



HAL
open science

Découverte du rôle crucial du résidu en position 2 des séquences MTS d'adressage mitochondrial

Houssam El Barbry

► **To cite this version:**

Houssam El Barbry. Découverte du rôle crucial du résidu en position 2 des séquences MTS d'adressage mitochondrial. Biologie cellulaire. Sorbonne Université, 2023. Français. NNT : 2023SORUS035 . tel-04062087

HAL Id: tel-04062087

<https://theses.hal.science/tel-04062087v1>

Submitted on 7 Apr 2023

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.

Sorbonne Université

ED515 Complexité du Vivant

Laboratoire de Biologie Computationnelle et Quantitative

Equipe Structure des Réseaux Génétiques

Découverte du rôle crucial du résidu en position 2 des séquences MTS d'adressage mitochondrial

Par Houssam El Barbry

Thèse de doctorat de Génétique et Génomique

Dirigée par DEVAUX Frédéric et co-encadrée par GARCIA Mathilde

Présentée et soutenue publiquement le 10 février 2023

Devant un jury composé de :

Hubert Dominique BECKER, Professeur d'université, Rapporteur

Carmela GIGLIONE, Directrice de Recherche, Rapportrice

Agnès RÖTIG, Directrice de recherche, Examinatrice

Benoit PALANCADE, Directeur de Recherche, Examineur

Hubert François BECKER, Maître de Conférences, Examineur

Frédéric DEVAUX, Professeur d'université, Directeur de thèse

Mathilde GARCIA, Maître de Conférences, Co-encadrante de thèse

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury de thèse Hubert Dominique Becker, Hubert François Becker, Carmela Giglione, Benoit Palancade et Agnès Rötig, qui ont eu la gentillesse d'accepter de prendre du temps pour évaluer mon travail.

Je tiens à remercier Alessandra Carbone de m'avoir permis d'intégrer le laboratoire dans lequel j'ai eu l'opportunité d'effectuer cette thèse.

Je remercie mon directeur de thèse et directeur d'équipe, Frédéric Devaux, pour son professionnalisme, sa pédagogie, sa bienveillance et la patience qu'il a eue à mon égard durant ces trois années de recherche. Je garderai en mémoire nos nombreuses discussions très sérieuses.

Je remercie Mathilde Garcia, ma co-encadrante de thèse, pour sa bienveillance et sa disponibilité. Sa rigueur, couplée à sa gentillesse, a été pour moi un moteur au quotidien. Je suis conscient d'avoir eu beaucoup de chance de pouvoir mener à bien mon projet dans un environnement agréable, empli d'ondes positives.

Je tiens aussi à remercier Thierry Delaveau, l'ingénieur de recherche de l'équipe, pour sa bonne humeur constante. Il a su partager avec moi son savoir infini, mais aussi de nombreux plaisirs sucrés.

Je veux remercier également les autres membres de mon équipe : Stéphane Le Crom et Pierre Louis Blaiseau.

Je remercie également Oana Ilioaia, Stéphane Delmas et Nicolas Agier, avec qui j'ai eu le plaisir de partager de succulents repas à la cantine. Oana, une personne profondément gentille et authentique, avec qui j'ai passé de très bons moments. Stéphane Delmas, un amoureux de football, qui a su partager ma peine lors des rares éliminations du PSG. Nicolas Agier, alias ultimate, avec qui j'ai eu plaisir à troller.

Merci à Frédéric Chaux, pour sa bonne humeur constante. Merci à Gilles Fisher et Zhou Xu, pour leurs conseils.

Je remercie l'ensemble des membres de mon comité de thèse Sandrine Lanfranchi, Naima Belgareh et Geneviève Dujardin qui ont pris de leur temps pour suivre l'évolution de ma thèse.

Merci à Jeanne Le Peillet, avec qui j'ai partagé mon bureau, pour sa joie de vivre et sa folie.

Thank you Lexuan Liu, I enjoyed learning new Chinese words with you.

Merci à Salomé Nashed, ma sœur de thèse, pour ton aide, ta bienveillance et ta bonne humeur.

