



HAL
open science

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 Meinel

► To cite this version:

Frédéric Rivière, Cyril Dian, Rémi F. Dutheil, Paul Monassa, Carmela Giglione, et al.. Structural and large-scale analysis unveil the intertwined paths promoting NMT-catalyzed lysine and glycine myristoylation. *Journal of Molecular Biology*, 2022, 434 (22), pp.67843. 10.1016/j.jmb.2022.167843 . hal-03793088

HAL Id: hal-03793088

<https://hal.science/hal-03793088>

Submitted on 30 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3

4 Frédéric Rivière^{iD,a,1}, Cyril Dian^{iD,a}, Rémi F. Dutheil^{iD,a}, Paul Monassa^{iD,a}, Carmela
5 Giglione^{iD,@,a,*}, Thierry Meinnel^{iD,@,a,*}

6

7 ^a Université Paris Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC),
8 91198 Gif-sur-Yvette cedex, France

9

10 **Running title:** NMT-mediated lysine myristoylation rules

11

12 ***Correspondence:** carmela.giglione@i2bc.paris-saclay.fr (C. Giglione) or

13 thierry.meinnel@i2bc.paris-saclay.fr (T. Meinnel)

14

15 ^{iD} ORCID 0000-0001-6366-1904 (F. Rivière); ORCID iD 0002-6349-3901 (C. Dian); 0002-
16 2910-6648 (R.F. Dutheil); ORCID iD 0000-0003-4713-8069 (P. Monassa); 0002-7475-1558
17 (C. Giglione); 0001-5642-8637 (T. Meinnel)

18 ^{@twitter} @giglionelab (C. Giglione); @meinnel (T. Meinnel)

19

20 Present address

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

22

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

25 **ABSTRACT**

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

41

42 INTRODUCTION

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

52
53 Myristoylation, as with other lipid acylations, is notoriously difficult to systematically map *in*
54 *vivo* due to its hydrophobicity, poor antigenicity, low abundance, and subcellular localization
55 [1]. Tools have been difficult to develop for these types of modifications, hampering progress
56 in the field for many years. Considering that NMT activity and specificity are mainly
57 influenced by the N-termini of substrate proteins, a combination of structural, proteomic, and
58 genome-wide approaches using short peptides has delivered an exhaustive list of substrates
59 undergoing co-translational G-myristoylation. The so-called G-myristoylated proteome, or
60 myristoylome [5], covers approximately 2% of the human proteome (~600 proteoforms [17,
61 18]). Additionally, high-resolution co-crystallography of human NMT1 with a number of
62 reactive substrates displaying an N-terminal Gly [19] has established a specific water channel
63 in NMT that provides a water-mediated bond (Wat2) between the N-terminal group of the G-
64 starting substrate and the C-terminal carboxy group (Q496) of NMT, which acts as the catalytic
65 base for deprotonation and further reactivity [1, 19]. These data have been instrumental in

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

77

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

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

95

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

105

106 **RESULTS**

107 **Nomenclature**

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

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

116 There is now compelling evidence that myristoylation studied *in vitro* with octapeptides readily
117 translates to the *in vivo* context [1, 18] (**Fig. S1**). For instance, G-myristoylation of a full-length
118 protein, the small RAB G-protein, is catalyzed with similar efficiency as the N-terminal
119 octapeptide [27]. This approach perfectly reflects the (i) co-translational and N-terminal nature
120 of the modification taking place on short, emerging polypeptide chains [28]; and (ii) seven
121 dedicated binding clefts in NMT (pockets 2-8 in **Fig. 1a**) accommodating peptide substrate
122 side-chains [17]. These peculiarities allow the definition of the N-terminal G-myristoylated
123 proteome based on the dogma that NMT-dependent myristoylation only occurs on glycines
124 [17]. However, there is a recent report of possible NMT-dependent myristoylation on a lysine
125 at position 3 [19, 20].

126
127 We previously reported a model peptide with a Lys3 (G2KSFSKPR; reference, GK) as an
128 efficient NMT substrate (**Fig. 1b** and **Table 1**). In light of this new NMT activity (K-
129 myristoylation), we assessed which type of myristoylation (G- and/or K-) *Homo sapiens*
130 (Hs)NMT1 catalyzes on this peptide by MALDI-Tof-Tof including MS-MS (MS2). G-
131 myristoylation was favored in the context of this model peptide, as evidenced by the
132 identification of y7 ions characteristic of unmodified Lys3 only in MS2 spectra of the
133 myristoylated product (**Data S1a** and **Table 1**). We next wondered if all natural substrates from
134 human proteins starting with N-terminal GKs behave like the reference peptide. Therefore, we
135 examined the myristoylation status of those proteins, including the six closely related members
136 of the HPCA/NCS1 family and ARF6 (GKVLISKIF). In all cases, we only unambiguously
137 detected G-myristoylation (**Data S1** and **Table 1**), with catalytic efficiencies in the same range
138 as reported for the other sequences of the G-myristoylated proteome, as exemplified by the

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

147

148 A new crystal structure of HsNMT1 with a peptide corresponding to the N-terminus of HPCA
149 (G2KQNSKLR) confirmed G-myristoylation as the reaction product (**Fig. 1a,c,d**), as
150 previously also observed with NCS1 (G2KSNSKLR; [17]). The crystal structure revealed that
151 the Lys3 side chain buried tightly in pocket 3 of NMT and that its positive charge bonded the
152 catalytic base (**Fig. 1c**). Full occupation of pocket 3 with Lys3 slightly displaced the reactive
153 alpha amino group of Gly2 from the MyrCoA thioester group whilst retaining the Wat2
154 interaction of the reactive N-terminal α -amino group required for G-myristoylation [19] (**Fig.**
155 **1c,d**).

156

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

164 starting with G^[^K] [19]. We next assessed whether modifying the side-chain length but
165 retaining the basic character of the primary amino group at position 7 impacted the
166 myristoylation type. Shortening the basic side chain of residue 7 to ornithine (Orn, reduced by
167 one CH₂) or di-amino butyrate (Dab, reduced by two CH₂s) had little impact on G-
168 myristoylation (**Table S1**), suggesting maintenance of the salt bridge with the Ab-loop. Further
169 shortening to di-amino propionate (Dap, short basic chain CH₂NH₂) or removal of the basic
170 character with homocysteine (Hcy) significantly reduced catalytic efficiency but still promoted
171 only G-myristoylation (**Table S1** and **Data S1**).

172

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

188 is an isomer of A) led to very poor G-myristoylation efficiency (**Fig. 1b, Table 1**). Furthermore,
189 the crystal structure of HsNMT1 in complex with meGN revealed that the methyl rotated the
190 side chain of Asn3 around C α 3, which was orientated towards the catalytic base. Conversely,
191 the methylated N-terminus was positioned in pocket 3 (**Fig. S2b,c**), explaining the low
192 performance of this substrate for G-myristoylation. Therefore, any subtle modification around
193 Gly2 protrudes aa2 into pocket 3, resulting in aa3 oriented in pocket 2 at the catalytic center
194 [19].

