Structural and large-scale analysis unveil the intertwined paths promoting NMT-catalyzed lysine and glycine myristoylation

Frédéric Rivière, Cyril Dian, Rémi F. Dutheil, Paul Monassa, Carmela Giglione, Thierry Meinnel

Structural and large-scale analysis unveil the intertwined paths promoting NMT-catalyzed lysine and glycine myristoylation

Frédéric Rivière[iD,a,1], Cyril Dian[iD,a], Rémi F. Dutheil[iD,a], Paul Monassa[iD,a], Carmela Giglione[iD,@,a,*], Thierry Meinnel[iD,@,a,*]

[a] Université Paris Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette cedex, France

Running title: NMT-mediated lysine myristoylation rules

***Correspondence:** carmela.giglione@i2bc.paris-saclay.fr (C. Giglione) or thierry.meinnel@i2bc.paris-saclay.fr (T. Meinnel)

[iD] ORCID 0000-0001-6366-1904 (F. Rivière); ORCID iD 0002-6349-3901 (C. Dian); 0002-2910-6648 (R.F. Dutheil); ORCID iD 0000-0003-4713-8069 (P. Monassa); 0002-7475-1558 (C. Giglione); 0001-5642-8637 (T. Meinnel)

[@]twitter @giglionelab (C. Giglione); @meinnel (T. Meinnel)

Present address

[1] Karolinska Institutet, Department of Biosciences and Nutrition, 14183 Huddinge, Sweden

Keywords: Acylation; lysine; myristoylation; *N*-myristoyltransferase; N-terminal modification

ABSTRACT

N-myristoyltransferases (NMTs) catalyze protein myristoylation, a lipid modification crucial for cell survival and a range of pathophysiological processes. Originally thought to modify only N-terminal glycine α-amino groups (G-myristoylation), NMTs were recently shown to also modify lysine ε-amino groups (K-myristoylation). However, the clues ruling NMT-dependent K-myristoylation and the full range of targets are currently unknown. Here we combine mass spectrometry, kinetic studies, *in silico* analysis, and crystallography to identify the specific features driving each modification. We show that direct interactions between the substrate's reactive amino group and the NMT catalytic base promote K-myristoylation but with poor efficiency compared to G-myristoylation, which instead uses a water-mediated interaction. We provide evidence of depletion of proteins with NMT-dependent K-myristoylation motifs in humans, suggesting evolutionary pressure to prevent this modification in favor of G-myristoylation. In turn, we reveal that K-myristoylation may only result from post-translational events. Our studies finally unravel the respective paths towards K-myristoylation or G-myristoylation, which rely on a very subtle tradeoff embracing the chemical landscape around the reactive group.

INTRODUCTION

Myristoylation is an essential lipidation that adds a C:14:0 fatty acid to proteins in all eukaryotes including pathogens [1-5]. The lipid moiety anchors soluble proteins to membranes, where they interact with partners to initiate signal transduction [6-8]. N-myristoyltransferases (NMTs; glycylpeptide N-tetradecanoyltransferases) are the only enzyme class known to catalyze myristoylation in eukaryotes and, given their central role in pathobiology, are promising therapeutic targets [9-13]. NMTs were long thought to exclusively target proteins with an N-terminal glycine (G-myristoylation) usually arising from co-translational methionine excision [1, 14] or, less frequently, post-translational cleavage exposing new N-terminal glycines [15, 16].

Myristoylation, as with other lipid acylations, is notoriously difficult to systematically map *in vivo* due to its hydrophobicity, poor antigenicity, low abundance, and subcellular localization [1]. Tools have been difficult to develop for these types of modifications, hampering progress in the field for many years. Considering that NMT activity and specificity are mainly influenced by the N-termini of substrate proteins, a combination of structural, proteomic, and genome-wide approaches using short peptides has delivered an exhaustive list of substrates undergoing co-translational G-myristoylation. The so-called G-myristoylated proteome, or myristoylome [5], covers approximately 2% of the human proteome (~600 proteoforms [17, 18]). Additionally, high-resolution co-crystallography of human NMT1 with a number of reactive substrates displaying an N-terminal Gly [19] has established a specific water channel in NMT that provides a water-mediated bond (Wat2) between the N-terminal group of the G-starting substrate and the C-terminal carboxy group (Q496) of NMT, which acts as the catalytic base for deprotonation and further reactivity [1, 19]. These data have been instrumental in

revealing that, surprisingly, NMTs further support myristoylation of lysine side chains (K-myristoylation), suggesting that the range of known substrates could be even larger [19, 20]. Of the two mammalian NMTs, NMT1 and NMT2, NMT1 appears to be the main driver of K-myristoylation, as its knockout (but not that of NMT2) inhibits the modification [20]. The mammalian small G-protein ARF6, which features a K at position 3 following an N-terminal G at position 2 unmasked after the cleavage of the first M (position 1), is the only protein currently known to undergo both N-terminal G- and K-myristoylation [20]. These modifications allow ARF6 to attach to plasma membranes when GTP is bound and permit K-myristoylation erasing by sirtuin deacylase 2 upon GTP hydrolysis. In this cycle, NMT ensures the novel ARF6 K-myristoylation after GTP refueling via an exclusive post-translational mechanism.

NMT-dependent G-myristoylation occurs on the α-amino group of the target protein via an amide bond, creating an extremely stable attachment *in vivo* and resistance to chemical cleavage *in vitro* [21-23]. Accordingly, the G-myristoylation tag is thought to be unmovable and permanent. However, the discovery of IpaJ, a pathogenic cysteine-dependent protease produced by several disease-causing bacteria including *Shigella flexneri*, challenged this four decade-old dogma [24]. In reality, IpaJ does not cleave the amide bond between myristate and G but rather cleaves the first peptide bond beyond the N-terminal myristoylated glycine from several proteins, leading to irreversible protein de-myristoylation. IpaJ relies on the presence of a G-myristoylated group embedded within a small dipeptide as a minimal chassis [25], with strong specificity for C:14 chains over longer (C:16) or shorter (C:10) fatty acyl chains [25]. Since this study predated the discovery of NMT-driven K-myristoylation, only a few IpaJ substrates have been characterized [25, 26]. Critically, it is unknown whether IpaJ contributes to the NMT-dependent K-myristoylated proteome during *Shigella* infection. Major challenges

in identifying and confirming the NMT-dependent K-myristoylated proteome include our lack of knowledge about its extent, the clues ruling NMT-dependent K-myristoylation, and its interplay with G-myristoylation. Whether there exist K-myristoylation targets not previously G-myristoylated (unlike ARF6) is unknown.

Here we exploited systematic structural and extensive biochemical analyses and a novel pipeline involving the *Shigella* IpaJ protease to discriminate K- and G-myristoylation and identify the specific features driving each modification. We provide evidence that direct interactions between the substrate's reactive amino group and the NMT catalytic base significantly slow K-myristoylation catalysis compared with G-myristoylation. Our studies also reveal that NMT-dependent K-myristoylation can only involve lysines occurring at positions 1, 2, or 3, likely arising from post-translational proteolytic cleavage. Our iterative analyses unravel the respective paths towards K-myristoylation or G-myristoylation, which rely on a very subtle tradeoff embracing the chemical landscape around the reactive group.

RESULTS

Nomenclature

In this study, amino acids are indicated using the three letter code (i.e.., Lys for K, Met for M, and Gly for G) when referring to a single residue. In peptide sequences, amino acids are referred to as their single letter codes. For consistency with previous works in which the initial M of any open reading frame (ORF) is the amino acid at position 1 (Met1), we number the subsequent amino acids of all peptides, including the G of myristoylated proteins, in ascending order (e.g., Gly2/G2).

G-myristoylation predominates over K-myristoylation on N-terminal GK-starting peptides

There is now compelling evidence that myristoylation studied *in vitro* with octapeptides readily translates to the *in vivo* context [1, 18] (**Fig. S1**). For instance, G-myristoylation of a full-length protein, the small RAB G-protein, is catalyzed with similar efficiency as the N-terminal octapeptide [27]. This approach perfectly reflects the (i) co-translational and N-terminal nature of the modification taking place on short, emerging polypeptide chains [28]; and (ii) seven dedicated binding clefts in NMT (pockets 2-8 in **Fig. 1a**) accommodating peptide substrate side-chains [17]. These peculiarities allow the definition of the N-terminal G-myristoylated proteome based on the dogma that NMT-dependent myristoylation only occurs on glycines [17]. However, there is a recent report of possible NMT-dependent myristoylation on a lysine at position 3 [19, 20].

We previously reported a model peptide with a Lys3 (G2KSFSKPR; reference, GK) as an efficient NMT substrate (**Fig. 1b** and **Table 1**). In light of this new NMT activity (K-myristoylation), we assessed which type of myristoylation (G- and/or K-) *Homo sapiens* (Hs)NMT1 catalyzes on this peptide by MALDI-Tof-Tof including MS-MS (MS2). G-myristoylation was favored in the context of this model peptide, as evidenced by the identification of y7 ions characteristic of unmodified Lys3 only in MS2 spectra of the myristoylated product (**Data S1a** and **Table 1**). We next wondered if all natural substrates from human proteins starting with N-terminal GKs behave like the reference peptide. Therefore, we examined the myristoylation status of those proteins, including the six closely related members of the HPCA/NCS1 family and ARF6 (GKVLSKIF). In all cases, we only unambiguously detected G-myristoylation (**Data S1** and **Table 1**), with catalytic efficiencies in the same range as reported for the other sequences of the G-myristoylated proteome, as exemplified by the

SOS3 peptide (GCSVSKKK, **Fig. 1b**). Moreover, the catalytic parameters measured with the ARF6 octapeptide were identical to those previously obtained with the full-length GTP-bound reactive protein [29], further validating the relevance of using short peptides. Remarkably, most N-terminal GK-starting proteins harbored the S6K7 motif, which is known to facilitate G-myristoylation [30]. Indeed, it was recently shown that the salt bridges of Lys7 with the three contiguous Ds of the NMT Ab-loop promote the movement of the Ab-loop from an open to a closed conformation (SuppFig.7 in [19]), favoring catalysis through distortion of the MyrCoA thioester bond in the closed conformation and facilitating transition to the tetrahedral state.

A new crystal structure of HsNMT1 with a peptide corresponding to the N-terminus of HPCA (G2KQNSKLR) confirmed G-myristoylation as the reaction product (**Fig. 1a,c,d**), as previously also observed with NCS1 (G2KSNSKLK; [17]). The crystal structure revealed that the Lys3 side chain buried tightly in pocket 3 of NMT and that its positive charge bonded the catalytic base (**Fig. 1c**). Full occupation of pocket 3 with Lys3 slightly displaced the reactive alpha amino group of Gly2 from the MyrCoA thioester group whilst retaining the Wat2 interaction of the reactive N-terminal α-amino group required for G-myristoylation [19] (**Fig. 1c,d**).

To examine whether GK-starting proteins have the same recognition rules as those in the G[^K] context (where ^K means any amino acid other than Lys) [17], we analyzed HsNMT1's capacity to acylate N-terminal peptide variants derived from NCS1. G-myristoylation only occurred when Lys7 was changed into Gly or Ser6 into Ala, although HsNMT1's catalytic efficiency significantly decreased (**Table S1** and **Data S1**). The roles of Ser6 in substrate binding and Lys7 in the MyrCoA binding-induced conformational switch of the conserved flexible acidic Ab-loop of NMT are expected to be similar to those already reported for proteins

starting with G[^K] [19]. We next assessed whether modifying the side-chain length but retaining the basic character of the primary amino group at position 7 impacted the myristoylation type. Shortening the basic side chain of residue 7 to ornithine (Orn, reduced by one CH_2) or di-amino butyrate (Dab, reduced by two CH_2s) had little impact on G-myristoylation (**Table S1**), suggesting maintenance of the salt bridge with the Ab-loop. Further shortening to di-amino propionate (Dap, short basic chain CH_2NH_2) or removal of the basic character with homocysteine (Hcy) significantly reduced catalytic efficiency but still promoted only G-myristoylation (**Table S1** and **Data S1**).

