# Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution

Thierry Meinnel, Cyril Dian, Carmela Giglione

***Review article***

# Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution

Thierry Meinnel*, Cyril Dian and Carmela Giglione*

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

**Correspondence:* thierry.meinnel@i2bc.paris-saclay.fr or carmela.giglione@i2bc.paris-saclay.fr

*Keywords:* acetylation; glycine; membrane; myristoylation; quality control; subcellular compartments

**Abstract**

N-myristoylation (MYR) is a crucial fatty acylation catalyzed by N-myristoyltransferases (NMTs) that is likely to have appeared over two billion years ago. Proteome-wide approaches have now delivered an exhaustive list of substrates undergoing MYR across approximately 2% of any proteome, with constituents, several unexpected, associated with different membrane compartments. A set of <10 proteins conserved in eukaryotes probably represents the original set of N-myristoylated targets, marking major changes occurring throughout eukaryogenesis. Recent findings have revealed unexpected mechanisms and reactivity, suggesting competition with other acylations that are likely to influence cellular homeostasis and the steady state of the modification landscape. Here, we review recent advances in NMT catalysis, substrate specificity, and MYR proteomics, and discuss concepts regarding MYR during evolution.

## Lipidated proteins

Plasma membranes (PMs) are composed of extrinsic and intrinsic proteins (52%) and lipids (40%), the latter sustaining the overall cellular architecture. Membrane-penetrating extrinsic proteins often possess covalently linked lipids, usually fatty acids, which allow the protein to contact other intra- and extracellular proteins [1]. Protein lipidation involves amides (i.e. **N-α-myristoylation**, **MYR**, **see Glossary** and glycosylphosphatidylinositol (GPI) anchors), thioesters (i.e. **S-palmitoylation**, **PAL**), and thioethers (i.e. isoprenylation and farnesylation) [2]. Of these, MYR is a frequent and conserved modification specific to eukaryotes that targets major cellular components. Mapping the proteins undergoing MYR has proven challenging due to their difficult handling characteristics and amphiphilic, chimeric nature. Recently, high-end technologies have allowed the first lipidated proteome, the **myristoylome**, to be described in detail in various organisms recapitulating the tree of life [3]. These and other studies on myristoylome composition and genesis have also revealed (i) an unexpected novel mechanism of action of **N-myristoyltransferase** (**NMT**), (ii) NMT substrate selectivity, and (iii) the capacity of NMT to act on N-terminal lysines (Lys) as well as the usual glycines (Gly).

This review highlights the series of ground-breaking discoveries that have recently significantly advanced our knowledge about this long-studied enzyme. This includes an overview of a new NMT catalytic mechanism, substrate specificity, and proteomics, and a discussion of how a reduced set of targets is closely related to **eukaryogenesis** and eukaryote **evolution.**

## How NMTs catalyze MYR with high selectivity

*NMTs are GNAT members closely related to Nα-acetyltransferases*

Seventy-four crystal structures have now revealed that the C-terminal 400 amino acid-long NMT catalytic core displays a conserved 3D **GCN5-related N-acetyltransferase** (**GNAT**) core. GNATs also include the catalytic subunits of **Nα-acetyltransferases** (**NAAs**) [4], with Naa10 being the closest to NMT as it modifies N-terminal Gly [5-7]. NMTs have two adjacent GNAT domains, most likely to have arisen through duplication of a Naa10-like GNAT domain [8]. Prokaryotes do not possess NMTs, which seem to have arisen as eukaryotes evolved from the **last archaeal**

**common ancestor** (**LACA**) to the **last eukaryotic common ancestor** (**LECA**). Therefore, NMTs are associated with events during eukaryogenesis.

*How NMTs promote high substrate specificity*

NMTs are classified as glycylpeptide N-tetradecanoyltransferases because MYR was thought to only modify N-terminal Gly residues. As Gly usually arises from **co-translational** methionine excision, the **N-myristoylated** (**MYRed**) Gly is Gly2 (Met1 being aa1, and aa*x* referring to the following amino acids at position *x*). Post-translational MYR has also been reported (see below)

Structural approaches

Structures of human NMT1 (HsNMT1) co-crystallized with reactive substrates, including polypeptide substrates, have recently become available [7]. These peptide-bound structures are different to those with inhibitors [9], especially at the level of the recognition of peptide side chains in dedicated cavities (**Figure 1A**). These cavities are formed with the second GNAT fold, which is adapted to host and select the first six residues (Gly2 to aa6). Peptides with a Lys or Arg at position aa7 form a salt bridge with conserved Asp residues at the cavity entrance. The peptide-substrate backbone adopts an extended structure and only promotes significant contacts with the enzyme at the outer side (aa7-9), explaining why octapeptides are optimal for NMT-driven MYR.

**Figure 1A** shows the various NMT pockets. Pocket 1 specifically recognizes the long linear hydrophobic chain of **myristate**. The first ten carbons are embedded within a linear pocket formed from the side chains borne by the βe-αC-βf adjacent secondary structures of the first domain [10]. Each side chain's hydrophobic character is conserved, optimizing the pocket for myristate recognition and excluding bulkier fatty acids (e.g., palmitate (C:16) or stearate (C:18)) and branched fatty acids from isoprenoids such as farnesyl (C:15). The use of a linear fatty acid, not synthesized in Archaea, might indicate that NMTs eventually acquired their lipid specificity after engulfing bacterial partners that introduced the pathway (LACA; [11]).

Narrow pocket 2 is involved in $NH_2$Gly2 recognition, pocket 3 favors small- and medium-sized residues [7], and large pockets 4/5 can accommodate almost any chain. The narrow constriction at the end of the peptide recognition pocket 6 favors small residues, especially Ser. While the

emerging peptide recognition pattern (**Figure 1**) defines global specificity, it does not fully define the actual substrate landscape in a given proteome.

Metabolic labeling

Chemical biology strategies coupled with mass spectrometry have been used to investigate myristoylomes. The approaches involve, for instance, metabolic strategies that combine fatty acid analogs as precursors of **myristoyl-CoA** (**MyrCoA**) bearing a bio-orthogonal alkyne with NMT-specific inhibitors [12] to identify NMT substrates in one shot in various organisms (**Table 1, Key Table**). These approaches were subsequently completed with subcellular fractionation approaches in plants. Combining the 364 sequences from these studies revealed a similar profile (**Figure 1B**) to those observed in each of the nine organisms [7, 13-19].

Peptide arrays and artificial intelligence complete the myristoylomes

MYR co-translational activity can be mimicked *in vitro* with short peptides recapitulating nascent chains combined with an NMT-catalyzed MYR assay [20-22]. Providing insights into NMT specificity based on a reduced, non-random dataset, these approaches are biased and poorly predictive [23]. Recent strategies have leveraged large arrays of peptides derived from the N-termini of open reading frames starting with Gly from *Arabidopsis thaliana* and *Homo sapiens* proteomes. These data have validated the quasi-entire set of positives in both organisms (**Table 1**) and provided crucial information on the negative set. About 20% of Gly-starting proteins were MYRed. Human NMT1 and NMT2 and the functional NMT from *A. thaliana* displayed identical substrate specificity. Nevertheless, *Saccharomyces cerevisiae* and *Plasmodium falciparum* NMTs had restricted specificities *in vitro* and *in vivo,* partly assigned to negative residues when occurring at aa8 and/or aa9 [22-24].

This sequence dataset was used to check NMT selectivity. The pattern (**Figure 1C**), although similar to that obtained with global approaches, was more robust due to the higher number of positive sequences and the 60% negative sequences. This allowed a better understanding of their selectivity and why some proteins, despite displaying favorable residues (e.g., N3/S6), are not substrates (**Box 1**).

*Revisiting NMT catalysis*

Early studies on NMT catalysis [8] revealed an ordered mechanism with initial MyrCoA binding. Although similar to all other GNATs, the NMT mechanism was distinct, with the peptide's ammonium group making direct contact with the carboxy-terminal group to trap the proton and allow the amino group to react. However, the widely-spaced reactive ammonium and thioester bond of MyrCoA suggested that it must fully flip to approach and bind to it. Recent 3D structures of HsNMT1 in complex with canonical substrates delivered sequential snapshots of the entire catalytic mechanism spanning the enzymatic pathway from substrates, intermediates, and products. In these models, the peptide substrate is differently positioned in pockets 2-4, while the ammonium creates a water-mediated bond with the base [7]. The reaction is thus easier to achieve without major conformational changes around the alpha carbon of Gly2. These structures reveal a conserved water channel connecting the deeply buried active site to the surface and the key role of the Ab-loop, on which NMT catalysis relies (**Figure 2**).

## Myristoylomes, MYR targets, and evolutionary boundaries

*The five main NMT subtypes in eukaryotes*

NMTs display an average of 50% identity, and phylogenetic clustering of NMTs recapitulate the eukaryotic tree (**Supplemental Figure 1**), falling into plant, animal, apicomplexa/diatom/ameba, and fungi/englenozoa/microsporidian classes. Unicellular organisms and invertebrates have one *NMT* gene, while land plants and vertebrates have two (*NMT1, NMT2*) [25]. The two land plant NMTs cannot be distinguished from each other and probably correspond to early gene duplication without true speciation. Vertebrate *NMT1/2* display very similar substrate specificity [7, 26]. Interestingly, caspase cleavage of the N-terminal extension, previously thought to favor ribosome binding, results in the targeting of human *NMT2* to membranes and *NMT1* to the cytosol [27].

