncoprotein by the ubiquitin-proteasome system: targeting via ubiquitination of the N-terminal residue. *Oncogene* **19,** 5944–5950 (2000).

94. Aviel, S., Winberg, G., Massucci, M. & Ciechanover, A. Degradation of the Epstein-Barr Virus Latent Membrane Protein 1 (LMP1) by the Ubiquitin-Proteasome Pathway. *J. Biol. Chem.* **275,** 23491–23499 (2000).

95. Ikeda, M., Ikeda, A. & Longnecker, R. Lysine-Independent Ubiquitination of Epstein–Barr Virus LMP2A. *Virology* **300,** 153–159 (2002).

96. Fajerman, I., Schwartz, A. L. & Ciechanover, A. Degradation of the Id2 developmental regulator: targeting via N-terminal ubiquitination. *Biochem. Biophys. Res. Commun.* **314,** 505–512 (2004).

97. Bloom, J., Amador, V., Bartolini, F., DeMartino, G. & Pagano, M. Proteasome-Mediated Degradation of p21 via N-Terminal Ubiquitinylation. *Cell* **115,** 71–82 (2003).

98. Coulombe, P., Rodier, G., Bonneil, E., Thibault, P. & Meloche, S. N-Terminal Ubiquitination of Extracellular Signal-Regulated Kinase 3 and p21 Directs Their Degradation by the Proteasome. *Mol. Cell. Biol.* **24,** 6140–6150 (2004).

99. Tatham, M. H., Plechanovová, A., Jaffray, E. G., Salmen, H. & Hay, R. T. Ube2W conjugates ubiquitin to α-amino groups of protein N-termini. *Biochem. J.* **453,** 137–45 (2013).

100. Scaglione, K. M., Basrur, V., Ashraf, N. S., Konen, J. R., Elenitoba-Johnson, K. S., Todi, S. V. & Paulson, H. L. The ubiquitin-conjugating enzyme (E2) ube2w ubiquitinates the N terminus of substrates. *J. Biol. Chem.* **288,** 18784–18788 (2013).

101. Noy, T., Suad, O., Taglicht, D. & Ciechanover, A. HUWE1 ubiquitinates MyoD and targets it for proteasomal degradation. *Biochem. Biophys. Res. Commun.* **418,** 408–413 (2012).

102. Akimov, V., Rigbolt, K. T. G., Nielsen, M. M. & Blagoev, B. Characterization of ubiquitination dependent dynamics in growth factor receptor signaling by quantitative proteomics. *Mol. Biosyst.* **7,** 3223 (2011).

103. Danielsen, J. M. R., Sylvestersen, K. B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J. W., Horn, H., Jensen, L. J., Mailand, N. & Nielsen, M. L. Mass Spectrometric Analysis of Lysine Ubiquitylation Reveals Promiscuity at Site Level. *Mol. Cell. Proteomics* **10** (2011).

104. Kliza, K., Taumer, C., Pinzuti, I., Franz-Wachtel, M., Kunzelmann, S., Stieglitz, B., Macek, B. & Husnjak, K. Internally tagged ubiquitin: a tool to identify linear polyubiquitin-modified proteins by mass spectrometry. *Nat. Methods* **14,** 504–512 (2017).

105. Akimov, V., Henningsen, J., Hallenborg, P., Rigbolt, K. T. G., Jensen, S. S., Nielsen, M. M., Kratchmarova, I. & Blagoev, B. StUbEx: Stable Tagged Ubiquitin Exchange System for the Global Investigation of Cellular Ubiquitination. *J. Proteome Res.* **13,** 4192–4204 (2014).

106. Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28,** 868–873 (2010).

107. Vittal, V., Shi, L., Wenzel, D. M., Scaglione, K. M., Duncan, E. D., Basrur, V., Elenitoba-Johnson, K. S. J., Baker, D., Paulson, H. L., Brzovic, P. S. & Klevit, R. E. Intrinsic disorder drives N-terminal ubiquitination by Ube2w. *Nat. Chem. Biol.* **11,** 83–89 (2015).

108. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science (80-.).* **234,** 179–186 (1986).

109. Koren, I., Timms, R. T., Kula, T., Xu, Q., Li, M. Z. & Elledge, S. J. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **173,** 1622–1635.e14 (2018).

110. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12142 (1996).

111. Kim, H.-k., Kim, R.-r., Oh, J.-h., Cho, H., Varshavsky, A. & Hwang, C.-s. The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell* **156,** 158–169 (2014).

112. Timms, R. T., Zhang, Z., Rhee, D. Y., Harper, J. W., Koren, I. & Elledge, S. J. A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science (80-.).* **365,** eaaw4912 (2019).

113. Hwang, C.-S., Shemorry, A., Auerbach, D. & Varshavsky, A. The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. *Nat. Cell Biol.* **12,** 1177–1185 (2010).

114. Tasaki, T., Sriram, S. M., Park, K. S. & Kwon, Y. T. The N-End Rule Pathway. *Annu. Rev. Biochem.* **81,** 261–289 (2012).

115. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20,** 1298 (2011).

116. Oh, J.-H., Hyun, J.-Y. & Varshavsky, A. Control of Hsp90 chaperone and its clients by N-terminal acetylation and the N-end rule pathway. *Proc. Natl. Acad. Sci.* **114,** E4370–E4379 (2017).

117. Turner, G. C., Du, F. & Varshavsky, A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature* **405,** 579–583 (2000).

118. Hu, R.-G., Wang, H., Xia, Z. & Varshavsky, A. The N-end rule pathway is a sensor of heme. *Proc. Natl. Acad. Sci.* **105,** 76–81 (2008).

119. Yoo, Y. D., Mun, S. R., Ji, C. H., Sung, K. W., Kang, K. Y., Heo, A. J., Lee, S. H., An, J. Y., Hwang, J., Xie, X.-Q., Ciechanover, A., Kim, B. Y. & Kwon, Y. T. N-terminal arginylation generates a bimodal degron that modulates autophagic proteolysis. *Proc. Natl. Acad. Sci.* **115,** E2716–E2724 (2018).

120. Chen, S.-J., Wu, X., Wadas, B., Oh, J.-H. & Varshavsky, A. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science (80-.).* **355** (2017).

121. Myasnikov, A. G., Simonetti, A., Marzi, S. & Klaholz, B. P. Structure–function insights into prokaryotic and eukaryotic translation initiation. *Curr. Opin. Struct. Biol.* **19,** 300–309 (2009).

122. Mazel, D, Pochet, S & Marlière, P. Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. *EMBO J.* **13,** 914 (1994).

123. Sandikci, A., Gloge, F., Martinez, M., Mayer, M. P., Wade, R., Bukau, B. & Kramer, G. Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding. *Nat. Struct. Mol. Biol.* **20,** 843–850 (2013).

124. Roderick, S. L. & Matthews, B. W. Structure of the cobalt-dependent methionine aminopeptidase from Escherichia coli: a new type of proteolytic enzyme. *Biochemistry* **32,** 3907–3912 (1993).

125. Walker, K. W. & Bradshaw, R. A. Yeast methionine aminopeptidase I can utilize either Zn2+ or Co2+as a cofactor: A case of mistaken identity? *Protein Sci.* **7,** 2684–2687 (1998).

126. D'souz, V. M. & Holz, R. C. The Methionyl Aminopeptidase from Escherichia coli Can Function as an Iron(II) Enzyme †. *Biochemistry* **38,** 11079–11085 (1999).

127. Wang, J., Sheppard, G. S., Lou, P., Kawai, M., Park, C., Egan, D. A., Schneider, A., Bouska, J., Lesniewski, R. & Henkin, J. Physiologically relevant metal cofactor for methionine aminopeptidase-2 is manganese. *Biochemistry* **42,** 5035–5042 (2003).

128. Bazan, J. F., Weaver, L. H., Roderick, S. L., Huber, R & Matthews, B. W. Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2473 (1994).

129. Arfin, S. M., Kendall, R. L., Hall, L., Weaver, L. H., Stewart, A. E., Matthews, B. W. & Bradshaw, R. A. Eukaryotic methionyl aminopeptidases: two classes of cobalt-dependent enzymes. *Proc. Natl. Acad. Sci.* **92,** 7714–7718 (1995).

130. Bradshaw, R. A., Brickey, W. W. & Walker, K. W. N-Terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends Biochem. Sci.* **23,** 263–267 (1998).

131. Chen, S., Vetro, J. A. & Chang, Y.-H. The Specificity in Vivo of Two Distinct Methionine Aminopeptidases in Saccharomyces cerevisiae. *Arch. Biochem. Biophys.* **398,** 87–93 (2002).

132. Li, X & Chang, Y. H. Amino-terminal protein processing in Saccharomyces cerevisiae is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 12357–61 (1995).

133. Liu, S., Widom, J., Kemp, C. W., Crews, C. M., Clardy, J., Parks, D. A., Williams, T. K., J, J. S. B. A., Andrew, M, Owen, M. J., Psychiatry, B. J., Wiley, E., Jonas, P, Bischofberger, J, Pfeiffer, F, Graham, D, Betz, H, Chem, J. B., Kirsch, J, Chem, J. B., Triller, A, Cluzeaud, F, Pfeiffer, F, Betz, H, Korn, H, Biol, J. C., Kirsch, J, Betz, H, Res, B., Todd, A. J., Spike, R. C., Neilson, M, Neurosci, E. J., Kirsch, J, Kuhse, J, Betz, H, Neurosci, M. C., Kirsch, J, Wolters, I, Triller, A, Betz, H, Kirsch, J, Meyer, G, Betz, H, Neurosci, M. C., Bechade, C, Triller, A, Dev, S. C., Liu, S., Widom, J., Kemp, C. W., Crews, C. M. & Clardy, J. Structure of Human Methionine Aminopeptidase-2 Complexed with Fumagillin. *Science (80-.).* **282,** 1324–1328 (1998).

134. Addlagatta, A., Hu, X., Liu, J. O. & Matthews, B. W. Structural basis for the functional differences between type I and type II human methionine aminopeptidases. *Biochemistry* **44,** 14741–14749 (2005).

135. Vetro, J. A., Dummitt, B., Micka, W. S. & Chang, Y. H. Evidence of a dominant negative mutant of yeast methionine aminopeptidase type 2 in Saccharomyces cerevisiae. *J. Cell. Biochem.* **94,** 656–668 (2005).

136. Vetro, J. A. & Chang, Y.-H. Yeast methionine aminopeptidase type 1 is ribosome-associated and requires its N-terminal zinc finger domain for normal function in vivo. *J. Cell. Biochem.* **85,** 678–688 (2002).

137. Datta, R., Choudhury, P., Bhattacharya, M., Soto Leon, F., Zhou, Y. & Datta, B. Protection of translation initiation factor eIF2 phosphorylation correlates with eIF2-associated glycoprotein p67 levels and requires the lysine-rich domain I of p67. *Biochimie* **83,** 919–931 (2001).

138. Addlagatta, A., Quillin, M. L., Omotoso, O., Liu, J. O. & Matthews, B. W. Identification of an SH3-Binding Motif in a New Class of Methionine Aminopeptidases from Mycobacterium tuberculosis Suggests a Mode of Interaction with the Ribosome. *Biochemistry* **44,** 7166–7174 (2005).

139. Serero, A., Giglione, C., Sardini, A., Martinez-Sanz, J. & Meinnel, T. An Unusual Peptide Deformylase Features in the Human Mitochondrial N-terminal Methionine Excision Pathway. *J. Biol. Chem.* **278,** 52953–52963 (2003).

140. Zuo, S., Guo, Q., Ling, C. & Chang, Y.-H. Evidence that two zinc fingers in the methionine aminopeptidase from Saccharomyces cerevisiae are important for normal growth. *Mol. Gen. Genet. MGG* **246,** 247–253 (1995).

141. Yang, G., Kirkpatrick, R. B., Ho, T., Zhang, G.-F., Liang, P.-H., Johanson, K. O., Casper, D. J., Doyle, M. L., Marino, J. P., Thompson, S. K., Chen, W., Tew, D. G. & Meek, T. D. Steady-State Kinetic Characterization of Substrates and Metal-Ion Specificities of the Full-Length and N-Terminally Truncated Recombinant Human Methionine Aminopeptidases (Type 2). *Biochemistry* **40,** 10645–54 (2001).

142. Datta, R., Choudhury, P., Ghosh, A. & Datta, B. A glycosylation site, 60SGTS63, of p67 is required for its ability to regulate the phosphorylation and activity of eukaryotic initiation factor 2α. *Biochemistry* **42,** 5453–5460 (2003).

143. Datta, B., Ghosh, A., Majumdar, A. & Datta, R. Autoproteolysis of Rat p67 Generates Several Peptide Fragments: The N-Terminal Fragment, p26, Is Required for the Protection of eIF2α from Phosphorylation†. *Biochemistry* **46,** 3465–3475 (2007).

144. Clinkinbeard, T., Ghoshal, S., Craddock, S., Pettigrew, L. C. & Guttmann, R. P. Calpain cleaves methionine aminopeptidase-2 in a rat model of ischemia/reperfusion. *Brain Res.* **1499,** 129 (2013).

145. Fujii, K., Susanto, T. T., Saurabh, S. & Barna, M. Decoding the Function of Expansion Segments in Ribosomes. *Mol. Cell* **72,** 1013–1020.e6 (2018).

146. Moerschells, R., Hosokawa, Y, Tsunasawa, S & Sherman, F. The Specificities of Aminoterminal of Yeast Methionine Aminopeptidase Methionine in Vivo. *J. Biol. Chem.* **265,** 19638–19643 (1990).

147. Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O. & Pei, D. Protein N-Terminal Processing: Substrate Specificity of Escherichia coli and Human Methionine Aminopeptidases. *Biochemistry* **49,** 5588–5599 (2010).

148. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C. & Meinnel, T. The Proteomics of N-terminal Methionine Cleavage. *Mol. Cell. Proteomics* **5,** 2336–2349 (2006).

149. Frottin, F., Bienvenut, W. V., Bignon, J., Jacquet, E., Vaca Jacome, A. S., Van Dorsselaer, A., Cianferani, S., Carapito, C., Meinnel, T. & Giglione, C. MetAP1 and MetAP2 drive cell selectivity for a potent anti-cancer agent in synergy, by controlling glutathione redox state. *Oncotarget* **7,** 63306–63323 (2016).

150. Jonckheere, V., Fijałkowska, D. & Van Damme, P. Omics Assisted N-terminal Proteoform and Protein Expression Profiling On Methionine Aminopeptidase 1 (MetAP1) Deletion. *Mol. Cell. Proteomics* **17,** 694–708 (2018).

151. Chang, S. Y., McGary, E. C. & Chang, S. Methionine aminopeptidase gene of Escherichia coli is essential for cell growth. *J. Bacteriol.* **171,** 4071 (1989).

152. Frottin, F., Espagne, C., Traverso, J. A., Mauve, C., Valot, B., Lelarge-Trouverie, C., Zivy, M., Noctor, G., Meinnel, T. & Giglione, C. Cotranslational Proteolysis Dominates Glutathione Homeostasis to Support Proper Growth and Development. *Plant Cell* **21,** 3296–3314 (2009).

153. Ross, S., Giglione, C., Pierre, M., Espagne, C. & Meinnel, T. Functional and Developmental Impact of Cytosolic Protein N-Terminal Methionine Excision in Arabidopsis. *Plant Physiol.* **137,** 623–637 (2005).

154. Bernier, S. G., Taghizadeh, N., Thompson, C. D., Westlin, W. F. & Hannig, G. Methionine aminopeptidases type I and type II are essential to control cell proliferation. *J. Cell. Biochem.* **95,** 1191–1203 (2005).

155. Yeh, J.-R. J., Ju, R., Brdlik, C. M., Zhang, W., Zhang, Y., Matyskiela, M. E., Shotwell, J. D. & Crews, C. M. Targeted gene disruption of methionine aminopeptidase 2 results in an embryonic gastrulation defect and endothelial cell growth arrest. *Proc. Natl. Acad. Sci.* **103,** 10379–10384 (2006).

156. Chang, Y. H., Teichert, U. & Smith, J. A. Molecular cloning, sequencing, deletion, and overexpression of a methionine aminopeptidase gene from Saccharomyces cerevisiae. *J. Biol. Chem.* **267,** 8007–8011 (1992).

157. Giglione, C., Serero, A., Pierre, M., Boisson, B. & Meinnel, T. Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. *EMBO J.* **19,** 5916 (2000).

158. Gupta, S., Bose, A., Chatterjee, N., Saha, D., Wu, S. & Gupta, N. K. p67 Transcription Regulates Translation in Serum-starved and Mitogen-activated KRC-7 Cells. *J. Biol. Chem.* **272,** 12699–12704 (1997).

159. Cutforth, T. & Gaul, U. A methionine aminopeptidase and putative regulator of translation initiation is required for cell growth and patterning in Drosophila. *Mech. Dev.* **82,** 23–28 (1999).

160. Dummitt, B., Micka, W. S. & Chang, Y. H. N-terminal methionine removal and methionine metabolism in Saccharomyces cerevisiae. *J. Cell. Biochem.* **89,** 964–974 (2003).

161. Chiu, J., Wong, J. W. H. & Hogg, P. J. Redox regulation of methionine aminopeptidase 2 activity. *J. Biol. Chem.* **289,** 15035–43 (2014).

162. Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A. & Hwang, C. S. The N-terminal methionine of cellular proteins as a degradation signal. *Cell* **156,** 158–169 (2014).

163. Shirakawa, R., Goto-Ito, S., Goto, K., Wakayama, S., Kubo, H., Sakata, N., Trinh, D. A., Yamagata, A., Sato, Y., Masumoto, H., Cheng, J., Fujimoto, T., Fukai, S. & Horiuchi, H. A SNARE geranylgeranyltransferase essential for the organization of the Golgi apparatus. *EMBO J.* (2020).

164. Ingber, D., Fujita, T., Kishimoto, S., Sudo, K., Kanamaru, T., Brem, H. & Folkman, J. Synthetic analogues of fumagillin that inhibit angiogenesis and suppress tumour growth. *Nature* **348,** 555–557 (1990).

165. Hanson, F. R. & Eble, T. E. An antiphage agent isolated from Aspergillus sp. *J. Bacteriol.* **58,** 527–9 (1949).

166. McCowen, M. C., Callender, M. E. & Lawlis, J. F. Fumagillin (H-3), a New Antibiotic with Amebicidal Properties. *Science (80-.).* **113,** 202–203 (1951).

167. Kusaka, M., Sudo, K., Fujita, T., Marui, S., Itoh, F., Ingber, D. & Folkman, J. Potent anti-angiogenic action of AGM-1470: comparison to the fumagillin parent. *Biochem. Biophys. Res. Commun.* **174,** 1070–1076 (1991).

168. Kusaka, M, Sudo, K, Matsutani, E, Kozai, Y, Marui, S, Fujita, T, Ingber, D & Folkman, J. Cytostatic inhibition of endothelial cell growth by the angiogenesis inhibitor TNP-470 (AGM-1470). *Br. J. Cancer* **69,** 212–216 (1994).

169. Kruger, E. A. & Figg, W. D. TNP-470: An angiogenesis inhibitor in clinical development for cancer. *Expert Opin. Investig. Drugs* **9,** 1383–1396 (2000).

170. O'Reilly, M. S., Brem, H. & Folkman, J. Treatment of murine hemangioendotheliomas with the angiogenesis inhibitor AGM-1470. *J. Pediatr. Surg.* **30,** 325–330 (1995).

171. Catalano, A., Romano, M., Robuffo, I., Strizzi, L. & Procopio, A. Methionine Aminopeptidase-2 Regulates Human Mesothelioma Cell Survival: Role of Bcl-2 Expression and Telomerase Activity. *Am. J. Pathol.* **159,** 721–731 (2001).

172. Shusterman, S., Grupp, S. A., Barr, R., Carpentieri, D., Zhao, H. & Maris, J. M. The angiogenesis inhibitor TNP-470 effectively inhibits human neuroblastoma xenograft growth, especially in the setting of subclinical disease. *Clin. Cancer Res.* **7,** 977–984 (2001).

173. Kidoikhammouan, S., Seubwai, W., Silsirivanit, A., Wongkham, S., Sawanyawisuth, K. & Wongkham, C. Blocking of methionine aminopeptidase-2 by TNP-470 induces apoptosis and increases chemosensitivity of cholangiocarcinoma. *J. Cancer Res. Ther.* **0,** 0 (2019).

174. Peacock, D. J., Banquerigo, M. L. & Brahn, E. A novel angiogenesis inhibitor suppresses rat adjuvant arthritis. *Cell. Immunol.* **160,** 178–184 (1995).

175. Sin, N., Meng, L., Wang, M. Q., Wen, J. J., Bornmann, W. G. & Crews, C. M. The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6099–6103 (1997).

176. Griffith, E. C., Su, Z., Turk, B. E., Chen, S., Chang, Y.-H., Wu, Z., Biemann, K. & Liu, J. O. Methionine aminopeptidase (type 2) is the common target for angiogenesis inhibitors AGM-1470 and ovalicin. *Chem. Biol.* **4,** 461–471 (1997).

177. Sigg, H. P. & Weber, H. P. Isolierung und Strukturaufklärung von Ovalicin (Isolation and structure elucidation of ovalicin). *Helv. Chim. Acta* **51,** 1395–1408 (1968).

178. Hartmann, G., Richter, H., Weiner, E. & Zimmermann, W. On the Mechanism of Action of the Cytostatic Drug Anguidine and of the Immunosuppressive Agent Ovalicin, two Sesquiterpenes from Fungi. *Planta Med.* **34,** 231–252 (1978).

179. Turk, B. E., Su, Z. & Liu, J. O. Synthetic analogues of TNP-470 and ovalicin reveal a common molecular basis for inhibition of angiogenesis and immunosuppression. *Bioorg. Med. Chem.* **6,** 1163–1169 (1998).

180. Griffith, E. C., Su, Z, Niwayama, S, Ramsay, C. A., Chang, Y. H. & Liu, J. O. Molecular recognition of angiogenesis inhibitors fumagillin and ovalicin by methionine aminopeptidase 2. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 15183–8 (1998).

181. Turk, B. E., Griffith, E. C., Wolf, S., Biemann, K., Chang, Y.-H. & Liu, J. O. Selective inhibition of amino-terminal methionine processing by TNP-470 and ovalicin in endothelial cells. *Chem. Biol.* **6,** 823–833 (1999).

182. Arico-Muendel, C. C., Blanchette, H., Benjamin, D. R., Caiazzo, T. M., Centrella, P. A., DeLorey, J., Doyle, E. G., Johnson, S. R., Labenski, M. T., Morgan, B. A., O'Donovan, G., Sarjeant, A. A., Skinner, S., Thompson, C. D., Griffin, S. T., Westlin, W. & White, K. F. Orally Active Fumagillin Analogues: Transformations of a Reactive Warhead in the Gastric Environment. *ACS Med. Chem. Lett.* **4,** 381–386 (2013).

183. Morgen, M., Jöst, C., Malz, M., Janowski, R., Niessing, D., Klein, C. D., Gunkel, N. & Miller, A. K. Spiroepoxytriazoles Are Fumagillin-like Irreversible Inhibitors of MetAP2 with Potent Cellular Activity. *ACS Chem. Biol.* **11,** 1001–1011 (2016).

184. Kudelka, A. P., Levy, T., Verschraegen, C. F., Edwards, C. L., Piamsomboon, S., Termrungruanglert, W., Freedman, R. S., Kaplan, A. L., Kieback, D. G., Meyers, C. A., Jaeckle, K. A., Loyer, E., Steger, M., Mante, R., Mavligit, G., Killian, A., Tang, R. A., Gutterman, J. U. & Kavanagh, J. J. A phase I study of TNP-470 administered to patients with advanced squamous cell cancer of the cervix. *Clin. Cancer Res.* **3,** 1501–1505 (1997).

185. Kudelka, A., Verschraegen, C. & Loyer, E. Complete Remission of Metastatic Cervical Cancer With the Angiogenesis Inhibitor TNP-470. *N. Engl. J. Med.* **338,** 991–2 (1998).

186. Tran, H. T., Blumenschein, G. R., Lu, C., Meyers, C. A., Papadimitrakopoulou, V., Fossella, F. V., Zinner, R., Madden, T., Smythe, L. G., Puduvalli, V. K., Munden, R., Truong, M. & Herbst, R. S. Clinical and pharmacokinetic study of TNP-470, an angiogenesis inhibitor, in combination with paclitaxel and carboplatin in patients with solid tumors. *Cancer Chemother. Pharmacol.* **54,** 308–314 (2004).

187. Herbst, R. S., Madden, T. L., Tran, H. T., Blumenschein, G. R., Meyers, C. A., Seabroake, L. F., Khuri, F. R., Puduvalli, V. K., Allgood, V., Fritsche, H. A., Hinton, L., Newman, R. A., Crane, E. A., Fossella, F. V., Dordal, M., Goodin, T. & Hong, W. K. Safety and pharmacokinetic effects of TNP-470, an angiogenesis inhibitor, combined with paclitaxel in patients with solid tumors: Evidence for activity in non-small-cell lung cancer. *J. Clin. Oncol.* **20,** 4440–4447 (2002).

188. Benny, O., Fainaru, O., Adini, A., Cassiola, F., Bazinet, L., Adini, I., Pravda, E., Nahmias, Y., Koirala, S., Corfas, G., D'Amato, R. J. & Folkman, J. An orally delivered small-molecule formulation with antiangiogenic and anticancer activity. *Nat. Biotechnol.* **26,** 799–807 (2008).

189. Satchi-Fainaro, R., Puder, M., Davies, J. W., Tran, H. T., Sampson, D. A., Greene, A. K., Corfas, G. & Folkman, J. Targeting angiogenesis with a conjugate of HPMA copolymer and TNP-470. *Nat. Med.* **10,** 255–261 (2004).

190. U.S. National Library of Medicine. Safety/Tolerance Study of PPI-2458 in Subjects With Non-Hodgkin's Lymphoma and Solid Tumors. *ClinicalTrials.gov,* NCT00100347 (2007).

191. Ashraf, S., Mapp, P. I. & Walsh, D. A. Angiogenesis and the persistence of inflammation in a rat model of proliferative synovitis. *Arthritis Rheum.* **62,** 1890–1898 (2010).

192. Ashraf, S., Mapp, P. I. & Walsh, D. A. Contributions of angiogenesis to inflammation, joint damage, and pain in a rat model of osteoarthritis. *Arthritis Rheum.* **63,** 2700–10 (2011).

193. Towbin, H., Bair, K. W., DeCaprio, J. A., Eck, M. J., Kim, S., Kinder, F. R., Morollo, A., Mueller, D. R., Schindler, P., Song, H. K., van Oostrum, J., Versace, R. W., Voshol, H., Wood, J., Zabludoff, S. & Phillips, P. E. Proteomics-based target identification: bengamides as a new class of methionine aminopeptidase inhibitors. *J. Biol. Chem.* **278,** 52964–71 (2003).

194. Dumez, H., Gall, H., Capdeville, R., Dutreix, C., Van Oosterom, A. T. & Giaccone, G. A phase I and pharmacokinetic study of LAF389 administered to patients with advanced cancer. *Anticancer. Drugs* **18,** 219–225 (2007).

195. Sheppard, G. S., Wang, J., Kawai, M., BaMaung, N. Y., Craig, R. A., Erickson, S. A., Lynch, L., Patel, J., Yang, F., Searle, X. B., Lou, P., Park, C., Kim, K. H., Henkin, J. & Lesniewski, R. 3-Amino-2-hydroxyamides and related compounds as inhibitors of methionine aminopeptidase-2. *Bioorg. Med. Chem. Lett.* **14,** 865–868 (2004).

196. Wang, J., Sheppard, G. S., Lou, P., Kawai, M., BaMaung, N., Erickson, S. A., Tucker-Garcia, L., Park, C., Bouska, J., Wang, Y.-C., Frost, D., Tapang, P., Albert, D. H., Morgan, S. J., Morowitz, M., Shusterman, S., Maris, J. M., Lesniewski, R. & Henkin, J. Tumor suppression by a rationally designed reversible inhibitor of methionine aminopeptidase-2. *Cancer Res.* **63,** 7861–9 (2003).

197. Morowitz, M. J., Barr, R., Wang, Q., King, R., Rhodin, N., Pawel, B., Zhao, H., Erickson, S. A., Sheppard, G. S., Wang, J., Maris, J. M. & Shusterman, S. Methionine aminopeptidase 2 inhibition is an effective treatment strategy for neuroblastoma in preclinical models. *Clin. Cancer Res.* **11,** 2680–5 (2005).

198. Sheppard, G. S., Wang, J., Kawai, M., Fidanze, S. D., BaMaung, N. Y., Erickson, S. A., Barnes, D. M., Tedrow, J. S., Kolaczkowski, L., Vasudevan, A., Park, D. C., Wang, G. T., Sanders, W. J., Mantei, R. A., Palazzo, F., Tucker-Garcia, L., Lou, P., Zhang, Q., Park, C. H., Kim, K. H., Petros, A., Olejniczak, E., Nettesheim, D., Hajduk, P., Henkin, J., Lesniewski, R., Davidsen, S. K. & Bell, R. L. Discovery and optimization of anthranilic acid sulfonamides as inhibitors of methionine aminopeptidase-2: A structural basis for the reduction of albumin binding. *J. Med. Chem.* **49,** 3832–3849 (2006).

199. Kawai, M., BaMaung, N. Y., Fidanze, S. D., Erickson, S. A., Tedrow, J. S., Sanders, W. J., Vasudevan, A., Park, C., Hutchins, C., Comess, K. M., Kalvin, D., Wang, J., Zhang, Q., Lou, P., Tucker-Garcia, L., Bouska, J., Bell, R. L., Lesniewski, R., Henkin, J. & Sheppard, G. S. Development of sulfonamide compounds as potent methionine aminopeptidase type II inhibitors with antiproliferative properties. *Bioorganic Med. Chem. Lett.* **16,** 3574–3577 (2006).

200. Wang, J., Tucker, L. A., Stavropoulos, J., Zhang, Q., Wang, Y.-C., Bukofzer, G., Niquette, A., Meulbroek, J. A., Barnes, D. M., Shen, J., Bouska, J., Donawho, C., Sheppard, G. S. & Bell, R. L. Correlation of tumor growth suppression and methionine aminopetidase-2 activity blockade using an orally active inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1838–43 (2008).

201. Tucker, L. A., Zhang, Q, Sheppard, G. S., Lou, P, Jiang, F, McKeegan, E, Lesniewski, R, Davidsen, S. K., Bell, R. L. & Wang, J. Ectopic expression of methionine aminopeptidase-2 causes cell transformation and stimulates proliferation. *Oncogene* **27,** 3967–3976 (2008).

202. Garrabrant, T., Tuman, R. W., Ludovici, D., Tominovich, R., Simoneaux, R. L., Galemmo, R. A. & Johnson, D. L. Small molecule inhibitors of methionine aminopeptidase type 2 (MetAP-2). *Angiogenesis* **7,** 91–6 (2004).

203. Marino, J. P., Fisher, P. W., Hofmann, G. A., Kirkpatrick, R. B., Janson, C. A., Johnson, R. K., Ma, C., Mattern, M., Meek, T. D., Ryan, M. D., Schulz, C., Smith, W. W., Tew, D. G., Tomazek, T. A., Veber, D. F., Xiong, W. C., Yamamoto, Y., Yamashita, K.,

Yang, G. & Thompson, S. K. Highly potent inhibitors of methionine aminopeptidase-2 based on a 1,2,4-triazole pharmacophore. *J. Med. Chem.* **50,** 3777–3785 (2007).

204. Heinrich, T., Seenisamy, J., Blume, B., Bomke, J., Calderini, M., Eckert, U., Friese-Hamim, M., Kohl, R., Lehmann, M., Leuthner, B., Musil, D., Rohdich, F. & Zenke, F. T. Discovery and Structure-Based Optimization of Next-Generation Reversible Methionine Aminopeptidase-2 (MetAP-2) Inhibitors. *J. Med. Chem.* **62,** 5025–5039 (2019).

205. Cheruvallath, Z., Tang, M., McBride, C., Komandla, M., Miura, J., Ton-Nu, T., Erikson, P., Feng, J., Farrell, P., Lawson, J. D., Vanderpool, D., Wu, Y., Dougan, D. R., Plonowski, A., Holub, C. & Larson, C. Discovery of potent, reversible MetAP2 inhibitors via fragment based drug discovery and structure based drug design—Part 1. *Bioorg. Med. Chem. Lett.* **26,** 2774–2778 (2016).

206. McBride, C., Cheruvallath, Z., Komandla, M., Tang, M., Farrell, P., Lawson, J. D., Vanderpool, D., Wu, Y., Dougan, D. R., Plonowski, A., Holub, C. & Larson, C. Discovery of potent, reversible MetAP2 inhibitors via fragment based drug discovery and structure based drug design—Part 2. *Bioorg. Med. Chem. Lett.* **26,** 2779–2783 (2016).

207. Huang, H.-J., Holub, C., Rolzin, P., Bilakovics, J., Fanjul, A., Satomi, Y., Plonowski, A., Larson, C. J. & Farrell, P. J. MetAP2 inhibition increases energy expenditure through direct action on brown adipocytes. *J. Biol. Chem.* **294,** 9567–9575 (2019).

208. Heinrich, T., Buchstaller, H.-P., Cezanne, B., Rohdich, F., Bomke, J., Friese-Hamim, M., Krier, M., Knöchel, T., Musil, D., Leuthner, B. & Zenke, F. Novel reversible methionine aminopeptidase-2 (MetAP-2) inhibitors based on purine and related bicyclic templates. *Bioorg. Med. Chem. Lett.* **27,** 551–556 (2017).

209. Heinrich, T., Seenisamy, J., Becker, F., Blume, B., Bomke, J., Dietz, M., Eckert, U., Friese-Hamim, M., Gunera, J., Hansen, K., Leuthner, B., Musil, D., Pfalzgraf, J., Rohdich, F., Siegl, C., Spuck, D., Wegener, A. & Zenke, F. T. Identification of Methionine Aminopeptidase-2 (MetAP-2) inhibitor M8891, a Clinical Compound for the Treatment of Cancer. *J. Med. Chem.,* acs.jmedchem.9b01070 (2019).