To obtain further insights into the overall rules governing NMT-mediated G-myristoylation independent of Gly at position 2, we substituted G2 with an Ala and Lys3 with an Asn (A2NCFSKPR; AN) in the reference peptide to prevent K-myristoylation. Surprisingly, AN was α-myristoylated, as shown by both MS and crystallographic analyses (**Fig. S2a**, **Data S1** and **Table 1**). However, catalytic data indicated that the substrate was very inefficiently modified by NMT (**Fig. 1b**), and the crystal structure of AN confirmed that the active site was mainly filled with a methyl addition, improperly orientating the reactive amino group, in agreement with the very reduced catalytic efficiency. To decipher whether an Ala resulted in significant A-myristoylation in another sequence context, we substituted Gly2 with Ala in G2SSVSKKK (SOS3 series), which was not myristoylated, clearly indicating that the S6K7 motif was insufficient for reactivity in the Ala2 context. In all cases, the Gly2Ala substitution dramatically reduced acylation but still led to myristoylation of the N-terminal α group of Ala, like in G-myristoylation. We therefore wondered whether the strong hindrance to the active site promoted by the additional methyl chain of Ala was similar if this group was grafted directly onto the N-terminus starting with G. Indeed, the N-methyl-GN peptide (meGN, meG

is an isomer of A) led to very poor G-myristoylation efficiency (**Fig. 1b**, **Table 1**). Furthermore, the crystal structure of HsNMT1 in complex with meGN revealed that the methyl rotated the side chain of Asn3 around Cα3, which was orientated towards the catalytic base. Conversely, the methylated N-terminus was positioned in pocket 3 (**Fig. S2b,c**), explaining the low performance of this substrate for G-myristoylation. Therefore, any subtle modification around Gly2 protrudes aa2 into pocket 3, resulting in aa3 oriented in pocket 2 at the catalytic center [19].

Together, these data indicated that GK-starting peptides led to G-myristoylation, which relied on Gly2 and the S6K7 motif. In the Lys3 context, S6K7 was crucial for G-myristoylation, consistent with previous reports suggesting strong dependence of G-myristoylation on Lys3 and Lys7 [20, 31]. However, considering that the N-termini of transmembrane proteins T106A/B (GKSLSHLP, GKTFSQLG) were unambiguously NMT substrates (**Fig. 1b** and **Table 1**) despite the absence of Lys7, the deduced motif for G-myristoylation with Lys3 was GKXXSX (**Fig. 1e**). We noticed that both T106A/B harbored a highly hydrophobic residue at position 5 (L and F, respectively), another feature favoring G-myristoylation [17, 30], probably compensating for the absence of Lys7 and leading to efficient G-myristoylation. In addition, both ARF6 and GK, efficiently myristoylated peptides, possessed a hydrophobic residue at position 5 (**Fig. 1b**). So, consistent with its positive role, permutation of positions 4 and 5 in the GK background (GKFS) significantly reduced myristoylation efficiency (**Fig. 1b**). Consistent with the aforementioned data, the N-terminal peptide (GKRGSRSQ) of the U3 small nucleolar RNA-associated protein 25 homolog UTP25, which displayed only Ser6 out of the three major determinants for GK G-myristoylation, was an extremely poor NMT substrate (**Table 1**). This dramatic effect observed despite the presence of a positively charged residue

(Arg) at position 7 indicated that absence of Lys7 could only be properly overcome by a hydrophobic residue at position 5, like in T106A/B (motif GKX[FLYW]S[HQR].

Lys3 myristoylation of N-acetylated peptides with the S6K7 motif

In apparent contradiction to the above results, the GK-starting protein ARF6 was recently shown to accept K-myristoylation *in vivo* [20] if the Gly2 α-amino group was blocked due to previous N-acetylation (*ac*) or G-myristoylation. We therefore examined any modification around the reactive G2 amino group allowing K-myristoylation of the ARF6 and/or GK-starting peptides. As expected, *ac* of the two G-starting peptides led to Lys3 side chain myristoylation (**Fig. 2a**, **Table 2** and **Data S1**). When the Lys side chain was shortened to Orn, in the context of N-acetylated peptides, side chain myristoylation was still observed at position 3 (**Fig. 2a**). The catalytic efficiencies of the *ac* series were two orders of magnitude lower than those of G-myristoylation (**Tables 1-2**), with the catalytic constant (k_{cat}) accounting for most of the k_{cat}/K_m decrease in both cases. This suggested that the affinity provided by the peptide core was similar but that the catalysis had a limiting step.

The crystal structures of HsNMT1 in complex with acG[Orn] or an acGN non-reactive derivative were obtained and compared with the acGK variant. In both cases, there was rotation around the Cα of the aa3 side chain (Cα3), as expected from the *ac* addition. The acG group moved to pocket 3, and the aa3 side chain contacted the catalytic base (**Fig. S3a-d**). The Lys side chain of acGK bonded directly with the C-terminal catalytic base of NMT, unlike the G-myristoylation mechanism, which involves a water-mediated bond (**Fig. 1c,d**). acG[Orn] led to a weaker salt bridge with the catalytic base than acGK as the distance was increased from 3.1 to 3.6 Å (**Fig. S3e**). Nevertheless, the associated side chain myristoylation mechanism still

involved direct interaction between the catalytic base and the ammonium group of Orn, like with acGK (**Fig. S3c-f**). Further reduction to Dab dramatically decreased side chain myristoylation (**Fig. 2a**, **Data S1** and **Table 2**). Therefore, tight direct binding of the side chain reactive amino group to the catalytic base most likely explains effective but slower K-myristoylation compared with G-myristoylation. Finally, we challenged the importance of the nature of the residues of the S6K7 motif by replacing each of them with an Ala in two different peptides in the acGK context (peptides A6 and A7). In each case, myristoylation was no longer measurable compared with acGK (**Table 2**). This indicates that the SK motif is a crucial motif allowing significant K-myristoylation. Though identical to the positive role displayed for G-myristoylation, the S6K7 motif becomes mandatory for K-myristoylation because of the poor reaction efficiency.

Side-chain myristoylation at Lys3 is preferred over N-terminal α-myristoylation when the N-terminal residue is not Gly

The above data indicate that the torsion angle psi (ψ) at Cα3 tends to rotate by 180° (**Fig. 1e**) and embed the N-terminal group into pocket 3 provided any change in Gly2. If aa3 displays an amino group (e.g., Lys3), it becomes reactive and promotes K-myristoylation. This suggests that the strong negative effect of the Gly2Ala substitution might favor K-myristoylation in the presence of Lys3. To test this hypothesis, we examined Gly2Ala N-terminus versions of the neuron-specific calcium-binding protein HPCA (A-HPCA), ARF6 (A-ARF6), and other GK variants (AK), which all feature a Lys3. Crystallographic and/or MS analysis of AK and A-HPCA showed K-myristoylation (**Fig. 2b-e** and **Table 2**), while data on A-ARF6 were less clear (**Data S1**). The K-myristoylation catalytic efficiencies with both AK and A-ARF6 were of the same order of magnitude as the acetylated versions of the corresponding variants. In

contrast, the efficiency associated to A-HPCA was two orders of magnitude lower (**Table 2**). Therefore, a Lys side chain was preferred over α-amino myristoylation provided that a side chain (e.g., methyl of Ala) was grafted onto the reactive residue at the first amino acid and it was not Gly. The crystal structure of AK complexed with HsNMT1 (**Fig. 2b,c**) showed that the Lys3 side chain bonded to the C-terminal catalytic base (Gln496) and to Thr282 and the Ala2 side chain buried in pocket 3. Again, direct bonding to the catalytic base likely explains the reduced K-myristoylation catalytic efficiency compared with that of the G-myristoylation, which involves water-mediated bonding of the different GK N-termini. The lower K_m of the acG derivatives are most likely related to additional interactions of the entire Ala2 chain in pocket 3 (**Fig. 2b**).

We next interrogated the human proteome to discover possible K-myristoylation targets featuring AKXXSK immediately following a Met start residue (Met1). Only tuberin (TSC2; P49815; **Data S2a**) was identified, but no myristoylation was observed with the corresponding peptide derivative (AKPTSKDS; **Table 2**), likely due to the Pro4 residue inducing a local conformational kink disfavoring both G- and K-myristoylation by misaligning S6K7 with the reactive amino upstream group. The absence of a hydrophobic side chain at residue 5 (similar to A-HPCA *vs* A-ARF6 or AK) might also explain this phenomenon, as also noted above for G-myristoylation.

We next examined whether non-Gly N-termini other than Ala might also lead to K-myristoylation by replacing the N-terminal Gly2 of the GK peptide with Met, Ser, and Pro. K-myristoylation occurred in all cases, with catalytic efficiencies inversely proportional to side chain length (**Fig. 2a**, **Table 2** and **Data S1**). We concluded that the nature of the N-terminal residue was a major determinant guiding N-terminal K-myristoylation, provided that aa3

harbored a proximal acceptor group (**Fig. 2f**). Due to improper positioning of the N-terminal amino group, any non-G-starting peptides with the S6K7 motif may be modified by NMT but with extremely poor catalytic efficiencies several orders of magnitude lower than the corresponding Gly derivative. A Lys3 downstream of any amino acid but Gly at position 2 permitted K-myristoylation with small but significant catalytic efficiencies (**Table 2**). Higher efficiencies were possible when the peptide context was favorable, as with AK and A-ARF6 peptides (this work and [20]). Remarkably, both peptides featured a bulky, hydrophobic, favorable residue at position 5 (Phe or Leu, respectively). Finally, interrogation of proteome data libraries indicated that no protein with A2 (i.e., originating from M1 excision) was likely to undergo K3-myristoylation in the human proteome.

Taken together, our *in vitro* data reveal that K-myristoylation does not require a given side chain of the first amino acid. The data also show significantly lower catalytic efficiencies compared with G-myristoylation (**Table 2**). This suggests that K-myristoylation relies on position 5-7 residues for optimal catalysis. Therefore, the [^G]KX[FL]SK motif emerges as the determinant of K-myristoylation (where ^X means any residue but X). Any non-Gly2 side chains appear to privilege pocket 3 binding, preventing their reactivity and ensuring high specificity for Gly2 undergoing G-myristoylation.

Mining the human proteome reveals that K-myristoylation may solely arise from a post-translational event

In the human proteome, 7548 hits harbored a KXXSK motif (**Data S2a**) corresponding to 2831 unique entries with 3615 sites in 6764 proteoforms (14-15% of the proteome). 572 sequences displayed Phe or Leu at position 3, like in ARF6 or the GK peptide. Among the 20 unique

sequences with translation start sites at positions 1 or 2 (i.e., arising from cotranslational N-Met removal rules), 10 of the 15 known G-myristoylated N-termini starting with GK were found [17] plus ten additional sequences, most not comprising Phe or Leu4. As a result, this second small subset featuring proteins starting with Ala (TSC2), Phe (EVPL), Lys (ANR26, FA71D), Met (CHD8, GXLT2, SIM1/2), Arg (RAB20), and Thr (LRC59) was very unlikely to undergo K-myristoylation. This was demonstrated for TSC2, which displays the most favorable context with its N-terminal Ala, the smallest residue of the subset (**Table 2**). In the G-starting subset undergoing G-myristoylation, only ARF6 displayed the important hydrophobic position at position 4, making it the best target for dual G- and K-myristoylation. The other 3595 sequences corresponded to internal sequences starting beyond codon 3, and 285 were preceded by a Gly, suggesting large scope for post-translational G-myristoylation in this subset. Post-translational myristoylation, however, requires prior unmasking by proteolytic cleavage provided before the first XK. Indeed, a number of K-myristoylation sites among the 543 [^G]KX[FL]SK internal sequences were preceded by a loose motif recapitulating known G-myristoylated caspase-induced substrates ([1]; **Data S2b-c**).