*NMT-catalyzed MYR is essential for cell compartmentation*

*In silico* analyses predicted 46 targets in the *S. cerevisiae* myristoylome, a dozen of which were essential during stationary phase when inactivated (**Table 1**) [8, 28]. In *A. thaliana*, three essential

targets were identified during development [24, 29, 30], and inhibition of Rpt2 MYR induced phenotypes indicating decreased proteasome activity in yeast [31]. Another important MYRed protein is the protein kinase Vps15, which is a PI3 kinase complex I subunit involved in macro-**autophagy** in eukaryotes. All these targets shuttle between the cytosol and other compartments including the nucleus (Rpt2/Sip2/Ptc2), the PM (Gpa1), and the endosome (Arf1/Arf2/Vps20).

The weak binding energy provided by a myristate moiety suggests that MYR cannot maintain stable protein-membrane interactions by itself [32]. MYR can, however, target proteins to endomembranes, and cytosolic partitioning was demonstrated to be correlated with NMT's catalytic efficiency in *A. thaliana* [32]. Therefore, MYR of less efficient NMT substrates is partial and varies between MYRed proteins [7]. While the hydrophobic myristate group is crucial for membrane localization, its N-terminus position plays a key role. For instance, C-terminal prenylation does not compensate for the endocytic function of a non-MYRed **ADP-ribosylation factor** (**ARF**) variant [33].

MYR's involvement in protein shuttling and reversible membrane binding is regulated through various signals which allow the protein to associate or dissociate from the lipid bilayer [34]. The switch may involve, for instance, (i) GTP hydrolysis inducing conformational changes in ARFs [35], (ii) $Ca^{2+}$ binding to EF-hand calcium-sensor proteins [36], (iii) phosphorylation of a polybasic motif involved in electrostatic interactions with membrane phospholipids, or (iv) depalmitoylation of vicinal cysteine residues. Such reversible myristoyl switches control MYR-dependent subcellular protein trafficking, distributing MYRed proteins to different compartments. For instance, the mammalian protein kinase A undergoes nuclear-cytoplasmic shuttling through significant deamination of N3 in muscles due to electronegative charges caused by the new D3 [37].

The PM localization induced by cysteine residue PAL in the vicinity of the MYR site was initially discovered for the mammalian homolog of Gpa1 [38, 39] and later in yeast [40]. This phenomenon was recently confirmed by hierarchical clustering of protein accumulation profiles, which revealed a clear shift in MYRed protein abundance toward PM fractions in the presence of second targeting signals including PAL or polybasic tracks [19]. Yeast only features four such proteins including Gpa2, and only Gpa1 is essential (**Table 1**). The proximal PAL of Gpa1 is therefore likely

to be an ancient function linking MYR, PAL, and PM localization, probably later in evolution than endosomes and relating to multi-cellularity and its required intercellular communication. Recent studies on PAL and MYR co-occurrence in plants and humans have shown that PAL on cysteine residues in the immediate vicinity of the MYRed Gly are readily detectable, with co-occurring cysteine codons at positions 3-7 significantly more frequent in MYRed proteins. 50% of MYRed targets are expected to be **palmitoylated** (**PALed**) in plants and 23% in humans [7].

*Perspectives with respect to eukaryogenesis*

The myristoylomes of nine organisms displaying NMTs of all classes (**Table 1**) allow us to draw conclusions about the specific features promoted by NMTs throughout evolution. **Figure 3A** shows that larger proteomes correspond to larger myristoylomes, which always constitute about 2% of the proteome. The yeast myristoylome is an outlier, probably due to the gene loss characterizing fungal evolution [3]. Nevertheless, Trypanosomatidae and Apicomplexa, which also lost genes through parasitism [3], do not have similarly small myristoylomes, so the small myristoylome observed in yeast might be partly linked to the restricted substrate specificity of yeast NMT [24]. This is similar to yeast N-acetylation patterns, which are similarly subtly reduced, especially as no N-acetylation Gly-starting proteins have been identified. Note that fungi still display branch-specific essential MYR targets such as the **GTPase** EGO (**Table 1**) involved in micro-autophagy.

**Table 1** shows that a subset of eight yeast myristoylome proteins are also present throughout all cell compartments of the myristoylomes of other organisms. The myristoylome has, therefore, exploited all compartments elaborated by eukaryogenesis (**Table 2**). In addition, the major eukaryotic sensing events also involve MYRed targets (**Table 2**). As a result, the myristoylome illustrates the various steps of eukaryogenesis and explains why myristoylome size closely mirrors that of the proteome (**Figure 3A**).

## Proteome-wide approaches reveal unexpected targets

These high-throughput data have revealed that MYR targets unexpected compartments and functions, especially catalytic processes, cell organization, and sensing (**Figure 3B,C**) but not gene

expression and processes characterizing the first **eukaryotic signature proteins** (**ESPs**) [41, 42]. Therefore, NMT probably appeared after these early ESP emerged in Archaea, likely in LACA. They are much more relevant to the compartmentalization events underpinning eukaryogenesis. Kinases, small GTPases, proteostasis-related phosphatases, hydrolases, transferases, and calcium sensors feature often as targets (**Figure 3B, C**). These unexpected processes are eukaryote specific and further illustrate the tight link between MYR and eukaryogenesis, as detailed below.

*The immune system and MHC presentation*

The mammalian immune system is equipped with virus detection machinery via virus-synthesized MYRed peptides [43]. This was first shown for HIV-Nef N-terminal pentapeptides, which add to the antigen repertoire presented by **major histocompatibility complex** (**MHC**) class I molecules [43]. The crystal structure of the complex reveals a groove suitable for 4-9 residue peptides [44, 45].

As MYRed peptides are derived from protein components associated with viral or bacterial pathogenicity, this raises the possibility of exploiting MYR for vaccines [43]. The relationship between MYR and immunity is now firmly established in plants and animals [46, 47]. When impaired in humans, such as in rheumatoid arthritis, the lack of MYR of AMPK eventually triggers inflammation due to failed lysosomal targeting [48].

*Cell motility*

Cilia

The cilium is a sensory organelle occurring in vertebrates and invertebrates, including euglenozoa. Protein trafficking to cilia implies the presence of lipid microdomains that help overcome the lipid diffusion barrier and relies on polarized trafficking of post-Golgi vesicles [49]. Lipid modification of the cilial proteome is widespread and includes MYRed proteins [50]. Of note is the MYRed cystin/nephrocystin. Cilial interactions are mediated by the UNC119 cargo, which appears to be a major coordinator of MYRed protein trafficking, including a **Src family kinase** (**SFK**) and nephrocystin [51]. Both are targeted into cilia via UNC119-mediated trafficking [52]. The X-linked human retinitis pigmentosa gene *XRP2* encodes a GTPase-activating protein involved in Golgi-cilia trafficking, leading to UNC119 release [53]. XRP2 has MYR and PAL modification sites

233 featuring Cys3 and Ser6, with deletion of the latter causing disease by inhibiting MYR and further
234 PM localization [54]. Finally, calflagin is a MYRed/PALed calcium-binding protein in kinetoplastids.
235 PAL on Cys3 is required for cilial sorting, while MYR alone is sufficient for membrane targeting
236 [55]

237 The glideosome of apicomplexa

238 The complex life cycles of human parasites like apicomplexa alternate between
239 asexual/sexual and parasitic/free stages in various hosts or organs. The parasitic stage involves
240 complex host cell interactions. Parasites elaborate dedicated secretory compartments such as
241 micronemes and rhoptries to promote the required motility. The gliding machinery allows
242 parasites to cross cell barriers to exit infected cells in a process analogous to but distinct from
243 that caused by cilia.

244 MYR is strongly involved in gliding motility [56, 57]. In the malaria parasite *P. falciparum*, this
245 involves GAP45, which is MYRed [58]. Of the rhoptry components, ARO and the cGMP-dependent
246 protein kinase PKG are retrieved, while the ISP1 and ISP3 are major MYRed proteins [59] that link
247 the machinery to actin and myosin to induce traction [60]. Finally, gliding motility signaling
248 requires the same calcium-dependent protein kinases found in plants, which are almost
249 systematically MYRed and PALed [60].

250 *Mitochondria*

251 There are >20 MYRed proteins in the human mitochondrion, either induced by Met removal
252 or caspase cleavage. Several target the cytosolic side of the **outer mitochondrial membrane**
253 (**OMM**). These include NADH-cytochrome $b_5$ reductase, Sam50 (a sorting and assembly
254 machinery complex component), MOSC1/2 through co-translational MYR [61], and several post-
255 translationally MYRed proteins (see below). TOM40 is a MYRed subunit of the mitochondrial
256 import receptor subunit lying within the OMM but with an N-terminus protruding at the outer
257 side [62, 63]. It was unexpected that MYRed proteins could exist within the intermembrane space,
258 including the MIC19 and MIC25 components of the mitochondrial contact site complex (**MICOS**,
259 also found in yeast; **Table 1**), which is crucial for maintaining crista integrity and mitochondrial
260 function [64], and other proteins of the inner membrane space such as complex I subunits of the

respiratory chain including B18/NDUFB7 and NDUF4 [7]. How they reach the outer membrane without a mitochondrial targeting signal is unknown. A similar feature was noted in plants with (i) the yeast homolog Mia40 (a mitochondrial intermembrane space assembly machinery component) lying at the inner membrane next to MICOS and interacting with TOM at the OMM [65], and (ii) two prohibins involved in crista morphogenesis [7, 19].

One hypothesis on how a MYRed protein might access the intermembrane space comes from studies of yeast MIC19, which undergoes MYR in the cytosol. By interacting with Tom20 of the TOM complex at the OMM, MIC19 would be imported into the inner mitochondrial membrane (IMM) by so-called entropy pushing [66]. By contrast, delivery of the MYRed Tom40, which features several transmembrane helices which insert into the OMM, involves Hsp90 [67].