Je suis fier d'avoir travaillé avec elle et sincèrement content que nous soyons devenus amis.

Je tiens particulièrement à remercier mes parents et mes sœurs qui ont toujours cru en moi et qui m'ont toujours soutenu. شكرا لك يا أبي ، هذه الدكتوراه لك

Merci à Audrey sans qui je n'aurais sûrement pas fait de doctorat. Merci pour son soutien, son aide, sa patience et son amour. J'espère qu'elle trouvera toujours en moi le soutien que j'ai trouvé en elle durant ma thèse. Enfin, merci à Oia ma petite princesse ! Le plus beau chat du monde.

Le plus important c'est de participer

Table des matières

Remerciements	2
Introduction	9
I. L'import des protéines mitochondriales codées par le génome nucléaire	13
1. Le complexe TOM	14
2. Les différents complexes TIM	18
a. Le complexe TIM23	18
b. Le complexe TIM22	20
3. Les mitochondrial targeting sequences (MTS)	20
a. Caractéristiques biochimiques et prédiction des MTS	21
b. Algorithmes de prédiction	22
c. L'import des protéines avec MTS	23
d. Devenir du MTS après import	26
4. Les autres signaux d'import	28
a. Protéines de la membrane interne	28
b. Protéines de la Membrane Externe	30
c. Protéines de l'espace intermembranaire	31
5. Import mitochondrial et cytotoxicité	34
6. Import mitochondrial et dynamique traductionnelle	36
a. Localisation des ARNm	37
b. Protéines chaperonnes capable d'interagir avec les protéines mitochondriales	39
II. Les modifications N-terminales co-traductionnelles jouent un rôle important dans le devenir des protéines	42
1. Les enzymes de modifications N-terminales co-traductionnelles	42
a. La Méthionine aminopeptidase	42
b. Les N-Acétyltransférases (NATs)	44

2.	Impact de ces modifications sur le devenir des protéines.....	50
a.	Dégradation (Voie N-End Rule)	50
b.	Adressage et interaction protéines/protéines.....	55
c.	Repliement des protéines	55
3.	Biais de composition à la position 2	57
III.	Problématique de la thèse : exploration du lien entre les biais de composition en position 2, NatC et l’import des protéines mitochondriales.....	60
1.	Identification d’un biais en position 2 des protéines mitochondriales.....	60
2.	Confirmation de l’interaction entre l’enzyme NatC et les protéines mitochondriales	62
3.	Problématique de la thèse	64
	Résultats	65
I.	L’article : Functional mapping of N-terminal residues in the yeast proteome uncovers novel determinants for mitochondrial protein import	65
	Résultat complémentaires non inclus dans l’article	118
II.	Exploration fonctionnelle de l’impact de NatC sur les précurseurs mitochondriaux.	119
1.	Préambule	119
2.	La perte du génome mitochondrial (rho0) est un facteur de confusion dans l’interprétation des mutants NatC	120
3.	Construction et caractérisation de souches rho+ mutées pour NatC.....	124
a)	Phénotype de croissance ralentie pour les souches <i>Δmak3</i>	124
b)	Phénotype moléculaire des souches délétées de NatC	127
4.	L’arg/ N End rule empêche-t-elle la détection du rôle de NatC ?	129
a)	Création de souches délétées d’Ubr1 et de doubles mutants Mak3/Ubr1	129
b)	Étude des souches délétées d’Ubr1 dans le fond <i>YPH-pam16Δ-hpam16(N76D)</i>	132
5.	Utilisation du système de criblage Hsp60-myc pour étudier l’effet de la déplétion de l’enzyme NatC.....	133
6.	De nouveaux résultats encourageants ?	135
	Discussion	139
I.	Le résidu en position 2 du MTS joue un rôle critique sur le devenir de Hsp60.....	139