210. Farrell, P. J., Zopf, C. J., Huang, H.-J., Balakrishna, D., Holub, C., Bilakovics, J., Fanjul, A., Matuszkiewicz, J., Plonowski, A., Rolzin, P., Banerjee, U., Ermolieff, J., Cheruvallath, Z. S., McBride, C., Bartkowski, D., Mazur, C., Pachori, A. & Larson, C. J. Using Target Engagement Biomarkers to Predict Clinical Efficacy of MetAP2 Inhibitors. *J. Pharmacol. Exp. Ther.* **371,** 299–308 (2019).

211. Chang, Y.-H. Common therapeutic target for both cancer and obesity. *World J. Biol. Chem.* **8,** 102–107 (2017).

212. Kim, E. J. & Shin, W. H. General pharmacology of CKD-732, a new anticancer agent: Effects on central nervous, cardiovascular, and respiratory system. *Biol. Pharm. Bull.* **28,** 217–223 (2005).

213. Shin, S. J., Jeung, H. C., Ahn, J. B., Rha, S. Y., Roh, J. K., Park, K. S., Kim, D. H., Kim, C. & Chung, H. C. A phase i pharmacokinetic and pharmacodynamic study of CKD-732, an antiangiogenic agent, in patients with refractory solid cancer. *Invest. New Drugs* **28,** 650–658 (2010).

214. Shin, S. J., Ahn, J. B., Park, K. S., Lee, Y. J., Hong, Y. S., Kim, T. W., Kim, H. R., Rha, S. Y., Roh, J. K., Kim, D.-H., Kim, C. & Chung, H. C. A Phase Ib pharmacokinetic study of the anti-angiogenic agent CKD-732 used in combination with capecitabine and oxaliplatin (XELOX) in metastatic colorectal cancer patients who progressed on irinotecan-based chemotherapy. *Invest. New Drugs* **30,** 672–680 (2012).

215. Hughes, T. E., Kim, D. D., Marjason, J., Proietto, J., Whitehead, J. P. & Vath, J. E. Ascending dose-controlled trial of beloranib, a novel obesity treatment for safety, tolerability, and weight loss in obese women. *Obesity* **21,** 1782–1788 (2013).

216. Bråkenhielm, E., Cao, R., Gao, B., Angelin, B., Cannon, B., Parini, P. & Cao, Y. Angiogenesis inhibitor, TNP-470, prevents diet-induced and genetic obesity in mice. *Circ. Res.* **94,** 1579–88 (2004).

217. Lijnen, H. R., Frederix, L. & Van Hoef, B. Fumagillin reduces adipose tissue formation in murine models of nutritionally induced obesity. *Obesity* **18,** 2241–2246 (2010).

218. Kim, Y. M., An, J. J., Jin, Y. J., Rhee, Y., Cha, B. S., Lee, H. C. & Lim, S. K. Assessment of the anti-obesity effects of the TNP-470 analog CKD-732. *J. Mol. Endocrinol.* **38,** 455–465 (2007).

219. Proietto, J., Malloy, J., Zhuang, D., Arya, M., Cohen, N. D., de Looze, F. J., Gilfillan, C., Griffin, P., Hall, S., Nathow, T., Oldfield, G. S., O'Neal, D. N., Roberts, A., Stuckey, B. G. A., Yue, D., Taylor, K. & Kim, D. Efficacy and safety of methionine aminopeptidase 2 inhibition in type 2 diabetes: a randomised, placebo-controlled clinical trial. *Diabetologia* **61,** 1918–1922 (2018).

220. McCandless, S. E., Yanovski, J. A., Miller, J., Fu, C., Bird, L. M., Salehi, P., Chan, C. L., Stafford, D., Abuzzahab, M. J., Viskochil, D., Barlow, S. E., Angulo, M., Myers, S. E., Whitman, B. Y., Styne, D., Roof, E., Dykens, E. M., Scheimann, A. O., Malloy, J., Zhuang, D., Taylor, K., Hughes, T. E., Kim, D. D. & Butler, M. G. Effects of MetAP2 inhibition on hyperphagia and body weight in Prader-Willi syndrome: A randomized, double-blind, placebo-controlled trial. *Diabetes. Obes. Metab.* **19,** 1751–1761 (2017).

221. Burkey, B. F., Hoglen, N. C., Inskeep, P., Wyman, M., Hughes, T. E. & Vath, J. E. Preclinical Efficacy and Safety of the Novel Antidiabetic, Antiobesity MetAP2 Inhibitor ZGN-1061. *J. Pharmacol. Exp. Ther.* **365,** 301–313 (2018).

222. Malloy, J., Zhuang, D., Kim, T., Inskeep, P., Kim, D. & Taylor, K. Single and multiple dose evaluation of a novel MetAP2 inhibitor: Results of a randomized, double-blind, placebo-controlled clinical trial. *Diabetes. Obes. Metab.* **20,** 1878 (2018).

223. Zafgen. *Zafgen Provides Update on Investigational New Drug Application for ZGN-1061* 2018.

224. U.S. National Library of Medicine. Study to Assess the Effects and Safety of ZGN-1061 in Overweight and Obese Participants With Type 2 Diabetes. *ClinicalTrials.gov,* NCT03254368 (2019).

225. Hori, A., Ikeyama, S. & Sudo, K. Suppression of Cyclin D1 mRNA Expression by the Angiogenesis Inhibitor TNP-470 (AGM-1470) in Vascular Endothelial Cells. *Biochem. Biophys. Res. Commun.* **204,** 1067–1073 (1994).

226. Abe, J, Zhou, W, Takuwa, N, Taguchi, J, Kurokawa, K, Kumada, M & Takuwa, Y. A fumagillin derivative angiogenesis inhibitor, AGM-1470, inhibits activation of cyclin-dependent kinases and phosphorylation of retinoblastoma gene product but not protein tyrosyl phosphorylation or protooncogene expression in vascular endothelial cells. *Cancer Res.* **54,** 3407–12 (1994).

227. Bernier, S. G., Lazarus, D. D., Clark, E., Doyle, B., Labenski, M. T., Thompson, C. D., Westlin, W. F. & Hannig, G. A methionine aminopeptidase-2 inhibitor, PPI-2458, for the treatment of rheumatoid arthritis. *Proc. Natl. Acad. Sci.* **101,** 10768–10773 (2004).

228. Lin, M., Zhang, X., Jia, B. & Guan, S. Suppression of glioblastoma growth and angiogenesis through molecular targeting of methionine aminopeptidase-2. *J. Neurooncol.* **136,** 243–254 (2018).

229. Wang, J., Lou, P. & Henkin, J. Selective inhibition of endothelial cell proliferation by fumagillin is not due to differential expression of methionine aminopeptidases. *J. Cell. Biochem.* **77,** 465–473 (2000).

230. Zhang, Y, Griffith, E. C., Sage, J, Jacks, T & Liu, J. O. Cell cycle inhibition by the anti-angiogenic agent TNP-470 is mediated by p53 and p21WAF1/CIP1. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6427–32 (2000).

231. Kanno, T., Endo, H., Takeuchi, K., Morishita, Y., Fukayama, M. & Mori, S. High Expression of Methionine Aminopeptidase Type 2 in Germinal Center B Cells and Their Neoplastic Counterparts. *Lab. Investig.* **82,** 893–901 (2002).

232. Priest, R. C., Spaull, J, Buckton, J, Grimley, R. L., Sims, M, Binks, M & Malhotra, R. Immunomodulatory activity of a methionine aminopeptidase-2 inhibitor on B cell differentiation. *Clin. Exp. Immunol.* **155,** 514 (2009).

233. Sakamoto, K. M., Kim, K. B., Kumagai, A., Mercurio, F., Crews, C. M. & Deshaies, R. J. Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8554–8559 (2001).

234. Zhang, Y., Yeh, J. R., Mara, A., Ju, R., Hines, J. F., Cirone, P., Griesbach, H. L., Schneider, I., Slusarski, D. C., Holley, S. A. & Crews, C. M. A Chemical and Genetic Approach to the Mode of Action of Fumagillin. *Chem. Biol.* **13,** 1001–1009 (2006).

235. Folkman, J. Angiogenesis and c-Jun. *JNCI J. Natl. Cancer Inst.* **96,** 644–644 (2004).

236. Hoang, M. V., Whelan, M. C. & Senger, D. R. Rho activity critically and selectively regulates endothelial cell organization during angiogenesis. *Proc. Natl. Acad. Sci.* **101,** 1874–1879 (2004).

237. Sundberg, T., Darricarrere, N., Cirone, P., Li, X., McDonald, L., Mei, X., Westlake, C., Slusarski, D., Beynon, R. & Crews, C. Disruption of Wnt Planar Cell Polarity Signaling by Aberrant Accumulation of the MetAP-2 Substrate Rab37. *Chem. Biol.* **18,** 1300–1311 (2011).

238. Ma, A. C. H., Fung, T. K., Lin, R. H. C., Chung, M. I. S., Yang, D., Ekker, S. C. & Leung, A. Y. H. Methionine aminopeptidase 2 is required for HSC initiation and proliferation. *Blood* **118,** 5448–5457 (2011).

239. Aznar, N., Sun, N., Dunkel, Y., Ear, J., Buschman, M. D. & Ghosh, P. A Daple-Akt feed-forward loop enhances noncanonical Wnt signals by compartmentalizing β-catenin. *Mol. Biol. Cell* **28** (ed Luo, K.) 3709–3723 (2017).

240. Chen, G. J., Weylie, B., Hu, C., Zhu, J. & Forough, R. FGFR1/PI3K/AKT signaling pathway is a novel target for antiangiogenic effects of the cancer drug fumagillin (TNP-470). *J. Cell. Biochem.* **101,** 1492–1504 (2007).

241. Resh, M. D. Covalent lipid modifications of proteins. *Curr. Biol.* **23,** R431–5 (2013).

242. Deichaite, I, Casson, L. P., Ling, H. P. & Resh, M. D. In vitro synthesis of pp60v-src: myristylation in a cell-free system. *Mol. Cell. Biol.* **8,** 4295 (1988).

243. Zha, J., Weiler, S., Oh, K. J., Wei, M. C. & Korsmeyer, S. J. Posttranslational N-Myristoylation of BID as a Molecular Switch for Targeting Mitochondria and Apoptosis. *Science (80-.).* **290,** 1761–1765 (2000).

244. Thinon, E., Serwa, R. A., Broncel, M., Brannigan, J. A., Brassat, U., Wright, M. H., Heal, W. P., Wilkinson, A. J., Mann, D. J. & Tate, E. W. Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* **5,** 4919 (2014).

245. Martinez, A., Traverso, J. A., Valot, B., Ferro, M., Espagne, C., Ephritikhine, G., Zivy, M., Giglione, C. & Meinnel, T. Extent of N-terminal modifications in cytosolic proteins from eukaryotes. *Proteomics* **8,** 2809–2831 (2008).

246. Wright, M. H., Clough, B., Rackham, M. D., Rangachari, K., Brannigan, J. A., Grainger, M., Moss, D. K., Bottrill, A. R., Heal, W. P., Broncel, M., Serwa, R. A., Brady, D., Mann, D. J., Leatherbarrow, R. J., Tewari, R., Wilkinson, A. J., Holder, A. A. & Tate, E. W. Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. *Nat. Chem.* **6,** 112–121 (2014).

247. Wright, M., Paape, D., Storck, E., Serwa, R., Smith, D. & Tate, E. Global Analysis of Protein N-Myristoylation and Exploration of N-Myristoyltransferase as a Drug Target in the Neglected Human Pathogen Leishmania donovani. *Chem. Biol.* **22,** 342–354 (2015).

248. Maurer-Stroh, S. & Eisenhaber, F. Myristoylation of viral and bacterial proteins. *Trends Microbiol.* **12,** 178–185 (2004).

249. Giang, D. K. & Cravatt, B. F. A second mammalian N-myristoyltransferase. *J. Biol. Chem.* **273,** 6595–8 (1998).

250. Ducker, C. E., Upson, J. J., French, K. J. & Smith, C. D. Two N-myristoyltransferase isozymes play unique roles in protein myristoylation, proliferation, and apoptosis. *Mol. Cancer Res.* **3,** 463–76 (2005).

251. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).

252. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M. & Lancet, D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinforma.* **54,** 1.30.1–1.30.33 (2016).

253. Ganesan, L. & Levental, I. Pharmacological Inhibition of Protein Lipidation. *J. Membr. Biol.* **248,** 929–941 (2015).

254. Yang, S. H., Shrivastav, A., Kosinski, C., Sharma, R. K., Chen, M. H., Berthiaume, L. G., Peters, L. L., Chuang, P. T., Young, S. G. & Bergo, M. O. N-myristoyltransferase 1 is essential in early mouse development. *J. Biol. Chem.* **280,** 18990–18995 (2005).

255. Shrivastav, A., Varma, S., Lawman, Z., Yang, S. H., Ritchie, S. A., Bonham, K., Singh, S. M., Saxena, A. & Sharma, R. K. Requirement of N-myristoyltransferase 1 in the development of monocytic lineage. *J. Immunol.* **180,** 1019–28 (2008).

256. Goncalves, V., Brannigan, J. A., Thinon, E., Olaleye, T. O., Serwa, R., Lanzarone, S., Wilkinson, A. J., Tate, E. W. & Leatherbarrow, R. J. A fluorescence-based assay for N-myristoyltransferase activity. *Anal. Biochem.* **421,** 342–4 (2012).

257. Rudnick, D. A., McWherter, C. A., Rocque, W. J., Lennon, P. J., Getman, D. P. & Gordon, J. I. Kinetic and structural evidence for a sequential ordered Bi Bi mechanism of catalysis by Saccharomyces cerevisiae myristoyl-CoA:protein N-myristoyltransferase. *J. Biol. Chem.* **266,** 9732–9 (1991).

258. Weston, S. A., Camble, R., Colls, J., Rosenbrock, G., Taylor, I., Egerton, M., Tucker, A. D., Tunnicliffe, A., Mistry, A., Mancia, F., de la Fortelle, E., Irwin, J., Bricogne, G. & Pauptit, R. A. Crystal structure of the anti-fungal target N-myristoyl transferase. *Nat. Struct. Biol.* **5,** 213–221 (1998).

259. Heuckeroth, R. O., Glaser, L & Gordon, J. I. Heteroatom-substituted fatty acid analogs as substrates for N-myristoyltransferase: an approach for studying both the enzymology and function of protein acylation. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8795–9 (1988).

260. Rudnick, D. A., Lu, T, Jackson-Machelski, E, Hernandez, J. C., Li, Q, Gokel, G. W. & Gordon, J. I. Analogs of palmitoyl-CoA that are substrates for myristoyl-CoA:protein N-myristoyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 10507–11 (1992).

261. Maurer-Stroh, S., Eisenhaber, B. & Eisenhaber, F. N-terminal N -myristoylation of proteins: prediction of substrate proteins from amino acid sequence 1 1Edited by J. Thornton. *J. Mol. Biol.* **317,** 541–557 (2002).

262. Maurer-Stroh, S., Eisenhaber, B. & Eisenhaber, F. N-terminal N -myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences 1 1Edited by J. Thornton. *J. Mol. Biol.* **317,** 523–540 (2002).

263. Utsumi, T., Nakano, K., Funakoshi, T., Kayano, Y., Nakao, S., Sakurai, N., Iwata, H. & Ishisaka, R. Vertical-scanning mutagenesis of amino acids in a model N-myristoylation motif reveals the major amino-terminal sequence requirements for protein N-myristoylation. *Eur. J. Biochem.* **271,** 863–874 (2004).

264. Johnson, D. R., Bhatnagar, R. S., Knoll, L. J. & Gordon, J. I. Genetic and Biochemical Studies of Protein N-Myristoylation. *Annu. Rev. Biochem.* **63,** 869–914 (1994).

265. Bereta, G. & Palczewski, K. Heterogeneous N-Terminal Acylation of Retinal Proteins Results from the Retina's Unusual Lipid Metabolism. *Biochemistry* **50,** 3764–3776 (2011).

266. Tang, H. & Han, M. Fatty Acids Regulate Germline Sex Determination through ACS-4-Dependent Myristoylation. *Cell* **169,** 457–469.e13 (2017).

267. Rioux, V. & Legrand, P. Saturated fatty acids: simple molecular structures with complex cellular functions. *Curr. Opin. Clin. Nutr. Metab. Care* **10,** 752–758 (2007).

268. Bhatnagar, R. S., Schall, O. F., Jackson-Machelski, E., Sikorski, J. A., Devadas, B., Gokel, G. W. & Gordon, J. I. Titration Calorimetric Analysis of AcylCoA Recognition by MyristoylCoA:Protein N-Myristoyltransferase †. *Biochemistry* **36,** 6700–6708 (1997).

269. Glover, C. J., Hartman, K. D. & Felsted, R. L. Human N-Myristoyltransferase Amino-terminal Domain Involved in Targeting the Enzyme to the Ribosomal Subcellular Fraction. *J. Biol. Chem.* **272,** 28680–28689 (1997).

270. Perinpanayagam, M. A., Beauchamp, E., Martin, D. D. O., Sim, J. Y. W., Yap, M. C. & Berthiaume, L. G. Regulation of co- and post-translational myristoylation of proteins during apoptosis: interplay of N-myristoyltransferases and caspases. *FASEB J.* **27,** 811–821 (2013).