Finally, it is interesting that ER-mitochondrial crosstalk involves both lipid exchange and $Ca^{2+}$ regulation, with the MYRed calcium sensor NCS1 homolog to the yeast form described in **Table 1** involved [68]. The link between mitochondrial contact sites, MICOS, calcium signaling, and MYRed protein import [65] needs further detailed examination, especially as these are some of the most conserved MYRed targets (**Table 1**).

*MYR and proteostasis*

The plant myristoylome reveals about 50 proteins annotated as RING finger, F-box, or ubiquitin domain proteins and E3 ubiquitin ligases [7, 19], some of which localize to the nucleus [19]. Among *Arabidopsis* proteins, the SKIP2 subunit of the SCF(ASK-cullin-F-box) E3 ubiquitin ligase complex was found to serve as a substrate receptor for E3 ligases. Interestingly, *Toxoplasma* TgFBO1 is an inner membrane complex MYRed component with an F-box domain that might bind SKIP1 [69]. In humans, ten such MYRed E3 cullins were identified [7, 70]. The identification of major MYRed components of protein degradation reveals MYR as an important feature of **proteostasis** (see also glossary and below).

*Apoptosis*

Induction of apoptosis leads to the activation of numerous caspase proteases. In some cases, caspase cleavage may expose internal cryptic MYR sites to the action of NMTs. MYR-dependent relocalization to the OMM of fragments of BID and actin [71, 72] has been reported. BID switching

contributes to enhanced cytochrome *c* release and cell death [71]. A Huntingtin fragment also relocalized to the autophagosome due to MYR [73], while the MYRed gelsolin fragment is cytosolic [74], and the MYRed C-terminal fragments of PAK2 and PKCε are redirected to internal membranes and that of filensin to the PM [75-77]. A global approach has identified another 40 proteins MYRed following caspase cleavage. The MYR consensus is oversimplified while the caspase consensus strongly emerges (**Figure 1D**), suggesting that post-translational MYR was acquired later in the course of eukaryotic evolution and has not yet converged.

Ferroptosis suppressor protein 1 (FSP1/AIFM2) is a MYRed protein that localizes to the PM and behaves as an oxidoreductase, converting ubiquinone to ubiquinol to inhibit ferroptosis [78, 79]. This programmed cell death mechanism is conserved in plants [80] and leads to lipid damage and membrane degradation. Glutathione peroxidase GPX4 in vertebrates together with glutathione prevent lipid peroxidation. While the MYRed FSP1/AIFM2 acts as a parallel system to GPX4, of note, two of the eight GPX isoforms in plants (GPX4/5) are MYRed [19] so might act similarly to FSP1/AIFM2 directly at the PM. Also, plants have a large array of thioredoxins and glutaredoxins featuring MYR [32].

## MYR and quality control

### *A MYR, Gly-specific N-degron*

Recent data show that two cullin-RING E3 ligases (CRL2 and CRL5) target Gly-starting peptides for degradation [81]. Globally, Gly substrates favored by CRL2 included those starting with a Gly followed by Phe, Gly, His, Leu, Met, or Tyr, while those followed by Asp, Glu, Ile, Asn, Pro, Arg, Ser, or Thr were disfavored. Interestingly, metazoan proteomes seem to be depleted of CRL2-mediated N-terminal Gly **degrons**, while CRL2 complexes most likely target protein fragments bearing N-terminal Gly degrons after caspase cleavage during apoptosis. 13% and 17% of known post-translationally MYRed proteins display a residue 3 favored and disfavored by CRL2, respectively. Loss of MYR destabilized Gly-starting polypeptides followed by Cys, Gln, or Lys, so the authors proposed that N-terminal Gly degrons might play a role in MYR quality control, allowing degradation of proteins failing to undergo MYR and avoiding accumulation of non-

MYRed “aberrant” species. However, partial stabilization of endogenous N-MYRed proteins after loss of MYR including CRL2 adaptors suggests a requirement for additional E3 ligases and/or other undetermined factors. Finally, it remains to be fully understood which non-MYRed Gly-starting proteins are related to Gly degrons, as there clearly remains a non-negligible number of such species that eventually accumulate in multicellular organisms as either Nα-acetylated or with an unmodified Gly [5, 7]. The favored CRL2 sequences indeed deal with <14% of the human myristoylome , while the disfavored sequences involve 51% (see SupDataset 2 in [7]).

*Competition with other modifications or processes*

There is increasing evidence that one unique **proteoform** may undergo various modifications at the same site within the same cell, including for MYRed proteins, which may be either partially modified or partially N-acetylated [7]. MYR’s pronounced effect on membrane binding is likely to have a major effect on protein localization, as observed for mammalian NADH-cytochrome $b_5$ reductase, which is targeted to the ER if Nα acetylated and the mitochondrion when MYRed [61].

*Roles of sirtuins and HDACs with Lys-MYR*

NMTs are thought to exclusively acylate N-terminal Gly residues. Nevertheless, the ε-amino groups of internal lysines of several proteins including tumor necrosis factor-α, interleukin-α1 precursor, and serine hydroxymethyltransferase 2 also undergo MYR [82, 83], although the corresponding catalyst(s) still remain unidentified.

Interestingly, it was recently demonstrated that human NMTs may transfer myristate onto the N-terminal epsilon amino groups of lysine side chains, expanding its known substrate range beyond the well-established N-terminal alpha moiety of Gly peptides [84]. Furthermore, Lys3 MYR allows ARF6 to remain on membranes during GTPase cycling [85]. In contrast to the irreversible Gly-MYR, several acyl hydrolases targeting Lys-MYR have recently been described including Sirtuin6, HDAC8, and HDAC11 [82], making Lys-MYR a reversible modification dependent on these specific hydrolases. This implies that Lys-MYR, as with other reversible modifications (i.e., phosphorylation, **N-terminal acetylation**), might be intimately related to signal transduction and tightly regulated by as yet unknown biotic or abiotic signals. In the case of ARF6, NMT prefers the GTP-bound form of ARF6 while SIRT2 prefers the GDP-bound form,

allowing an NMT/SIRT2-ARF6 regulatory axis of the GTPase cycle [85]. The N-terminal Lys-MYR might be part of an N-degron, so an important role of Lys-MYR deacylases could be to ensure protein quality control by preventing intracellular Lys-MYR accumulation [84].

**Concluding remarks**

Four decades of seminal contributions have now contributed to knowledge about the key MYR process, with recent studies on MYR and NMT biology leading to revisiting and rethinking old dogmas. Multiscale approaches have proven pivotal in expanding our knowledge of myristoylomes. Recent efforts to precisely annotate all proteoforms in several proteomes have been valuable for comprehensive MYR mapping [7]. Future studies will increase the actual myristoylome with all possible substrates resulting from post-translational MYR. This new pool of post-translational MYRed substrates can be unraveled taking into account protein substrate specificity ensured by the newly identified specific NMT pockets, the amino acid pattern of the identified MYRed proteins, and protease cleavage specificity. This will reveal whether the pools of co- and post-translational Gly-Myr share the same N-terminal features.

Gly-myristoylome characterization has also revealed that a non-negligible number of MYRed substrates are partially modified by MYR and that the same N-terminal Gly can undergo different modifications, completely altering the protein's properties. The factors influencing specific modifications and the impact of quality control involving Gly N-degrons and E3 ligases are still unknown, but future elucidation will prove crucial for understanding both the involved factors and their biological roles.

MYR was long considered an irreversible modification. However, myristoyl-glycine cleavage by *Shigella* protease IpaJ was recently described [86, 87], opening up new directions on the dynamic regulation of MYR. The plant hormone ABA was proposed to inhibit MYR, in turn triggering relocalization of several protein targets [88]. In this context, we anticipate that NMT activity may be insufficient to modify all protein substrates and MyrCoA donor availability may also be insufficient, while dysregulation of *de novo* synthesis might accelerate the growth of tumors that express high levels of activated Src kinase [89].

Finally, the recent discovery of a novel NMT catalytic mechanism suggesting MYR of N-terminal Lys in addition to Gly takes the field into an exciting era where mechanisms of MYR regulation, biological effects, and pathological consequences become central aspects.

**Acknowledgements**

The team was funded by Agence Nationale de la Recherche (ANR-2010-BLAN-1611-01) and Fondation ARC (SFI2011120111203841) and benefitted from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

**Glossary**

**26S proteasome**: a 2.5 MDa molecular machine controlling ubiquitin-dependent proteolysis in eukaryotes made of two (19S-regulatory and 20S-core proteolytic) complexes.

**ADP-ribosylation factors (ARFs)**:, small **GTPases** essential in eukaryotes that function in vesicular trafficking and actin remodeling.

**Autophagy:** mechanism allowing degradation or recycling of cellular components by fusion to lysosomes.

**Co-translational:** an event occurring while the nascent protein chain is not yet complete and still bound to the ribosome

**Degron:** specific sequence in a protein that directs its degradation *via* the **26S proteasome** or **autophagy**.

**Endosomal sorting complex required for transport (ESCRT)**:, a set of three complexes. ESCRT-III consists of 8-12 subunits including Vps20/CHMP6.

**ESP:** eukaryotic signature proteins, early markers of **eukaryogenesis** and including DNA, RNA, and protein synthesis proteins but also **ESCRT**, small **GTPases**, cytoskeletal proteins, and ubiquitin signaling proteins.

**Eukaryogenesis:** approximately 2-billion-year process by which eukaryotes evolved from an Archaeon. **LACA** and **FECA** witnessed progressive emergence of endomembranes, nuclei, and organelles.

**Evolution:** all life on Earth has evolved from common ancestors in an unbroken chain since its origin, approximately 3.8 billion years ago.

**FECA**: first eukaryotic common ancestor, a protoeukaryote with endomembranes (approx. 2.4 billion years ago).

**GNATs**: GCN5-related N-acetyltransferases, a large enzyme superfamily that transfer acyl derivatives from acyl-CoA donors to a variety of acceptors.