1. Le résidu en position 2 de Hsp60 semble avoir un rôle sur son import mitochondrial...	139
2. ... mais certains de mes résultats suggèrent une réalité plus complexe.	141
II. Exploration globale du rôle de NatC	143
1. Un lien avéré entre NatC et les mitochondries	143
2. Les mutants de NatC ont un phénotype respiratoire mais leurs mitochondries respirent normalement	144
3. L'impact de la délétion de NatC sur Hsp60 est difficile à caractériser par Western Blot	144
4. L'acétylation par NatC joue peut-être un rôle mineur sur le devenir de Hsp60 et n'est peut-être pas la raison de la conservation de la leucine en position 2	145
III. Conclusion générale	148
Matériels et Méthodes	150
1. Souches disponibles au laboratoire et milieux de croissance	150
2. Amplification par PCR de cassettes de délétion par recombinaison homologue	150
3. Transformation et sélection phénotypique des levures en vue de l'obtention des mutants	151
4. Vérification des mutants : PCR et séquençage	152
5. Tests de croissance des souches de levures en milieux solide et liquide	153
6. Analyse des contenus protéiques par Western Blot	154
7. Test de complémentation fonctionnelle pour l'analyse de la perte potentielle du génome mitochondrial	154
Références	156
Table des figures :	178
Table des tableaux	181
Résumé en anglais :	182
Résumé en français :	183

Introduction

Les mitochondries sont des organites qui ont été décrits pour la première fois en 1857 par Rudolf Albert Von Kölliker lorsqu'il a visualisé des granules dans des fibres musculaires striées (Amar-Costesec, n.d., chap. 4). Après des études plus approfondies de cet organite nouvellement identifié, la communauté scientifique décide de l'appeler « mitochondrie », à cause de sa forme en bâtonnet. Le terme est en effet formé des mots grecs *mitos*, qui veut dire « fil » et *chondros*, signifiant « grain ». Les mitochondries sont présentes dans quasiment toutes les cellules eucaryotes. Elles jouent un rôle clé pour la cellule puisqu'elles sont à l'origine de la production d'une grande partie de l'énergie cellulaire sous forme d'ATP. Cependant, il est aujourd'hui admis qu'elles sont impliquées dans de nombreux autres processus essentiels comme la croissance cellulaire, la différenciation, l'apoptose (Osellame et al., 2012).

Les mitochondries sont délimitées par deux membranes, et comprennent donc quatre sous-compartiments : la membrane externe, l'espace intermembranaire, la membrane interne et la matrice (Frey and Mannella, 2000) (Figure 1).

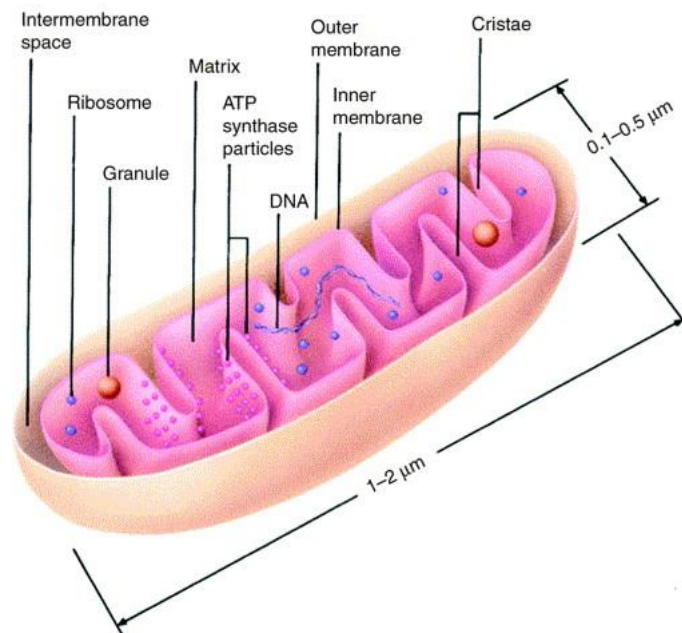


Figure 1 : Schéma simplifié d'une mitochondrie avec ses 4 sous-compartiments (Frey and Mannella, 2000).