195

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

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

214

215 **Lys3 myristoylation of N-acetylated peptides with the S6K7 motif**

216 In apparent contradiction to the above results, the GK-starting protein ARF6 was recently
217 shown to accept K-myristoylation *in vivo* [20] if the Gly2 α -amino group was blocked due to
218 previous N-acetylation (*ac*) or G-myristoylation. We therefore examined any modification
219 around the reactive G2 amino group allowing K-myristoylation of the ARF6 and/or GK-
220 starting peptides. As expected, *ac* of the two G-starting peptides led to Lys3 side chain
221 myristoylation (**Fig. 2a, Table 2** and **Data S1**). When the Lys side chain was shortened to Orn,
222 in the context of N-acetylated peptides, side chain myristoylation was still observed at position
223 3 (**Fig. 2a**). The catalytic efficiencies of the *ac* series were two orders of magnitude lower than
224 those of G-myristoylation (**Tables 1-2**), with the catalytic constant (k_{cat}) accounting for most
225 of the k_{cat}/K_m decrease in both cases. This suggested that the affinity provided by the peptide
226 core was similar but that the catalysis had a limiting step.

227

228 The crystal structures of HsNMT1 in complex with acG[Orn] or an acGN non-reactive
229 derivative were obtained and compared with the acGK variant. In both cases, there was rotation
230 around the C α of the aa3 side chain (C α 3), as expected from the *ac* addition. The acG group
231 moved to pocket 3, and the aa3 side chain contacted the catalytic base (**Fig. S3a-d**). The Lys
232 side chain of acGK bonded directly with the C-terminal catalytic base of NMT, unlike the G-
233 myristoylation mechanism, which involves a water-mediated bond (**Fig. 1c,d**). acG[Orn] led
234 to a weaker salt bridge with the catalytic base than acGK as the distance was increased from
235 3.1 to 3.6 Å (**Fig. S3e**). Nevertheless, the associated side chain myristoylation mechanism still

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

247

248 **Side-chain myristoylation at Lys3 is preferred over N-terminal α -myristoylation when** 249 **the N-terminal residue is not Gly**

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

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

270

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

279

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

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

295

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

303

304 **Mining the human proteome reveals that K-myristoylation may solely arise from a post-** 305 **translational event**

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

309 sequences with translation start sites at positions 1 or 2 (i.e., arising from cotranslational N-
310 Met removal rules), 10 of the 15 known G-myristoylated N-termini starting with GK were
311 found [17] plus ten additional sequences, most not comprising Phe or Leu4. As a result, this
312 second small subset featuring proteins starting with Ala (TSC2), Phe (EVPL), Lys (ANR26,
313 FA71D), Met (CHD8, GXLT2, SIM1/2), Arg (RAB20), and Thr (LRC59) was very unlikely
314 to undergo K-myristoylation. This was demonstrated for TSC2, which displays the most
315 favorable context with its N-terminal Ala, the smallest residue of the subset (**Table 2**). In the
316 G-starting subset undergoing G-myristoylation, only ARF6 displayed the important
317 hydrophobic position at position 4, making it the best target for dual G- and K-myristoylation.
318 The other 3595 sequences corresponded to internal sequences starting beyond codon 3, and
319 285 were preceded by a Gly, suggesting large scope for post-translational G-myristoylation in
320 this subset. Post-translational myristoylation, however, requires prior unmasking by proteolytic
321 cleavage provided before the first XK. Indeed, a number of K-myristoylation sites among the
322 543 [^G]KX[FL]SK internal sequences were preceded by a loose motif recapitulating known
323 G-myristoylated caspase-induced substrates ([1]; **Data S2b-c**).

324

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

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

336

337 Of all the known myristoylated proteins resulting from post-translational myristoylation, the
338 only GK-starting G-myristoylated protein was RNA-binding protein FUS, which displays a
339 non-canonical caspase 3 motif (DWFDGK). When extracting a subset with the motif including
340 a Phe at position 2 behind GK, none of the 5 retrieved proteins - including FUS - displayed the
341 optimal motif for K-myristoylation (**Data S3e**). A similar survey performed with the caspase
342 6 motif ([VLIT][EDQ][^DENQRKAPGS]D[ASGMP]) - which retrieved a couple of
343 additional G-myristoylated post-translational proteins including the pro-apoptotic protein BID
344 and the interleukin enhancer-binding factor 3 (ILF3) - led to the same conclusion (**Data S4**).

345

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

351

352 **Myristoylation may occur on Lys4**

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

358 (Table 2,e, Supplementary Fig 4, and Supplementary Figure 5a). K-myristoylation was as
359 efficient with these A/G-grafted peptides as with the acG peptides (Fig. 2a and Table 2). We
360 concluded that the pocket usually hosting residue 3 could host a dipeptide moiety as large as
361 AG, provided that a close and available Lys-reactive side chain was properly spaced from the
362 crucial S6K7 motif. We noted that, with five bonds and a positive charge, a Lys side chain was
363 the most bulky moiety tolerated in NMT pocket 3 for G-myristoylation. With seven bonds, the
364 AG backbone was much bulkier than K but still promoted K-myristoylation (Fig. 2f). To
365 establish whether larger chains would make K-myristoylation possible, we examined large,
366 poorly-branched N-terminal grafts larger than GG or AG (AcGGK, AcGGGK, AcGGGGK,
367 and AcGGGGGK with 9, 12, 15, and 18 bonds, respectively) upstream of Lys3 in the GK
368 series. With AcGGK and AcGGGK, there was unequivocal K-myristoylation by MALDI
369 analysis (Data S1). K-myristoylation catalytic efficiency progressively decreased with
370 increasing chain length to very low values, and K-myristoylation was not detected for
371 AcGGGGK or AcGGGGGK (Table 2). The crystal structure obtained with the GGK derivative
372 showed that the GG chain was fully buried in pocket 3 and propelled the main backbone peptide
373 chain into a more remote location at C α 3 (Fig. S5a,b); there was also a greater distance between
374 the reactive K amino group and both the catalytic base and T282. The 7-bond GG or AG chain
375 corresponded to the bulkiest chains accepted by pocket 3 for efficient K-myristoylation (Fig.
376 2f).

377

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

383 recognition particle complex, SRP54, displayed a putative K-myristoylation site upon caspase
384 3 cleavage (**Data S3h**).