We next mined the human proteome for caspase 3 cleavage sites (see motifs as defined at http://elm.eu.org/elms/CLV_C14_Caspase3-7 [32]; **Data S3**). This more restrictive and robust motif [DSTE][^P][^DEWHFYC]D[GSAN] described and encompassed most known caspase-induced cleavages including those generating G-myristoylation sites (**Data S3a**). In addition, caspase 3 leads to neo-protein N-termini showing not only G but also A, S, and T, which are compatible with either G- or K-myristoylation (**Figs 1e, 2f**). The best candidates for G-myristoylation are displayed in **Data S3b**, and all possible K-myristoylation sites are displayed in **Data S3c**. There were only a few putative primary K-myristoylation sites following non-G residues and none displayed the optimal motif (**Data S3d**). Further data analysis and filtering

(**Data S3e**) revealed only a couple of putative secondary K-myristoylation sites following G-starting sequences as in ARF6 (NUF2, TCOF; **Data S3f**).

Of all the known myristoylated proteins resulting from post-translational myristoylation, the only GK-starting G-myristoylated protein was RNA-binding protein FUS, which displays a non-canonical caspase 3 motif (DWFDGK). When extracting a subset with the motif including a Phe at position 2 behind GK, none of the 5 retrieved proteins - including FUS - displayed the optimal motif for K-myristoylation (**Data S3e**). A similar survey performed with the caspase 6 motif ([VLIT][EDQ][^DENQRKAPGS]D[ASGMP]) - which retrieved a couple of additional G-myristoylated post-translational proteins including the pro-apoptotic protein BID and the interleukin enhancer-binding factor 3 (ILF3) - led to the same conclusion (**Data S4**).

We conclude that K-myristoylation catalyzed by NMT most likely results from a (i) secondary myristoylation, as in the case of ARF6 and at most affecting some of the 10 G-myristoylated proteins displaying Lys3 and Ser6Lys7, and/or (ii) cleavage by proteases other than caspases. In both cases, the associated K-myristoylation mechanism is post-translational and we obtained no evidence for cotranslational K-myristoylation in the human proteome.

Myristoylation may occur on Lys4

We next wondered whether extending the amino acid sequence upstream of the reactive Lys might lead to K-myristoylation, provided that a two-amino acid spacer between the accepting K and the S6K7 motif was retained. We therefore grafted an extra N-terminal Ala or Gly to the GK peptide (A- or G-GKSFSKP; AGK or GGK) and checked whether HsNMT1 could acylate the Lys at position 4. K-myristoylation of both was assessed by either MS or crystallography

(**Table 2,e**, **Supplementary Fig 4,** and **Supplementary Figure 5a**). K-myristoylation was as efficient with these A/G-grafted peptides as with the acG peptides (**Fig. 2a** and **Table 2**). We concluded that the pocket usually hosting residue 3 could host a dipeptide moiety as large as AG, provided that a close and available Lys-reactive side chain was properly spaced from the crucial S6K7 motif. We noted that, with five bonds and a positive charge, a Lys side chain was the most bulky moiety tolerated in NMT pocket 3 for G-myristoylation. With seven bonds, the AG backbone was much bulkier than K but still promoted K-myristoylation (**Fig. 2f**). To establish whether larger chains would make K-myristoylation possible, we examined large, poorly-branched N-terminal grafts larger than GG or AG (AcGGK, AcGGGK, AcGGGGK, and AcGGGGGK with 9, 12, 15, and 18 bonds, respectively) upstream of Lys3 in the GK series. With AcGGK and AcGGGK, there was unequivocal K-myristoylation by MALDI analysis (**Data S1**). K-myristoylation catalytic efficiency progressively decreased with increasing chain length to very low values, and K-myristoylation was not detected for AcGGGGK or AcGGGGGK (**Table 2**). The crystal structure obtained with the GGK derivative showed that the GG chain was fully buried in pocket 3 and propelled the main backbone peptide chain into a more remote location at Cα3 (**Fig. S5a,b**); there was also a greater distance between the reactive K amino group and both the catalytic base and T282. The 7-bond GG or AG chain corresponded to the bulkiest chains accepted by pocket 3 for efficient K-myristoylation (**Fig. 2f)**.

We concluded that to allow K-myristoylation, the [^G]KX[FL]SK motif may be associated with a one amino acid extension or *ac* at the N-terminus, provided that an *ad hoc* proteolytic cellular machinery can produce such extremities. In addition, one dozen internal XXKXXSK K-myristoylation motifs were noted in the human proteome displaying an upstream classic caspase cleavage site (**Data S2d**). Only one of the six protein components of the signal

recognition particle complex, SRP54, displayed a putative K-myristoylation site upon caspase 3 cleavage (**Data S3h**).

With the knowledge that Lys4 may be reactive in the context of a G2XK4 derivative, we next assessed whether Gly2- and Lys4-myristoylation could compete within the same protein. For this, we first altered the G-myristoylation S6K7 motif of the GK peptide by introducing an additional Ser between Ser6 and Lys7, pushing the Lys back to position 8 (G2G3K4F5S6S7K8PR). The resulting peptide therefore featured two overlapping myristoylation motifs separated by a two amino acid spacer, one for modification of the α-group of G2, relying on motif S6S7, and a second on the ε-amino of Lys4, relying on motif S7K8. This modification reduced the catalytic efficiency compared with the GK peptide (**Table 1**). Furthermore, the MS2 spectra of the 1173 Da myristoylated product evidenced only a major G-myristoylation product, as unambiguous characteristic b2, y7, and y8 ions were identified (**Fig. S6a,b**; **Data S1**). Nevertheless, we could not exclude the possibility of K-myristoylation, as associated prototypic K-myristoylation ions could be identified at intensities lower than the threshold. Therefore, other experimental approaches were required to assess for simultaneous K-myristoylation.

The IpaJ protease as a tool to unravel ε-myristoylation

The above data indicate that, whatever the context, G-myristoylation appears to be systematically favored over K-myristoylation when in competition. This implies that K-myristoylation is unlikely to originate from a co-translational event, as myristoylation of human proteins on residue 2 appears to be promoted solely on sequences featuring Gly2. To date, two physiological proteolytic processes have been shown to impact myristoylation,

caspases and IpaJ proteolytic cleavage, and indeed several internal K-myristoylation sites like in SRP54 may arise from caspase cleavage. Therefore, we decided to assess conditions favoring possible post-translational K-myristoylation.

The virulence factor IpaJ from *Shigella flexneri* is a cysteine protease that specifically recognizes the myristoyl (Myr)-G moiety of several human proteins *in vivo* [26]. As a result of its activity on Myr-G-peptides (**Fig. S7a**), IpaJ is predicted to induce a 267 Da shift (Myr-Gly product) in MALDI MS spectra. Since it cleaves off G-myristoylation sites, we anticipated that IpaJ - provided it specifically processes any G-myristoylated peptide – could be used to unravel unambiguous and undiscovered K-myristoylation modification sites.

The *S. flexneri IpaJ* gene was cloned and the protein overexpressed and purified to homogeneity (**Fig. S7b,c**). An inactive variant (IpaJ-C64S) was also produced as a control. Under specific conditions (**Fig. S7d**), IpaJ was sufficient to cleave off the G-myristoylated moiety from any tested peptide *in vitro*, unlike IpaJ-C64S (**Fig. S8a,b,c**). Additionally, IpaJ was inactive on an N-acetylated G-peptide, an Ala2-myristoylated peptide, and a Lys3-myristoylated peptide derived from ARF6 (**Fig. S8d** and **Tables S2-3**). Due to its unique and high specificity for G-myristoylation and the observed complete cleavage of any G-myristoylated peptide, we concluded that IpaJ could be used as a powerful tool to distinguish G-myristoylation from K-myristoylation and, additionally, investigate IpaJ-dependent post-translational K-myristoylation.

We therefore established an IpaJ pipeline that determined the myristoylation type of any peptide subjected to NMT (**Fig. 3a**). The workflow faithfully reproduced the sequence of events occurring in the context of cellular IpaJ activity. As proof-of-concept, a dozen peptides

were carefully chosen for their sequence diversity and ability to undergo G-myristoylation (**Fig. S9a,b**; see ARF6 in **Fig. S10**). Combined with previous data [25], IpaJ efficiently cleaved any G-myristoylated peptide regardless of the sequence downstream of the Gly residue (**Fig. S9c**).

Taking the double myristoylation of ARF6 at both the N-terminus of Gly2 and the side chain of Lys3 into account [20], we used the IpaJ pipeline to assess double myristoylation of other GK-starting NMT substrates. We identified dual myristoylation of only a peptide derived from the calcium sensor family HPCA/HPCL1/NCALD at T2 (**Fig. 3b**; **Fig. S11**). Finally, double myristoylation was not observed with other closely related substrates of HPCA in the same NCS1 family of calcium sensors sharing the GK motif (**Data S1**). We cannot exclude that the extremely high hydrophobicity of such acylated products prevented their crystallization in the matrix and consequently their identification. Myristoyl-Gly cleavage by IpaJ at T2 contributes to decrease the overall hydrophobicity and may help to reveal double myristoylation.

Competition between G- and K-myristoylation

With the IpaJ/NMT pipeline in hand, we next examined the ambiguous myristoylation state of the GGKFSSKPR peptide. At T1, MS2 analysis unambiguously revealed G-myristoylation as a product (**Fig. S6b**). At T2 (i.e., after incubation with NMT and IpaJ), both an uncleaved myristoylated peptide and a shorter, 267 Da peptide resulting from G-myristoylation cleavage were identified (T2, 1173 and 906 Da ions **Fig. S6c**). There was also unambiguous and unique K-myristoylation at Lys4 of the GGKFSSKPR peptide at T3 (1059 Da ion, **Extended Data 4c,d**). IpaJ, therefore, decomplexed the 1173 Da ion observed at T1 and evidenced K-myristoylation within this ion, which was otherwise only barely visible (**Fig. S6b, Data S1a**).

Interestingly, the IpaJ product of G-myristoylation of this peptide started with GK to form GKFSSKPR (GKFS), which displayed reduced myristoylation efficiency (**Fig. 1b**). This peptide could be further myristoylated at T3 to produce a 1116 Da form. MS2 analysis revealed K-myristoylation at the N-terminal Lys3 of this peptide (**Fig. S6d**). To obtain a simpler set of products for the pipeline, we further characterized the myristoylation profile of GKFS in the IpaJ workflow (**Fig. 3c,d,e**). Though clearly myristoylated at T1 (1116 Da ion), the type of myristoylation could not be deciphered by MS2 (**Data S1**). The 849 Da ion at T2 revealed that part of the 1116 ion corresponded to G-myristoylation of the peptide. MS/MS data of the 1116 ion remaining on part of the peptide at T3 showed that it had undergone K-myristoylation. We speculated that the GKFS-starting peptide displayed a double identity for G- and K-myristoylation due to the bulky residue at position 4 (Phe) and a small residue at position 5 (Ser). We hypothesized that while a spacer is mandatory for both myristoylation types, either myristoylation might be favored depending on as yet unknown amino acid requirements occurring in that spacer (i.e., aa4-5, the two residues switched in the GK series). This is likely, as both pockets 4 and 5 are large enough to accommodate all side chains [17], so some combinations could favor K-myristoylation in a GK-starting context.