L'origine des mitochondries est expliquée par l'hypothèse dite endosymbiotique. D'après cette hypothèse, les mitochondries seraient le résultat de l'endosymbiose de l'ancêtre de la cellule eucaryote actuelle et d'une α -protéobactérie à gram négatif (Gray, 2012; Roger et al., 2017). L'un des arguments en faveur de cette hypothèse est le fait que la mitochondrie possède son propre patrimoine génétique, proche de celui des α -protéobactéries, mais qui, aujourd'hui, ne contient qu'un nombre très réduit de gènes. Le génome mitochondrial de la levure *S. cerevisiae* par exemple présente 35 gènes, dont 8 codant des protéines (D'après la revue (Malina et al., 2018)). En effet, suite à un transfert horizontal massif, la quasi-totalité des gènes mitochondriaux a été transférée au noyau des cellules eucaryotes. Ainsi la très grande majorité des protéines mitochondriales est désormais codée dans le génome nucléaire, ce qui pose la question de la définition du protéome dit mitochondrial.

D'après la théorie de l'endosymbiose, on pourrait envisager que l'identification de la majorité des protéines constituant le protéome mitochondrial puisse se faire par simple étude d'homologie avec le protéome d'une α -protéobactérie. Or, cette démarche n'est pas suffisante, car elle ne prend pas en compte le fait qu'au cours de l'évolution de nombreuses protéines eucaryotes sont devenues mitochondriales, ni l'apparition de novo de gènes codant des protéines mitochondriales (Takeda et al., 2022). D'après une revue publiée en 2012, seules 10 à 15% des protéines mitochondriales de levure proviennent clairement de la lignée α -protéobactérienne. De 20 à 30% des protéines mitochondriales n'ont aucun homologue évident chez les bactéries tandis qu'environ 20% d'entre elles n'ont aucun homologue identifiable chez d'autres eucaryotes ou procaryotes (Gray, 2012). Par ailleurs, on peut imaginer que des gènes codant des protéines mitochondriales ont perdu au cours de l'évolution leur fonctionnalité mitochondriale pour en acquérir une nouvelle.

L'étude du protéome mitochondrial est un enjeu majeur pour mieux comprendre et caractériser le rôle de cet organite. La technique la plus utilisée pour définir le protéome mitochondrial est l'étude par spectrométrie de masse des protéines présentes dans les mitochondries après purification. Grâce à cette technique, en 2006, une équipe a identifié 851 protéines mitochondriales chez la levure (Reinders et al., 2006). Une étude plus récente publiée en 2017 avait pour objectif d'identifier un maximum de protéines mitochondriales chez *S. cerevisiae* et de préciser leur sous-localisation mitochondriale. Pour cela, les auteurs ont utilisé différentes techniques : le marquage des protéines avec des acides aminés présentant des isotopes lourds, diverses stratégies d'enrichissement et d'extraction de protéines et spectrométrie de masse quantitative. Comme présenté sur la figure 2, cette étude a permis l'identification de 986

protéines mitochondriales et les auteurs ont pu trier 818 protéines en fonction de leur sous-compartiment mitochondrial (Vögtle et al., 2017) : 481 protéines sont destinées à la membrane interne, 95 à la membrane externe, 49 à l'espace intermembranaire et 193 à la matrice. Cette étude a également permis l'identification de 206 protéines qui n'étaient jusqu'alors pas considérées comme étant mitochondriales (Figure 2). Pour 14 d'entre elles, la localisation mitochondriale a été confirmée par un test d'import *in vitro*.

A l'heure actuelle, les techniques de biochimie ne peuvent pas fournir une liste exhaustive des protéines mitochondriales. On ne peut exclure la présence de faux positifs liés à des contaminations par des protéines non mitochondriales lors des étapes de purification, ou de faux négatifs liés au contraire à des pertes de protéines mitochondriales durant ces expériences. C'est pourquoi des approches *in silico* complémentaires aux approches biochimiques ont été mises en place pour estimer le nombre de protéines mitochondriales. Il s'agit du développement d'algorithmes de prédiction basés pour la plupart sur la détection des séquences d'import mitochondrial (voir chapitre I.2.d). C'est le cas par exemple de l'algorithme Predator, capable de prédire la présence d'une séquence d'adressage N-terminale portée par certaines protéines mitochondriales (Small et al., 2004).