**GTPases**: hydrolyze GTP into GDP, include small proteins like **ARFs** or large ones such as heterotrimeric G proteins, major molecular switches of signal transduction pathways located at the PM.

**LACA**: last archaeal common ancestor (approx. 2.8 billion years ago), an archaeon featuring ESPs such as cytoskeletal proteins and membrane trafficking systems.

**LECA**: last eukaryotic common ancestor, featuring fully differentiated internal structures (approx. 1 billion years ago).

**Methionine aminopeptidases**: enzymes cleaving the first methionine of approx. 60% proteins to unmask small residues such as glycines.

**Major histocompatibility complex**: cell surface proteins essential for vertebrate acquired immunity.

**MICOS**: mitochondrial contact site complex, maintains crista integrity and mitochondrial function.

**MYR**: N-α-myristoylation, a lipid acylation of proteins using MyrCoA as a myristate donor; makes amide bonds with free amino groups, usually from N-terminal glycines.

**MYRed**: N-myristoylated, refers to glycine MYR in large-scale analyses.

**Myristate,** a saturated (i.e., without double bonds) 14 carbon fatty acid (C14:0).

**Myristoylome**: complete set of MYRed proteins in a proteome.

**NAA**: enzymes of the **GNAT** family, which use acetyl CoA as a donor to acetylate the alpha amino-termini of proteins. The catalytic subunit may be associated with auxiliary subunits to form N-α-acetyltransferase complexes (NATs).

**Naa10:** catalytic subunit of N-$\alpha$-acetyltransferase A (NatA), an **NAA** operating on small N-terminal residues such as glycine.

**NMT**: N-myristoyltransferase, the MYR catalyst, part of the GNAT family, uses MyrCoA as donor.

**N-terminal acetylation**: a modification occurring on the N-termini of proteins. Various complexes have been described, such as NatA and NatB, depending their substrate specificity. Involves 80% of proteins of multicellular eukaryotes.

**PAL**: S-palmitoylation next to MYR; catalyzed by membrane enzymes, occurs anywhere on cysteine residues provided a membrane-binding motif (transmembrane helix/helices, lipid modification) brings the residue to the interface.

**Proteoform**: protein derivative from one gene, may differ by one modification.

**Proteostasis:** the proteome steady-state homeostasis as guaranteed by a network of quality control components ensuring proteins integrity.

**SFK**: Src family kinase, non-receptor Src family of protein tyrosine kinases, nine members, none in plants.

## Text Boxes

### Box 1: Protein modification probability and frequency

Common representations of the substrate selectivity of enzymes involved in protein modifications such as proteases use extremely useful logos and powerful visualizations that integrate massive data into a single, striking image [90]. Nevertheless, care must be taken when drawing conclusions given the complexity associated with the recognized pattern, which takes into account: (i) various positive and negative effects of the enzyme; (ii) evolutionary constraints from which the initial non-MYRed target starts; (iii) functional and structural constraints of the N-terminal domain, which may be required for *in vivo* activity (interaction, folding, catalysis, other modifications); and (iv) the fact that there are other constraints from the genetic message and its interpretation by a specific organism (codon usage, amino acid availability, RNA folding, completion with other enzymes including **methionine aminopeptidases** and NAAs). **Figure I** exemplifies this feature. From **Figure 1**, one can easily identify the main features of the NMT recognition site (for instance C, N, or Q at position 3). If one compares the positives from global analysis in blue and macroarray results in green in **Figure I**, the overall frequency is clearly identical. Still, considering the most obvious element of selectivity at position 6 (with S being a feature) (**Figures 1** & **2**), they only represent 60% of the complete set. With the help of the negative microarray set, one can analyze the probability that an extracted genomic sequence is MYRed (orange). The data show that there is no single amino acid which leads to a frequency higher than 82% with N3 (**Figure I**). Still, such proteins only represent 25% of the entire set. Combination increases the probability to 93% (N3/S6), but those proteins are only 5% of all positives.

The difficulty in drawing decision trees from amino acid patterns explains why artificial intelligence approaches including machine learning strategies are required for prediction [7]. Such tools need not only positive but negative sets to ideally delineate, learn, and distinguish a positive from a negative. It is important to remember that the theoretical diversity with 7 amino acids is huge ($10^9$), and current approaches still deal only with hundreds of data points or two thousand in the case of the macroarray. Obtaining negative sets with MS approaches based on

positive selection is only feasible if combined strategies involving diverse selection procedures are used. A strategy involving the identification of proteins with non-modified or Nα-acetylated N-termini was used to obtain information on proteins starting with non-MYRed glycines *in vivo* [7].

***Legend to Figure I in Box 1: relationships between the probability and frequency of MYR***

Data are derived from the peptide array [7] of 2,066 sequences including 837 positives and all data from click chemistry (365 sequences). % positives click refers to the click chemistry set. % positives macroarray (green bars) refers to the frequency of positives in the positive set of 837 sequences. Probability of being positive or true positives (orange bars) refers to the frequency of positives in the complete array of 2,066, defining the probability of being MYRed.

## Figure Legends

**Figure 1**. **Specific pockets of NMTs ensure protein substrate specificity**

*Panel A*. Various structures of human NMT1 crystallized with MyrCoA and one of 5 different peptide substrates (PDB IDs: 5O9S, 5O9V, 5O9U, 6EHJ, 6QRM) superimposed to show how the various side chains fill each pocket. Each pocket is delineated with a full line and labeled (see text).

*Panel B*. Corresponds to 365 octapeptides extracted from N-terminal sequences from MYRed proteins identified by global approaches with a negative set against Swiss-Prot (*Homo sapiens* was used as the proteome as very similar results were obtained with other proteomes).

*Panel C* A data set of a macroarray made of >2,000 octapeptides assayed *in vitro* with NMT is available in [7]. The panel displays the positive data (836) against the negatives (1,229). Gly2 does not appear as it represents >99% of the entire dataset, but "G" was added for the sake of easier comparison with panel A.

*Panel D*. Post-translational MYR recognition pattern. The panel displays residues around the modified glycine (residue 5 before cleavage and MYR). The data are compiled from SupData4 in [13]. The sequence of the filensin C-domain was added to build the pattern. Panels B-D are independent and were built with iceLogo 1.2 [90].

**Figure 2: Pre-organized water molecules in a dedicated conserved channel contribute to the catalytic mechanism**

*Top*, the NMT water channel is schematized. Flux directions of water and protons are indicated in blue and red, respectively. The two proton donors are indicated on the right side. Water molecules are numbered from the outer to the inner side of the active site (W12 to W1).

*Center,* superimposition of the water molecules of the channel and active site in HsNMT1:MyrCoA and HsNMT1:MyrCoA:peptide states. The organized water molecules in the peptide-free state shown in purple and those in the peptide-binding, tetrahedral intermediate state in red. The

volume of the water channel is drawn in grey. The MyrCoA is shown as ball and sticks and colored in orange. Hydrogen bonds are shown as dashed lines.

*Middle box*, left, organization of the exit of the water channel with three residues contributing to a hydrogen bond network including the conserved Y356; right, close up of the active site in an alternative view. W1-3 are displayed at the two states to show how a new water molecule (W3) is pushed into the active site while entering through the water channel at TI.

*Bottom box*, right. Close-up of the active site around T282 and Q496. Detailed hydrogen bond network around W1-W5.

*Bottom box*, left. Operating principle of a proton wire semiconductor: the protons hop along chains of hydrogen bonds between water molecules making intermediary hydronium ions (adapted from [91]). Four steps of proton transfers figured in this cartoon of a wire made of 5 water molecules. A wire can be made of two and more than ten water molecules.

**Figure 3: Myristoylome sizes and the categories of MYRed proteins in multicellular organisms**

*Panel A*. Linear relationship between proteome and myristoylome sizes. The data from **Table 1** were plotted. The dot in grey corresponds to the yeast data, which is an outlier featuring a reduced myristoylome (see text).

*Panel B & C*. Categories of MYRed proteins in multicellular organisms. Only those proteins for which gene ontology information is available were analyzed. 439 proteins sorted from *A. thaliana* (*Panel B*). 221 proteins sorted from *H. sapiens* (*Panel C*).

## Tables

**Table 1, Key Table: Myristoylome sizes and conservation of major MYR targets throughout species**

Proteomes were extracted from Uniprot (https://www.uniprot.org/); the Gly-ome corresponds to all ORFs displaying a Gly2 residue. Green boxes refer to conserved targets throughout species.

[£] data from GT1 strain consolidated and confirmed with data from *T. gondii* strains ME49 and FOU. All data are from ToxoDB 36 release.

[$] library likely cumulating all proteoforms.

[£] value calculated from data obtained with the machine learning predictor described in [7]. The values can be compared to those published with another predictor dating to 2002 with far less robust genomic annotation (see Table 6 in Ref.[92]). The performances of various predictors available online are compared in Table 1 of Ref.[23].

[&] data from individual or large-scale analysis (with exception of *T. gondii*, see [93]); data in parentheses refer to those obtained with peptide libraries derived from genuine protein sequences, see references [7, 13-19]. For *A. thaliana* and *H. sapiens*, see Supplementary Datasets 1 & 2 in Ref.[7].

[#] data taking into account all validated proteoforms from NextProt2017 (https://www.nextprot.org/) and AraProt2017 (https://www.arabidopsis.org/index.jsp).

[*] means that an ORF with strong homology, Gly2, and a likely MYR consensus is retrieved in the proteome.