385

386 With the knowledge that Lys4 may be reactive in the context of a G2XK4 derivative, we next
387 assessed whether Gly2- and Lys4-myristoylation could compete within the same protein. For
388 this, we first altered the G-myristoylation S6K7 motif of the GK peptide by introducing an
389 additional Ser between Ser6 and Lys7, pushing the Lys back to position 8
390 (G2G3K4F5S6S7K8PR). The resulting peptide therefore featured two overlapping
391 myristoylation motifs separated by a two amino acid spacer, one for modification of the α -
392 group of G2, relying on motif S6S7, and a second on the ϵ -amino of Lys4, relying on motif
393 S7K8. This modification reduced the catalytic efficiency compared with the GK peptide (**Table**
394 **1**). Furthermore, the MS2 spectra of the 1173 Da myristoylated product evidenced only a major
395 G-myristoylation product, as unambiguous characteristic b2, y7, and y8 ions were identified
396 (**Fig. S6a,b; Data S1**). Nevertheless, we could not exclude the possibility of K-myristoylation,
397 as associated prototypic K-myristoylation ions could be identified at intensities lower than the
398 threshold. Therefore, other experimental approaches were required to assess for simultaneous
399 K-myristoylation.

400

401 **The IpaJ protease as a tool to unravel ϵ -myristoylation**

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

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

410

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

417

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

428

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

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

436

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

446

447 **Competition between G- and K-myristoylation**

448 With the IpaJ/NMT pipeline in hand, we next examined the ambiguous myristoylation state of
449 the GGKFSSKPR peptide. At T1, MS2 analysis unambiguously revealed G-myristoylation as
450 a product (**Fig. S6b**). At T2 (i.e., after incubation with NMT and IpaJ), both an uncleaved
451 myristoylated peptide and a shorter, 267 Da peptide resulting from G-myristoylation cleavage
452 were identified (T2, 1173 and 906 Da ions **Fig. S6c**). There was also unambiguous and unique
453 K-myristoylation at Lys4 of the GGKFSSKPR peptide at T3 (1059 Da ion, **Extended Data**
454 **4c,d**). IpaJ, therefore, decomplexed the 1173 Da ion observed at T1 and evidenced K-
455 myristoylation within this ion, which was otherwise only barely visible (**Fig. S6b, Data S1a**).

456

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

473

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

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

489

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

494

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

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

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

521

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

527

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

535

536

537 **DISCUSSION**

538 N-myristoyltransferases (NMT), also known as glycylopeptide N-tetradecanoyltransferases (EC
539 2.3.1.97; see BRENDA resource [33]), are the only enzymes known to catalyze N-
540 myristoylation in eukaryotes [5]. This modification mostly occurs co-translationally and
541 involves the transfer of a lipid (myristate) to the α -amino group of the N-terminal amino group
542 of G-starting proteins exposed mainly by the prior action of methionine aminopeptidases
543 (MetAP) at the ribosome. NMT, if unbound to the ribosome, can also act post-translationally
544 on proteolytically-processed folded proteins. For a long time, all known post-translational
545 targets of NMTs were associated with proteolytic cleavages induced by caspases during
546 apoptosis [16, 34-39]. Both human NMTs (NMT1 and NMT2) undergo caspase-3 cleavage,
547 which removes their N-terminal ribosome binding motif [40], decreasing co-translational
548 activity and strengthening their post-translational properties. To date, caspases - mostly caspase
549 3 and to a lesser extent caspase 6 [18] - are the only proteases known to promote post-
550 translational myristoylation. This apoptosis-induced post-translational G-myristoylation subset
551 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].

577

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

586

587 Here, we reveal the unique and common specificities underlying NMT-dependent G- and K-
588 myristoylation. We found that the main specificity rules for NMT-dependent G- and K-
589 myristoylation partially overlap, leading to competition between the two types of
590 myristoylation in GK-starting contexts. The myristoylation rules are summarized in **Figs. 1e**
591 **and 2f**. The very narrow shape of pocket 2 ensures reactivity of a primary amino borne only on
592 unbranched moieties such as N-terminal glycine or side chains like lysine, which facilitates the
593 70° rotation of the amino-reactive group from the catalytic base towards the thio-reacting group
594 of MyrCoA. This motion allows Wat3 to enter from the water channel and play a crucial role
595 at the transition state [19]. The addition of a methyl group such as in Ala2 or meGly variants
596 makes this motion unlikely and dramatically decreases catalysis. The rules directing G or side
597 chain myristoylation are predominantly determined by the R₁ and R₂ chemical groups branched
598 at C α 3. NMT usually favors G- over K-myristoylation on proteins starting with Gly, while K-
599 myristoylation dominates on any other N-terminal chain.

600

601 Overall, NMT-dependent K-myristoylation is less efficient than G-myristoylation whatever the
602 context. Although the two myristoylation types may directly act on an N-terminal residue, K-
603 myristoylation appears less strict than G-myristoylation in this regard; indeed, K-
604 myristoylation does not require a given side chain of the first amino acid but relies on optimized
605 residues at positions 5-7. A major difference between the two myristoylation types is that K-
606 myristoylation needs the KXXSK-restricted pattern. This is unlike G-myristoylation, which
607 allows the modification of many proteins not exhibiting the S6K7 motif by NMT on their N-
608 terminal G (see [17] and **Table 1**). The KXXSK motif of K-myristoylation can be associated
609 with a one amino acid extension or acetylation at the N-terminus provided that an *ad hoc*
610 proteolytic machinery can produce these extremities. The specificity of MetAPs - which
611 remove the first Met from proteins - favors new N-terminal residues with short side chains like
612 Ala, Gly, Pro, or Ser. However, a survey of the human proteome reveals no K-myristoylation
613 on Lys2, Lys3, or Lys4 and 15 G-myristoylation substrates in the subset of proteins arising
614 from MetAP-guided co-translational cleavage rules. Among MetAP unprocessed substrates,
615 four MK-starting proteins were identified, but we ruled out that they might undergo K-
616 myristoylation. A very reduced number of internal KXXSK K-myristoylation motifs are
617 preceded by caspase cleavage sites in the human proteome, suggesting that K-myristoylation -
618 in contrast to mostly co-translational G-myristoylation – might not result from primary post-
619 translational addition, for instance initiated by cleavage by specific caspases during apoptosis
620 [16, 34]. K-myristoylation motifs resulting from cotranslational Met removal or
621 posttranslational caspase cleavage are therefore apparently significantly depleted from the
622 human proteome. This is unlike G-myristoylation motifs, which are enriched and represent 2%
623 of the proteome [18].

624

625 Our current data suggest that post-translational myristoylation might occur after initial G-
626 myristoylation, possibly for few additional proteins such as HPCA exhibiting the G2K3 motif,
627 in addition to ARF6, causing double myristoylation. Alternatively, it might occur upon
628 proteolytic cleavage caused by a protease with a specificity different from MetAP or any
629 caspase. While such proteases in the human proteome have yet to be discovered, we show here
630 that the bacterial protease IpaJ - a virulent factor expressed in human cells during *Shigella*
631 infection - might correspond to such a protease. In this report, we used IpaJ both as a case study
632 and a unique tool to discriminate between G- and K-myristoylation. Our data confirm that
633 ARF6 is myristoylated before and after IpaJ cleavage [19, 20], in contrast to the other ARFs
634 (ARF1/3/4/5) or ARL1, which do not feature a K3 like ARF6. The insensitivity of ARF6 – but
635 not the other ARFs - to IpaJ was initially proposed to be due to ARF6 not being located at the
636 ER/Golgi, unlike the other members of the ARF/ARL family, but rather at the plasma
637 membrane. ARF6 translocates to endosomes, where it acts as a major regulator of endocytosis
638 [42, 43]. Our data indicate that ARF6 resistance to IpaJ might also result from the unmasking
639 of a myristoylatable Lys3 upon G-myristoylation cleavage. As a result, we surmise that ARF6
640 would still be myristoylated and retained on myristoylation affinity azido-biotin prior to and
641 after IpaJ cleavage. IpaJ would therefore not act as a myristoyl eraser of G-myristoylated
642 proteins displaying a Lys3. Indeed, NMT would be still able of adding a myristoyl group on
643 the side-chain of the neo N-terminal K.