271. Duronio, R. J., Towler, D. A., Heuckeroth, R. O. & Gordon, J. I. Disruption of the yeast N-myristoyl transferase gene causes recessive lethality. *Science (80-.).* **243,** 796–800 (1989).

272. Lodge, J. K., Jackson-Machelski, E, Toffaletti, D. L., Perfect, J. R. & Gordon, J. I. Targeted gene replacement demonstrates that myristoyl-CoA: protein N-myristoyltransferase is essential for viability of Cryptococcus neoformans. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12008–12 (1994).

273. Price, H. P., Menon, M. R., Panethymitaki, C., Goulding, D., McKean, P. G. & Smith, D. F. Myristoyl-CoA:Protein N-Myristoyltransferase, an Essential Enzyme and Potential Drug Target in Kinetoplastid Parasites. *J. Biol. Chem.* **278,** 7206–7214 (2003).

274. Farazi, T. A., Waksman, G & Gordon, J. I. The biology and enzymology of protein N-myristoylation. *J. Biol. Chem.* **276,** 39501–4 (2001).

275. Peitzsch, R. M. & McLaughlin, S. Binding of acylated peptides and fatty acids to phospholipid vesicles: Pertinence to myristoylated proteins. *Biochemistry* **32,** 10436–10443 (1993).

276. Traverso, J. A., Micalella, C., Martinez, A., Brown, S. C., Satiat-Jeunemaître, B., Meinnel, T. & Giglione, C. Roles of N-Terminal Fatty Acid Acylations in Membrane Compartment Partitioning: Arabidopsis H-Type Thioredoxins as a Case Study. *Plant Cell* **25,** 1056–1077 (2013).

277. Resh, M. D. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1451,** 1–16 (1999).

278. Resh, M. D. Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat. Chem. Biol.* **2,** 584–590 (2006).

279. Resh, M. D. Palmitoylation of Ligands, Receptors, and Intracellular Signaling Molecules. *Sci. STKE,* re14 (2006).

280. Ames, J. B., Porumb, T., Tanaka, T., Ikura, M. & Stryer, L. Amino-terminal Myristoylation Induces Cooperative Calcium Binding to Recoverin. *J. Biol. Chem.* **270,** 4526–4533 (1995).

281. Patwardhan, P. & Resh, M. D. Myristoylation and membrane binding regulate c-Src stability and kinase activity. *Mol. Cell. Biol.* **30,** 4094–107 (2010).

282. Sigal, C. T., Zhou, W, Buser, C. A., McLaughlin, S & Resh, M. D. Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12253–7 (1994).

283. Resh, M. D. A myristoyl switch regulates membrane binding of HIV-1 Gag. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 417–8 (2004).

284. Swierczynski, S. L. & Blackshear, P. J. Myristoylation-dependent and electrostatic interactions exert independent effects on the membrane association of the myristoylated alanine-rich protein kinase C substrate protein in intact cells. *J. Biol. Chem.* **271,** 23424–30 (1996).

285. McLaughlin, S & Aderem, A. The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* **20,** 272–6 (1995).

286. Zheng, J., Knighton, D. R., Taylor, S. S., Xuong, N.-H., Sowadski, J. M. & Eyck, L. F. T. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci.* **2,** 1559–1573 (1993).

287. Hantschel, O., Nagar, B., Guettler, S., Kretzschmar, J., Dorey, K., Kuriyan, J. & Superti-Furga, G. A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112,** 845–57 (2003).

288. Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G. & Kuriyan, J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112,** 859–71 (2003).

289. Matsubara, M, Titani, K, Taniguchi, H & Hayashi, N. Direct Involvement of Protein Myristoylation in Myristoylated Alanine-Rich C Kinase Substrate (MARCKS)-calmodulin Interaction. *J. Biol. Chem.* **278** (2003).

290. Hayashi, N., Izumi, Y., Titani, K. & Matsushima, N. The binding of myristoylated N-terminal nonapeptide from neuron-specific protein CAP-23/NAP-22 to calmodulin does not induce the globular structure observed for the calmodulin-nonmyristoylated peptide complex. *Protein Sci.* **9,** 1905–1913 (2000).

291. Sharma, R. K. Potential role of N-myristoyltransferase in pathogenic conditions. *Can. J. Physiol. Pharmacol.* **82,** 849–859 (2004).

292. Schlott, A. C., Holder, A. A. & Tate, E. W. N-Myristoylation as a Drug Target in Malaria: Exploring the Role of N-Myristoyltransferase Substrates in the Inhibitor Mode of Action. *ACS Infect. Dis.* **4,** 449–457 (2018).

293. Rackham, M. D., Yu, Z., Brannigan, J. A., Heal, W. P., Paape, D., Barker, K. V., Wilkinson, A. J., Smith, D. F., Leatherbarrow, R. J. & Tate, E. W. Discovery of high affinity inhibitors of Leishmania donovani N-myristoyltransferase. *Medchemcomm* **6,** 1761 (2015).

294. Herrera, L. J., Brand, S., Santos, A., Nohara, L. L., Harrison, J., Norcross, N. R., Thompson, S., Smith, V., Lema, C., Varela-Ramirez, A., Gilbert, I. H., Almeida, I. C. & Maldonado, R. A. Validation of N-myristoyltransferase as Potential Chemotherapeutic Target in Mammal-Dwelling Stages of Trypanosoma cruzi. *PLoS Negl. Trop. Dis.* **10** (ed Franco-Paredes, C.) e0004540 (2016).

295. Weinberg, R. A., McWherter, C. A., Freeman, S. K., Wood, D. C., Gordon, J. I. & Lee, S. C. Genetic studies reveal that myristoylCoA:protein N-myristoyltransferase is an essential enzyme in Candida albicans. *Mol. Microbiol.* **16,** 241–250 (1995).

296. Chow, M., Newman, J. F. E., Filman, D., Hogle, J. M., Rowlands, D. J. & Brown, F. Myristylation of picornavirus capsid protein VP4 and its structural significance. *Nature* **327,** 482–486 (1987).

297. Corbic Ramljak, I., Stanger, J., Real-Hohn, A., Dreier, D., Wimmer, L., Redlberger-Fritz, M., Fischl, W., Klingel, K., Mihovilovic, M. D., Blaas, D. & Kowalski, H. Cellular N-myristoyltransferases play a crucial picornavirus genus-specific role in viral assembly, virion maturation, and infectivity. *PLoS Pathog.* **14,** e1007203 (2018).

298. Mousnier, A., Bell, A. S., Swieboda, D. P., Morales-Sanfrutos, J., Pérez-Dorado, I., Brannigan, J. A., Newman, J., Ritzefeld, M., Hutton, J. A., Guedán, A., Asfor, A. S., Robinson, S. W., Hopkins-Navratilova, I., Wilkinson, A. J., Johnston, S. L., Leatherbarrow, R. J., Tuthill, T. J., Solari, R. & Tate, E. W. Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat. Chem.* **10,** 599–606 (2018).

299. Felsted, R. L., Glover, C. J. & Hartman, K. *Protein N-myristoylation as a chemotherapeutic target for cancer* 1995.

300. Wheeler, D. L., Iida, M. & Dunn, E. F. The Role of Src in Solid Tumors. *Oncologist* **14,** 667 (2009).

301. Kamps, M. P., Buss, J. E. & Sefton, B. M. Rous sarcoma virus transforming protein lacking myristic acid phosphorylates known polypeptide substrates without inducing transformation. *Cell* **45,** 105–112 (1986).

302. Bagrodia, S, Taylor, S. J. & Shalloway, D. Myristylation is required for Tyr-527 dephosphorylation and activation of pp60c-src in mitosis. *Mol. Cell. Biol.* **13,** 1464–70 (1993).

303. Ureshino, H, Shindo, T & Kimura, K. Role of Cancer Immunology in Chronic Myelogenous Leukemia. *Leuk. Res.* **88** (2020).

304. Skora, L, Mestan, J, Fabbro, D, Jahnke, W & Grzesiek, S. NMR Reveals the Allosteric Opening and Closing of Abelson Tyrosine Kinase by ATP-site and Myristoyl Pocket Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **110** (2013).

305. Fong, L. W. R., Yang, D. C. & Chen, C.-H. Myristoylated alanine-rich C kinase substrate (MARCKS): a multirole signaling protein in cancers. *Cancer Metastasis Rev.* **36,** 737–747 (2017).

306. Uno, F., Sasaki, J., Nishizaki, M., Carboni, G., Xu, K., Atkinson, E. N., Kondo, M., Minna, J. D., Roth, J. A. & Ji, L. Myristoylation of the fus1 protein is required for tumor suppression in human lung cancer cells. *Cancer Res.* **64,** 2969–76 (2004).

307. Rajala, R. V., Radhi, J. M., Kakkar, R, Datla, R. S. & Sharma, R. K. Increased expression of N-myristoyltransferase in gallbladder carcinomas. *Cancer* **88,** 1992–9 (2000).

308. Lu, Y., Selvakumar, P., Ali, K., Shrivastav, A., Bajaj, G., Resch, L., Griebel, R., Fourney, D., Meguro, K. & Sharma, R. K. Expression of N-myristoyltransferase in human brain tumors. *Neurochem. Res.* **30,** 9–13 (2005).

309. Shrivastav, A., Sharma, A. R., Bajaj, G., Charavaryamath, C., Ezzat, W., Spafford, P., Gore-Hickman, R., Singh, B., Copete, M. A. & Sharma, R. K. Elevated N-myristoyltransferase activity and expression in oral squamous cell carcinoma. *Oncol. Rep.* **18,** 93–7 (2007).

310. Kumar, S., Dimmock, J. R. & Sharma, R. K. The Potential Use of N-Myristoyltransferase as a Biomarker in the Early Diagnosis of Colon Cancer. *Cancers (Basel).* **3,** 1372–1382 (2011).

311. Bergqvist, M., Brattström, D., Brodin, D., Lindkvist, A., Dahlman-Wright, K., Dreilich, M., Wagenius, G. & Paulsson-Karlsson, Y. Genes associated with telomerase activity levels in esophageal carcinoma cell lines. *Dis. Esophagus* **19,** 20–23 (2006).

312. Thinon, E., Morales-Sanfrutos, J., Mann, D. J. & Tate, E. W. N-Myristoyltransferase Inhibition Induces ER-Stress, Cell Cycle Arrest, and Apoptosis in Cancer Cells. *ACS Chem. Biol.,* acschembio.6b00371 (2016).

313. Schlott, A. C., Mayclin, S., Reers, A. R., Coburn-Flynn, O., Bell, A. S., Green, J., Knuepfer, E., Charter, D., Bonnert, R., Campo, B., Burrows, J., Lyons-Abbott, S., Staker, B. L., Chung, C.-W., Myler, P. J., Fidock, D. A., Tate, E. W. & Holder, A. A. Structure-Guided Identification of Resistance Breaking Antimalarial N-Myristoyltransferase Inhibitors. *Cell Chem. Biol.* **26,** 991–1000.e7 (2019).

314. Resh, M. D. Targeting Protein Lipidation in Disease. *Trends Mol. Med.* **18,** 206 (2012).

315. Sogabe, S., Masubuchi, M., Sakata, K., Fukami, T. A., Morikami, K., Shiratori, Y., Ebiike, H., Kawasaki, K., Aoki, Y., Shimma, N., D'Arcy, A., Winkler, F. K., Banner, D. W. & Ohtsuka, T. Crystal Structures of Candida albicans N-Myristoyltransferase with Two Distinct Inhibitors. *Chem. Biol.* **9,** 1119–1128 (2002).

316. Yamazaki, K., Kaneko, Y., Suwa, K., Ebara, S., Nakazawa, K. & Yasuno, K. Synthesis of potent and selective inhibitors of Candida albicans N-myristoyltransferase based on the benzothiazole structure. *Bioorg. Med. Chem.* **13,** 2509–2522 (2005).

317. Frearson, J. A., Brand, S., McElroy, S. P., Cleghorn, L. A. T., Smid, O., Stojanovski, L., Price, H. P., Guther, M. L. S., Torrie, L. S., Robinson, D. A., Hallyburton, I., Mpamhanga, C. P., Brannigan, J. A., Wilkinson, A. J., Hodgkinson, M., Hui, R., Qiu, W., Raimi, O. G., van Aalten, D. M. F., Brenk, R., Gilbert, I. H., Read, K. D., Fairlamb, A. H., Ferguson, M. A. J., Smith, D. F. & Wyatt, P. G. N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature* **464,** 728–732 (2010).

318. Brand, S., Cleghorn, L. A. T., McElroy, S. P., Robinson, D. A., Smith, V. C., Hallyburton, I., Harrison, J. R., Norcross, N. R., Spinks, D., Bayliss, T., Norval, S., Stojanovski, L., Torrie, L. S., Frearson, J. A., Brenk, R., Fairlamb, A. H., Ferguson, M. A. J., Read, K. D., Wyatt, P. G. & Gilbert, I. H. Discovery of a Novel Class of Orally Active Trypanocidal N-Myristoyltransferase Inhibitors. *J. Med. Chem.* **55,** 140–152 (2012).

319. Pino, P., Sebastian, S., Kim, E., Bush, E., Brochet, M., Volkmann, K., Kozlowski, E., Llinás, M., Billker, O. & Soldati-Favre, D. A Tetracycline-Repressible Transactivator System to Study Essential Genes in Malaria Parasites. *Cell Host Microbe* **12,** 824–834 (2012).

320. Bell, A., Mills, J., Williams, G., Brannigan, J., Wilkinson, A., Parkinson, T, Leatherbarrow, R., Tate, E., Holder, A. & Smith, D. Selective inhibitors of protozoan protein N-myristoyltransferases as starting points for tropical disease medicinal chemistry programs. *PLoS Negl. Trop. Dis.* **6,** e1625–e1625 (2012).

321. Goncalves, V., Brannigan, J. A., Whalley, D., Ansell, K. H., Saxty, B., Holder, A. A., Wilkinson, A. J., Tate, E. W. & Leatherbarrow, R. J. Discovery of Plasmodium vivax N-Myristoyltransferase Inhibitors: Screening, Synthesis, and Structural Characterization of their Binding Mode. *J. Med. Chem.* **55,** 3578–3582 (2012).

322. Yu, Z., Brannigan, J. A., Moss, D. K., Brzozowski, A. M., Wilkinson, A. J., Holder, A. A., Tate, E. W. & Leatherbarrow, R. J. Design and Synthesis of Inhibitors of Plasmodium falciparum N-Myristoyltransferase, A Promising Target for Antimalarial Drug Discovery. *J. Med. Chem.* **55,** 8879–8890 (2012).

323. Rackham, M. D., Brannigan, J. A., Moss, D. K., Yu, Z., Wilkinson, A. J., Holder, A. A., Tate, E. W. & Leatherbarrow, R. J. Discovery of Novel and Ligand-Efficient Inhibitors of Plasmodium falciparum and Plasmodium vivax N-Myristoyltransferase. *J. Med. Chem.* **56,** 371–375 (2013).

324. Rackham, M. D., Brannigan, J. A., Rangachari, K., Meister, S., Wilkinson, A. J., Holder, A. A., Leatherbarrow, R. J. & Tate, E. W. Design and Synthesis of High Affinity Inhibitors of Plasmodium falciparum and Plasmodium vivax N-Myristoyltransferases Directed by Ligand Efficiency Dependent Lipophilicity (LELP). *J. Med. Chem.* **57,** 2773 (2014).

325. Kallemeijn, W. W., Lueg, G. A., Faronato, M., Hadavizadeh, K., Goya Grocin, A., Song, O.-R. R., Howell, M., Calado, D. P. & Tate, E. W. Validation and Invalidation of Chemical Probes for the Human N-myristoyltransferases. *Cell Chem. Biol.* **26,** 892–900.e4 (2019).

326. Zhou, W, Parent, L. J., Wills, J. W. & Resh, M. D. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* **68,** 2556–2569 (1994).

327. Furuishi, K., Matsuoka, H., Takama, M., Takahashi, I., Misumi, S. & Shoji, S. Blockage ofN-Myristoylation of HIV-1 Gag Induces the Production of Impotent Progeny Virus. *Biochem. Biophys. Res. Commun.* **237,** 504–511 (1997).

328. Knorr, A. G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B. & Beckmann, R. Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.* **26,** 35–39 (2019).

329. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520,** 640–645 (2015).

330. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13,** 355–369 (2012).

331. Guo, H. Specialized ribosomes and the control of translation. *Biochem. Soc. Trans.* **46,** 855–869 (2018).

332. Simsek, D., Tiu, G. C., Flynn, R. A., Byeon, G. W., Leppek, K., Xu, A. F., Chang, H. Y. & Barna, M. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **169,** 1051–1065.e18 (2017).

333. McIlhinney, R. A. & McGlone, K. Evidence for a non-myristoylated pool of the 80 kDa protein kinase C substrate of rat brain. *Biochem. J.* **271,** 681–685 (1990).

334. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Mol. Cell. Proteomics* **13,** 3497–3506 (2014).

335. Heal, W. P., Wright, M. H., Thinon, E. & Tate, E. W. Multifunctional protein labeling via enzymatic N-terminal tagging and elaboration by click chemistry. *Nat. Protoc.* **7,** 105–117 (2012).