[**] the protein exists but not with a Gly2; homologs of the apicomplexan phylum (*Babesia* spp.) display a Gly2 followed by a good MYR consensus.

| Phylum<br>*Organism/species* | | | Unicellular organisms | | | | | | Multicellular organisms | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Fungi | Trypanosomatidae | | | Apicomplexa | | Vertebrates | | | Vascular plants | |
| | | | *S. cerevisiae* | *L. donovani* | *T. cruzi* | *T. brucei* | *P. falciparum* | *T. gondii* | *D. rerio*[$] | *H. sapiens*[#] | | *A. thaliana*[#] | |
| | | | | | | | | | *Proteoforms* | *Proteo forms* | *N-termini* | *Proteo forms* | *N-termini* |
| Protein number | Proteome | | 6 701 | 7 960 | 19 242 | 8 587 | 5 548 | 8 460 | 44 132 | 42 204 | 20 399 | 48 359 | 27 655 |
| | Gly-ome | | 311 | 412 | 867 | 466 | 238 | 546 | 2 495 | 3 303 | 1 980 | 4 469 | 3 083 |
| | Myristoylome | Predicted[£] | 46 | 126 | - | 136 | 79 | 157 | 553 | 589 | 376 | 935 | 556 |
| | | Verified[&] | 15 (6) | 46 | 43 | 46 | 33 | 0 | 57 | 157 (293) | | 115 (522) | |
| **Example of MYRed proteins** | | | | | | | | | | | | | |
| **Localization** | **Type** | **Protein family** | **Essential in yeast** | ***Occurrence of MYRed isoforms (exact number only indicated with S. cerevisiae)*** | | | | | | | | | |
| Golgi | GTPases | ARF | Yes | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | | ARL | No | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1 | 1 |
| PM | | Heterotrimeric G protein α | Yes | 2 | - | - | - | - | - | 1 | 1 | 1 | 1 |
| Vacuole | | EGO/GSE complexes component | Yes | 1 | - | - | - | - | - | - | - | - | - |
| ER | Phosphatases | Phosphatase PPM, PP2C | Yes | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| PM | Calcium binding | EF-hand calcium binding protein; NCS1; CBL; calcineurin | Yes | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Nucleus | ATPase | 26S proteasome RPT2 subunit | Yes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Cytosol | Kinases | AMP kinase beta subunit | Yes | 2 | - | - | - | - | 1 | 1 | 1 | 1 | 1 |
| ER/ Endosome | | VPS15; component of VPS34 PI3-kinase complex I | No | 1 | 1* | 1* | 1* | 1* | 1* | 1 | 1 | 1 | 1 |
| Endosome | Complex | ESCRT-III subunit VPS20/CHMP6 | Yes | 1 | 1* | 1* | 1* | 1** | 1** | 1 | 1 | 1 | 1 |
| Mitochondrion | Complex | MICOS 19/25 | No | 1 | - | - | - | - | - | 1 | 1 | - | - |
| Vacuole | Complex | Vac8, protein with Armadillo repeats | No | 1 | - | - | - | 1 | 1 | 1 | 1 | 1 | 1 |

550

551 **Table 2: NMT, its targets, and the major events in eukaryotic evolution**

| **Stage of eukaryogenesis** | **MYR event** |
|---|---|
| Protoeukaryote | Early **NMT evolution** from a prokaryotic **GNAT** acetyltransferase |
| Emergence of endomembranes (ER, Golgi) | **ARFs** act as early NMT targets |
| Early cell division | **ESCRT**III Vps20/CHMP6 involved |
| Creation of nuclear endomembrane | **26S proteasome** regulatory subunits are shuttled together with Rpt2 |
| Aerobic bacterium engulfment, allowing oxygen accommodation and energy sensing | AMK/Sip1/SnRK kinase assemble as heterotrimeric complexes |
| Calcium as major signal transducer | Calcium sensing with calcineurin-like proteins |
| ER stress, regulation of the unfolded protein response | PP2C phosphatases involved |
| Vacuole sorting, macroautophagy | Protein kinase Vps15 |
| Mitochondria become symbiotic organelles, gene transfer to the nucleus | **MICOS** complex ensures link between ER and mitochondria |
| Early cell differentiation allows cell communication through membrane receptors and allow sex mating in fungi, for instance; leads to multicellularity | Heterotrimeric G alpha subunit, interaction with PI3 kinase/phosphodiesterase activation |
| Animal cell differentiation | Tyrosine kinases of the **SFK** family appear |
| Plastids become symbiotic organelles | Fusion of MYRed calcium sensors to kinases to create $Ca^{2+}$-dependent kinases |

552

## References


1. Meinnel, T. and Giglione, C. (2008) Protein lipidation meets proteomics. Front Biosci 13, 6326-40.

2. Resh, M.D. (2016) Fatty acylation of proteins: The long and the short of it. Prog Lipid Res 63, 120-31.

3. Dacks, J.B. and Field, M.C. (2018) Evolutionary origins and specialisation of membrane transport. Curr Opin Cell Biol 53, 70-76.

4. Aksnes, H. et al. (2019) Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. Mol Cell 73 (6), 1097-1114.

5. Arnesen, T. et al. (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci U S A 106 (20), 8157-62.

6. Bienvenut, W.V. et al. (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. Mol Cell Proteomics 11 (6), M111 015131.

7. Castrec, B. et al. (2018) Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. Nat Chem Biol 14 (7), 671-679.

8. Bhatnagar, R.S. et al. (2001) Biology and enzymology of protein N-myristoylation. In The enzymes (Tamanoi, F. and Sigman, D.S. eds), pp. 241-286, Academic Press.

572 9. Farazi, T.A. et al. (2001) Structures of Saccharomyces cerevisiae N-myristoyltransferase with
573 bound myristoylCoA and peptide provide insights about substrate recognition and catalysis.
574 Biochemistry 40 (21), 6335-43.

575 10. Bhatnagar, R.S. et al. (1999) The structure of myristoyl-CoA:protein N-myristoyltransferase.
576 Biochim Biophys Acta 1441 (2-3), 162-72.

577 11. Koonin, E.V. (2015) Origin of eukaryotes from within archaea, archaeal eukaryome and bursts
578 of gene gain: eukaryogenesis just made easier? Philos Trans R Soc Lond B Biol Sci 370 (1678),
579 20140333.

580 12. Lanyon-Hogg, T. et al. (2017) Dynamic protein acylation: new substrates, mechanisms, and
581 drug targets. Trends Biochem Sci 42 (7), 566-581.

582 13. Thinon, E. et al. (2014) Global profiling of co- and post-translationally N-myristoylated
583 proteomes in human cells. Nat Commun 5, 4919.

584 14. Wright, M.H. et al. (2014) Validation of N-myristoyltransferase as an antimalarial drug target
585 using an integrated chemical biology approach. Nat Chem 6 (2), 112-21.

586 15. Broncel, M. et al. (2015) Multifunctional reagents for quantitative proteome-wide analysis of
587 protein modification in human cells and dynamic profiling of protein lipidation during vertebrate
588 development. Angew Chem Int Ed Engl 54 (20), 5948-51.

589 16. Wright, M.H. et al. (2015) Global analysis of protein N-myristoylation and exploration of N-
590 myristoyltransferase as a drug target in the neglected human pathogen Leishmania donovani.
591 Chem Biol 22 (3), 342-54.

592 17. Wright, M.H. et al. (2016) Global profiling and inhibition of protein lipidation in vector and
593 host stages of the sleeping sickness parasite Trypanosoma brucei. ACS Infect Dis 2 (6), 427-441.

594 18. Roberts, A.J. and Fairlamb, A.H. (2016) The N-myristoylome of Trypanosoma cruzi. Sci Rep 6,
595 31078.

596 19. Majeran, W. et al. (2018) Targeted profiling of A. thaliana sub-proteomes illuminates new co-
597 and post-translationally N-terminal Myristoylated proteins. Plant Cell 30 (3), 543-62.

598 20. Towler, D.A. et al. (1988) The biology and enzymology of eukaryotic protein acylation. Annu.
599 Rev. Biochem. 57, 69-99.

600 21. Boutin, J.A. (1997) Myristoylation. Cell Signal 9 (1), 15-35.

601 22. Boisson, B. et al. (2003) Unexpected protein families including cell defense components
602 feature in the N-myristoylome of a higher eukaryote. J Biol Chem 278 (44), 43418-43429.

603 23. Traverso, J.A. et al. (2013) High-throughput profiling of N-myristoylation substrate specificity
604 across species including pathogens. Proteomics 13 (1), 25-36.

605 24. Pierre, M. et al. (2007) N-Myristoylation regulates the SnRK1 pathway in Arabidopsis. Plant
606 Cell 19 (9), 2804-21.

607 25. Giang, D.K. and Cravatt, B.F. (1998) A second mammalian N-myristoyltransferase. J. Biol.
608 Chem. 273 (12), 6595-6598.

609 26. Rundle, D.R. et al. (2002) Characterization of type I and type II myristoyl-CoA:protein N-
610 myristoyltransferases with the Acyl-CoAs found on heterogeneously acylated retinal proteins. Exp
611 Eye Res 75 (1), 87-97.

612 27. Perinpanayagam, M.A. et al. (2013) Regulation of co- and post-translational myristoylation of
613 proteins during apoptosis: interplay of N-myristoyltransferases and caspases. FASEB J 27 (2), 811-
614 21.

615 28. Ashrafi, K. et al. (1998) A role for *Saccharomyces cerevisiae* fatty acid activation protein E in
616 regulating *N*-myristoylation during entry into stationary phase. J Biol Chem 273 (40), 25864-
617 25874.

618 29. Bayer, M. et al. (2009) Paternal control of embryonic patterning in Arabidopsis thaliana.
619 Science 323 (5920), 1485-8.

620 30. Renna, L. et al. (2013) Golgi traffic and integrity depend on N-myristoyl transferase-1 in
621 Arabidopsis. Plant Cell 25 (5), 1756-73.