644

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

649

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

659 MATERIALS AND METHODS

660 Chemicals

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

666 *IpaJ* cloning

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

680 Enzyme production and purification

681 *HsNMT1*. The *HsNMT1* isoform containing residues 81-496 corresponds to the major isoform
682 [3]. *HsNMT1* 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 HsNMT1s
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 OD₆₀₀=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-HisTrapTM FF 5 mL column (GE Healthcare, Chicago, IL) at 3.5 mL/min; elution
696 was achieved with a linear 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.* HsNMT1s 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

707 by co-crystallization using the hanging-drop vapor diffusion method at 20°C in the
708 crystallization conditions previously described [17, 19]. Briefly, crystallization droplets were
709 formed by mixing 2 μ L of the of HsNMT1:MyrCoA:peptide complex (ratio 1:1.5:1.5) at 6-9
710 mg/mL with 2 μ L of the precipitant solution containing either 0.1 M MgCl₂, 0.2 M NaCl, 0.1
711 M sodium citrate pH 5.6, and 18-24% (w/v) PEG 6K or 0.1 M sodium acetate pH 4.6, and 18-
712 24% (w/v) PEG 6K. Crystals were cryoprotected in the reservoir solution supplemented with
713 15% (v/v) glycerol and flash cooled in liquid nitrogen. Complete X-ray datasets of complex
714 were collected at 100K a single wavelength from a single crystal at the French National
715 Synchrotron Facility (SOLEIL) PROXIMA1 (for HPCA and AcG[Orn] at $\lambda=0.98400$ Å and
716 AcK at $\lambda=0.97856$ Å) or PROXIMA2 (for AK and GGK at $\lambda= 0.98012$ Å) beamlines and at
717 European Radiation Synchrotron Facility (ESRF) ID30A1 (for AN at $\lambda=0.96600$ Å) and
718 ID30A3 (for me-GN $\lambda=0.96775$ Å) beamlines. Datasets were integrated with XDS [44] and
719 scaled and reduced using AIMLESS from the CCP4 package [45]. For crystals that suffered
720 from anisotropic diffraction (HPCA, AcK and GGK), data were processed with STARANISO
721 on unmerged data [46] before AIMLESS data reduction. Crystals belonged essentially to the
722 space group P2₁2₁2 with similar unit cell parameters with the exception of GGK and meGN
723 belonging both to C2 with similar unit cell parameters (summarized in **Table S5**). In both space
724 groups identified, unit cells contained two NMT molecules per asymmetric unit. Structure
725 resolution was accomplished in all cases using the molecular replacement method and solved
726 using PHASER [47] and the HsNMT1:MyrCoA:peptide ternary complex (PDB entry 5O9T or
727 6SK2) as a search model. The structure of AN was solved using MOLREP [48] and protein
728 coordinates of HsNMT1:MyrCoA (PDB entry 5O9T) as a search model. Structures were
729 subjected to alternating refinement cycles using PHENIX and manual model building using
730 COOT [49-51]. The good quality of the electron density maps also enabled the refinement of

731 substrate peptide, reaction intermediate, and reaction product molecules bound to HsNMT1 in
732 each complex. Chemical compound libraries were generated using PRODRG server [52] in
733 combination with eLBOW from the PHENIX suite. The geometry of the final models was
734 validated using MOLPROBITY [53]. Figures were generated using PYMOL (DeLano
735 Scientific LLC, <http://pymol.sourceforge.net/>). X-ray data collection and refinement statistics
736 are summarized in **Table S5**. Omit electron density maps were calculated using PolderMap
737 [54] from the Phenix suite after omitting the peptide of each active site and excluding the bulk
738 solvent around the omitted region. Omit maps of the nine crystal structures are displayed in
739 **Fig. S12**.

740 **Measurements of activity and associated parameters**

741 HsNMT1 activity was assayed at 30°C in a coupled assay using an updated version of the
742 previously described protocol [27]. The reaction mixture contained 50 mM Tris-HCl (pH 8.0),
743 1 mM MgCl₂, 0.193 mM EGTA, 0.32 mM DTT, 0.2 mM thiamine pyrophosphate, 2 mM
744 pyruvate, 0.1 mg/mL of BSA, 0.1% Triton X-100, 2.5 mM NAD⁺, 0.125 units/mL of porcine
745 heart PDH (0.33 units/mg), 40 μM myristoyl-CoA, and 1-2000 μM peptides. Unless otherwise
746 stated for tight binding studies, the reaction mixture was pre-incubated for 3 min at 30°C before
747 starting the reaction by adding MyrCoA. A final volume of 200 μL was used in 96-well black
748 plates (Grenier Bio One and Dominique Dutscher, Brumath, France; the optical path for 0.2
749 mL is 0.55 cm). A value of 6300 M⁻¹.s⁻¹ was used as the molecular extinction coefficient of
750 NADH at 340 nm. An Infinite M Nano⁺ plate reader equipped with micro-injectors (Tecan,
751 Lyon, France) was set at 340 nm to monitor the absorbance over time at 30°C. Briefly, a
752 reaction mixture containing HsNMT1/ at different concentrations of peptide acceptors was pre-
753 incubated for 3 min at 30°C before starting the reaction with MyrCoA.

754 Myristoylation kinetics were monitored continuously for 30 min, and the data were fitted as to
755 obtain the initial velocities associated to each peptide concentration. Curve fits to obtain kinetic
756 parameters were achieved by non-linear regression with GraphPad Prism 9.1 (GraphPad
757 Software, La Jolla, CA). Parameters with standard errors were computed for all parameters
758 using the complete dataset including replicates. Both k_{cat} and K_m kinetic parameters were
759 obtained by fitting to the Michaelis–Menten equation. k_{cat}/K_m values and the associated
760 standard deviations were obtained by taking advantage of the k_{SP} approach [55] with
761 $k_{SP}=k_{cat}/K_m$ and $v_0/[E]=k_{SP}[S]/(1+[S]/K_m)$, where v_0 is the reaction rate measured at NMT
762 concentration $[E_0]$.