336. Broncel, M., Serwa, R. A., Ciepla, P., Krause, E., Dallman, M. J., Magee, A. I. & Tate, E. W. Myristoylation profiling in human cells and zebrafish. *Data Br.* **4,** 379–383 (2015).

337. Yuanwu, M., Lianfeng, Z. & Xingxu, H. Genome Modification by CRISPR/Cas9. *FEBS J.* **281,** 5186–93 (2014).

338. Thompson, S. L. & Compton, D. A. Chromosomes and cancer cells. *Chromosome Res.* **19,** 433–44 (2011).

339. Aguirre, A. J., Meyers, R. M., Weir, B. A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W. F., Doshi, M. B., Kost-Alimova, M., Gill, S., Xu, H., Ali, L. D., Jiang, G., Pantel, S., Lee, Y., Goodale, A., Cherniack, A. D., Oh, C., Kryukov, G., Cowley, G. S., Garraway, L. A., Stegmaier, K., Roberts, C. W., Golub, T. R., Meyerson, M., Root, D. E., Tsherniak, A. & Hahn, W. C. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* **6,** 914–29 (2016).

340. Delitto, D., Judge, S. M., Delitto, A. E., Nosacka, R. L., Rocha, F. G., DiVita, B. B., Gerber, M. H., George, T. J., Behrns, K. E., Hughes, S. J., Wallet, S. M., Judge, A. R., Trevino, J. G. & Trevino, J. G. Human pancreatic cancer xenografts recapitulate key aspects of cancer cachexia. *Oncotarget* **8,** 1177–1189 (2017).

341. Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E. G., Van Drogen, A., Borel, C., Frank, M., Germain, P.-L., Bludau, I., Mehnert, M., Seifert, M., Emmenlauer, M., Sorg, I., Bezrukov, F., Bena, F. S., Zhou, H., Dehio, C., Testa, G., Saez-Rodriguez, J., Antonarakis, S. E., Hardt, W.-D. & Aebersold, R. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.* **37,** 314–322 (2019).

342. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. & Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

343. Ding, Q., Regan, S. N., Xia, Y., Oostrom, L. A., Cowan, C. A. & Musunuru, K. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12,** 393–4 (2013).

344. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J. & Joung, J. K. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

345. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids* **4,** e264 (2015).

346. Neurohr, G. E., Terry, R. L., Lengefeld, J., Bonney, M., Brittingham, G. P., Moretto, F., Miettinen, T. P., Vaites, L. P., Soares, L. M., Paulo, J. A., Harper, J. W., Buratowski, S., Manalis, S., van Werven, F. J., Holt, L. J. & Amon, A. Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. *Cell* **176,** 1083–1097.e18 (2019).

347. Riss, T. L., Moravec, R. A., Niles, A. L., Duellman, S., Benink, H. A., Worzella, T. J. & Minor, L. in *Assay Guid. Man. [Internet]* July 1, 20. Chap. Cell Viabi (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2013).

348. Ahmad, Y. & Lamond, A. I. A perspective on proteomics in cell biology. *Trends Cell Biol.* **24,** 257–64 (2014).

349. Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13,** S12 (2012).

350. Goya Grocin, A., Serwa, R. A., Morales Sanfrutos, J., Ritzefeld, M. & Tate, E. W. Whole proteome profiling of N-myristoyltransferase activity and inhibition using sortase A. *Mol. Cell. Proteomics* **18,** 115–126 (2019).

351. Broncel, M., Serwa, R. A., Ciepla, P., Krause, E., Dallman, M. J., Magee, A. I. & Tate, E. W. Multifunctional Reagents for Quantitative Proteome-Wide Analysis of Protein Modification in Human Cells and Dynamic Profiling of Protein Lipidation During Vertebrate Development. *Angew. Chemie Int. Ed.* **54,** 5948–5951 (2015).

352. Proft, T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol. Lett.* **32,** 1–10 (2010).

353. Mazmanian, S. K., Liu, G, Ton-That, H & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285,** 760–3 (1999).

354. Wu, Z. & Guo, Z. Sortase-Mediated Transpeptidation for Site-Specific Modification of Peptides, Glycopeptides, and Proteins. *J. Carbohydr. Chem.* **31,** 48–66 (2012).

355. Antos, J. M., Truttmann, M. C. & Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38,** 111–8 (2016).

356. Sarpong, K. & Bose, R. Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins. *Anal. Biochem.* **521,** 55–58 (2017).

357. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11399–404 (2011).

358. Hirakawa, H., Ishikawa, S. & Nagamune, T. Design of Ca 2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **109,** 2955–2961 (2012).

359. Hirakawa, H., Ishikawa, S. & Nagamune, T. Ca2+ -independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of Escherichia coli. *Biotechnol. J.* **10,** 1487–1492 (2015).

360. Williamson, D. J., Fascione, M. A., Webb, M. E. & Turnbull, W. B. Efficient N-terminal labeling of proteins by use of sortase. *Angew. Chem. Int. Ed. Engl.* **51,** 9377–80 (2012).

361. Williamson, D. J., Webb, M. E. & Turnbull, W. B. Depsipeptide substrates for sortase-mediated N-terminal protein ligation. *Nat. Protoc.* **9,** 253–262 (2014).

362. Schmohl, L. & Schwarzer, D. Sortase-mediated ligations for the site-specific modification of proteins. *Curr. Opin. Chem. Biol.* **22,** 122–128 (2014).

363. McClatchy, D. B., Ma, Y., Liu, C., Stein, B. D., Martínez-Bartolomé, S., Vasquez, D., Hellberg, K., Shaw, R. J., Yates, J. R. & III. Pulsed Azidohomoalanine Labeling in Mammals (PALM) Detects Changes in Liver-Specific LKB1 Knockout Mice. *J. Proteome Res.* **14,** 4815–22 (2015).

364. Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K. & Ellestad, G. Kinetic Mechanism of Staphylococcus aureus Sortase SrtA. *Biochemistry* **42,** 11307–11315 (2003).

365. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **6,** 786–787 (2009).

366. Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., Kolippakkam, D., Mintseris, J., Obar, R. A., Harris, T., Artavanis-Tsakonas, S., Sowa, M. E., De Camilli, P., Paulo, J. A., Harper, J. W. & Gygi, S. P. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162,** 425–440 (2015).

367. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266,** 171–84 (2015).

368. Lowther, W. T., McMillen, D. A., Orville, A. M. & Matthews, B. W. The anti-angiogenic agent fumagillin covalently modifies a conserved active-site histidine in the Escherichia coli methionine aminopeptidase. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12153–7 (1998).

369. Yin, S, Wang, J, Zhang, C & Liu, Z. The Development of MetAP-2 Inhibitors in Cancer Treatment. *Curr. Med. Chem.* **19,** 1021–1035 (2012).

370. Warder, S. E., Tucker, L. A., McLoughlin, S. M., Strelitzer, T. J., Meuth, J. L., Zhang, Q., Sheppard, G. S., Richardson, P. L., Lesniewski, R., Davidsen, S. K., Bell, R. L., Rogers, J. C. & Wang, J. Discovery, Identification, and Characterization of Candidate Pharmacodynamic Markers of Methionine Aminopeptidase-2 Inhibition. *J. Proteome Res.* **7,** 4807–4820 (2008).

371. Hu, X., Addlagatta, A., Lu, J., Matthews, B. W. & Liu, J. O. Elucidation of the function of type 1 human methionine aminopeptidase during cell cycle progression. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 18148–53 (2006).

372. Martinez Molina, D. & Nordlund, P. The Cellular Thermal Shift Assay: A Novel Biophysical Assay for In Situ Drug Target Engagement and Mechanistic Biomarker Studies. *Annu. Rev. Pharmacol. Toxicol.* **56,** 141–161 (2016).

373. Savitski, M. M., Reinhard, F. B. M., Franken, H., Werner, T., Savitski, M. F., Eberhard, D., Molina, D. M., Jafari, R., Dovega, R. B., Klaeger, S., Kuster, B., Nordlund, P., Bantscheff, M. & Drewes, G. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science (80-.).* **346,** 1255784 (2014).

374. Dziekan, J. M., Yu, H., Chen, D., Dai, L., Wirjanata, G., Larsson, A., Prabhu, N., Sobota, R. M., Bozdech, Z. & Nordlund, P. Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay. *Sci. Transl. Med.* **11** (2019).

375. Marino, G., Eckhard, U. & Overall, C. M. Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chem. Biol.* **10,** 1754–1764 (2015).

376. Xu, G., Shin, S. B. Y. & Jaffrey, S. R. Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19310–5 (2009).

377. Mahrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlingame, A. L. & Wells, J. A. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134,** 866–876 (2008).

378. Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. & Vandekerckhove, J. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* **21,** 566–569 (2003).

379. Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J. & Gevaert, K. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **8,** 1362–70 (2008).

380. Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J. & Gevaert, K. Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat. Protoc.* **6,** 1130–1141 (2011).

381. Venne, A. S., Vögtle, F.-N., Meisinger, C., Sickmann, A. & Zahedi, R. P. Novel Highly Sensitive, Specific, and Straightforward Strategy for Comprehensive N-Terminal Proteomics Reveals Unknown Substrates of the Mitochondrial Peptidase Icp55. *J. Proteome Res.* **12,** 3823–3830 (2013).

382. Bienvenut, W. V., Giglione, C. & Meinnel, T. Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition. *Proteomics* **15,** 2503–2518 (2015).

383. Bienvenut, W. V., Giglione, C. & Meinnel, T. in *Methods Mol. Biol.* 17–34 (Humana Press, New York, NY, 2017).

384. Shen, P. T., Hsu, J. L. & Chen, S. H. Dimethyl isotope-coded affinity selection for the analysis of free and blocked N-termini of proteins using LC-MS/MS. *Anal. Chem.* **79,** 9520–9530 (2007).

385. Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N. & Overall, C. M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28,** 281–288 (2010).

386. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J. N. & Overall, C. M. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* **6,** 1578–1611 (2011).

387. Mommen, G. P. M., van de Waterbeemd, B., Meiring, H. D., Kersten, G., Heck, A. J. R. & de Jong, A. P. J. M. Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO(2)-based depletion. *Mol. Cell. Proteomics* **11,** 832–42 (2012).

388. McDonald, L., Robertson, D. H. L., Hurst, J. L. & Beynon, R. J. Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat. Methods* **2,** 955–957 (2005).

389. Chen, L., Shan, Y., Weng, Y., Sui, Z., Zhang, X., Liang, Z., Zhang, L. & Zhang, Y. Hydrophobic Tagging-Assisted N-Termini Enrichment for In-Depth N-Terminome Analysis. *Anal. Chem.* **88,** 8390–8395 (2016).

390. Bagert, J. D., Xie, Y. J., Sweredoski, M. J., Qi, Y., Hess, S., Schuman, E. M. & Tirrell, D. A. Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol. Cell. Proteomics* **13,** 1352–1358 (2014).

391. Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates, J. R. Quantitative analysis of newly synthesized proteins. *Nat. Protoc.* **13,** 1744–1762 (2018).

392. Rothenberg, D. A., Taliaferro, J. M., Huber, S. M., Begley, T. J., Dedon, P. C. & White, F. M. A Proteomics Approach to Profiling the Temporal Translational Response to Stress and Growth. *iScience* **9,** 367–381 (2018).

393. Truong, T. H., Garcia, F. J., Seo, Y. H. & Carroll, K. S. Isotope-coded chemical reporter and acid-cleavable affinity reagents for monitoring protein sulfenic acids. *Bioorg. Med. Chem. Lett.* **21,** 5015–5020 (2011).

394. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X. & Thomas, P. D. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14,** 703–721 (2019).

395. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. & Vilo, J. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47,** W191–W198 (2019).

396. Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C. & Apweiler, R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25,** 3045–6 (2009).

397. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).

398. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–70 (2003).

399. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324 (2014).

400. Werner, T., Sweetman, G., Savitski, M. F., Mathieson, T., Bantscheff, M. & Savitski, M. M. Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. *Anal. Chem.* **86,** 3594–3601 (2014).

401. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).

402. R Core Team. *R: A Language and Environment for Statistical Computing* Viena, Austria, 2016.

403. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M. & Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13,** 731–740 (2016).

A

Appendix

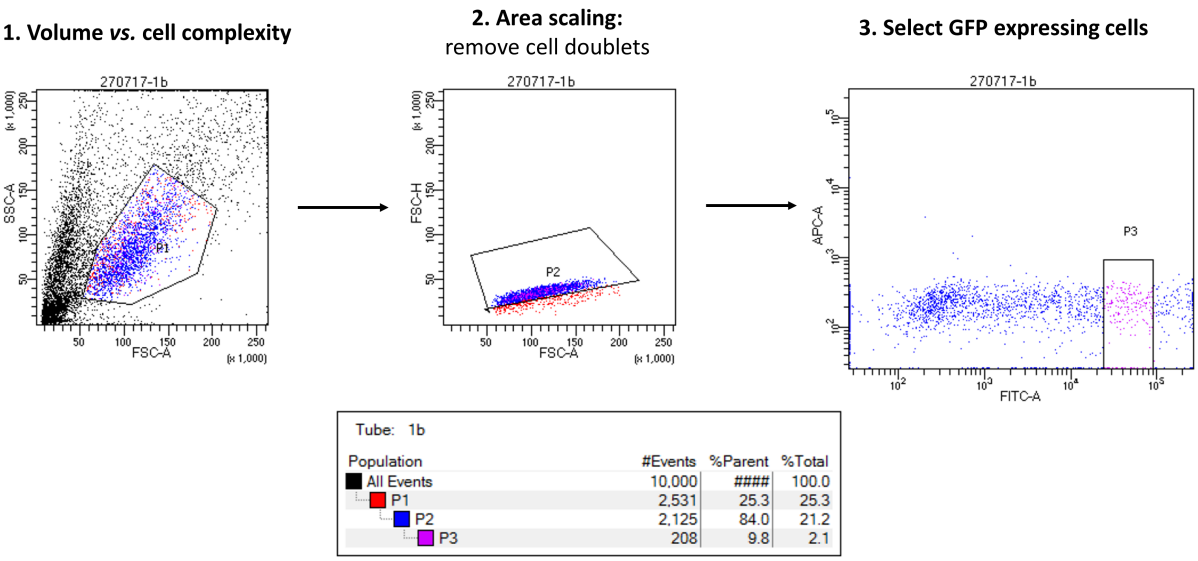


Figure A.1: Example of the gating settings used to select the GFP-positive cells during FACS procedure. This example corresponds to Panc-1 cells transfected with AG1 gRNA, bur the exact same conditions were used to sort the rest of the clones.

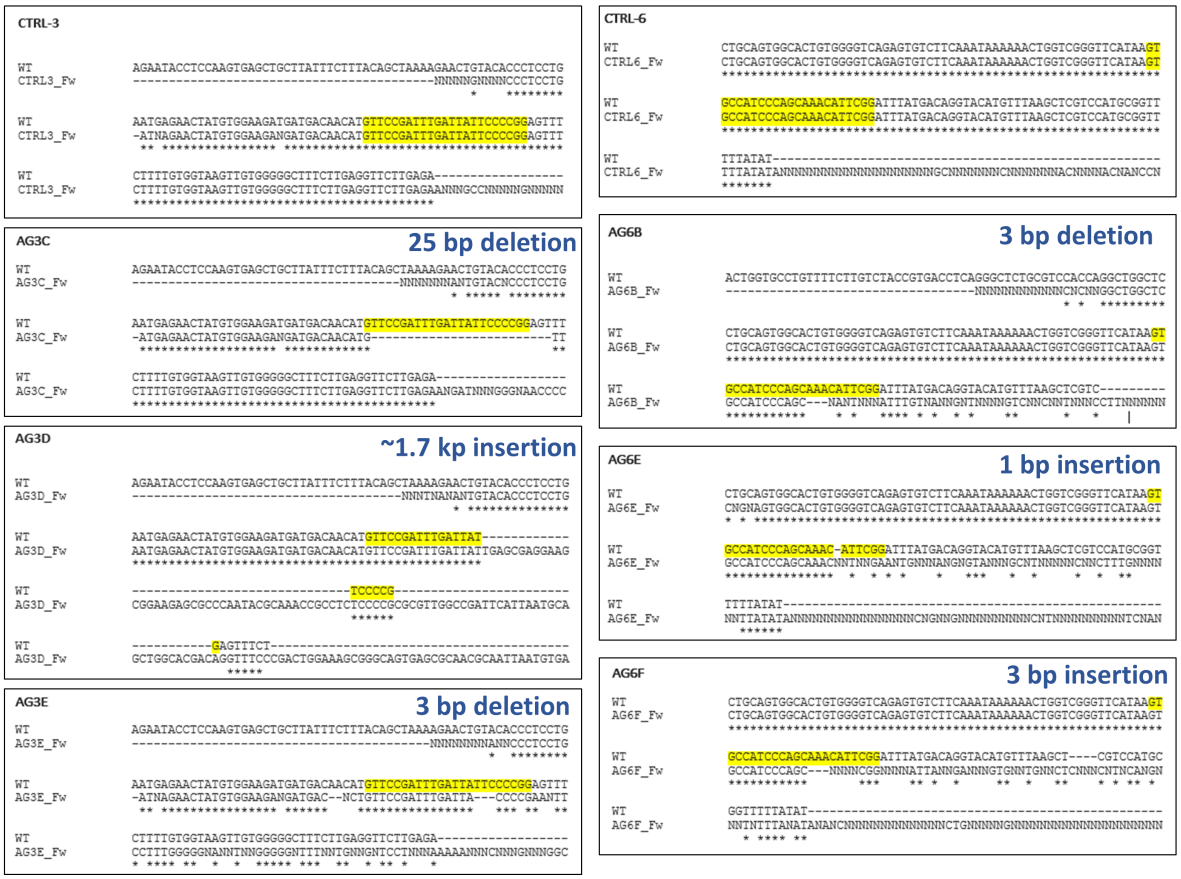


Figure A.2: Sequence alignment of sequencing results of CTRL, AG3 and AG6 clones against WT sequence. The genomic region targeted by each gRNA was amplified by PCR, purified and sent for sequencing. Sequences were then aligned to their corresponding WT sequences using the MUSCLE multiple sequence alignment online tool (https://www.ebi.ac.uk/Tools/msa/muscle/)

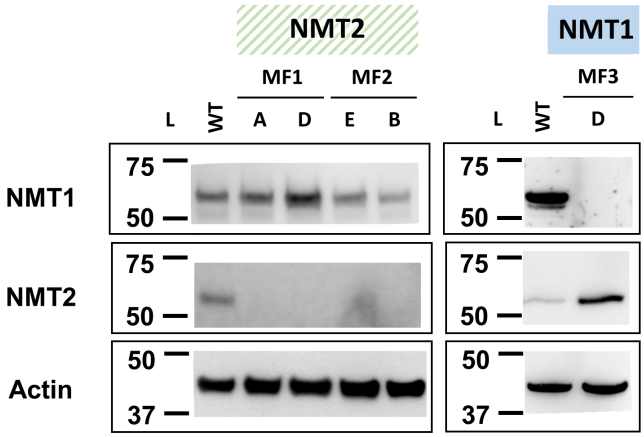


Figure A.3: Western blot showing full ablation of NMT1 and NMT2 in HeLa knock outs. HeLa clones and this western blot were generated by Dr. Monica Faronato.