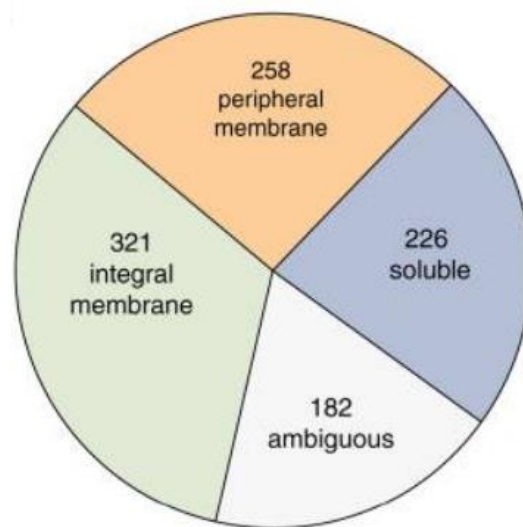


Figure 2 : Répartition des 818 protéines identifiées en fonction de leur compartiment dans une étude réalisée par Vogtle en 2017 en purifiant les mitochondries (Vögtle et al., 2017).

En résumé, en fonction des publications et des différentes bases de données on estime le nombre de protéines mitochondriales chez *S. cerevisiae* comme variant entre 800 et 1200. D'après le GO consortium, en 2022, 1191 protéines sont annotées comme appartenant à la catégorie fonctionnelle « Mitochondrion » chez la levure. Cependant, mon directeur de thèse a estimé à partir de nombreuses recherches bibliographiques que seules 726 protéines sur ces 1191 peuvent être considérées comme mitochondriales avec un risque relativement faible de faux positifs, car leur localisation a été validée par plusieurs approches fournissant des résultats solides.

Comme le protéome mitochondrial est en grande majorité composé de protéines codées dans le génome nucléaire et synthétisées dans le cytoplasme, cela pose la problématique de la coordination de l'import de ces protéines vers les mitochondries. En effet, la cellule a développé au cours de l'évolution plusieurs systèmes d'adressage de ces protéines vers ces organites.

Dans la première partie de cette introduction, je m'intéresserai aux différents systèmes d'import mitochondriaux et je me focaliserai en particulier sur celui impliquant la reconnaissance d'une séquence d'adressage mitochondriale N-terminale appelée MTS pour Mitochondrial Targeting Sequence. Dans une deuxième partie, je présenterai les principales enzymes reconnaissant le résidu en position 2 des protéines et modifiant leur extrémité N-terminale de façon co-translationnelle. Enfin, dans une troisième partie, je présenterai la problématique de ma thèse qui fait le lien entre import mitochondrial et modifications N-terminales des chaînes naissantes.

I. L'import des protéines mitochondriales codées par le génome nucléaire

Seule une fraction des nombreuses protéines présentes dans le cytosol doit être importée dans la mitochondrie. De plus, les protéines mitochondriales n'ont pas toutes la même destination dans la mitochondrie. C'est pourquoi il existe différentes voies d'import qui vont impliquer différents signaux portés par les protéines à destination des mitochondries (Chacinska et al., 2009)(Figure 3). On peut distinguer 3 groupes de signaux. Le plus répandu est situé à l'extrémité N-terminale des protéines et est clivé après import. Ce signal est appelé « Mitochondrial Targeting Sequence » ou MTS. Il y a également les signaux qui ne sont pas clivés et dont la localisation dans la séquence de la protéine est très variable. Enfin, il existe un signal interne riche en cystéines qui est spécifique aux protéines ayant pour destination la membrane interne. Quel que soit le signal porté par les protéines destinées aux mitochondries, il doit pouvoir être reconnu à la surface externe de ces dernières. Ainsi l'import des protéines mitochondriales met en jeu différents complexes protéiques reconnaissant ces signaux. Dans cette partie, je vais commencer par présenter les deux principaux complexes d'import : le complexe de membrane externe (TOM) et le complexe de membrane interne (TIM). Je décrirai ensuite dans le détail les différents signaux d'import et je finirai en exposant le lien entre l'import mitochondrial des protéines, la dynamique de traduction et l'homéostasie du protéome cellulaire.