622 31. Kimura, A. et al. (2012) N-myristoylation of the Rpt2 subunit regulates intracellular localization
623 of the yeast 26S proteasome. Biochemistry 51 (44), 8856-66.

624 32. Traverso, J.A. et al. (2013) Roles of N-Terminal fatty acid acylations in membrane
625 compartment partitioning: Arabidopsis h-type thioredoxins as a case study. Plant Cell 25 (3),
626 1056-77.

627 33. D'Souza-Schorey, C. and Stahl, P.D. (1995) Myristoylation is required for the intracellular
628 localization and endocytic function of ARF6. Exp Cell Res 221 (1), 153-9.

629 34. Resh, M.D. (2006) Trafficking and signaling by fatty-acylated and prenylated proteins. Nat
630 Chem Biol 2 (11), 584-90.

631 35. Goldberg, J. (1998) Structural basis for activation of ARF GTPase: mechanism of guanine
632 nucleotide exchange and GTP-myristoyl switching. Cell 95, 237-248.

633 36. Ames, J.B. and Lim, S. (2012) Molecular structure and target recognition of neuronal calcium
634 sensor proteins. Biochim Biophys Acta 1820 (8), 1205-13.

635 37. Pepperkok, R. et al. (2000) Intracellular distribution of mammalian protein kinase A catalytic
636 subunit altered by conserved Asn2 deamidation. J Cell Biol 148 (4), 715-26.

637 38. Parenti, M. et al. (1993) A novel N-terminal motif for palmitoylation of G-protein α subunits.
638 Biochem J 291 (2), 349-353.

639 39. Degtyarev, M.Y. et al. (1993) The G protein .alpha.s subunit incorporates [3H]palmitic acid and
640 mutation of cysteine-3 prevents this modification. Biochemistry 32 (32), 8057-8061.

641 40. Song, J. and Dohlman, H.G. (1996) Partial constitutive activation of pheromone responses by
642 a palmitoylation-site mutant of a G protein alpha subunit in yeast. Biochemistry 35 (47), 14806-
643 17.

644 41. Koumandou, V.L. et al. (2013) Molecular paleontology and complexity in the last eukaryotic
645 common ancestor. Crit Rev Biochem Mol Biol 48 (4), 373-396.

646 42. Spang, A. et al. (2015) Complex archaea that bridge the gap between prokaryotes and
647 eukaryotes. Nature 521 (7551), 173-179.

648 43. Morita, D. and Sugita, M. (2016) Lipopeptides: a novel antigen repertoire presented by major
649 histocompatibility complex class I molecules. Immunology 149 (2), 139-145.

650 44. Morita, D. et al. (2016) Crystal structure of the N-myristoylated lipopeptide-bound MHC class
651 I complex. Nat Commun 7, 10356.

652 45. Yamamoto, Y. et al. (2019) Identification and structure of an MHC class I–encoded protein
653 with the potential to present N-myristoylated 4-mer peptides to T cells. J Immunol 202 (12), 3349-
654 3358.

655 46. Rowe, D.C. et al. (2006) The myristoylation of TRIF-related adaptor molecule is essential for
656 Toll-like receptor 4 signal transduction. Proc Natl Acad Sci U S A 103 (16), 6299-6304.

657 47. Udenwobele, D.I. et al. (2017) Myristoylation: an important protein modification in the
658 immune response. Front Immunol 8, 751.

659 48. Wen, Z. et al. (2019) N-myristoyltransferase deficiency impairs activation of kinase AMPK and
660 promotes synovial tissue inflammation. Nat Immunol 20, 313–325.

661 49. Nachury, M.V. et al. (2010) Trafficking to the ciliary membrane: how to get across the
662 periciliary diffusion barrier? Ann Rev Cell Dev Biol 26 (1), 59-87.

663 50. Roy, K. and Marin, E.P. (2019) Lipid modifications in cilia biology. J Clin Med 8 (7), 921.

664 51. Constantine, R. et al. (2012) Uncoordinated (UNC)119: coordinating the trafficking of
665 myristoylated proteins. Vision Res 75, 26-32.

666 52. Stephen, L.A. et al. (2018) The ciliary machinery is repurposed for T cell immune synapse
667 trafficking of LCK. Dev Cell 47 (1), 122-132.e4.

668 53. Wright, K.J. et al. (2011) An ARL3–UNC119–RP2 GTPase cycle targets myristoylated NPHP3 to
669 the primary cilium. Genes Dev 25 (22), 2347-2360.

670 54. Chapple, J.P. et al. (2000) Mutations in the N-terminus of the X-linked retinitis pigmentosa
671 protein RP2 interfere with the normal targeting of the protein to the plasma membrane. Hum
672 Mol Genet 9 (13), 1919-1926.

673 55. Emmer, B.T. et al. (2009) Identification of a palmitoyl acyltransferase required for protein
674 sorting to the flagellar membrane. J Cell Sci 122 (6), 867-874.

675 56. Doerig, C. et al. (2015) Post-translational protein modifications in malaria parasites. Nat Rev
676 Microbiol 13 (3), 160-72.

677 57. Schlott, A.C. et al. (2018) N-Myristoylation as a drug target in malaria: exploring the role of N-
678 myristoyltransferase substrates in the inhibitor mode of action. ACS Infect Dis 4 (4), 449-457.

679 58. Rees-Channer, R.R. et al. (2006) Dual acylation of the 45 kDa gliding-associated protein
680 (GAP45) in Plasmodium falciparum merozoites. Mol Biochem Parasitol 149 (1), 113-6.

681 59. Beck, J.R. et al. (2010) A novel family of Toxoplasma IMC proteins displays a hierarchical
682 organization and functions in coordinating parasite division. PLoS Pathog 6 (9), e1001094.

683 60. Frénal, K. et al. (2017) Gliding motility powers invasion and egress in Apicomplexa. Nat Rev
684 Microbiol 15, 645-660.

685 61. Colombo, S. et al. (2005) N-myristoylation determines dual targeting of mammalian NADH-
686 cytochrome b5 reductase to ER and mitochondrial outer membranes by a mechanism of kinetic
687 partitioning. J Cell Biol 168 (5), 735-45.

688 62. Utsumi, T. et al. (2018) Identification and characterization of protein N-myristoylation
689 occurring on four human mitochondrial proteins, SAMM50, TOMM40, MIC19, and MIC25. PLoS
690 One 13 (11), e0206355.

691 63. Araiso, Y. et al. (2019) Structure of the mitochondrial import gate reveals distinct preprotein
692 paths. Nature 575 (7782), 395-401.

693 64. Schorr, S. and van der Laan, M. (2018) Integrative functions of the mitochondrial contact site
694 and cristae organizing system. Semin Cell Dev Biol 76, 191-200.

695 65. van der Laan, M. et al. (2016) Mitochondrial contact site and cristae organizing system. Curr
696 Opin Cell Biol 41, 33-42.

697 66. Ueda, E. et al. (2019) Myristoyl group-aided protein import into the mitochondrial
698 intermembrane space. Sci Rep 9 (1), 1185.

699 67. Humphries, A.D. et al. (2005) Dissection of the mitochondrial import and assembly pathway
700 for human Tom40. J Biol Chem 280 (12), 11535-43.

701 68. Angebault, C. et al. (2018) ER-mitochondria cross-talk is regulated by the Ca(2+) sensor NCS1
702 and is impaired in Wolfram syndrome. Sci Signal 11 (553).

703 69. Baptista, C.G. et al. (2019) Toxoplasma F-box protein 1 is required for daughter cell scaffold
704 function during parasite replication. PLoS Pathog 15 (7), e1007946.

705 70. Thinon, E. et al. (2016) N-Myristoyltransferase Inhibition Induces ER-Stress, Cell Cycle Arrest,
706 and Apoptosis in Cancer Cells. ACS Chem Biol 11 (8), 2165-2176.

707 71. Zha, J. et al. (2000) Posttranslational N-myristoylation of BID as a molecular switch for
708 targeting mitochondria and apoptosis. Science 290 (5497), 1761-1765.

709 72. Utsumi, T. et al. (2003) C-terminal 15 kDa fragment of cytoskeletal actin is posttranslationally
710 N-myristoylated upon caspase-mediated cleavage and targeted to mitochondria. FEBS Lett 539
711 (1-3), 37-44.

712 73. Martin, D.D. et al. (2014) Identification of a post-translationally myristoylated autophagy-
713 inducing domain released by caspase cleavage of huntingtin. Hum Mol Genet 23 (12), 3166-79.

714 74. Sakurai, N. and Utsumi, T. (2006) Posttranslational N-myristoylation is required for the anti-
715 apoptotic activity of human tGelsolin, the C-terminal caspase cleavage product of human gelsolin.
716 J Biol Chem 281 (20), 14288-95.

717 75. Vilas, G.L. et al. (2006) Posttranslational myristoylation of caspase-activated p21-activated
718 protein kinase 2 (PAK2) potentiates late apoptotic events. Proc Natl Acad Sci USA 103 (17), 6542-
719 6547.

720 76. Martin, D.D. et al. (2012) Tandem reporter assay for myristoylated proteins post-
721 translationally (TRAMPP) identifies novel substrates for post-translational myristoylation:
722 PKCepsilon, a case study. FASEB J 26 (1), 13-28.

723 77. Tapodi, A. et al. (2019) BFSP1 C-terminal domains released by post-translational processing
724 events can alter significantly the calcium regulation of AQP0 water permeability. Exp Eye Res 185,
725 107585.

726 78. Bersuker, K. et al. (2019) The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit
727 ferroptosis. Nature 575 (7784), 688-692.

728 79. Doll, S. et al. (2019) FSP1 is a glutathione-independent ferroptosis suppressor. Nature 575
729 (7784), 693-698.