We found no straightforward natural chassis in human sequences starting with a Gly2 that led to competition between K- and G-myristoylation and a systematic preference for G-myristoylation in sequences originating from co-translational myristoylation. When we applied the IpaJ pipeline to GK-starting peptides derived from the human G-myristoylated proteome and exploited its enhanced sensitivity to detect K-myristoylation, we observed that most displayed the myristoylated peptide at T3 after IpaJ cleavage (**Table S3**). Therefore, there was competition between G- and K-myristoylation on GK-starting peptides. Furthermore, none of

these peptides displayed a hydrophobic residue at position 4, explaining why the K-myristoylation/G-myristoylation ratio was smaller than with the GKFS peptide and why prototypic K-myristoylation ions were not detected in MS2 spectra. Nevertheless, as myristoylation can also arise from post-translational addition, provided that a cleavage generates a new K–accepting site, competition favoring K-myristoylation is likely. Indeed, as noted above, hundreds of internal proteins displayed a Phe/Leu4 profile (**Data S2a**; see columns aa3, Phe3, and Leu3 corresponding to this position, as all sequences start at the cleavage site and not the unprocessed M1).

In conclusion, a Gly at the N-terminus of peptides with GKFXSK motifs might sustain competition between K- and G-myristoylation. These data also indicate that the KXXSK motif – not only [^G]KX[FL]SK – is indicative of K-myristoylation independent of the residue preceding the K.

IpaJ reveals that K-myristoylation may also occur on free or acetylated Lys1

When the IpaJ/NMT pipeline was applied to either the myristoylated GGKFSSKPR or the GKFSSKPR peptides (see above and **Fig. 4a**), IpaJ-induced cleavage of this reaction product also led to myristoylation of an unexpected K-starting peptide resulting from reiterated G-myristoylation IpaJ cleavage of each new G-myristoylation product. Although we could not distinguish alpha and epsilon myristoylation on this K-starting peptide, its side-chain myristoylation was expected. Indeed, taking into account the S6K7 clamping of the peptide at the peptide-binding site, 3D molecular modeling indicated that the N-terminal free amino group would be too distant from the catalytic base to promote α-myristoylation. To test the hypothesis of K-myristoylation, we first prepared an acetylated (i.e., N-α-blocked) derivative

of the GK peptide starting with K3 (acK). In the presence of MyrCoA and NMT, this peptide was myristoylated with a catalytic efficiency similar to the previously tested peptides (**Fig. 2a**, **Table 2**). An acetylated (i.e., N-alpha-blocked) derivative of a peptide with the aa4-5 spacer shortened to only one residue was no longer myristoylated, as expected (acKΔS, **Table 2**). Finally, we checked that such Lys-starting peptides derived from the GK (K) and HPCA chassis (K-HPCA) underwent K-myristoylation (**Fig. 2a**, **Table 2**). The crystal structure of the complex between NMT and an acK derivative revealed two alternative structures, both of which supported a K-myristoylation mechanism involving direct interaction between the Lys side chain and the catalytic base (**Fig. 4b**). The group around Cα3 could be retrieved in two opposite directions: the first showed the *ac* group entering either pocket 3 mimicking an Asn side chain (**Fig. 4c**) while the overall active site positioning mimicked that observed with GK; and the second revealed the *ac* moiety occupying part of pocket 4, which was large enough to host two moieties (i.e., *ac* and Ser4 in acK) from the same peptide (**Fig. 4b,c,** right). We concluded that side chain myristoylation can also occur on an N-terminal lysine (Lys1) regardless of acetylation status. Lys1 myristoylation might result from post-translational proteolytic cleavages such as those induced by IpaJ during *Shigella* infection.

Substitution of the Ser6 into Ala of the IpaJ cleavage product of NCS1 prevented K-myristoylation. This reinforces the conclusion of the strict requirement for the S6K7 motif to achieve K-myristoylation (**Table 2**). The similar change with the Gly-starting NCS1 variant also revealed a dramatic decrease but not fully abolished G-myristoylation (**Table S1**). This confirmed the overlap of the recognition motifs for both myristoylation types.

We next interrogated the 15 GK starting peptides derived from different human proteins using the IpaJ pipeline (**Table S3**). As noted above, because they mostly resulted in G-myristoylation, all of them are likely to generate Lys1 N-termini after IpaJ cleavage. **Fig. S11** summarizes such behavior with HPCA. As expected, there was evidence of K-myristoylation at T3, further indicating that all such peptides might undergo K-myristoylation after IpaJ cleavage (**Data S1**). We conclude that all GK-starting proteins might undergo post-translational K-myristoylation after IpaJ cleavage.

DISCUSSION

N-myristoyltransferases (NMT), also known as glycylpeptide N-tetradecanoyltransferases (EC 2.3.1.97; see BRENDA resource [33]), are the only enzymes known to catalyze N-myristoylation in eukaryotes [5]. This modification mostly occurs co-translationally and involves the transfer of a lipid (myristate) to the α-amino group of the N-terminal amino group of G-starting proteins exposed mainly by the prior action of methionine aminopeptidases (MetAP) at the ribosome. NMT, if unbound to the ribosome, can also act post-translationally on proteolytically-processed folded proteins. For a long time, all known post-translational targets of NMTs were associated with proteolytic cleavages induced by caspases during apoptosis [16, 34-39]. Both human NMTs (NMT1 and NMT2) undergo caspase-3 cleavage, which removes their N-terminal ribosome binding motif [40], decreasing co-translational activity and strengthening their post-translational properties. To date, caspases - mostly caspase 3 and to a lesser extent caspase 6 [18] - are the only proteases known to promote post-translational myristoylation. This apoptosis-induced post-translational G-myristoylation subset now comprises a couple of dozen protein targets, while there are hundreds of putative sites in

552 the human proteome [5, 18]. The physiological relevance of apoptosis-induced post-
553 translational myristoylation of these targets is still unclear, except for the BH3-interacting
554 domain death agonist cBID [34]. Whether post-translational myristoylation exists under
555 constitutive or inducible physiological conditions other than apoptosis is unknown, as are the
556 possible proteases other than caspases involved. If this non-apoptotic post-translational
557 myristoylation exists, the major NMT (NMT1), the only NMT expressed in cells in different
558 truncated versions (some devoid of the ribosome binding domain) under standard physiological
559 conditions, likely fulfills this housekeeping post-translational activity [5].
560
561 Recent crystal structures of NMTs co-crystallized with G-starting substrates have revealed how
562 the side chains of these substrates are recognized and lie within dedicated cavities (pockets 3-
563 8; [1, 17, 19]), characteristics that contribute to the G-starting recognition pattern. Although
564 this G-starting recognition pattern appeared to clearly define a global specificity, it did not
565 explain the entire NMT-dependent myristoylation landscape. This was even more evident when
566 we and others reported the possibility that NMTs catalyze K-myristoylation *in vitro* and *in vivo*
567 on a couple of targets, including ARF6 [19, 20]. Although this K-myristoylation function
568 should expand the range of NMT substrates, the extent and clues ruling NMT-dependent K-
569 myristoylation are unknown. It was recently shown that both NMTs similarly achieve the
570 second K-myristoylation step *in vitro*, in keeping with their overlapping substrate specificity
571 [17]. However, when overexpressed in HEK293T cells, NMT2 appears more efficient than
572 NMT1 for K-myristoylation of overexpressed ARF6 [20]. This may reflect either (i) a distinct
573 subcellular localization of NMT2, which its slightly different N-terminal extension compared
574 with NMT1 might promote (there is currently no evidence to support this hypothesis) or (ii)
575 preferential binding of NMT2 over NMT1 to ARF6, though NMT2 and NMT1 are both part
576 of the ARF6 interactome [18, 41].

So far, NMT-catalyzed K-myristoylation can be described as a post-translational secondary event acting on G-myristoylated, fully folded, and active ARF6 forms. Secondary ARF6 myristoylation is therefore post-translational but mechanistically very different from the aforementioned caspase-induced post-translational G-myristoylation. In as much as K-myristoylation is reversible *in vivo* [5, 19], definitive proof of NMT-induced K-myristoylation under standard physiological conditions (i.e., without either overexpression of an NMT or one of its targets such as ARF6, a G2A variant, or a sirtuin deacetylase knockout) is still lacking, as are the targets undergoing this modification.

Here, we reveal the unique and common specificities underlying NMT-dependent G- and K-myristoylation. We found that the main specificity rules for NMT-dependent G- and K-myristoylation partially overlap, leading to competition between the two types of myristoylation in GK-starting contexts. The myristoylation rules are summarized in **Figs. 1e** and **2f**. The very narrow shape of pocket 2 ensures reactivity of a primary amino borne only on unbranched moieties such as N-terminal glycine or side chains like lysine, which facilitates the 70° rotation of the amino-reactive group from the catalytic base towards the thio-reacting group of MyrCoA. This motion allows Wat3 to enter from the water channel and play a crucial role at the transition state [19]. The addition of a methyl group such as in Ala2 or meGly variants makes this motion unlikely and dramatically decreases catalysis. The rules directing G or side chain myristoylation are predominantly determined by the R_1 and R_2 chemical groups branched at Cα3. NMT usually favors G- over K-myristoylation on proteins starting with Gly, while K-myristoylation dominates on any other N-terminal chain.

Overall, NMT-dependent K-myristoylation is less efficient than G-myristoylation whatever the context. Although the two myristoylation types may directly act on an N-terminal residue, K-myristoylation appears less strict than G-myristoylation in this regard; indeed, K-myristoylation does not require a given side chain of the first amino acid but relies on optimized residues at positions 5-7. A major difference between the two myristoylation types is that K-myristoylation needs the KXXSK-restricted pattern. This is unlike G-myristoylation, which allows the modification of many proteins not exhibiting the S6K7 motif by NMT on their N-terminal G (see [17] and **Table 1**). The KXXSK motif of K-myristoylation can be associated with a one amino acid extension or acetylation at the N-terminus provided that an *ad hoc* proteolytic machinery can produce these extremities. The specificity of MetAPs - which remove the first Met from proteins - favors new N-terminal residues with short side chains like Ala, Gly, Pro, or Ser. However, a survey of the human proteome reveals no K-myristoylation on Lys2, Lys3, or Lys4 and 15 G-myristoylation substrates in the subset of proteins arising from MetAP-guided co-translational cleavage rules. Among MetAP unprocessed substrates, four MK-starting proteins were identified, but we ruled out that they might undergo K-myristoylation. A very reduced number of internal KXXSK K-myristoylation motifs are preceded by caspase cleavage sites in the human proteome, suggesting that K-myristoylation - in contrast to mostly co-translational G-myristoylation – might not result from primary post-translational addition, for instance initiated by cleavage by specific caspases during apoptosis [16, 34]. K-myristoylation motifs resulting from cotranslational Met removal or posttranslational caspase cleavage are therefore apparently significantly depleted from the human proteome. This is unlike G-myristoylation motifs, which are enriched and represent 2% of the proteome [18].

Our current data suggest that post-translational myristoylation might occur after initial G-myristoylation, possibly for few additional proteins such as HPCA exhibiting the G2K3 motif, in addition to ARF6, causing double myristoylation. Alternatively, it might occur upon proteolytic cleavage caused by a protease with a specificity different from MetAP or any caspase. While such proteases in the human proteome have yet to be discovered, we show here that the bacterial protease IpaJ - a virulent factor expressed in human cells during *Shigella* infection - might correspond to such a protease. In this report, we used IpaJ both as a case study and a unique tool to discriminate between G- and K-myristoylation. Our data confirm that ARF6 is myristoylated before and after IpaJ cleavage [19, 20], in contrast to the other ARFs (ARF1/3/4/5) or ARL1, which do not feature a K3 like ARF6. The insensitivity of ARF6 – but not the other ARFs - to IpaJ was initially proposed to be due to ARF6 not being located at the ER/Golgi, unlike the other members of the ARF/ARL family, but rather at the plasma membrane. ARF6 translocates to endosomes, where it acts as a major regulator of endocytosis [42, 43]. Our data indicate that ARF6 resistance to IpaJ might also result from the unmasking of a myristoylatable Lys3 upon G-myristoylation cleavage. As a result, we surmise that ARF6 would still be myristoylated and retained on myristoylation affinity azido-biotin prior to and after IpaJ cleavage. IpaJ would therefore not act as a myristoyl eraser of G-myristoylated proteins displaying a Lys3. Indeed, NMT would be still able of adding a myristoyl group on the side-chain of the neo N-terminal K.