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

770

771 **Mass spectrometry**

772 300 μ L of a mixture containing 50 mM Tris (pH 8), 0.193 mM EGTA, 1 mM $MgCl_2$, 1 mM
773 DTT, 5 μ M sodium cholate, 40 μ M Myr-CoA solution (stock solution 0.2 mM Myr-CoA, 10
774 mM sodium acetate, 2.5 μ M sodium cholate), 0.5 μ M NMT, and 100 μ M of synthetic peptide
775 were pre-incubated at 30°C. The myristoylation reaction was followed over time by collection
776 of 10 μ L samples further diluted in 90 μ L of water/acetonitrile (90/10) solution. The different
777 samples were then diluted five times in the matrix solution made of 5 mg/mL of α -cyano-4-
778 hydroxycinnamic acid solubilized in water/formic acid/acetonitrile (50/50/0.1%). 1 μ L of each
779 dilution was spotted on a metal target and dried. MS and MS/MS spectra of each sample were
780 acquired with an AB SCIEX 5800 MALDI-Tof-Tof instrument in positive ion mode. Survey
781 scans were performed using delayed extraction (390 ns) in reflector mode for a total of 15,000
782 shots. MS/MS scans were operated with a collision energy of 1 kV. Peptide and fragment mass
783 tolerances were set at 10 ppm and 0.8 Da, respectively. Mass spectra were analyzed with
784 PeakView® 2.2 software (AB Sciex, Macclesfield, UK). The default threshold in MS/MS peak
785 labelling and finding was 5% (displayed with a black dot on the intensity axis) and centroid
786 height percentage was 50% as recommended by the constructor. MS/MS deviations from
787 theoretical values were on average less than 0.03 Da. MS/MS analysis and assignments of
788 peptides with non-natural aminoacids was also performed with the additional help of software
789 ProteinProspector v6.2.2 (<https://prospector.ucsf.edu/prospector/mshome.htm>) [56].
790 Theoretical mass values of all fragments were double-checked at
791 https://web.expasy.org/peptide_mass/. All spectra are available in **Data S1**.

792

793 **ACCESSION NUMBERS**

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

797

798 **ACKNOWLEDGMENTS**

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

804

805 **FUNDING**

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

815

816 **AUTHOR CONTRIBUTIONS**

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

822

823 **COMPETING INTERESTS**

824 The authors declare that they have no competing interests.

825

826 **SUPPLEMENTAL INFORMATION**

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

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

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

833

834 **REFERENCES**

835 [1] Meinnel T, Dian C, Giglione C. Myristoylation, an ancient protein modification mirroring
836 eukaryogenesis and evolution. *Trends Biochem Sci.* 2020;45:619-32.

837 [2] Lodge JK, Johnson RL, Weinberg RA, Gordon JI. Comparison of myristoyl-CoA:protein
838 N-myristoyltransferases from three pathogenic fungi: *Cryptococcus neoformans*, *Histoplasma*
839 *capsulatum*, and *Candida albicans*. *J Biol Chem*. 1994;269:2996-3009.

840 [3] Pierre M, Traverso JA, Boisson B, Domenichini S, Bouchez D, Giglione C, et al. N-
841 Myristoylation regulates the SnRK1 pathway in *Arabidopsis*. *Plant Cell*. 2007;19:2804-21.

842 [4] Price HP, Menon MR, Panethymitaki C, Goulding D, McKean PG, Smith DF. Myristoyl-
843 CoA:protein N-myristoyltransferase, an essential enzyme and potential drug target in
844 kinetoplastid parasites. *J Biol Chem*. 2003;278:7206-14.

845 [5] Giglione C, Meinnel T. Mapping the myristoylome through a complete understanding of
846 protein myristoylation biochemistry. *Prog Lipid Res*. 2022;85:101139.

847 [6] Bhatnagar RS, Ashrafi K, Futterer K, Waksman G, Gordon JI. Biology and enzymology of
848 protein N-myristoylation. In: Tamanoi F, Sigman DS, editors. *The Enzymes*. San Diego:
849 Academic Press; 2001. p. 241-86.

850 [7] Resh MD. Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat Chem*
851 *Biol*. 2006;2:584-90.

852 [8] Traverso JA, Micalella C, Martinez A, Brown SC, Satiat-Jeunemaitre B, Meinnel T, et al.
853 Roles of N-Terminal fatty acid acylations in membrane compartment partitioning: *Arabidopsis*
854 h-type thioredoxins as a case study. *Plant Cell*. 2013;25:1056-77.

855 [9] Frearson JA, Brand S, McElroy SP, Cleghorn LA, Smid O, Stojanovski L, et al. N-
856 myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature*. 2010;464:728-
857 32.

858 [10] Wright MH, Clough B, Rackham MD, Rangachari K, Brannigan JA, Grainger M, et al.
859 Validation of N-myristoyltransferase as an antimalarial drug target using an integrated
860 chemical biology approach. *Nat Chem*. 2014;6:112-21.

861 [11] Wright MH, Paape D, Price HP, Smith DF, Tate EW. Global profiling and inhibition of
862 protein lipidation in vector and host stages of the sleeping sickness parasite *Trypanosoma*
863 *brucei*. *ACS Infect Dis*. 2016;2:427-41.

864 [12] Mousnier A, Bell AS, Swieboda DP, Morales-Sanfrutos J, Perez-Dorado I, Brannigan JA,
865 et al. Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and
866 replication of the common cold virus. *Nat Chem*. 2018;10:599-606.

867 [13] Beauchamp E, Yap MC, Iyer A, Perinpanayagam MA, Gamma JM, Vincent KM, et al.
868 Targeting N-myristoylation for therapy of B-cell lymphomas. *Nat Commun*. 2020;11:5348.

869 [14] Kosciuk T, Lin H. N-Myristoyltransferase as a glycine and lysine myristoyltransferase in
870 cancer, immunity, and infections. *ACS chemical biology*. 2020;15:1747-58.

871 [15] Martin DD, Beauchamp E, Berthiaume LG. Post-translational myristoylation: Fat matters
872 in cellular life and death. *Biochimie*. 2011;93:18-31.

873 [16] Thinon E, Serwa RA, Broncel M, Brannigan JA, Brassat U, Wright MH, et al. Global
874 profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat*
875 *Commun*. 2014;5:4919.

876 [17] Castrec B, Dian C, Ciccone S, Ebert CL, Bienvenut WV, Le Caer J-P, et al. Structural and
877 genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern.
878 *Nat Chem Biol*. 2018;14:671-9.

879 [18] Meinnel T. Comment on “Binding affinity determines substrate specificity and enables
880 discovery of substrates for N-Myristoyltransferases”. *ACS Catal*. 2022;12:8195-201.

881 [19] Dian C, Pérez-Dorado I, Rivière F, Asensio T, Legrand P, Ritzefeld M, et al. High-
882 resolution snapshots of human N-myristoyltransferase in action illuminate a mechanism
883 promoting N-terminal Lys and Gly myristoylation. *Nat Commun*. 2020;11:1132.