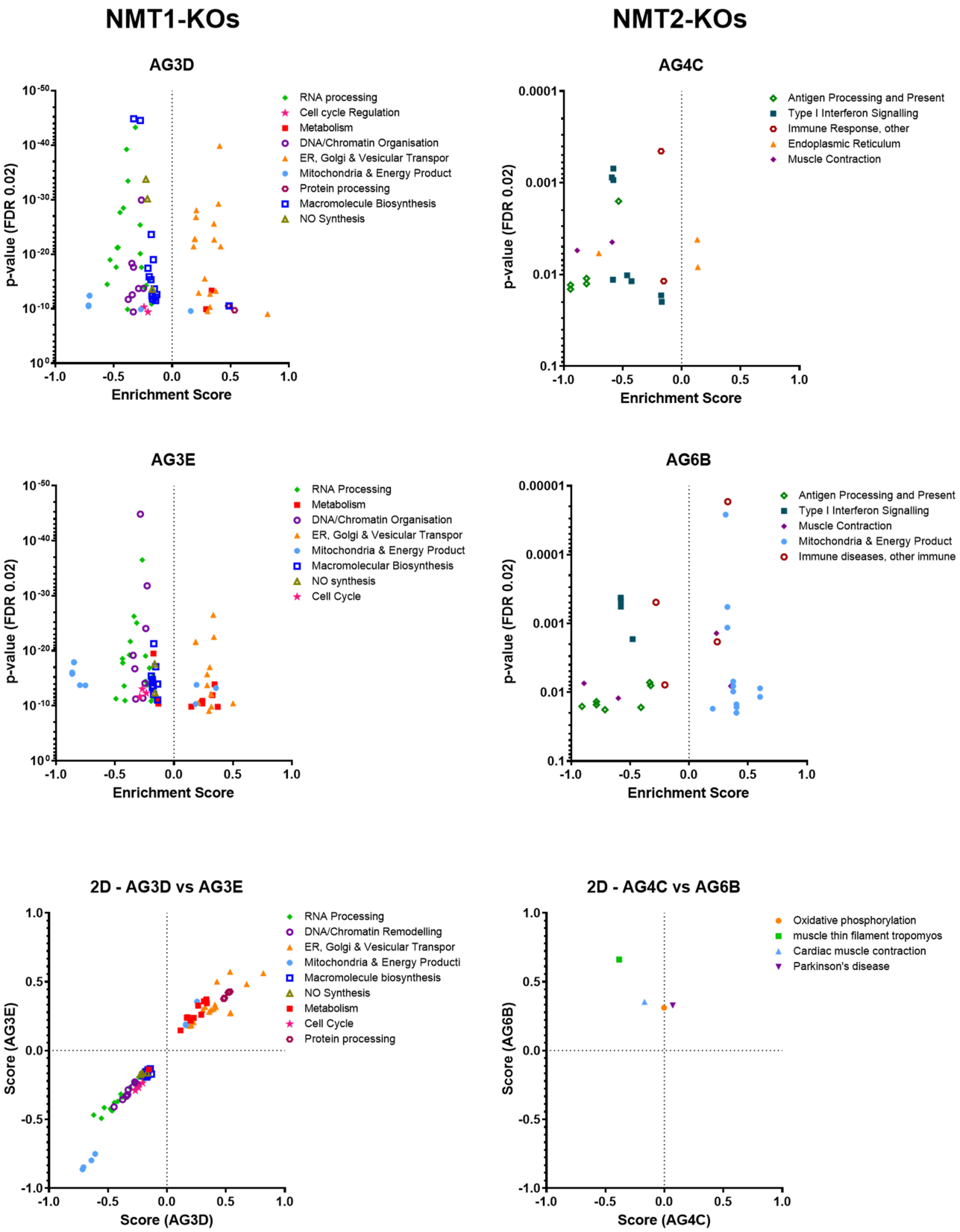


Figure A.4: 1D and 2D annotation enrichment analysis of individual Panc-1 clones. Protein groups were annotated with KEGG, GOBP, GOMF and GOCC terms before annotation enrichment analysis (FDR<0.02). For NMT1-KO clones (AG3D and AG3E), only terms showing a p-value$<10^{-9}$ are shown for ease of visualisation. Annotation enrichment analysis was performed in Perseus (version 1.6.0.2).

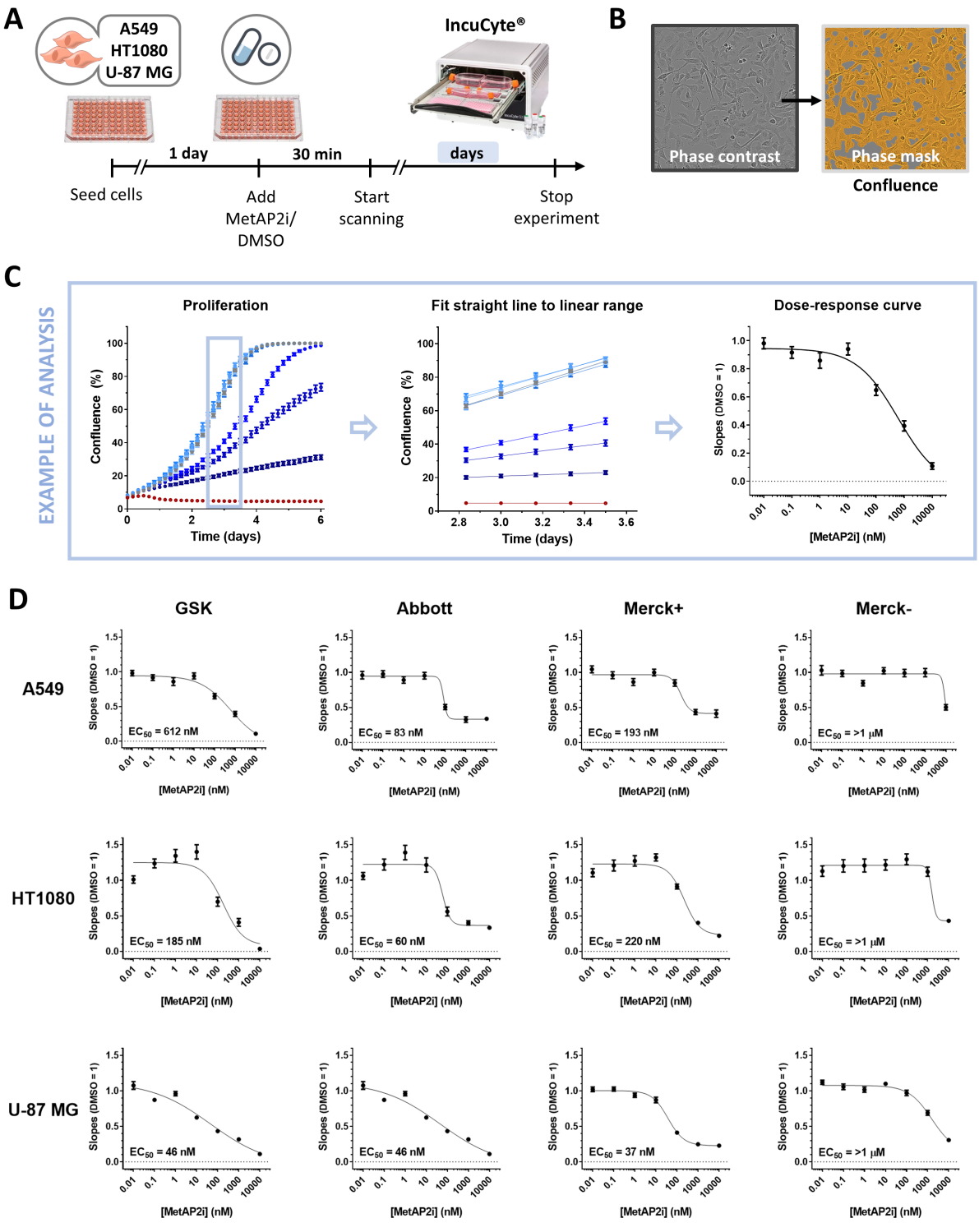


Figure A.5: Real-time proliferation assay in A549, HT1080 and U-87 MG cells. **(A)** Schematic representation of the experimental procedures for this real-time proliferation and cytotoxicity assay using the Incucyte. **(B)** Representative image taken with the Incucyte, showing the image taken by the phase-contrast microscope and the corresponding 'Phase mask' built with the integrated image-recognition software for this specific experiment that allows for conversion of image data into quantitative data. **(C)** Representation of the analysis performed on the proliferation data for each inhibitor. A common linear range for all inhibitors was manually selected for each cell line based on their growth-curve shape (A549 2.8-3.5 days; HT1080 1.5-2.5 days; U-87 MG 4.5-6 days) and fitted to a straight line. The slopes were then plotted against the $\log_{10}$(MetAP2i concentration) for each inhibitor and a 4-parameter curve fitted to determine the EC_{50} value. **(D)** Dose-response curves for each MetAP2 inhibitor in each of the three cell lines and their respective EC_{50}s.

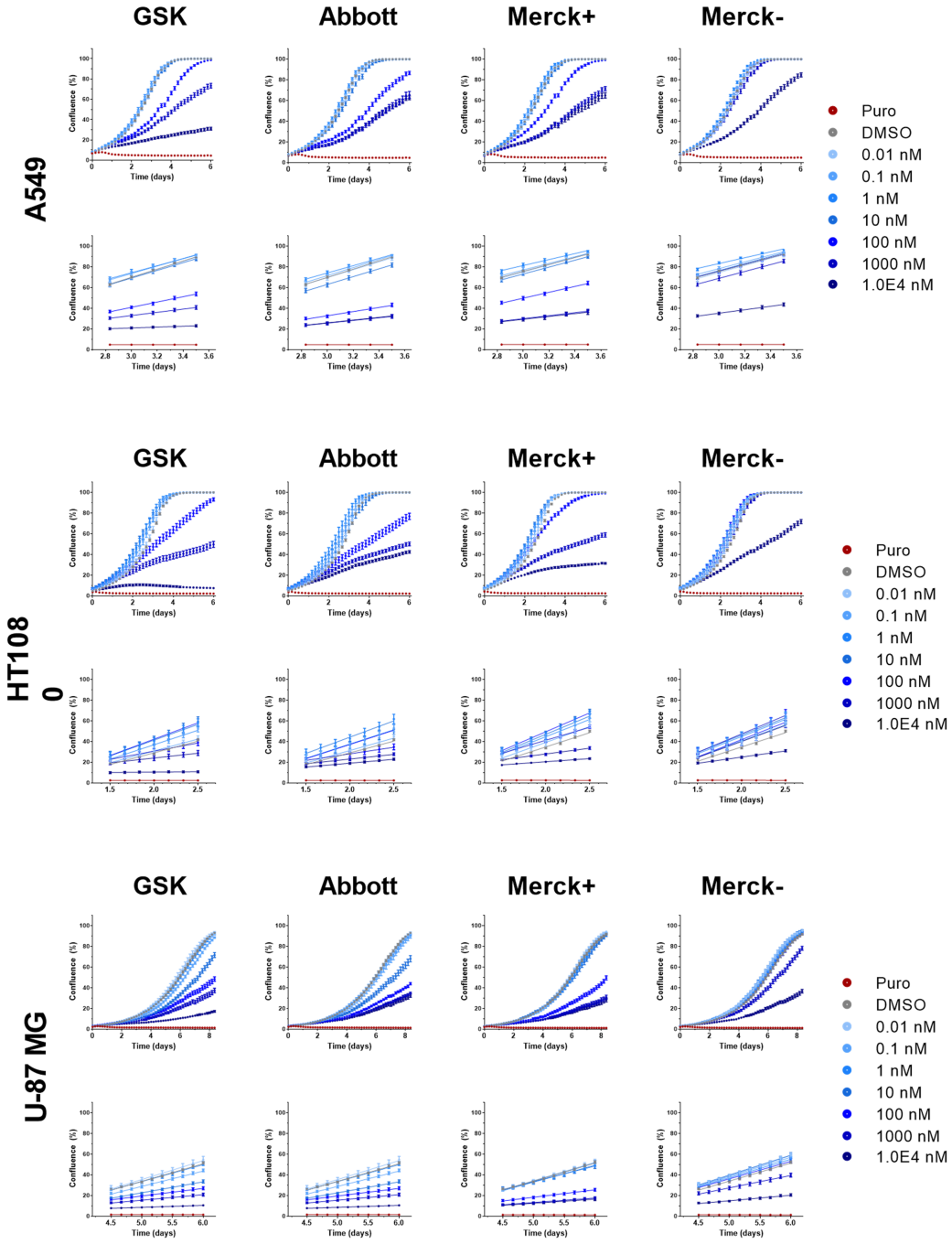


Figure A.6: Individual proliferation plots, linear ranges and dose-response curves for MetAP2i tested in A549, HT1080 and U-87 MG cells.

This data shows individual proliferation plots and selected linear ranges for all the dose-response curves shown in Figure **??**D. Confluence, represented by the percentage of phase-contrast area covered by cells was plotted against elapsed time. Linear ranges (A549 2.8-3.5 days; HT1080 1.5-2.5 days; U-87 MG 4.5-6 days) were fitted to a straight line and the slopes calculated. Slopes were then used to plot dose-response curves and calculate the EC_{50}s (shown in Figure **??**D).