Overall, our data indicate that, unlike G-myristoylation, NMT-catalyzed K-myristoylation should be rare in humans, as it may only occur as a secondary myristoylation on a few already G-myristoylated proteins or induced after IpaJ cleavage of the same subset. So far, ARF6 remains the only K-myristoylation clear-cut target of the human proteome.

To conclude, this study extents our knowledge of the subproteome, which is sensitive to NMT-guided myristoylation. Unlike G-myristoylation, which may arise by NMT-catalyzed addition after caspase cleavage, our current data suggest that K-myristoylation in the GK context might arise *in vivo* from secondary myristoylation after primary myristoylation of a Gly residue, mostly unmasked by Met cleavage, or after the action of proteases distinct from caspases able to unravel new K-myristoylation sites. Such proteases - yet to be discovered - might be induced under defined physiological conditions and, like caspases during apoptosis, not be constitutively active like the methionine aminopeptidases that trigger cotranslational G-myristoylation.

MATERIALS AND METHODS

Chemicals

All peptides (see all sequences in **Tables 1-2** and **S1-S3**) were purchased at 95% purity (Genscript, Leiden, Netherlands). NAD was purchased from Roche (Basel, Switzerland). All other chemicals were purchased from Sigma-Aldrich (St. Quentin, France). Stock solutions of myristoyl-CoA (0.2 mM) were prepared in sodium acetate, pH 5.6, and 1% Triton X-100, except for MALDI analysis, where cholate was used to reduce background.

IpaJ cloning

Wild-type full-length IpaJ numbering is used throughout the text. The nucleotide sequence encoding residues including the final TGA stop codon of *S. flexneri IpaJ* ORF (UniProt code Q54150) was optimized for *E. coli* expression (**Table S4**) and synthesized *in vitro* (Genscript). For protein expression, the 30-259 fragment was cloned between the *Nco*I and *Hin*dIII restriction sites of expression vector pETM30 as a C-terminal fusion with glutathione S-transferase (GST, **Fig. S7b**). A His-tag was placed at the N-terminus of the fusion to facilitate protein purification (see below). A TEV cleavage sequence inserted between the two ORFs allowed release of IpaJ from GST with only an additional N-terminal Gly-Ala-Met-Ala tetrapeptide sequence upstream of Arg30, the first residue of IpaJ. Two variants were created by site-directed mutagenesis using a QuikChange site-directed mutagenesis kit (Stratagene, San Diego, CA) with the primer pairs displayed in **Table S4**. The first variant was inactive (C64S), while the other (3M) was fully active with improved solubility and stability over time (Leu97Asn/Leu99Asn/Cys103Ser).

Enzyme production and purification

HsNMT1. The HsNMT1*l* isoform containing residues 81-496 corresponds to the major isoform [3]. HsNMT1*l* was cloned into pET16b as an N-terminal His-tag fusion [3], and the

683 recombinant protein was expressed and purified as described [17]. Isoform HsNMT1*s*
684 containing residues 99-496 was cloned into pET28 and purified as described [19].
685
686 The three soluble 30-259 variants of the protease IpaJ were produced as follows.
687 Rosetta2pLysS cells expressing the pETM30 derivative were grown in 1 L of 2xTY medium
688 supplemented with kanamycine (50 µg/mL) and chloramphenicol (34 µg/mL) at 37°C under
689 vigorous shaking. Protein overexpression was induced with 0.5 mM IPTG at OD_{600}=0.4. Cells
690 were transferred at 20°C and grown for 20 hours. Cells were centrifuged at 10,000 x g, and the
691 pellet was resuspended (10 mL/g) in lysis buffer (20 mM Tris pH=8.0, 0.2 M NaCl, 5 mM 2-
692 mercaptoethanol, 5 mM imidazole, 5% glycerol). Cells were lysed with a sonicator Q700
693 (amplitude 50%, 10 s on and 20 s off) for 3 min at 4°C. The sample was then centrifuged at
694 40,000 x g (Rotor JA20) for 30 min and the pellet discarded. The supernatant was loaded on a
695 Ni-IMAC-HisTrap^{TM} FF 5 mL column (GE Healthcare, Chicago, IL) at 3.5 mL/min; elution
696 was achieved with a liner 0-0.5M imidazole gradient run over 100 ml. Purified proteins were
697 dialyzed in Spectra-Por7 semi-permeable dialysis membranes (8 kDa cut-off; Thermo Fisher
698 Scientific, Waltham, MA) for 48 hours against conservation buffer (20 mM Tris pH=8.0, 0.2
699 M NaCl, 5 mM dithiothreitol, 55% glycerol) at 4°C. The sample was stored at -20°C.
700 Protein concentrations were measured with the Bio-Rad Protein Assay Kit using bovine serum
701 albumin (Sigma-Aldrich) as reference.

702 **Protein crystallization and structure determination**

703 *HsNMT1:MyrCoA:X structure.* HsNMT1*s* was used to solve the structures of the complexes
704 with seven different peptides in the course of this study (HPCA, AK, AcK, AcG[Orn], AN,
705 meGN and GGK) with PDB entries 7OWM, 7OWN, 7OWO, 7OWP, 7OWU, 7OWQ and
706 7OWR, respectively. Suitable crystals of HsNMT1:MyrCoA:peptide substrates were obtained

by co-crystallization using the hanging-drop vapor diffusion method at 20°C in the crystallization conditions previously described [17, 19]. Briefly, crystallization droplets were formed by mixing 2 µL of the of HsNMT1:MyrCoA:peptide complex (ratio 1:1.5:1.5) at 6-9 mg/mL with 2 µL of the precipitant solution containing either 0.1 M $MgCl_2$, 0.2 M NaCl, 0.1 M sodium citrate pH 5.6, and 18-24% (w/v) PEG 6K or 0.1 M sodium acetate pH 4.6, and 18-24% (w/v) PEG 6K. Crystals were cryoprotected in the reservoir solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. Complete X-ray datasets of complex were collected at 100*K* a single wavelength from a single crystal at the French National Synchrotron Facility (SOLEIL) PROXIMA1 (for HPCA and AcG[Orn] at λ=0.98400 Å and AcK at λ=0.97856 Å) or PROXIMA2 (for AK and GGK at λ= 0.98012 Å) beamlines and at European Radiation Synchrotron Facility (ESRF) ID30A1 (for AN at λ=0.96600 Å) and ID30A3 (for me-GN λ=0.96775 Å) beamlines. Datasets were integrated with XDS [44] and scaled and reduced using AIMLESS from the CCP4 package [45]. For crystals that suffered from anisotropic diffraction (HPCA, AcK and GGK), data were processed with STARANISO on unmerged data [46] before AIMLESS data reduction. Crystals belonged essentially to the space group $P2_12_12$ with similar unit cell parameters with the exception of GGK and meGN belonging both to C2 with similar unit cell parameters (summarized in **Table S5**). In both space groups identified, unit cells contained two NMT molecules per asymmetric unit. Structure resolution was accomplished in all cases using the molecular replacement method and solved using PHASER [47] and the HsNMT1:MyrCoA:peptide ternary complex (PDB entry 5O9T or 6SK2) as a search model. The structure of AN was solved using MOLREP [48] and protein coordinates of HsNMT1:MyrCoA (PDB entry 5O9T) as a search model. Structures were subjected to alternating refinement cycles using PHENIX and manual model building using COOT [49-51]. The good quality of the electron density maps also enabled the refinement of

substrate peptide, reaction intermediate, and reaction product molecules bound to HsNMT1 in each complex. Chemical compound libraries were generated using PRODRG server [52] in combination with eLBOW from the PHENIX suite. The geometry of the final models was validated using MOLPROBITY [53]. Figures were generated using PYMOL (DeLano Scientific LLC, http://pymol.sourceforge.net/). X-ray data collection and refinement statistics are summarized in **Table S5**. Omit electron density maps were calculated using PolderMap [54] from the Phenix suite after omitting the peptide of each active site and excluding the bulk solvent around the omitted region. Omit maps of the nine crystal structures are displayed in **Fig. S12**.

Measurements of activity and associated parameters

HsNMT1 activity was assayed at 30°C in a coupled assay using an updated version of the previously described protocol [27]. The reaction mixture contained 50 mM Tris-HCl (pH 8.0), 1 mM $MgCl_2$, 0.193 mM EGTA, 0.32 mM DTT, 0.2 mM thiamine pyrophosphate, 2 mM pyruvate, 0.1 mg/mL of BSA, 0.1% Triton X-100, 2.5 mM NAD^+, 0.125 units/mL of porcine heart PDH (0.33 units/mg), 40 µM myristoyl-CoA, and 1-2000 µM peptides. Unless otherwise stated for tight binding studies, the reaction mixture was pre-incubated for 3 min at 30°C before starting the reaction by adding MyrCoA. A final volume of 200 µL was used in 96-well black plates (Grenier Bio One and Dominique Dutscher, Brumath, France; the optical path for 0.2 mL is 0.55 cm). A value of 6300 $M^{-1}.s^{-1}$ was used as the molecular extinction coefficient of NADH at 340 nm. An Infinite M $Nano^+$ plate reader equipped with micro-injectors (Tecan, Lyon, France) was set at 340 nm to monitor the absorbance over time at 30°C. Briefly, a reaction mixture containing HsNMT1*l* at different concentrations of peptide acceptors was pre-incubated for 3 min at 30°C before starting the reaction with MyrCoA.

Myristoylation kinetics were monitored continuously for 30 min, and the data were fitted as to obtain the initial velocities associated to each peptide concentration. Curve fits to obtain kinetic parameters were achieved by non-linear regression with GraphPad Prism 9.1 (GraphPad Software, La Jolla, CA). Parameters with standard errors were computed for all parameters using the complete dataset including replicates. Both k_{cat} and K_m kinetic parameters were obtained by fitting to the Michaelis–Menten equation. k_{cat}/K_m values and the associated standard deviations were obtained by taking advantage of the k_{SP} approach [55] with $k_{SP}=k_{cat}/K_m$ and $v_0/[E]= k_{SP}[S]/(1+[S]/K_m)$, where v_0 is the reaction rate measured at NMT concentration $[E_0]$.

The NMT/IpaJ pipeline described in **Fig. 5b** first involved incubation of the peptide (100 µM) for 1 hour in the presence of 0.5 µM HsNMT1 or HsNMT2 (T1). The buffer was the same as for NMT activity measurement but contained cholate (5 µM) instead of Triton X-100. Full IpaJ cleavage conditions involved further incubation at 20°C for 1 hour in the presence of 10 µM IpaJ-3M (T2). 5 µM HsNMT1 was finally added for 60 min at 30°C (T3). T0 corresponded to the T3 time (3 hours) with the peptide diluted in the incubation buffer but in the absence of any enzyme.

Mass spectrometry

300 µL of a mixture containing 50 mM Tris (pH 8), 0.193 mM EGTA, 1 mM $MgCl_2$, 1 mM DTT, 5 µM sodium cholate, 40 µM Myr-CoA solution (stock solution 0.2 mM Myr-CoA, 10 mM sodium acetate, 2.5 µM sodium cholate), 0.5 µM NMT, and 100 µM of synthetic peptide were pre-incubated at 30°C. The myristoylation reaction was followed over time by collection of 10 µL samples further diluted in 90 µL of water/acetonitrile (90/10) solution. The different samples were then diluted five times in the matrix solution made of 5 mg/mL of α-cyano-4-hydroxycinnamic acid solubilized in water/formic acid/acetonitrile (50/50/0.1%). 1 µL of each dilution was spotted on a metal target and dried. MS and MS/MS spectra of each sample were acquired with an AB SCIEX 5800 MALDI-Tof-Tof instrument in positive ion mode. Survey scans were performed using delayed extraction (390 ns) in reflector mode for a total of 15,000 shots. MS/MS scans were operated with a collision energy of 1 kV. Peptide and fragment mass tolerances were set at 10 ppm and 0.8 Da, respectively. Mass spectra were analyzed with PeakView® 2.2 software (AB Sciex, Macclesfield, UK). The default threshold in MS/MS peak labelling and finding was 5% (displayed with a black dot on the intensity axis) and centroid height percentage was 50% as recommended by the constructor. MS/MS deviations from theoretical values were on average less than 0.03 Da. MS/MS analysis and assignments of peptides with non-natural aminoacids was also performed with the additional help of software ProteinProspector v6.2.2 (https://prospector.ucsf.edu/prospector/mshome.htm) [56]. Theoretical mass values of all fragments were double-checked at https://web.expasy.org/peptide_mass/. All spectra are available in **Data S1**.