884 [20] Kosciuk T, Price IR, Zhang X, Zhu C, Johnson KN, Zhang S, et al. NMT1 and NMT2 are
885 lysine myristoyltransferases regulating the ARF6 GTPase cycle. *Nat Commun*. 2020;11:1067.

- 886 [21] Carr SA, Biemann K, Shoji S, Parmelee DC, Titani K. n-Tetradecanoyl is the NH₂-
887 terminal blocking group of the catalytic subunit of cyclic AMP-dependent protein kinase from
888 bovine cardiac muscle. *Proc Natl Acad Sci U S A*. 1982;79:6128-31.
- 889 [22] Gheorghe MT, Bergman T. Deacetylation and internal cleavage of polypeptides for N-
890 Terminal sequence analysis. In: Atassi MZ, Appella E, editors. *Methods in Protein Structure*
891 *Analysis*. Boston, MA: Springer; 1995. p. 81-6.
- 892 [23] Leone JW, Hampton B, Fowler E, Moyer M, Krishna RG, Chin CCQ. Removal of N-
893 terminal blocking groups from proteins. *Current protocols in protein science*. 2011;63:11.7.1-
894 .7.20.
- 895 [24] Mattock E, Blocker AJ. How do the virulence factors of *Shigella* work together to cause
896 disease? *Front Cell Infect Microbiol*. 2017;7:64.
- 897 [25] Burnaevskiy N, Peng T, Reddick LE, Hang HC, Alto NM. Myristoylome profiling reveals
898 a concerted mechanism of ARF GTPase deacylation by the bacterial protease IpaJ. *Mol Cell*.
899 2015;58:110-22.
- 900 [26] Burnaevskiy N, Fox TG, Plymire DA, Ertelt JM, Weigele BA, Selyunin AS, et al.
901 Proteolytic elimination of N-myristoyl modifications by the *Shigella* virulence factor IpaJ.
902 *Nature*. 2013;496:106-9.
- 903 [27] Boisson B, Meinnel T. A continuous assay of myristoyl-CoA:protein N-
904 myristoyltransferase for proteomic analysis. *Anal Biochem*. 2003;322:116-23.
- 905 [28] Giglione C, Fieulaine S, Meinnel T. N-terminal protein modifications: bringing back into
906 play the ribosome. *Biochimie*. 2015;114:134-46.
- 907 [29] Padovani D, Zeghouf M, Traverso JA, Giglione C, Cherfils J. High yield production of
908 myristoylated Arf6 small GTPase by recombinant N-myristoyl transferase. *Small GTPases*.
909 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.

934 [38] Tapodi A, Clemens DM, Uwineza A, Jarrin M, Goldberg MW, Thinon E, et al. BFSP1 C-
935 terminal domains released by post-translational processing events can alter significantly the
936 calcium regulation of AQP0 water permeability. *Exp Eye Res.* 2019;185:107585.

937 [39] Martin DD, Ahpin CY, Heit RJ, Perinpanayagam MA, Yap MC, Veldhoen RA, et al.
938 Tandem reporter assay for myristoylated proteins post-translationally (TRAMPP) identifies
939 novel substrates for post-translational myristoylation: PKCepsilon, a case study. *FASEB J.*
940 2012;26:13-28.

941 [40] Perinpanayagam MA, Beauchamp E, Martin DD, Sim JY, Yap MC, Berthiaume LG.
942 Regulation of co- and post-translational myristoylation of proteins during apoptosis: interplay
943 of N-myristoyltransferases and caspases. *FASEB J.* 2013;27:811-21.

944 [41] Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, et al. Dual
945 proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell.*
946 2021;184:3022-40.e28.

947 [42] Donaldson JG. Multiple roles for Arf6: sorting, structuring, and signaling at the plasma
948 membrane. *J Biol Chem.* 2003;278:41573-6.

949 [43] Donaldson JG, Jackson CL. ARF family G proteins and their regulators: roles in
950 membrane transport, development and disease. *Nat Rev Mol Cell Biol.* 2011;12:362-75.

951 [44] Kabsch W. Automatic processing of rotation diffraction data from crystals of initially
952 unknown symmetry and cell constants. *J Appl Cryst.* 1993;26:795-800.

953 [45] Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta*
954 *crystallographica Section D, Biological crystallography.* 2013;69:1204-14.

955 [46] Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, et al. STARANISO
956 (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>). Cambridge, United Kingdom:
957 Global Phasing Ltd. ; 2018.

958 [47] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser
959 crystallographic software. *J Appl Crystallogr.* 2007;40:658-74.

960 [48] Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J*
961 *Appl Cryst.* 1997;30:1022-225.

962 [49] Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al.
963 REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica*
964 *Section D, Biological crystallography.* 2011;67:355-67.

965 [50] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a
966 comprehensive Python-based system for macromolecular structure solution. *Acta*
967 *crystallographica Section D, Biological crystallography.* 2010;66:213-21.

968 [51] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta*
969 *crystallographica Section D, Biological crystallography.* 2010;66:486-501.

970 [52] Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography
971 of protein-ligand complexes. *Acta crystallographica Section D, Biological crystallography.*
972 2004;60:1355-63.

973 [53] Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al.
974 MolProbity: all-atom structure validation for macromolecular crystallography. *Acta*
975 *crystallographica Section D, Biological crystallography.* 2010;66:12-21.

976 [54] Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, et al.
977 Polder maps: improving OMIT maps by excluding bulk solvent. *Acta crystallographica Section*
978 *D, Structural biology.* 2017;73:148-57.

979 [55] Johnson KA. New standards for collecting and fitting steady state kinetic data. *Beilstein*
980 *journal of organic chemistry.* 2019;15:16-29.