730 80. Conrad, M. et al. (2018) Regulation of lipid peroxidation and ferroptosis in diverse species.
731 Genes & Development 32 (9-10), 602-619.

732 81. Timms, R.T. et al. (2019) A glycine-specific N-degron pathway mediates the quality control of
733 protein N-myristoylation. Science 365 (6448), eaaw4912.

734 82. Jiang, H. et al. (2018) Protein lipidation: occurrence, mechanisms, biological Functions, and
735 enabling technologies. Chem Rev 118 (3), 919-988.

736 83. Cao, J. et al. (2019) HDAC11 regulates type I interferon signaling through defatty-acylation of
737 SHMT2. Proc Natl Acad Sci U S A 116 (12), 5487-5492.

738 84. Dian, C. et al. (2020) High-resolution snapshots of human N-myristoyltransferase in action
739 illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. Nat Commun 11 (1),
740 1132.

741 85. Kosciuk, T. et al. (2020) NMT1 and NMT2 are lysine myristoyltransferases regulating the ARF6
742 GTPase cycle. Nat Commun 11 (1), 1067.

743 86. Burnaevskiy, N. et al. (2013) Proteolytic elimination of N-myristoyl modifications by the
744 Shigella virulence factor IpaJ. Nature 496 (7443), 106-9.

745 87. Burnaevskiy, N. et al. (2015) Myristoylome profiling reveals a concerted mechanism of ARF
746 GTPase deacylation by the bacterial protease IpaJ. Mol Cell 58 (1), 110-22.

747 88. Belda-Palazon, B. et al. (2019) ABA inhibits myristoylation and induces shuttling of the RGLG1
748 E3 ligase to promote nuclear degradation of PP2CA. Plant J 98 (5), 813-825.

749 89. Kim, S. et al. (2017) Myristoylation of Src kinase mediates Src-induced and high-fat diet-
750 accelerated prostate tumor progression in mice. J Biol Chem 292 (45), 18422-18433.

751 90. Colaert, N. et al. (2009) Improved visualization of protein consensus sequences by iceLogo.
752 Nat Methods 6 (11), 786-7.

753 91. Deng, Y. et al. (2013) H+-type and OH- -type biological protonic semiconductors and
754 complementary devices. Sci Rep 3, 2481.

755 92. Maurer-Stroh, S. et al. (2002) N-terminal N-myristoylation of proteins: prediction of substrate
756 proteins from amino acid sequence. J Mol Biol 317 (4), 541-57.

757 93. Alonso, A.M. et al. (2019) Exploring protein myristoylation in Toxoplasma gondii. Exp Parasitol
758 203, 8-18.

759

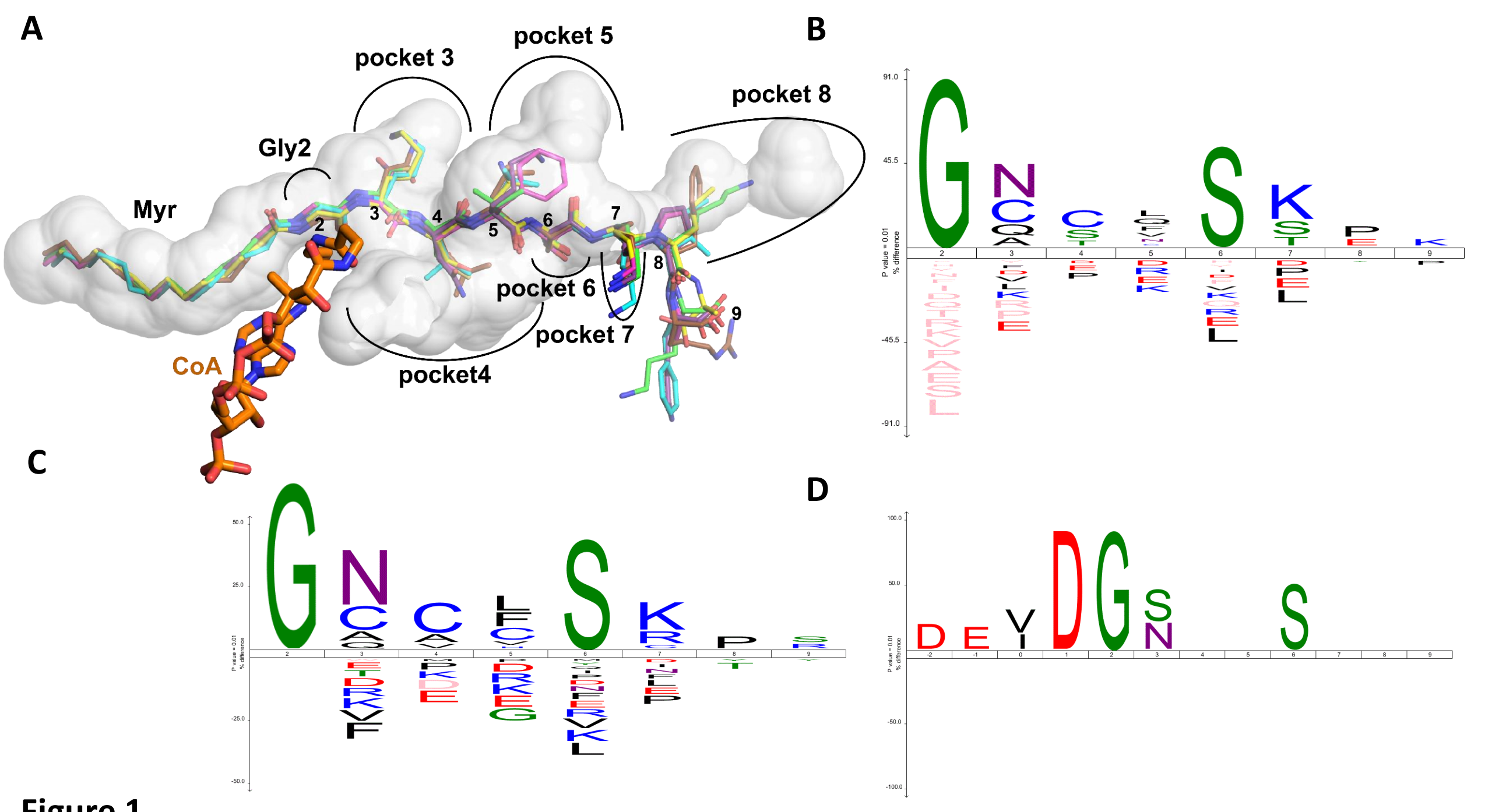
A
pocket 3
pocket 5
pocket 8
Gly2
Myr
2
3
4
5
6
7
8
9
pocket 6
pocket 7
CoA
pocket4
B
91.0
45.5
-45.5
-91.0
P value = 0.01
% difference
2
3
4
5
6
7
8
9
C
50.0
25.0
-25.0
-50.0
D
100.0
50.0
-50.0
-100.0
-2
-1
0
1
2
3
4
5
6
7
8
9

Figure 1

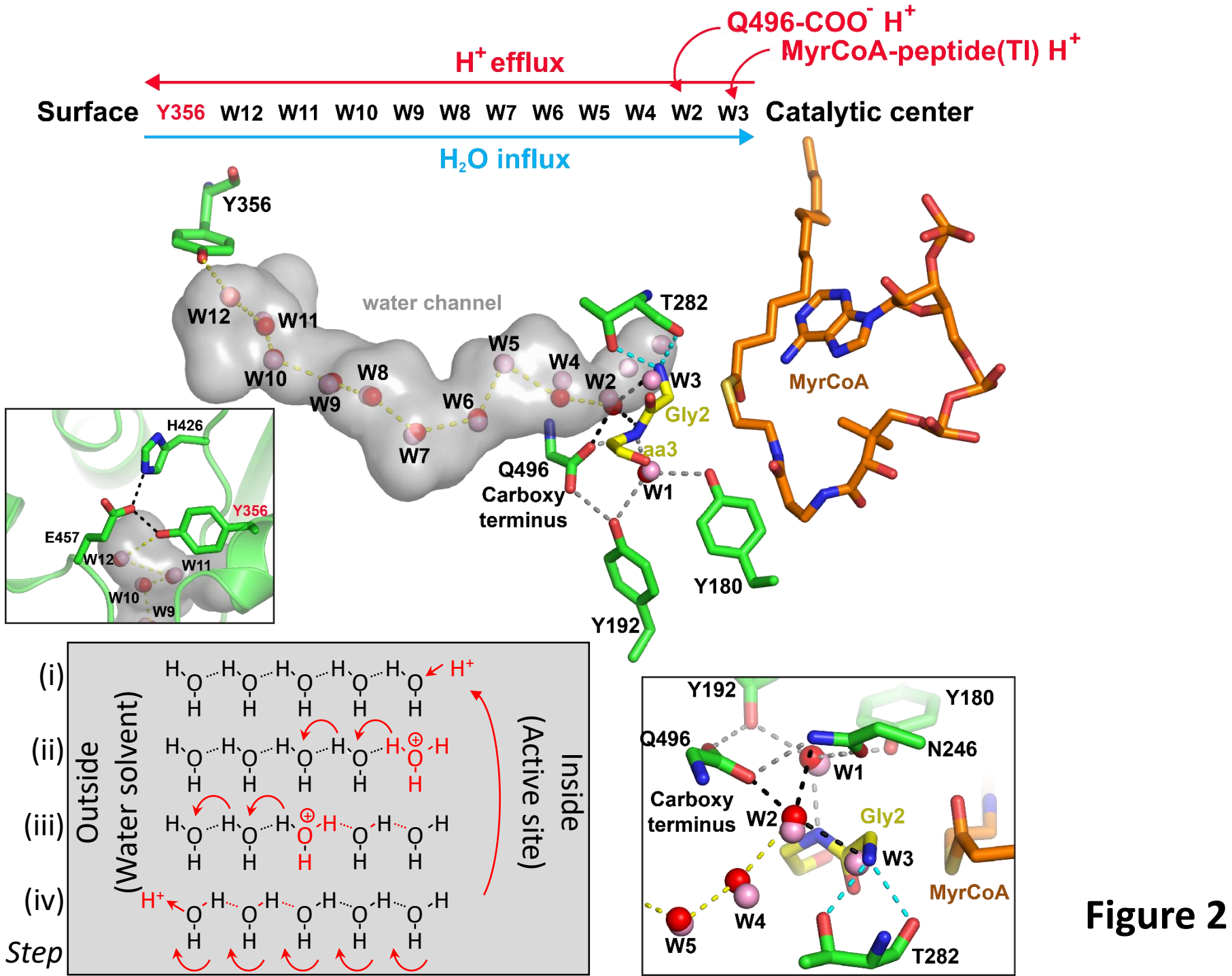
Q496-COO⁻ H⁺
MyrCoA-peptide(TI) H⁺
H⁺ efflux
Surface
Y356 W12 W11 W10 W9 W8 W7 W6 W5 W4 W2 W3
Catalytic center
H2O influx
Y356
W12
W11
W10
W9
W8
W7
W6
W5
W4
W2
W3
water channel
T282
MyrCoA
Gly2
aa3
W1
Q496
Carboxy
terminus
Y180
Y192
H426
E457
Y356
W12
W11
W10
W9
(i)
(ii)
(iii)
(iv)
Step
Outside
(Water solvent)
Inside
(Active site)
H+
Y192
Y180
Q496
N246
W1
Carboxy
terminus
W2
Gly2
W3
MyrCoA
W4
W5
T282