ACCESSION NUMBERS

The seven crystal structures of NMT in complex with MyrCoA and the peptides reported here have been deposited at the PDB under codes 7OWM, 7OWN, 7OWO, 7OWP, 7OWQ, 7OWR, and 7OWU.

ACKNOWLEDGMENTS

We warmly acknowledge Odile Schiltz (IPBS, Toulouse), Virginie Redeker, Jean-Pierre Le Caer, Laila Sago and David Cornu (all at SICaPS, Gif/Yvette) for their extensive support with mass spectrometry analyses. We thank the French National Synchrotron Facility (SOLEIL) for provision of synchrotron radiation facilities (proposal IDs 20191181 and 20170872) and the staff of the Proxima 1 & 2 beamlines.

FUNDING

This work was supported by French National Research Agency (ANR) DynaMYT (ANR-20-CE44-0013) and Fondation ARC (ARCPJA32020060002137) grants. This work has benefited from the support of a French State grant (ANR-17-EUR-0007, EUR SPS-GSR) managed by the ANR under an Investments for the Future program (ANR-11-IDEX-0003-02), from the I2BC crystallization platform supported by FRISBI (ANR-10-INSB-05-01), from the facilities and expertise of the I2BC proteomic platform SICaPS, supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Sud University, and the COST Action CA20113 ProteoCure. FR is supported by grants from Région Ile-de-France (17012695) and Fondation pour la Recherche Médicale (FDT202001010779).

AUTHOR CONTRIBUTIONS

FR conducted all MS experiments and kinetic analyses, characterized IpaJ variants, and set up the IpaJ/NMT pipeline. CD undertook cloning, mutagenesis, and purification of IpaJ and performed all NMT structural analyses. RFD and PM completed and consolidated the kinetic analysis. TM and CG conceived the project, supervised the experiments and analyzed the data. CG and TM, wrote the manuscript with contributions of FR and CD.

COMPETING INTERESTS

The authors declare that they have no competing interests.

SUPPLEMENTAL INFORMATION

Supplementary data to this article can be found online at https://doi.org/xxxx.

The Supplementary Materials PDF file contains all Supplementary Figures (**Figs S1-12**) and Tables (**Tables S1 to S5**).

The three Supplementary Data files correspond to one PDF file to mass spectrometry spectra (**Data S1**), and two Microsoft Excel format files (**Data S2-4**) for proteome-wide analysis and predictions.

REFERENCES

[1] Meinnel T, Dian C, Giglione C. Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution. Trends Biochem Sci. 2020;45:619-32.

[2] Lodge JK, Johnson RL, Weinberg RA, Gordon JI. Comparison of myristoyl-CoA:protein N-myristoyltransferases from three pathogenic fungi: Cryptococcus neoformans, Histoplasma capsulatum, and Candida albicans. J Biol Chem. 1994;269:2996-3009.

[3] Pierre M, Traverso JA, Boisson B, Domenichini S, Bouchez D, Giglione C, et al. N-Myristoylation regulates the SnRK1 pathway in Arabidopsis. Plant Cell. 2007;19:2804-21.

[4] Price HP, Menon MR, Panethymitaki C, Goulding D, McKean PG, Smith DF. Myristoyl-CoA:protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. J Biol Chem. 2003;278:7206-14.

[5] Giglione C, Meinnel T. Mapping the myristoylome through a complete understanding of protein myristoylation biochemistry. Prog Lipid Res. 2022;85:101139.

[6] Bhatnagar RS, Ashrafi K, Futterer K, Waksman G, Gordon JI. Biology and enzymology of protein N-myristoylation. In: Tamanoi F, Sigman DS, editors. The Enzymes. San Diego: Academic Press; 2001. p. 241-86.

[7] Resh MD. Trafficking and signaling by fatty-acylated and prenylated proteins. Nat Chem Biol. 2006;2:584-90.

[8] Traverso JA, Micalella C, Martinez A, Brown SC, Satiat-Jeunemaitre B, Meinnel T, et al. Roles of N-Terminal fatty acid acylations in membrane compartment partitioning: Arabidopsis h-type thioredoxins as a case study. Plant Cell. 2013;25:1056-77.

[9] Frearson JA, Brand S, McElroy SP, Cleghorn LA, Smid O, Stojanovski L, et al. N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. Nature. 2010;464:728-32.

[10] Wright MH, Clough B, Rackham MD, Rangachari K, Brannigan JA, Grainger M, et al. Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. Nat Chem. 2014;6:112-21.

[11] Wright MH, Paape D, Price HP, Smith DF, Tate EW. Global profiling and inhibition of protein lipidation in vector and host stages of the sleeping sickness parasite Trypanosoma brucei. ACS Infect Dis. 2016;2:427-41.

[12] Mousnier A, Bell AS, Swieboda DP, Morales-Sanfrutos J, Perez-Dorado I, Brannigan JA, et al. Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. Nat Chem. 2018;10:599-606.

[13] Beauchamp E, Yap MC, Iyer A, Perinpanayagam MA, Gamma JM, Vincent KM, et al. Targeting N-myristoylation for therapy of B-cell lymphomas. Nat Commun. 2020;11:5348.

[14] Kosciuk T, Lin H. N-Myristoyltransferase as a glycine and lysine myristoyltransferase in cancer, immunity, and infections. ACS chemical biology. 2020;15:1747-58.

[15] Martin DD, Beauchamp E, Berthiaume LG. Post-translational myristoylation: Fat matters in cellular life and death. Biochimie. 2011;93:18-31.

[16] Thinon E, Serwa RA, Broncel M, Brannigan JA, Brassat U, Wright MH, et al. Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. Nat Commun. 2014;5:4919.

[17] Castrec B, Dian C, Ciccone S, Ebert CL, Bienvenut WV, Le Caer J-P, et al. Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. Nat Chem Biol. 2018;14:671-9.

[18] Meinnel T. Comment on “Binding affinity determines substrate specificity and enables discovery of substrates for N-Myristoyltransferases”. ACS Catal. 2022;12:8195-201.

[19] Dian C, Pérez-Dorado I, Rivière F, Asensio T, Legrand P, Ritzefeld M, et al. High-resolution snapshots of human N-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. Nat Commun. 2020;11:1132.

[20] Kosciuk T, Price IR, Zhang X, Zhu C, Johnson KN, Zhang S, et al. NMT1 and NMT2 are lysine myristoyltransferases regulating the ARF6 GTPase cycle. Nat Commun. 2020;11:1067.

[21] Carr SA, Biemann K, Shoji S, Parmelee DC, Titani K. n-Tetradecanoyl is the NH2-terminal blocking group of the catalytic subunit of cyclic AMP-dependent protein kinase from bovine cardiac muscle. Proc Natl Acad Sci U S A. 1982;79:6128-31.

[22] Gheorghe MT, Bergman T. Deacetylation and internal cleavage of polypeptides for N-Terminal sequence analysis. In: Atassi MZ, Appella E, editors. Methods in Protein Structure Analysis. Boston, MA: Springer; 1995. p. 81-6.

[23] Leone JW, Hampton B, Fowler E, Moyer M, Krishna RG, Chin CCQ. Removal of N-terminal blocking groups from proteins. Current protocols in protein science. 2011;63:11.7.1-.7.20.

[24] Mattock E, Blocker AJ. How do the virulence factors of Shigella work together to cause disease? Front Cell Infect Microbiol. 2017;7:64.

[25] Burnaevskiy N, Peng T, Reddick LE, Hang HC, Alto NM. Myristoylome profiling reveals a concerted mechanism of ARF GTPase deacylation by the bacterial protease IpaJ. Mol Cell. 2015;58:110-22.

[26] Burnaevskiy N, Fox TG, Plymire DA, Ertelt JM, Weigele BA, Selyunin AS, et al. Proteolytic elimination of N-myristoyl modifications by the Shigella virulence factor IpaJ. Nature. 2013;496:106-9.

[27] Boisson B, Meinnel T. A continuous assay of myristoyl-CoA:protein N-myristoyltransferase for proteomic analysis. Anal Biochem. 2003;322:116-23.

[28] Giglione C, Fieulaine S, Meinnel T. N-terminal protein modifications: bringing back into play the ribosome. Biochimie. 2015;114:134-46.

[29] Padovani D, Zeghouf M, Traverso JA, Giglione C, Cherfils J. High yield production of myristoylated Arf6 small GTPase by recombinant N-myristoyl transferase. Small GTPases. 2013;4:3-8.

910 [30] Boisson B, Giglione C, Meinnel T. Unexpected protein families including cell defense
911 components feature in the N-myristoylome of a higher eukaryote. J Biol Chem.
912 2003;278:43418-29.
913 [31] Utsumi T, Nakano K, Funakoshi T, Kayano Y, Nakao S, Sakurai N, et al. Vertical-
914 scanning mutagenesis of amino acids in a model N-myristoylation motif reveals the major
915 amino-terminal sequence requirements for protein N-myristoylation. European journal of
916 biochemistry / FEBS. 2004;271:863-74.
917 [32] Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, et al. ELM-the
918 eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020;48:D296-d306.
919 [33] Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, et al. BRENDA, the
920 ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res.
921 2021;49:D498-D508.
922 [34] Zha J, Weiler S, Oh KJ, Wei MC, Korsmeyer SJ. Posttranslational N-myristoylation of
923 BID as a molecular switch for targeting mitochondria and apoptosis. Science. 2000;290:1761-
924 5.
925 [35] Utsumi T, Sakurai N, Nakano K, Ishisaka R. C-terminal 15 kDa fragment of cytoskeletal
926 actin is posttranslationally N-myristoylated upon caspase-mediated cleavage and targeted to
927 mitochondria. FEBS Lett. 2003;539:37-44.
928 [36] Vilas GL, Corvi MM, Plummer GJ, Seime AM, Lambkin GR, Berthiaume LG.
929 Posttranslational myristoylation of caspase-activated p21-activated protein kinase 2 (PAK2)
930 potentiates late apoptotic events. Proc Natl Acad Sci USA. 2006;103:6542-7.
931 [37] Martin DD, Heit RJ, Yap MC, Davidson MW, Hayden MR, Berthiaume LG. Identification
932 of a post-translationally myristoylated autophagy-inducing domain released by caspase
933 cleavage of huntingtin. Human molecular genetics. 2014;23:3166-79.

[38] Tapodi A, Clemens DM, Uwineza A, Jarrin M, Goldberg MW, Thinon E, et al. BFSP1 C-terminal domains released by post-translational processing events can alter significantly the calcium regulation of AQP0 water permeability. Exp Eye Res. 2019;185:107585.

[39] Martin DD, Ahpin CY, Heit RJ, Perinpanayagam MA, Yap MC, Veldhoen RA, et al. Tandem reporter assay for myristoylated proteins post-translationally (TRAMPP) identifies novel substrates for post-translational myristoylation: PKCepsilon, a case study. FASEB J. 2012;26:13-28.