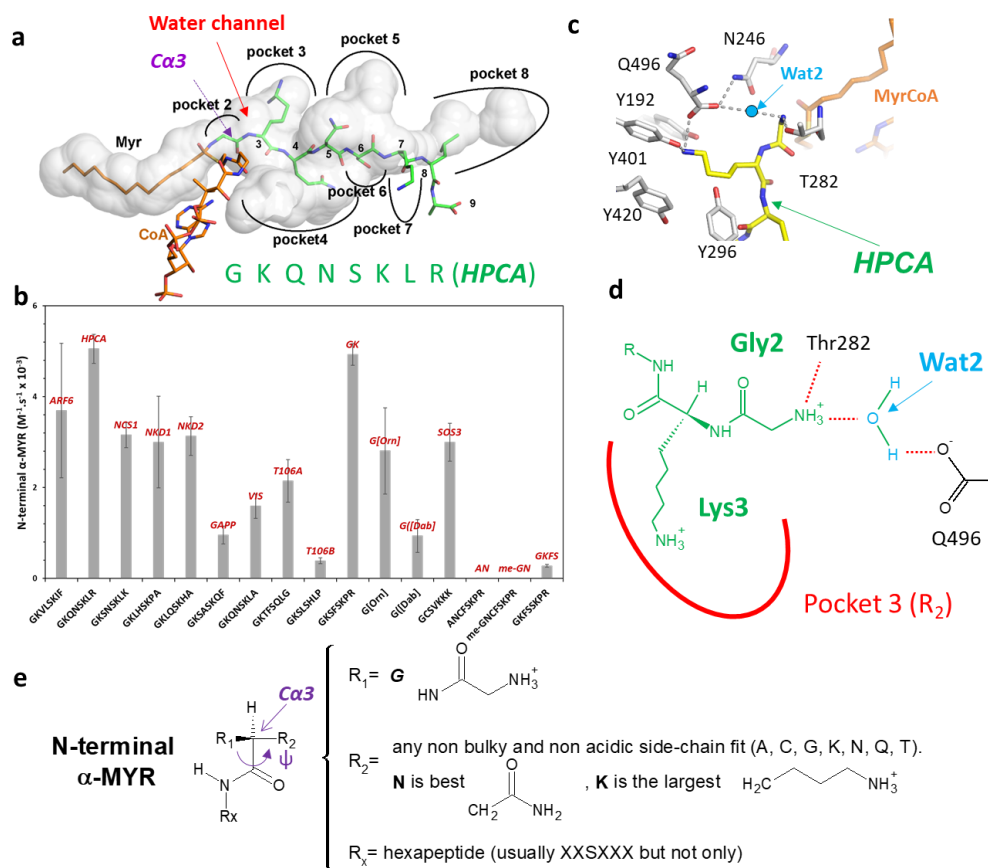
981 [56] Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth analysis
982 of tandem mass spectrometry data from disparate instrument types. Mol Cell Proteomics.
983 2008;7:2386-98.