Targeting and sorting signals of mitochondrial precursor proteins			
A	Presequences and variations	Imported via	Destination of proteins
	Presequence Amphipathic α helix Typically cleaved after import 	TOM TIM23 PAM	Matrix
	Presequence + uncleaved hydrophobic anchor (sorting signal) 	TOM TIM23 (PAM)	Inner membrane
	Presequence + cleaved hydrophobic sorting signal (bipartite presequence) 	TOM TIM23 (PAM)	Intermembrane space
B	Noncleavable signals of hydrophobic proteins		
	β signal 	TOM SAM	Outer membrane (β -barrel)
	Signal-anchor sequence 	Mim1	Outer membrane (α -helical)
	C tail anchor 	?	
	Internal signal 	(TOM) (SAM)	Inner membrane (metabolite carrier)
	Multiple internal signals 	TOM Tim9-Tim10 TIM22	
	Presequence-like internal signal (after hydrophobic region) 	TOM TIM23	Inner membrane
C	Internal signal for intermembrane space		
	Cysteine-containing signal 	TOM MIA	Intermembrane space

Figure 3 : Les différents signaux d’adressage des protéines mitochondriales. (A) Les signaux clivables d’adressage présents à l’extrémité N-terminale des protéines (B) les signaux internes et non clivables (C) le signal interne spécifique à l’adressage des protéines dans l’espace intermembranaire. (d’après (Chacinska et al., 2009))

1. Le complexe TOM

Le complexe TOM est un complexe protéique composé de 7 protéines : Tom40, Tom22, Tom20, Tom70, Tom5, Tom6 et Tom7 (Neupert and Herrmann, 2007). D’après une revue publiée en 2020, le complexe TOM est responsable de près de 90% de l’import mitochondrial (Palmer et al., 2021). Ce complexe est en fait constitué de protéines qui vont jouer le rôle de récepteurs et d’autres qui vont être directement impliquées dans la translocation des protéines vers l’intérieur des mitochondries.

L’acteur principal de ce complexe est Tom40. Cette protéine est chargée de transloquer les précurseurs mitochondriaux via le tonneau qu’elle forme. C’est une protéine en barillet β qui s’associe à Tom22, Tom5, Tom6 et Tom7 par leurs hélices transmembranaires pour former un complexe stable (Neupert and Herrmann, 2007) (Figure 4).

Dans le complexe TOM, trois sous-unités jouent le rôle de protéines réceptrices : Tom22, Tom20 et Tom70. Elles ont pour rôle de sélectionner et de guider les protéines mitochondriales

vers le canal d'import. Tom22 joue un rôle essentiel dans la stabilisation et la formation du complexe TOM mature (Shiota et al., 2015; Wiedemann et al., 2003). Sa région N-terminale est exposée au cytosol et sa région C-terminale à l'espace intermembranaire. En plus de jouer son rôle stabilisateur pour le complexe TOM, cette protéine est capable de reconnaître les séquences d'adressage et ainsi de favoriser l'import des protéines portant ces signaux via Tom40 (Brix et al., 1999; Yamano et al., 2008). Sa région située dans l'espace intermembranaire permet de faciliter le transfert de la protéine en cours d'import de Tom40 vers Tim50, la protéine réceptrice du complexe TIM associé à la membrane interne (Shiota et al., 2011).