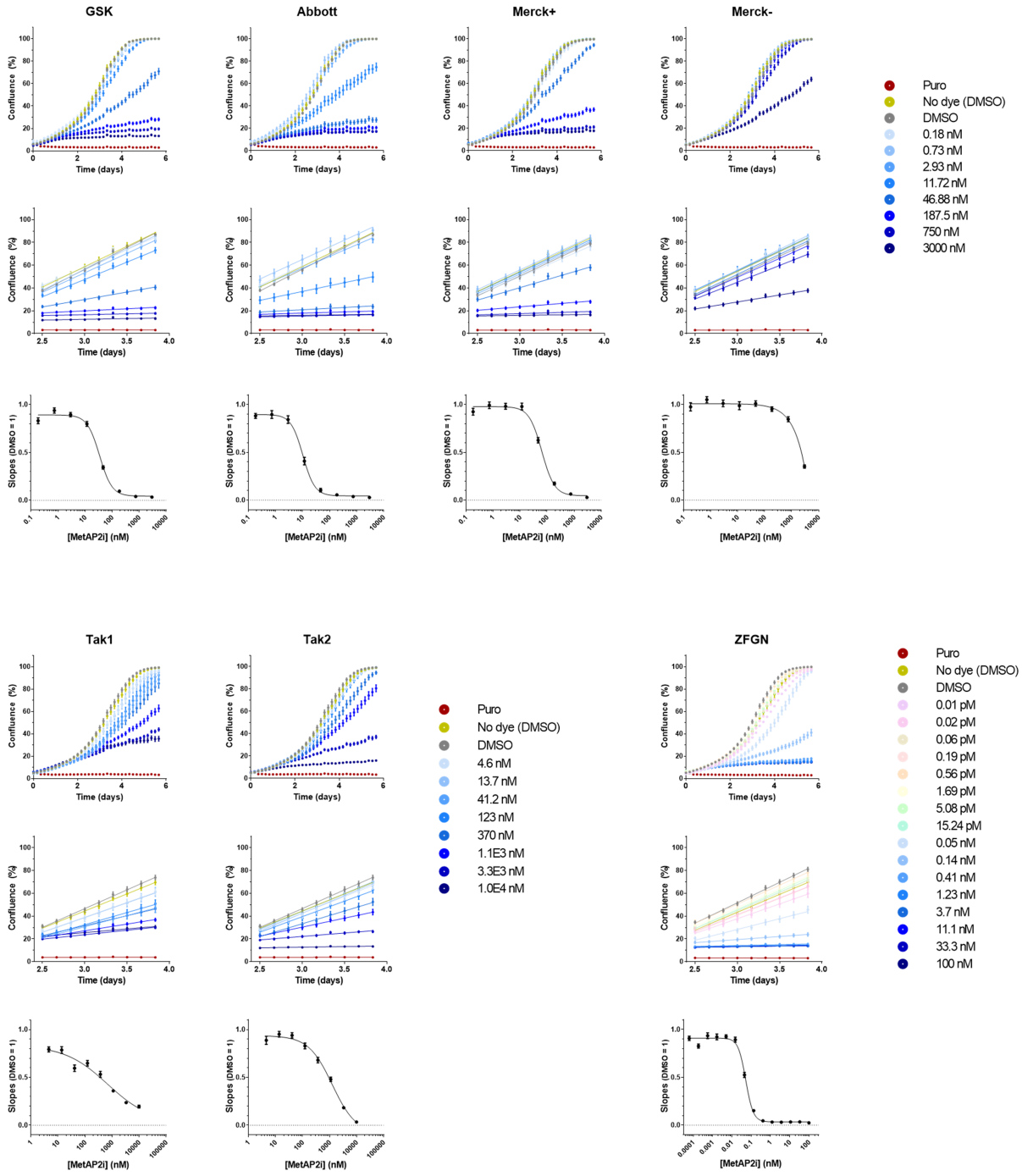


Figure A.7: Individual proliferation plots, linear ranges and dose-response curves for all MetAP2 inhibitors tested in HT1080 cells. This data shows individual plots representing the step by step analysis performed to obtain all the dose-response curves shown in Figure 4.4E. Confluence, represented by the percentage of phase-contrast area covered by cells was plotted against elapsed time. Linear ranges (2.5-4 days) were fitted to a straight line and the slopes calculated. Slopes were then plotted against the $\log_{10}$ MetAP2i concentration and fitted to a 4-parameter curve.

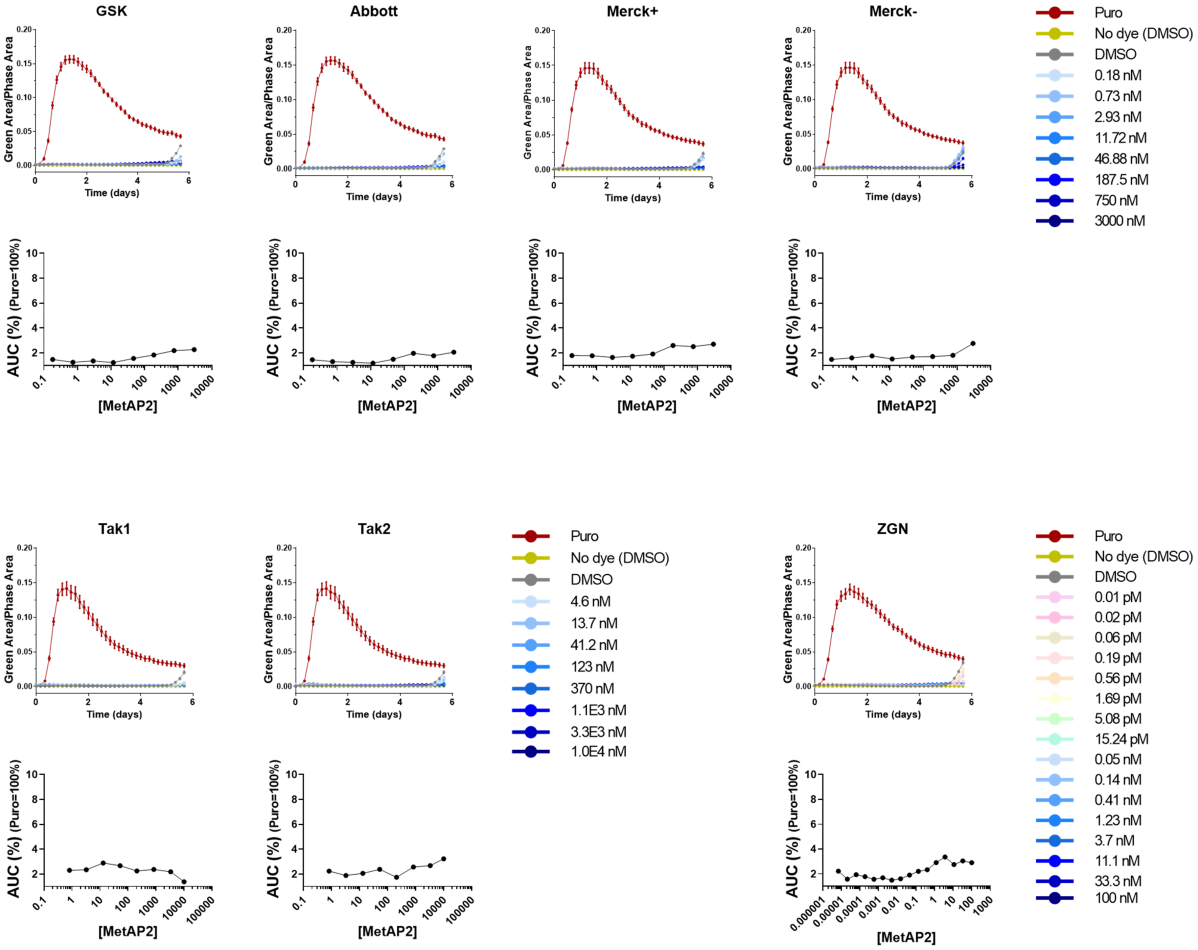


Figure A.8: Individual cytotoxicity plots for all MetAP2 inhibitors tested in HT1080 cells. This data shows the step by step analysis results for the data shown in Figure 4.4F. Cytotoxicity (Green Area/Phase Area) was plotted against elapsed time. Area under the curve (AUC) was calculated for each compound concentration, normalised against positive control (puromycin=100%) and plotted against the $\log_{10}$ concentration of MetAP2i. AUC was calculated within the interval of 0-5 days to avoid cytotoxicity from over-confluence in control and lower MetAP2i concentrations.

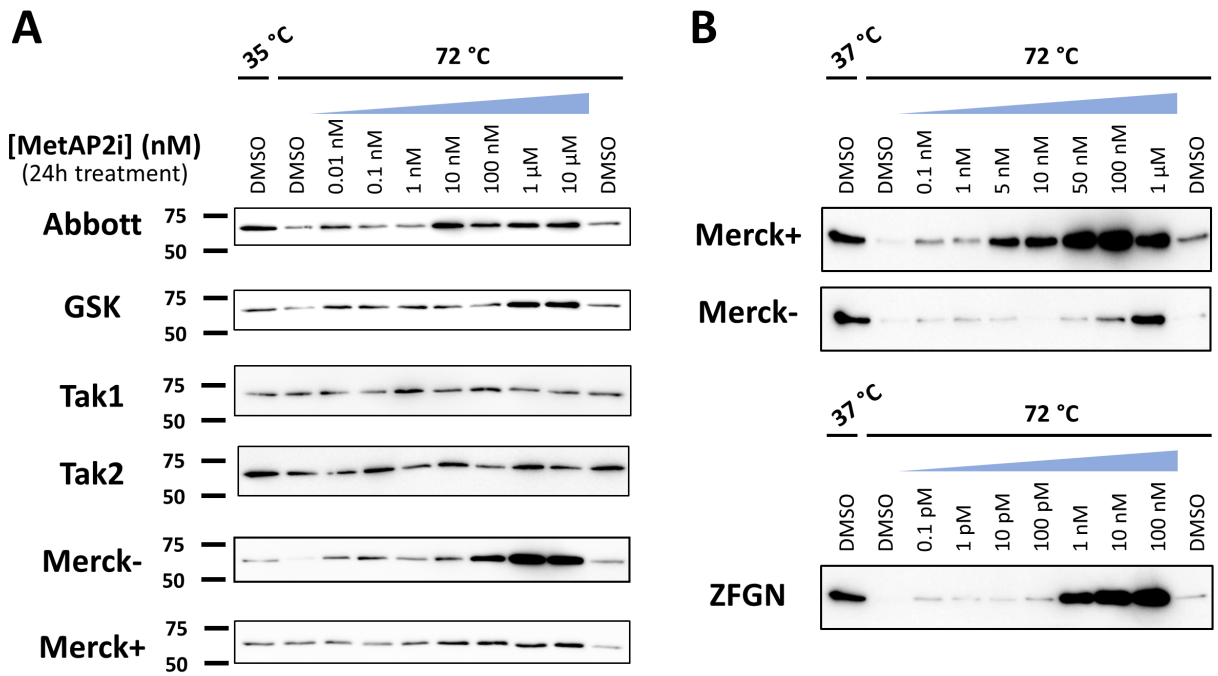


Figure A.9: CETSA-WB Supplementary Figures. (A) Technical replicate of results shown in Figure 4.5D. **(B)** Refined concentration range for Merck+/Merck- enantiomer pair and addition of ZGN, performed by Cassandra Kennedy

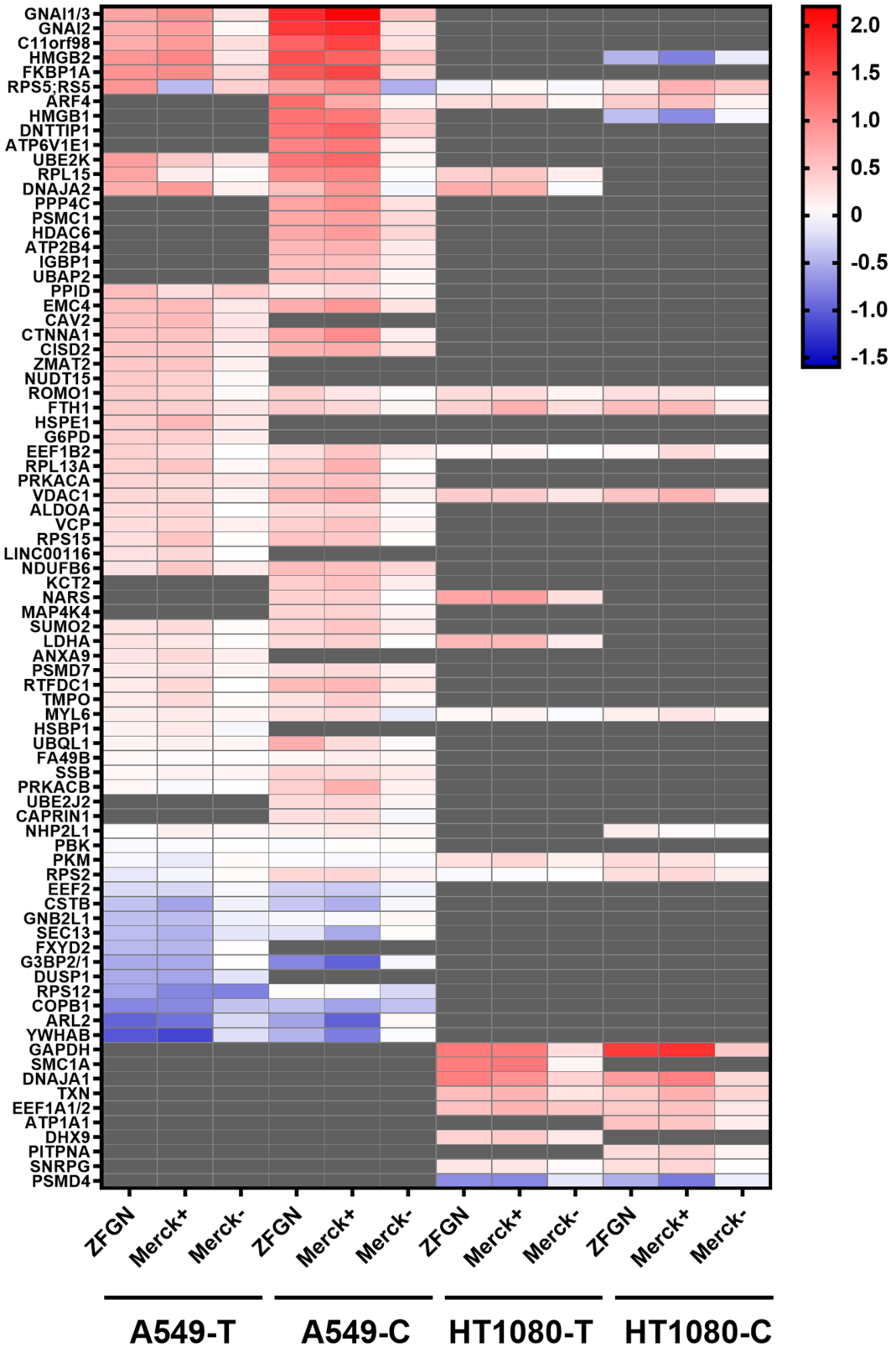


Figure A.10: Heatmap representing the same data as in Figure 4.14 but showing the full overlap across experiments. Heatmap displaying the levels of enrichment of all N-terminal peptides that showed significant enrichment in one experiment (candidate MetAP2 candidate substrates) across the rest of experiments, independently of their statistical significance status in the rest of experiments. Overlap in identification was strong across cell lines but not shared between proteases.

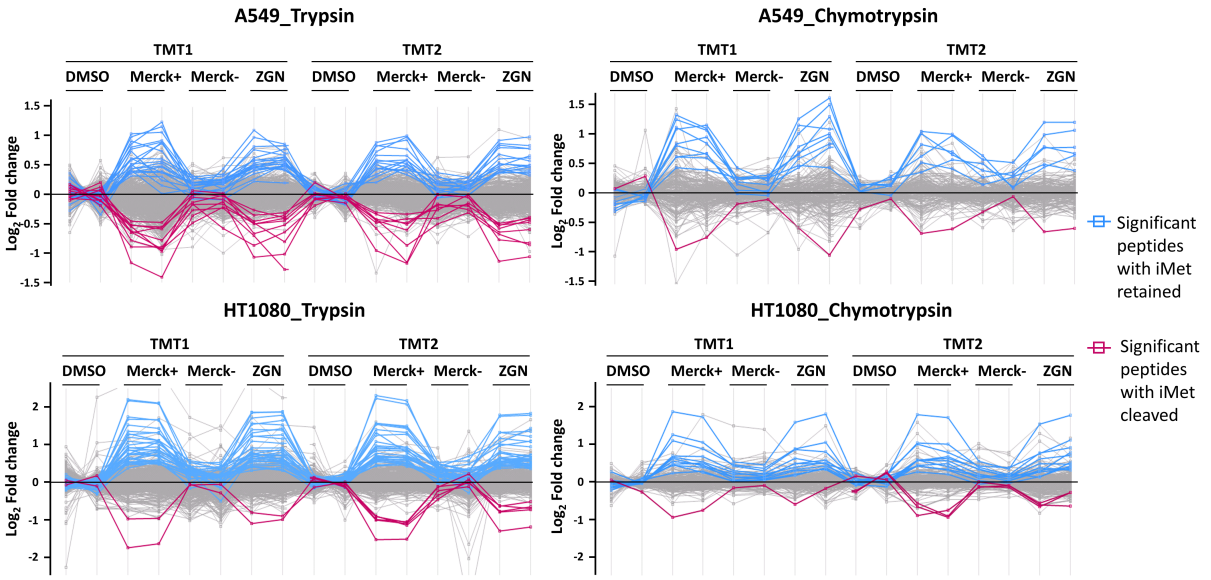


Figure A.11: Profile plots of AHA peptide pull down experiment in HT1080 and A549 cells using trypsin and chymotrypsin. These data correspond to the non-processed data shown in Figure 4.14. Each line in the profile plot represents one N-terminal peptide. Highlighted in blue are N-terminal peptides showing iMet retained and a significant increase in enrichment upon MetAP2i. Highlighted in magenta are N-terminal peptides showing iMet cleaved and a significant decrease in enrichment upon MetAP2i.