984

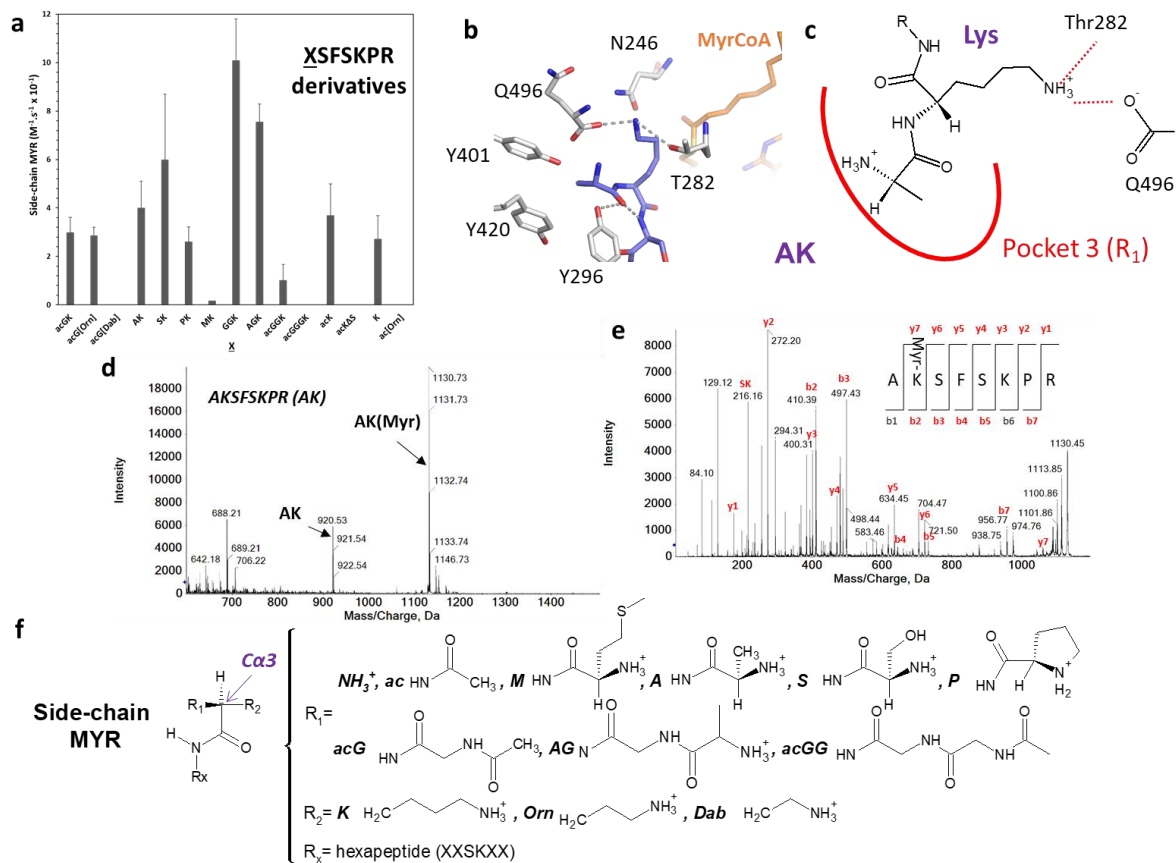
985 **FIGURES AND TABLES**

986 See next pages

987



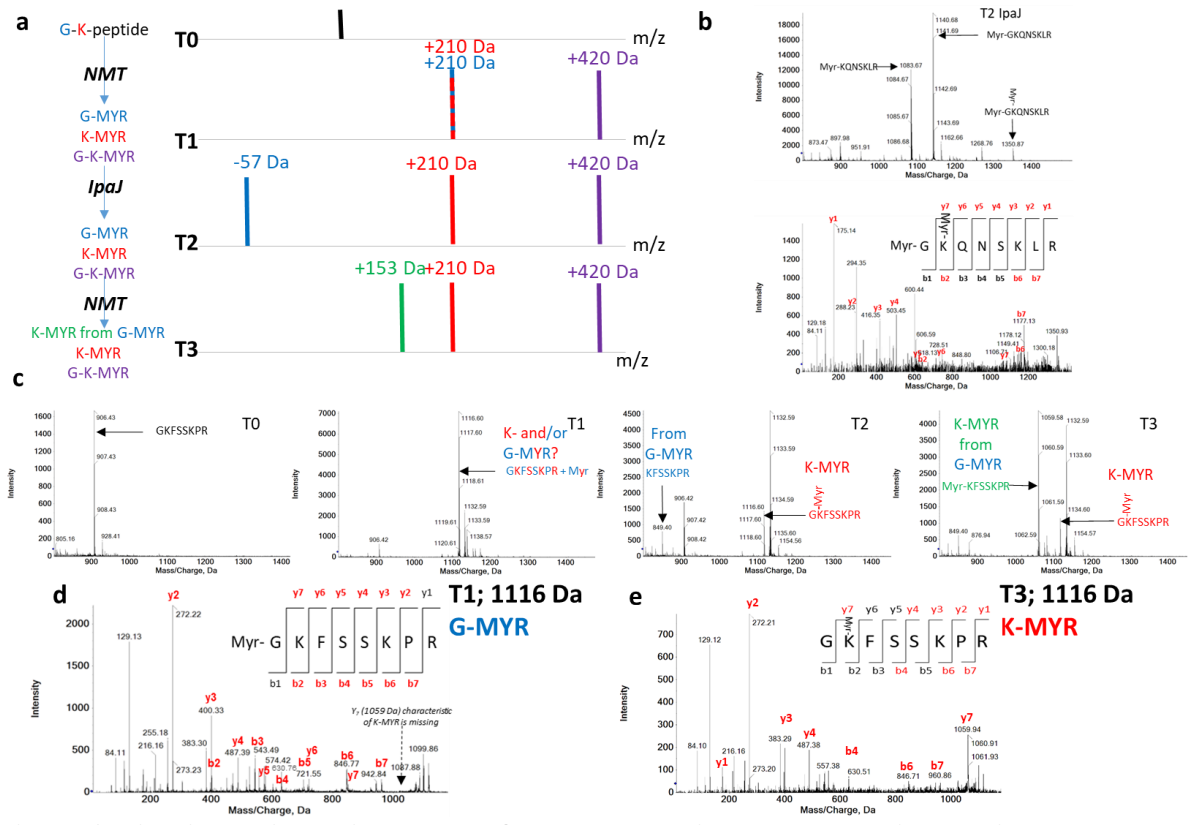
989
 990 All peptides are derived from the GK peptide unless otherwise stated (see short name and the
 991 associated sequence in **Table 1**). **(a)** Crystal structure of the HPCA peptide with HsNMT; the
 992 various pockets and their extents are displayed and $\text{Ca}3$ is indicated. The peptide sequence of
 993 HPCA is indicated in green. **(b)** Catalytic efficiencies (k_{cat}/K_m) values of GK-starting peptides
 994 (see all parameters in **Table 1**). **(c)** 3D details of the active site organization of the HPCA
 995 complex. The myristate moiety of MyrCoA is shown in orange, the HPCA backbone peptide
 996 is colored in yellow and indicated with a green arrow, residues of NMT in the immediate
 997 vicinity are in displayed grey, and the water 2 molecule (Wat2) is shown as a turquoise sphere.
 998 **(d)** Planar representation of panel c, R₂ refers to the group defined in panel e while pocket 3 is
 999 displayed in panel a. The color code is the same as in panel c. **(e)** Types of chemical groups
 1000 allowed around $\text{Ca}3$ leading to classic N-terminal myristoylation with the G acceptor residue
 1001 at R₁ (G-myristoylation). The ψ torsion angle $\text{Ca}3$ is indicated.



1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015

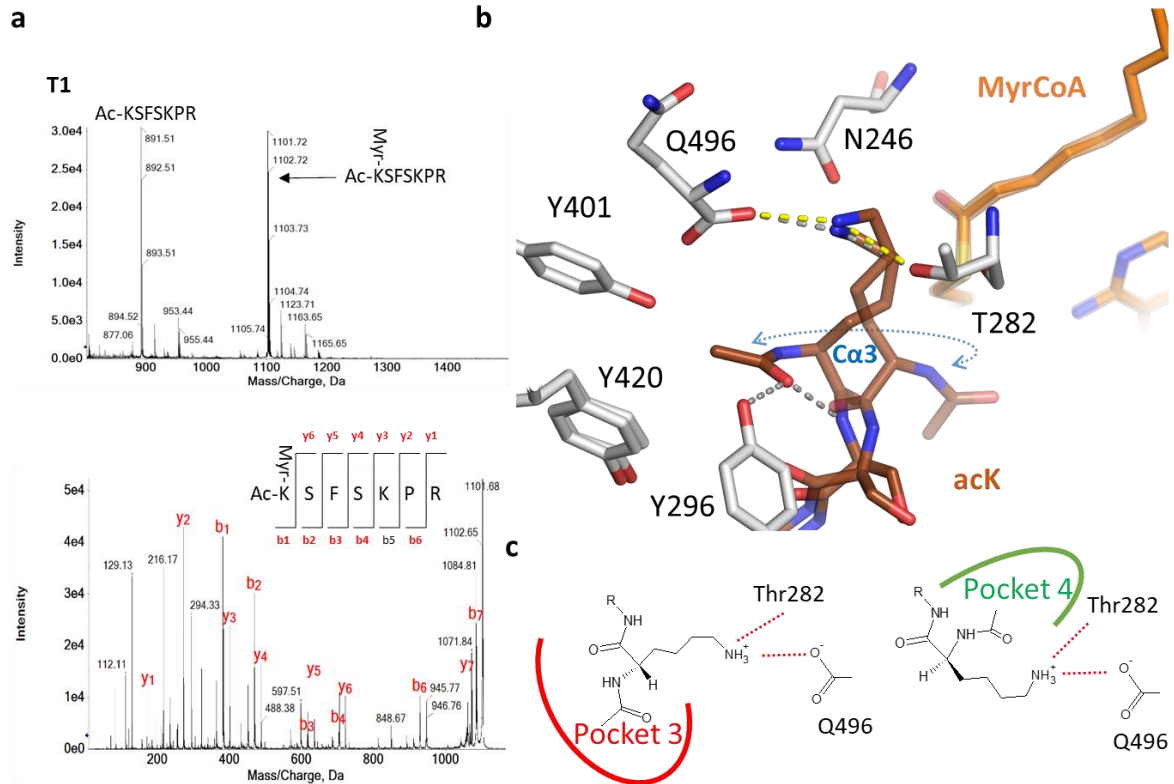
Fig. 2. K-myristoylation and G-myristoylation rules

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 R₂, 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].



1016
 1017
 1018
 1019
 1020
 1021
 1022
 1023
 1024
 1025
 1026
 1027
 1028

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) HPLC peptide at T2 (top) and MS2 spectrum of the double myristoylation product (bottom). (c) GKFSSK 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.



1029
1030
1031
1032
1033
1034
1035

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 Ca3 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).

1036
1037

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	$k_{cat}(s^{-1})$	$K_m(\mu M)$	$k_{cat}/K_m (s^{-1}\cdot M^{-1})$	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
	ANCFKPR	AN	5	Fig. S2a	0.0004±0.0001	>300	1.2±1.1	0.05
	me-GNCFKPR	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	GKVLKIF	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
	GKSNSKLLK	NCS1	14	PDB 5O9S	0.047±0.001	15±1	3157±277	105
	GKLSLHLP	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
	GKLHSPKA	NKD1	18	-	0.10±0.01	40±2	2500±200	83
	GKLQSKHA	NKD2	19	-	0.10±0.01	41±2	2501±200	83
	GKRGRSQ	UTP25	-	-	>0.01	>1000	7±2	0.2
GKX	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

1038 MS spectra are displayed in **Data S1a**. The corresponding page number is indicated in the
 1039 corresponding column.
 1040
 1041

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 (μ M)	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-GKVLKIF	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
	AKPTSKDGLK	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
	KSNAKLP		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

1043

1044

1045

1046

1047

1048

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 μ M were taken into account as substrate concentration inhibition was observed at higher concentrations