[40] Perinpanayagam MA, Beauchamp E, Martin DD, Sim JY, Yap MC, Berthiaume LG. Regulation of co- and post-translational myristoylation of proteins during apoptosis: interplay of N-myristoyltransferases and caspases. FASEB J. 2013;27:811-21.

[41] Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell. 2021;184:3022-40.e28.

[42] Donaldson JG. Multiple roles for Arf6: sorting, structuring, and signaling at the plasma membrane. J Biol Chem. 2003;278:41573-6.

[43] Donaldson JG, Jackson CL. ARF family G proteins and their regulators: roles in membrane transport, development and disease. Nat Rev Mol Cell Biol. 2011;12:362-75.

[44] Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst. 1993;26:795-800.

[45] Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta crystallographica Section D, Biological crystallography. 2013;69:1204-14.

[46] Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, et al. STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). Cambridge, United Kingdom: Global Phasing Ltd. ; 2018.

[47] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658-74.

[48] Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Cryst. 1997;30:1022-225.

[49] Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta crystallographica Section D, Biological crystallography. 2011;67:355-67.

[50] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography. 2010;66:213-21.

[51] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography. 2010;66:486-501.

[52] Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta crystallographica Section D, Biological crystallography. 2004;60:1355-63.

[53] Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological crystallography. 2010;66:12-21.

[54] Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, et al. Polder maps: improving OMIT maps by excluding bulk solvent. Acta crystallographica Section D, Structural biology. 2017;73:148-57.

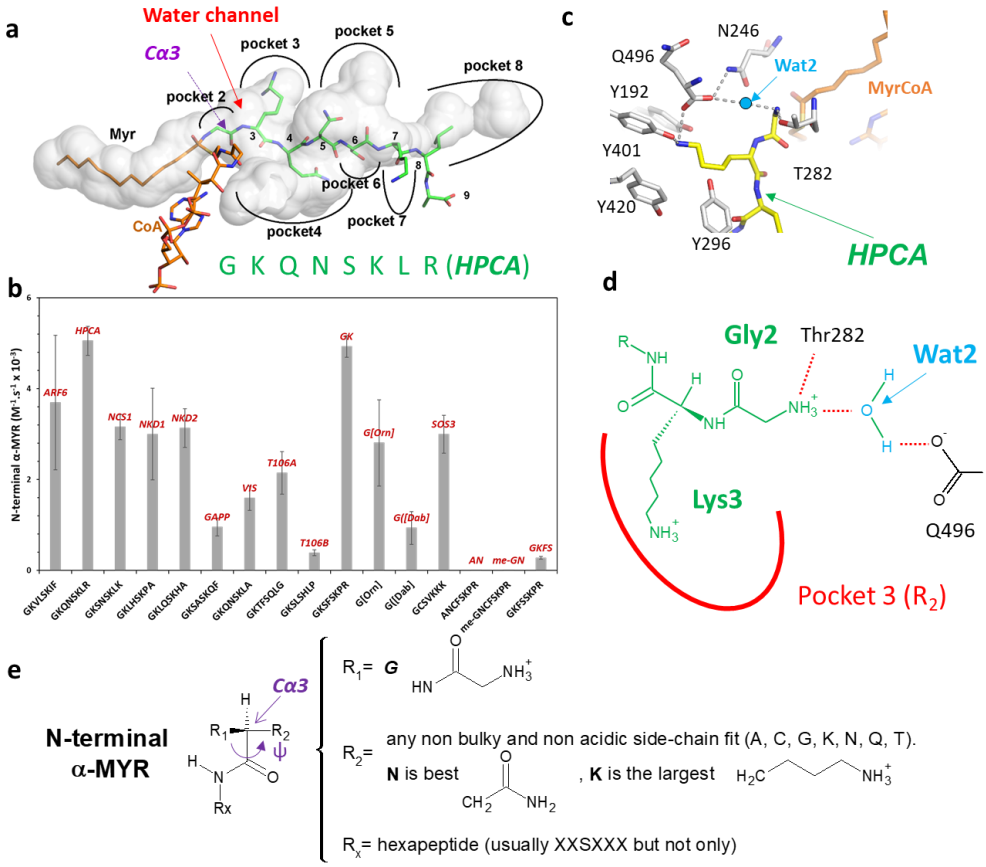
[55] Johnson KA. New standards for collecting and fitting steady state kinetic data. Beilstein journal of organic chemistry. 2019;15:16-29.

[56] Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol Cell Proteomics. 2008;7:2386-98.

FIGURES AND TABLES

See next pages

Fig. 1. G-myristoylation rules in the context of GK-starting substrates



All peptides are derived from the GK peptide unless otherwise stated (see short name and the associated sequence in **Table 1**). **(a)** Crystal structure of the HPCA peptide with HsNMT; the various pockets and their extents are displayed and Cα3 is indicated. The peptide sequence of HPCA is indicated in green. (**b**) Catalytic efficiencies (k_{cat}/K_m) values of GK-starting peptides (see all parameters in **Table 1**). (**c**) 3D details of the active site organization of the HPCA complex. The myristate moiety of MyrCoA is shown in orange, the HPCA backbone peptide is colored in yellow and indicated with a green arrow, residues of NMT in the immediate vicinity are in displayed grey, and the water 2 molecule (Wat2) is shown as a turquoise sphere. (**d**) Planar representation of panel c, R_2 refers to the group defined in panel **e** while pocket 3 is displayed in panel **a**. The color code is the same as in panel c. (**e**) Types of chemical groups allowed around Cα3 leading to classic N-terminal myristoylation with the G acceptor residue at R_1 (G-myristoylation). The ψ torsion angle Cα3 is indicated.

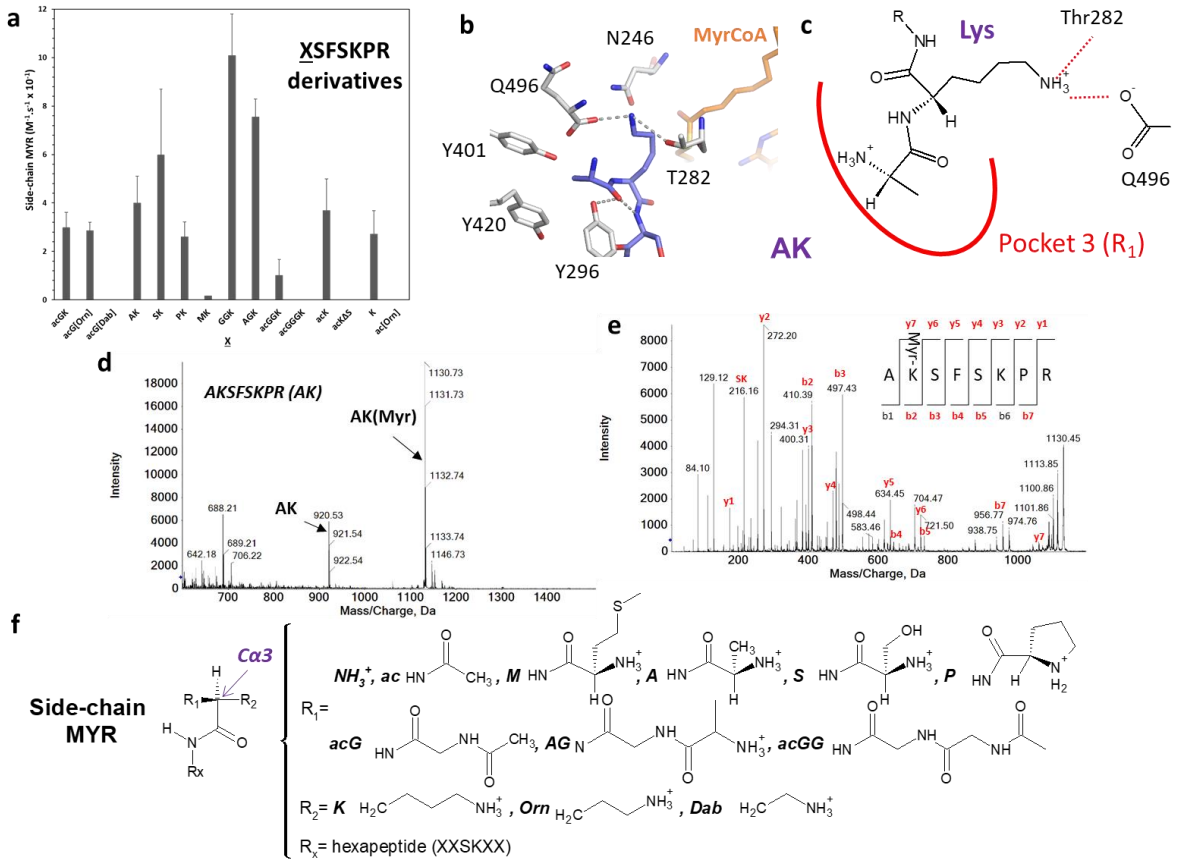


Fig. 2. K-myristoylation and G-myristoylation rules

The peptides are derived from the GK peptide. (**a**) Main K-myristoylation catalytic efficiencies (k_{cat}/K_m) values) measured (see details in **Table 2**). (**b**) Crystal structures of HsNMT1 in complex with MyrCoA and the AK variant. The panel displays a close-up of the structural environment around the catalytic base, the C-terminus of the C-terminal residue Q496. (**c**) Plane representation of panel **b**. (**d**) MS1 analysis of the AK derivative after incubation with HsNMT1. (**e**) MS2 analysis of the 1131 Da myristoylated peptide in panel d showing K-myristoylation; SK refers to an internal dipeptide fragment. (**f**) New knowledge on the type of acceptor amino group leading to K-myristoylation. The panel shows that if a lysine side chain or a non-natural basic side-chain with reduced length occurs at R_2, and depending on the substitutions made around the N-terminus, i.e., while blocking it (*ac* derivatives) or slightly modifying its plane orientation (e.g., if G is substituted for A, S, or P), then K-myristoylation (ε-myristoylation) is favored over N-terminal myristoylation (α-myristoylation) [19, 20].