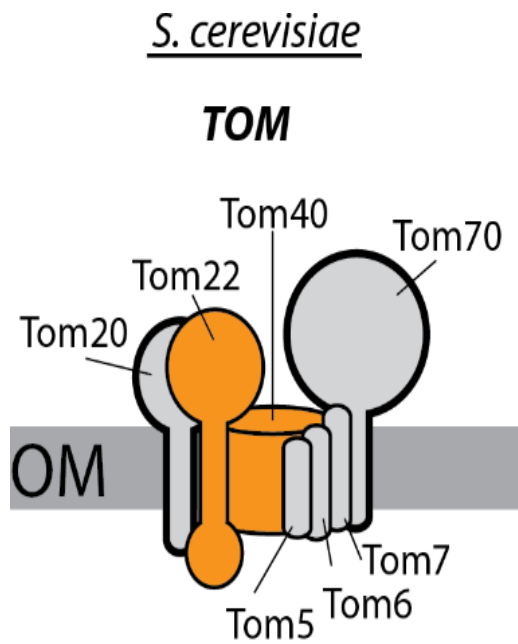
La protéine Tom20 est une protéine réceptrice située dans la membrane externe de la mitochondrie. Elle possède un domaine alpha hélicoïdal fiché dans la membrane et sa région C-terminale se trouve dans le cytosol. Elle possède également un domaine tétratricopeptide (TPR) qui lui permet d'interagir avec des protéines chaperonnes comme HSP70. Il a par ailleurs été démontré que cette protéine est capable d'interagir directement avec les séquences d'adressage N-terminales de certaines protéines mitochondriales, les MTS, qui ont pour principale caractéristique de former des hélices alpha amphiphiles. En effet, deux études structurales de Tom20 (Abe et al., 2000; Saitoh et al., 2007) ont montré que les résidus hydrophobes composant un côté de l'hélice amphipathique des MTS pénètrent dans le sillon de liaison hydrophobe de Tom20, tandis que les charges positives de ces MTS interagissent avec les résidus chargés négativement à la limite du sillon de liaison (Brix et al., 1997). Ces caractéristiques font de Tom20 une protéine capable de recruter un large panel de précurseurs mitochondriaux.

Tout comme Tom20, Tom70 possède un domaine transmembranaire formant une hélice alpha amphiphile, et sa région C-terminale baigne dans le cytosol. Grâce à des sites TPR elle interagit avec des protéines chaperonnes des précurseurs mitochondriaux et des domaines hydrophobes de ces précurseurs, et elle est capable de les transmettre à la sous-unité Tom22 lors de l'import. Cependant, Tom70 a la particularité de pouvoir reconnaître des signaux internes. Ces signaux, dont je détaillerai les caractéristiques plus loin, sont appelés (i-MTS) (Backes et al., 2018). L'intérêt de ce recrutement serait de faciliter l'import des protéines concernées en les maintenant dépliées. En effet les protéines mitochondriales doivent être dépliées pour être importées et leur repliement prématuré perturberait leur import (Backes et al., 2018). Tom20 et Tom70 interagissent avec des protéines communes et il existe une certaine redondance fonctionnelle entre ces sous-unités puisque la délétion de l'une d'elles n'est pas létale, contrairement à une double délétion. Cependant, une publication récente indique que le rôle

majeur de Tom70 est surtout de lier les chaperonnes à la membrane externe et ainsi d'empêcher l'accumulation de certaines protéines dans le cytosol, ce qui serait toxique (Backes et al., 2021). En effet, il a été mis en évidence, parmi les cibles de Tom70, un enrichissement en protéines qui ont une forte propension à s'agréger. C'est le cas notamment des petites protéines de la membrane interne. Il a été démontré qu'une sur-expression de ces protéines en absence de *TOM70* est hautement toxique pour la cellule. De manière intéressante, les auteurs de cette étude ont également montré que la délétion du site responsable de l'interaction de Tom70 avec les chaperonnes comme Hsp70, rend Tom70 inefficace, ce qui a confirmé le rôle de ce site de reconnaissance. D'après ces éléments, Tom70 serait donc une interface entre le système de chaperonnes cytosoliques et la machinerie d'import mitochondrial. Cette sous-unité facilite l'import de certaines protéines mais son rôle principal serait la protection du cytosol dans des conditions de stress protéotoxique causé par les précurseurs mitochondriaux.

Chez *S. cerevisiae*, des études ont été réalisées pour caractériser la structure de l'ensemble du complexe. Une étude récente a permis de visualiser la structure du complexe TOM avec une résolution quasi atomique grâce à la technique de cryomicroscopie électronique (Palmer et al., 2021). Les auteurs ont constaté que le complexe peut se trouver dans différentes conformations. On peut trouver le complexe sous forme de dimère, trimère et tétramère (Figure 4) (Palmer et al., 2021; Shiota et al., 2015). Le complexe central TOM est dépourvu de Tom70 et Tom20 et a une masse moléculaire totale d'environ 150 kDa. En effet, Tom20 et Tom70 sont situées à la périphérie du complexe et s'en dissocient facilement, contrairement à Tom22 (Dekker et al., 1998).