Figure 2

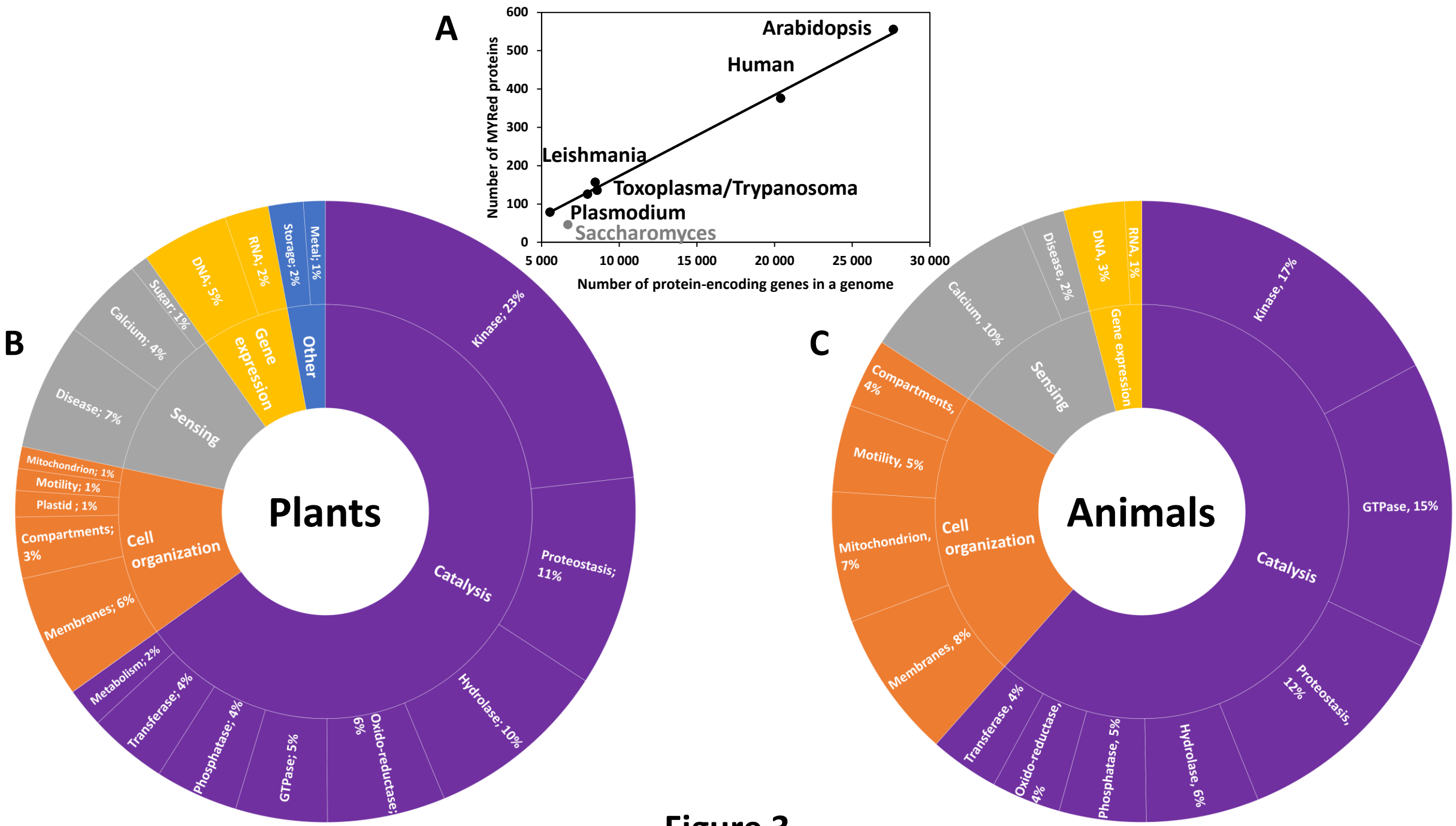
A
Number of MYRed proteins
600
500
400
300
200
100
0
5 000
10 000
15 000
20 000
25 000
30 000
Number of protein-encoding genes in a genome
Arabidopsis
Human
Leishmania
Toxoplasma/Trypanosoma
Plasmodium
Saccharomyces
B
Plants
Kinase; 23%
Proteostasis; 11%
Hydrolase; 10%
Oxido-reductase; 6%
GTPase; 5%
Phosphatase; 4%
Transferase; 4%
Metabolism; 2%
Membranes; 6%
Compartments; 3%
Plastid ; 1%
Motility; 1%
Mitochondrion; 1%
Catalysis
Cell organization
Disease; 7%
Calcium; 4%
Sugar; 1%
Sensing
DNA; 5%
RNA; 2%
Gene expression
Storage; 2%
Metal; 1%
Other
C
Animals
Kinase, 17%
GTPase, 15%
Proteostasis, 12%
Hydrolase, 6%
Phosphatase, 5%
Oxido-reductase, 4%
Transferase, 4%
Membranes, 8%
Mitochondrion, 7%
Motility, 5%
Compartments, 4%
Catalysis
Cell organization
Calcium, 10%
Disease, 2%
Sensing
DNA, 3%
RNA, 1%
Gene expression

Figure 3

## Box 1

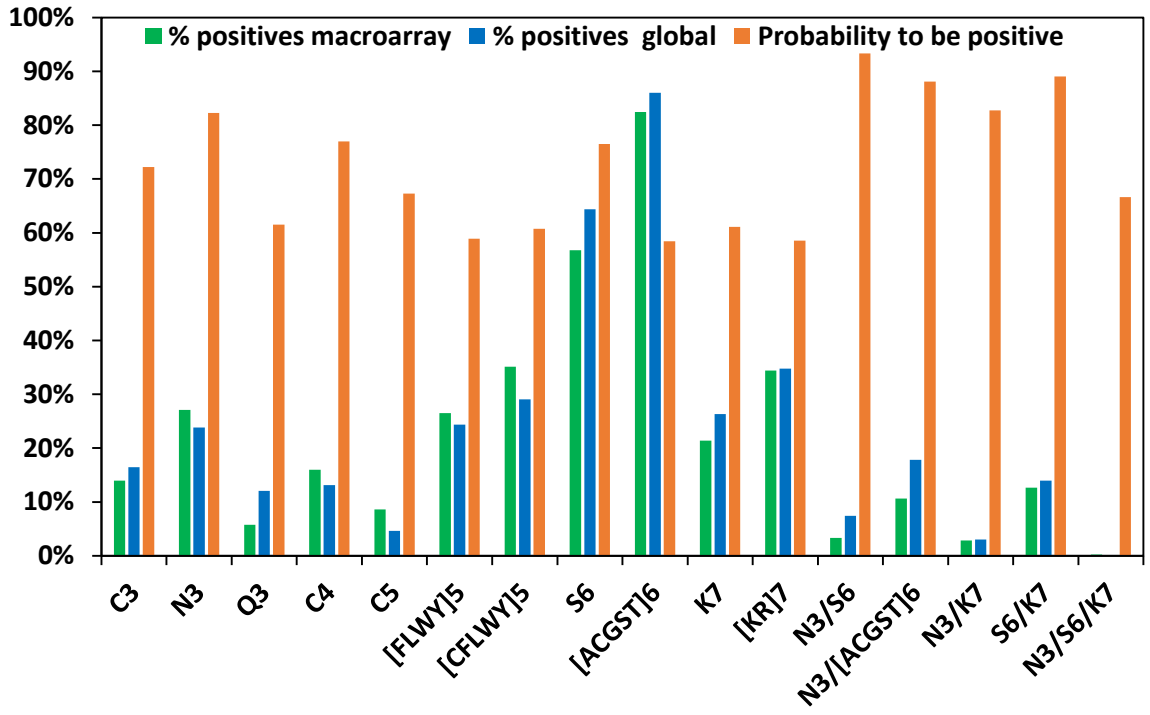


**Figure I**