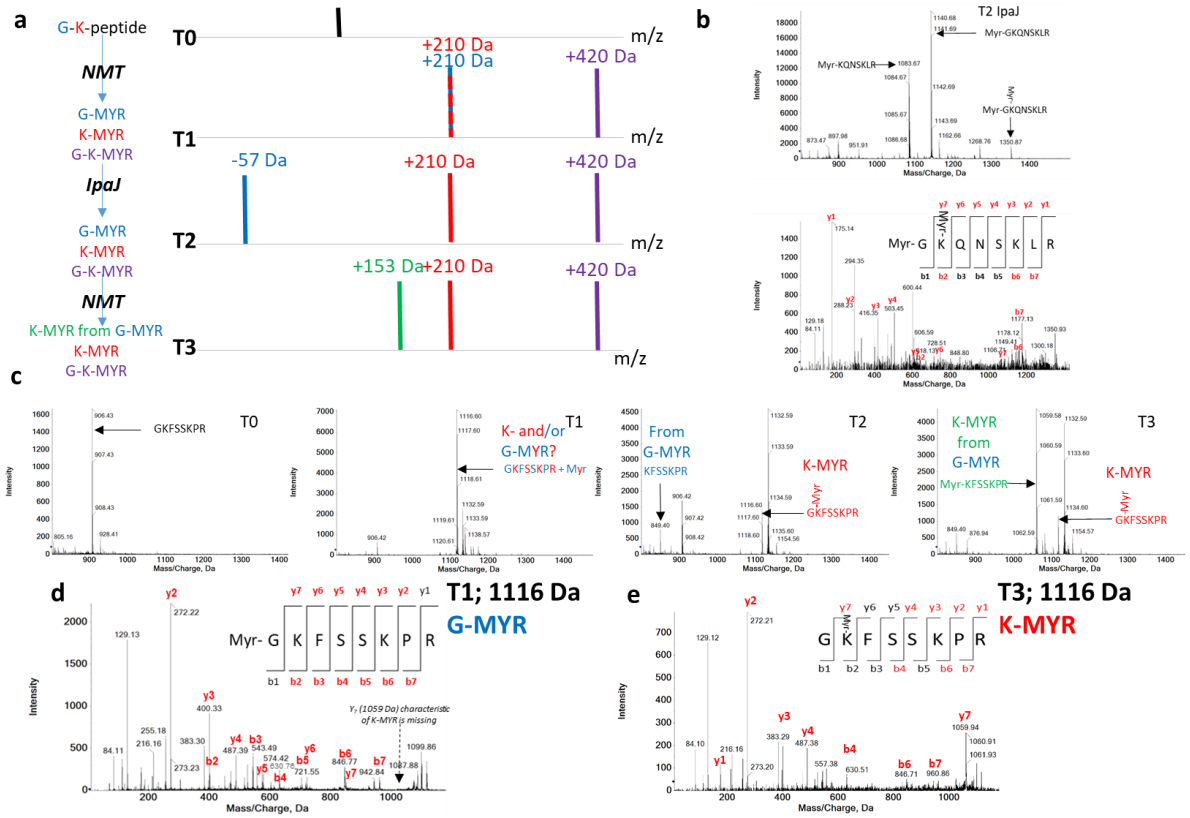


Fig. 3. A pipeline to investigate NMT/IpaJ-coupled impact on myristoylation

(**a**) The schematized NMT/IpaJ pipeline, expected m/z at each step, and corresponding isoforms. In brief, peptides are first incubated with NMT to allow myristoylation and visualization of the +210 Da addition (T1). Next, IpaJ is added at low temperature to allow complete removal of the G-myristoylated moiety, which results in a -57 Da shortening compared with the starting peptide and a -267 Da signature of the product compared with the myristoylated peptide (T2). Finally, high NMT concentrations are added to check whether substrate myristoylation results from the action of the first two steps (T3). (**b**) HPCA peptide at T2 (top) and MS2 spectrum of the double myristoylation product (bottom). (**c**) GKFS peptide (GKFSSKPR) during the different steps of the pipeline. (**d**) MS2 spectrum of the 1116 Da peptide at T1; the missing characteristic y_7 (1,059 Da) ion of K-myristoylation is indicated with an arrow. (**e**) MS2 spectrum of the 1116 Da peptide at T3 with the 1,059 Da ion now visible.

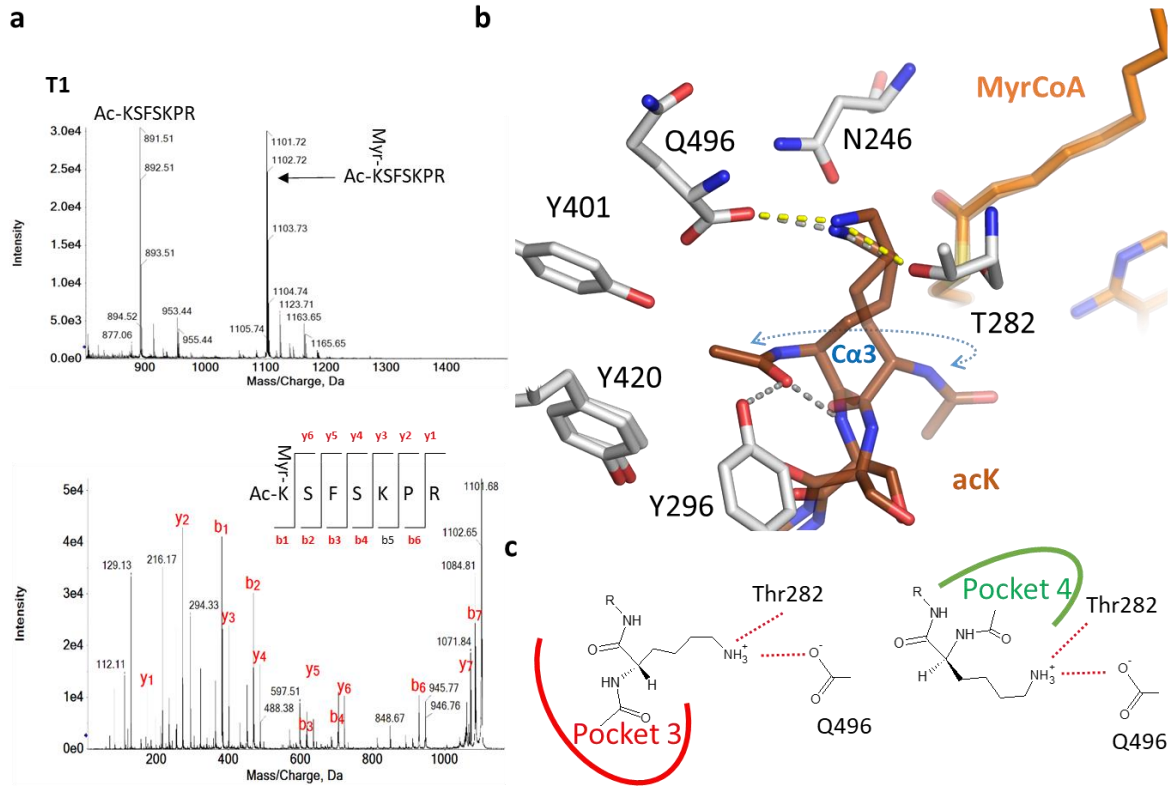


Fig. 4. K-myristoylation mechanism on an N-terminal side chain

The crystal structure of the complex made between acK and HsNMT1 is displayed. (**a**) MS analysis showing myristoylation of the acK; top MS1, bottom MS2. (**b**) The two conformations are displayed, and the rotation point around Cα3 is indicated. (**c**) Planar representation showing how each conformation may position the *ac* chain within pocket 3 (conformation A, left hand side) or pocket 4 (conformation B, right hand side).

Table 1. GK-starting peptides derived from the human proteome reveal N-terminal alpha-MYR as the major modification made by NMT

Series[a]	Peptide sequence	Short name (text & Figures)	Data S1a page	Crystal structure	k_{cat}(s^{-1})	K_m(µM)	k_{cat}/K_m ($s^{-1}.M^{-1}$)	Relative k_{cat}/K_m (%)
Reference GK-peptide variants	GKSFSKPR	GK	2	-	0.050±0.0002	10±2	4931±626	165
	G[Orn]SFSKPR		3	-	0.018±0.001	6±2	2808±951	94
	G[Dab]SFSKPR		4	-	0.015±0.002	16±7	935±362	31
	ANCFSKPR	AN	5	Fig. S2a	0.0004±0.0001	>300	1.2±1.1	0.05
	me-GNCFSKPR	meGN	6	Fig. S2b	0.0018±0.0001	1068±51	1.65±0.05	0.06
	GKSWSKGR		7	-	0.090±0.002	9±1	10,000	334
	GKFSSKPR[b]	GKFS	8/9	-	0.11±0.02	346±75	330±33	11
Human GKX	GKVLSKIF	ARF6	10	-	0.036±0.002	8±2	4320±897	144
	GKQNSKLR	HPCA	11	Fig.1a,c	-	-	5055±320	169
	GKTNSKLA	HPCL4	12	-	0.58±0.06	48±18	12026±3560	401
	GKQNSKLA	VIS	13	-	0.092±0.004	58±11	1589±265	53
	GKSNSKLK	NCS1	14	PDB 5O9S	0.047±0.001	15±1	3157±277	105
	GKSLSHLP	T106B	15	-	0.31±0.04	811±224	386±66	13
	GKTFSQLG	T106A	16	-	0.43±0.03	200±53	2147±467	72
	GKSASKQF	GAPP	17		0.065±0.04	69±17	952±198	32
	GKLHSKPA	NKD1	18	-	0.10±0.01	40±2	2500±200	83
	GKLQSKHA	NKD2	19	-	0.10±0.01	41±2	2501±200	83
	GKRGSRSQ	UTP25	-	-	>0.01	>1000	7±2	0.2
GXK	GAKQSGPA	ZNRF2	20	-	-	-	103	3
	GGKQSTAA	ZNRF1	21		1.45±0.23	2320±573	623	21
	GGKFSSKPR		22/23	-	0.14±0.01	85±32	1647	21
	GGKLSKKK	BASP	24	-	-	-	3900	130
SOS3	GCSVSKKK	SOS3	25	-	0.055±0.002	18±3	2997±415	**100**
	ASSVSKKK		-		-	-	-	<<0.01

MS spectra are displayed in **Data S1a**. The corresponding page number is indicated in the corresponding column.

Table 2. Substitutions around the N-terminus of peptides promoting side chain MYR

Series	Peptide sequence	Shortname (text & Figures)	Series sequence	Data S1c page	Crystal structure	Kinetic parameters		
						k_{cat} (s^{-1})	K_m(µM)	k_{cat}/K_m ($s^{-1}.M^{-1}$)
acGK	ac-GKSFSKPR	acGK	GK	2	Fig. S3e,f	0.0017±0.0002	58±15	30±6
	ac-G[Orn]SFSKPR	acG[Orn]	GK	3	1.81, Fig. S3cd	0.0006±0.0001	22±3	29±3
	ac-G[Dab]SFSKPR	acG[Dab][c]	GK	4	-	NM	NM	<<0.01
	ac-GNCFSKPR	acGN	GK	5	Fig. S3a,b	NM	NM	<<0.01
	ac-GKSFAKPR	A6	GK	-		NM	NM	<<0.01
	ac-GKSFSAPR	A7	GK	-		NM	NM	<<0.01
	ac-GKVLSKIF	acARF6	ARF6	6	-	0.0019±0.0001	37±4	51±6
XK	AKQNSKLR	A-HPCA	HPCA	7	-	>0.0004	>1000	0.4
	AKVLSKIF	A-ARF6	ARF6	8	-	0.0005±0.0001*	8±3*	64±18
	AKSFSKPR	AK	GK	9	Fig.2b,c	0.0006±0.0002*	14±5*	40±11
	AKPTSKDSGLK	TSC2	TSC2	10	-	NM	NM	<<0.01
	SKSFSKPR	SK	GK	11	-	0.0025±0.0003	43±23	60±27
	MKSFSKPR	MK	GK	12	-	>0.0001	>100	1.7
	PKSFSKPR	PK	GK	13	-	0.00068±0.00005	26±8	26±6
	ac-KSFSKPR	acK	GK	14	1.7, Fig.4a,b	0.0009±0.0001	24±10	37±13
	ac-KFSKPR	acKΔS	GK	15	-	NM	NM	<<0.01
K	KSFSKPR	K	GK	16	-	0.00043±0.00004	16±6	27±6
	KQNSKLRP	K-HPCA	HPCA	17	-	0.0013±0.0002	133±49	9
	KSFS[Orn]PR		GK	18	-	0.0029±0.0002	59±15	49
	KSFS[Dab]PR		GK	19	-	0.0024±0.0009	145±109	17
	KSFS[Dap)PR		GK	20	-	0.0012±0.00003	39±4	31
	KSNAKLKP		NCS1	21	-	NM	NM	<<0.01
ZXGK	GGKSFSKP	GGK	GK	22/23	Fig.2	0.089±0.012	882±263	101±17
	AGKSFSKPR	AGK	GK	24	-	0.0024±0.0004	31±3	76±7
	ac-GGKSFSKPR	acGGK	GK	25	-	0.00062±0.00005	82±24	8±2
	ac-GGGKSFSKPR	acGGGK	GK	26	-	NM	NM	<<0.01
	ac-GGGGKSFSKPR		GK	27	-	NM	NM	<<0.01
	ac-GGGGGKSFSKPR		GK	28	-	NM	NM	<<0.01

MS spectra are displayed in **Data S1c**. The corresponding page number is indicated in the corresponding column. NM, not measurable; -, not measured or obtained. Residues colored in red are those modified in the context of a series (see the corresponding fourth column).

* only concentrations lower than 50 µM were taken into account as substrate concentration inhibition was observed at higher concentrations