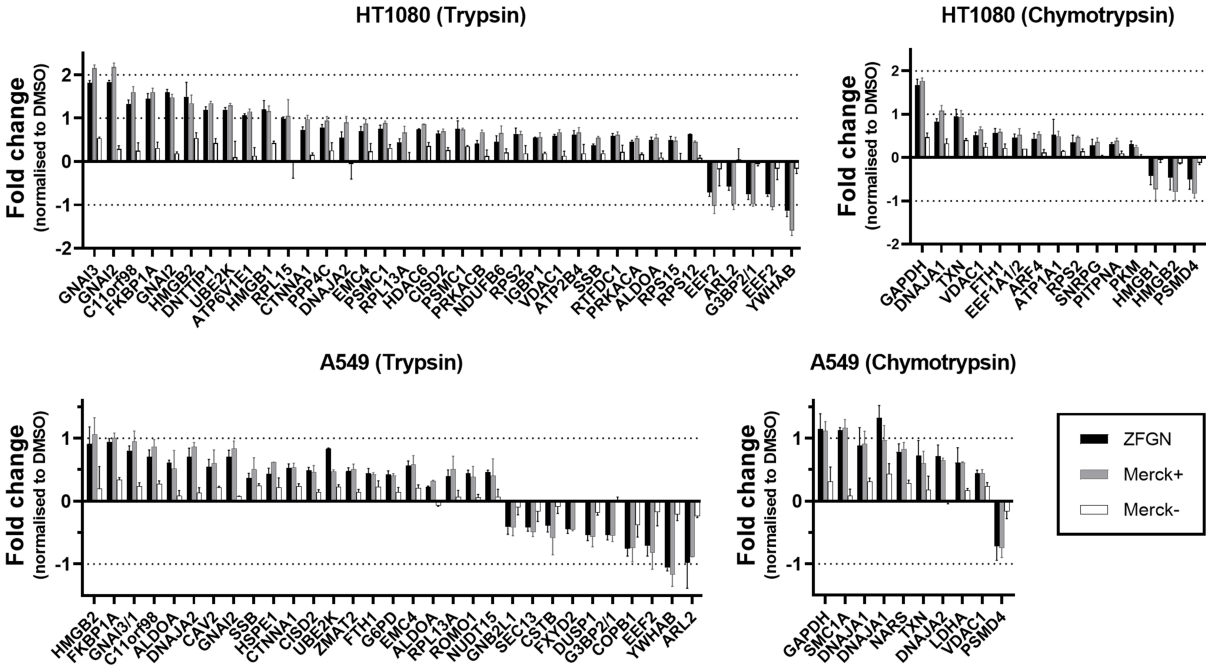


Figure A.12: Fold change enrichment of candidate MetAP2 substrates identified in HT1080 and A549. Barplots showing the fold change enrichments of significant N-terminal peptides represented in the heatmap of Figure 4.14D.

Table A.1: Full list of candidate MetAP2 substrates identified across all the experiments listed in Chapter 4. "+" indicates that N-terminal peptide was found significantly changing its enrichment levels in that experiment. The N-terminal peptide sequence identified is also provided. N-terminal peptides corresponding to known NMT substrates are annotated as "Co-Myr".

Gene names	Myr	Exp 61	Exp 68C	Exp 68T	Exp 69C	Exp 69T	Leading Prot ID	Sequence(s) identified
ALDOA	-	+		+		+	P04075	MPYQYPALTPEQK; MPYQYPALTPEQKK; MPYQYPALTPEQKKELSDIAHRIVAPGK
ANXA9	-	+					O76027	MSVTGGK
ARF4	Co-Myr				+		P18085	MGLTISSLF
ARL2	Co-Myr			+		+	P36404-2	GLLTILKKMK
ATP1A1	-				+		P05023	MGKGVGRDKY
ATP2B4	-					+	P23634	MTNPSDR
ATP6V1E1	-	+				+	P36543-2	MALSDADVQK
C11orf98	-			+		+	E9PRG8	MGAPGGK
CAPRIN1	-	+					Q14444-2	MPSATSHSGSGSK
CAV2	-			+			P51636-3	MGLETEK
CISD2	-	+		+		+	Q8N5K1	MVLESVAR
COPB1	-	+		+			P53618	TAAENVCYTLINVPMDSEPPSEISLK
CSTB	-	+		+			P04080	MCGAPSATQPATAETQHIADQVR
CTNNA1	-	+		+		+	P35221	MTAVHAGNINFK
DNAJA1	-		+		+		P31689-2	MVKETTY;MVKETTYY
DNAJA2	-	+	+	+		+	O60884	MANVADTK;MANVADTKLY
DNTTIP1	-	+				+	Q9H147	MGATGDAEQPR
DUSP1	-			+			P28562	VMEVGTLDAGGLR
EEF1A1; EEF1A1P5; EEF1A2	-				+		P68104; Q5VTE0; Q05639	MGKEKTHINIVVIGHVDSGKSTTTGHLIY
EEF1B2	-	+					A0A024R3W7	MGFGDLK
EEF2	-			+		+	P13639	MVNFTVDQIR; VNFTVDQIRAIMDKK; VNFTVDQIRAIMDKKANIR
EMC4	-	+		+		+	Q5J8M3-3	MTAQGGLVANR
FKBP1A	-	+		+		+	P62942	MGVQVETISPGDGR
FTH1	-	+		+	+		P02794	MTTASTSQVR;MTTASTSQVRQNY
FXYD2	-			+			P54710	TGLSMDGGGSPK
G3BP2; G3BP1	-			+		+	Q9UN86; Q13283	VMEKPSPLLVGR
G6PD	-			+			P11413	MAEQVALSR
GAPDH	-		+		+		P04406	MGKVKVGVNGF
GNAI1; GNAI3	Co-Myr	+		+		+	P63096; P08754	MGCTLSAEDK; MGCTLSAEDKAAVER
GNAI2	Co-Myr	+		+		+	P04899	MGCTVSAEDK; MGCTVSAEDKAAAER
GNAS	-	+					P63092-3	MGCLGNSK
GNB2L1	-			+			P63244	TEQMTLR
HDAC6	-	+				+	Q9UBN7	MTSTGQDSTTTR
HMGB1; HMGB1P1	-				+	+	P09429; B2RPK0	MGKGDPK; GKGDPKKPRGKMSSY
HMGB2	-			+	+	+	P26583	MGKGDPNKPR; GKGDPNKPRGKMSSY
HSBP1	-	+					O75506	MAETDPK
HSPE1	-			+			P61604	MAGQAFR
IGBP1	-					+	P78318	MAAEDELQLPR
KCT2	-	+					Q8NC54	MAAAVPK
LDHA	-		+				P00338	MATLKDQLIY
LINC00116	-	+					Q8NCU8	MADVSER
MAP4K4	-	+					O95819-4	MANDSPAK
MYL6	-	+					P60660-2	MCDFTEDQTAEFK; MCDFTEDQTAEFKEAFQLFDR
NARS	-		+				O43776	MVLAELY; MVLAELYVSDREGSDATGDGTK
NDUFB6	-					+	O95139	MTGYTPDEK
NUDT15	-	+		+			Q9NV35	MTASAQPR
PITPNA	-				+		Q00169	MVLLKEY
PKM	-				+		P14618	MSKPHSEAGTAF
PPID (CYPD)	-	+					Q08752	MSHPSPQAKPSNPSNPR
PPP4C	-					+	P60510	MAEISDLDR
PRKACA	Co-Myr	+				+	P17612	MGNAAAAK
PRKACB	Co-Myr					+	P22694	MGNAATAK
PSMC1	Co-Myr	+				+	P62191	MGQSQSGGHGPGGGK; MGQSQSGGHGPGGGKK
PSMD4	-		+		+		P55036	VLESTMVCVDNSEY
ROMO1	-	+		+			P60602	MPVAVGPYGQSQPSCFDR
RPL13A	-			+		+	P40429	MAEVQVLVLDGR
RPL15	-					+	P61313	MGAYKYIQELWRKK
RPS12	-					+	P25398	MAEEGIAAGGVMDVNTALQEVLK
RPS15	-					+	P62841	MAEVEQK
RPS2	-				+	+	P15880	MADDAGAAGGPGGPGGPGMGNR; MADDAGAAGGPGGPGGPGMGNRGGF
RTFDC1	-	+				+	Q9BY42	MGCDGGTIPK
SEC13	-	+		+			P55735	VSVINTVDTSHEDMIHDAQMDYYGTR
SEC22B	-	+					O75396	VLLTMIAR
SELK	-	+					Q9Y6D0	MVYISNGQVLDSR
SMC1A	-		+				Q14683	MGFLKLIEIENF
SNRPG; SNRPGP15	-				+		P62308; A8MWD9	MSKAHPPELKKF
SSB	-	+				+	P05455	MAENGDNEK
SUMO2	-	+					A0A024R8S3	MADEKPK
TXN	-		+		+		P10599	MVKQIESKTAF
UBAP2	-					+	Q5T6F2	MTSVSSDHCR
UBE2J2	-	+					Q8N2K1	MSSTSSK
UBE2K	-	+		+		+	P61086-2	MANIAVQR
UBQLN2	-	+					Q9UHD9	MAENGESSGPPRPSR
VCP	-	+					P55072	MASGADSK
VDAC1	-		+		+	+	P21796	MAVPPTYADLGK; MAVPPTYADLGKSARDVF
YWHAB	-	+		+		+	P31946	TMDKSELVQK
ZC3H7A	-	+					Q8IWR0	MSNVSEER
ZMAT2	-	+		+			Q96NC0	MASGSGTK

Table A.2: Protein Class classification for all 78 MetAP2 substrate candidates. The initial classification was performed using PANTHER [394]. Non classified proteins (marked with as asterisk, *) were then manually annotated based on the function listed in Uniprot [70] and their most prevalent GO Molecular Function annotations in QuickGO [396].

Gene Name	Prot ID	Protein Class	
ANXA9	O76027	Other	
DNAJA1	P31689-2	Chaperone	*
DNAJA2	O60884	Chaperone	*
FKBP1A	P62942	Chaperone	
HSPE1	P61604	Chaperone	
IGBP1	P78318	Chaperone	
PPID	Q08752	Chaperone	*
ATP6V1E1	P36543-2	chromatin-binding or regulatory protein	
HDAC6	Q9UBN7	chromatin-binding or regulatory protein	*
HMGB1	P09429	chromatin-binding or regulatory protein	*
HMGB2	P26583	chromatin-binding or regulatory protein	*
RTFDC1	Q9BY42	chromatin-binding or regulatory protein	
SMC1A	Q14683	chromatin-binding or regulatory protein	*
CTNNA1	P35221	Cytoskeletal protein	
MYL6	P60660-2	Cytoskeletal protein	
DNTTIP1	Q9H147	Gene-specific transcriptional regulator	
HSBP1	O75506	Gene-specific transcriptional regulator	*
COPB1	P53618	membrane traffic protein	
SEC22B	O75396	membrane traffic protein	
ALDOA	P04075	Metabolite interconversion enzyme	
G6PD	P11413	Metabolite interconversion enzyme	
GAPDH	P04406	Metabolite interconversion enzyme	
LDHA	P00338	Metabolite interconversion enzyme	
PKM	P14618	Metabolite interconversion enzyme	*
C11orf98	E9PRG8	Not known	
KCT2	Q8NC54	Not known	
NUDT15	Q9NV35	Nucleotide metabolism	*
CAPRIN1	Q14444-2	Nulceic acid-binding protein	
G3BP2	Q9UN86-2	Nulceic acid-binding protein	
SNRPG	P62308	Nulceic acid-binding protein	
SSB	P05455	Nulceic acid-binding protein	
ZC3H7A	Q8IWR0	Nulceic acid-binding protein	
ZMAT2	Q96NC0	Nulceic acid-binding protein	
LINC00116/MTLN	Q8NCU8	Other	
DUSP1	P28562	Phosphorylation signalling	
MAP4K4	O95819-4	Phosphorylation signalling	*
PPP4C	P60510	Phosphorylation signalling	
PRKACA	P17612	Phosphorylation signalling	
PRKACB	P22694	Phosphorylation signalling	
ARF4	P18085	Protein-binding, activity modulator	
ARL2	P36404-2	Protein-binding, activity modulator	
CISD2	Q8N5K1	Protein-binding, activity modulator	*
CSTB	P04080	Protein-binding, activity modulator	
EMC4	Q5J8M3-3	Protein-binding, activity modulator	*
GNAI1	P63096	Protein-binding, activity modulator	
GNAI2	P04899	Protein-binding, activity modulator	
GNAS	P63092-3	Protein-binding, activity modulator	*
PSMC1	P62191	Protein-modifying enzyme	
PSMD4	P55036	Protein-modifying enzyme	
SUMO2	A0A024R8S3	Protein-modifying enzyme	
UBE2J2	Q8N2K1	Protein-modifying enzyme	
UBE2K	P61086-2	Protein-modifying enzyme	
UBQLN2	Q9UHD9	Protein-modifying enzyme	
ROMO1	P60602	Redox regulatory protein	*
NDUFB6	O95139	Redox regulatory protein	*
TXN	P10599	Redox regulatory protein	*
CAV2	P51636-3	scaffold/adaptor protein	
UBAP2	Q5T6F2	scaffold/adaptor protein	*
YWHAB	P31946	scaffold/adaptor protein	
FTH1	P02794	Other	
EEF1A1	P68104	translational protein	
EEF1B2	A0A024R3W7	translational protein	*
EEF2	P13639	translational protein	
GNB2L1/RACK1	P63244	translational protein	*
NARS	O43776	translational protein	*
RPL13A	P40429	translational protein	
RPL15	P61313	translational protein	
RPS12	P25398	translational protein	
RPS15	P62841	translational protein	
RPS2	P15880	translational protein	
ATP1A1	P05023	transporter	
ATP2B4	P23634	transporter	
FXYD2	P54710	transporter	*
PITPNA	Q00169	transporter	*
SEC13	P55735	transporter	
SELK	Q9Y6D0	transporter	*
VCP	P55072	transporter	*
VDAC1	P21796	transporter	

* Manually annotated

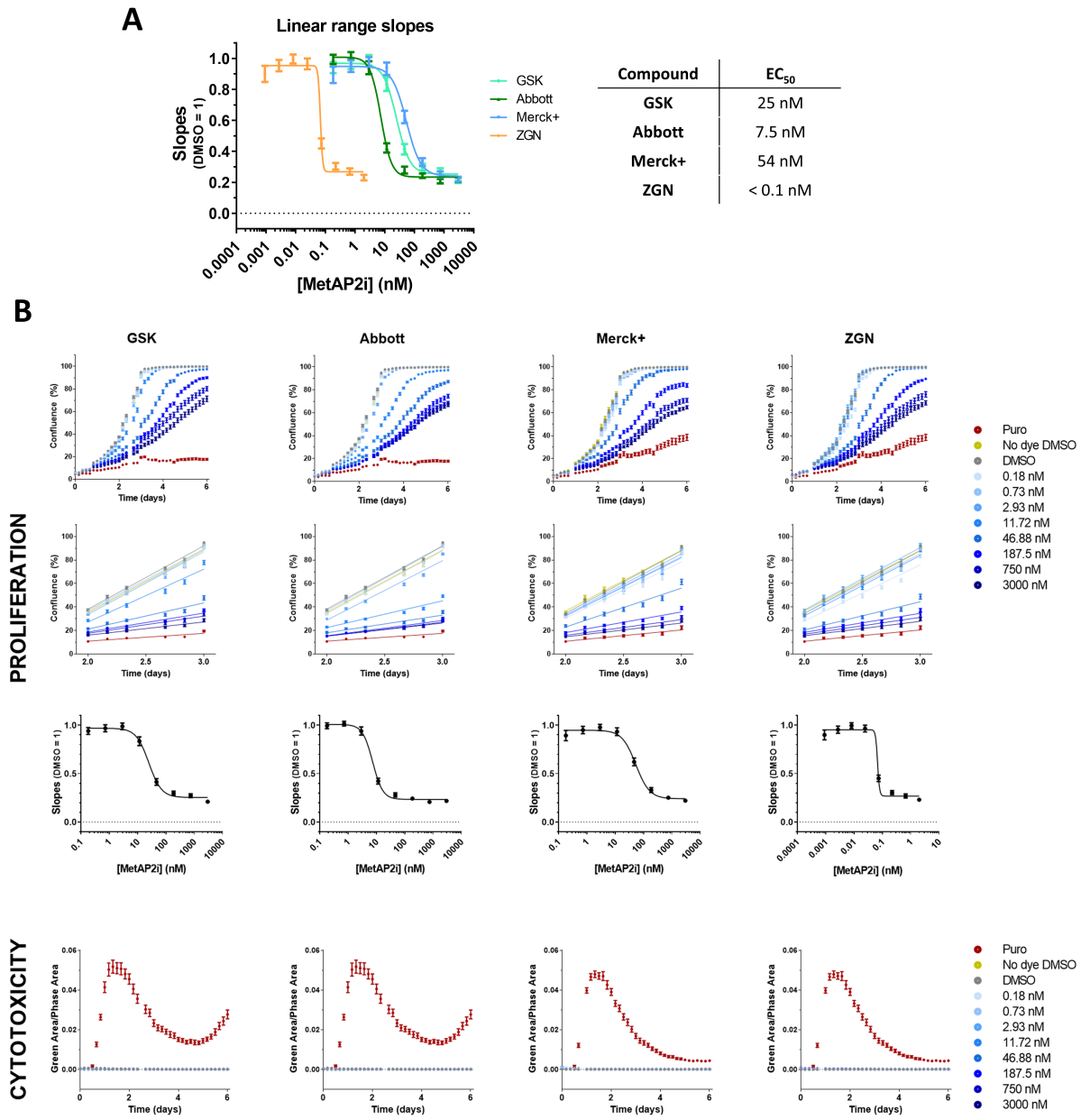


Figure A.13: Dose-response curves and EC_{50}s of proliferation inhibition in RPE1 cells. (A) Dose-response curves and EC_{50}s of GSK, Abbott, Merck+ and ZGN MetAP2i in RPE1 cells. **(B)** Proliferation plots, linear ranges and individual dose-response curves broken down for each inhibitor in RPE1 cells. These cells show resistance to puromycin since it was used for establishing this cell line.