A



B

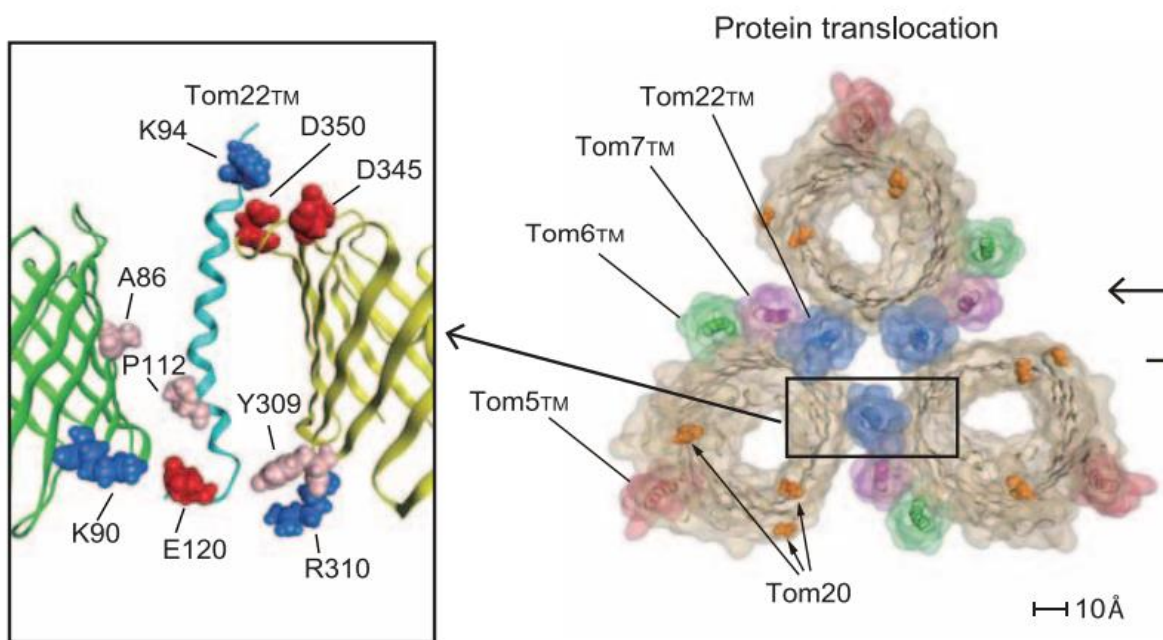


Figure 4 : A) Schéma simplifié de l'organisation des différentes protéines du complexe TOM (Schneider, 2018) (B) Organisation des sous-unités du complexe TOM les unes par rapport aux autres. La structure a été obtenue avec la technique de Cryomicroscopie électronique (Shiota et al., 2015).

2. Les différents complexes TIM

Les informations sur le fonctionnement et la composition du complexe TIM répertoriées ci-dessous sont essentiellement issues de deux revues récentes (Günsel and Mokranjac, 2019; Palmer et al., 2021). Le complexe TIM est présent dans la membrane interne de la mitochondrie. Il est divisé en deux complexes : le complexe appelé TIM23 qui a pour rôle de transporter les protéines de l'espace intermembranaire vers la matrice et le complexe TIM22, exclusivement chargé de l'adressage de protéines vers la membrane interne.

a. Le complexe TIM23

Le complexe TIM23 est constitué de 11 protéines que l'on peut classer selon trois fonctions différentes (Figure 5) : les protéines réceptrices, Tim50 et Tim23, les protéines formant le canal de translocation Tim23 et Tim17, et les protéines jouant le rôle moteur dans l'import des protéines mitochondriales mtHSP70 (Ssc1), Tim14 (Pam18), Tim16 (Pam16) et Mge1 (Figure 5).

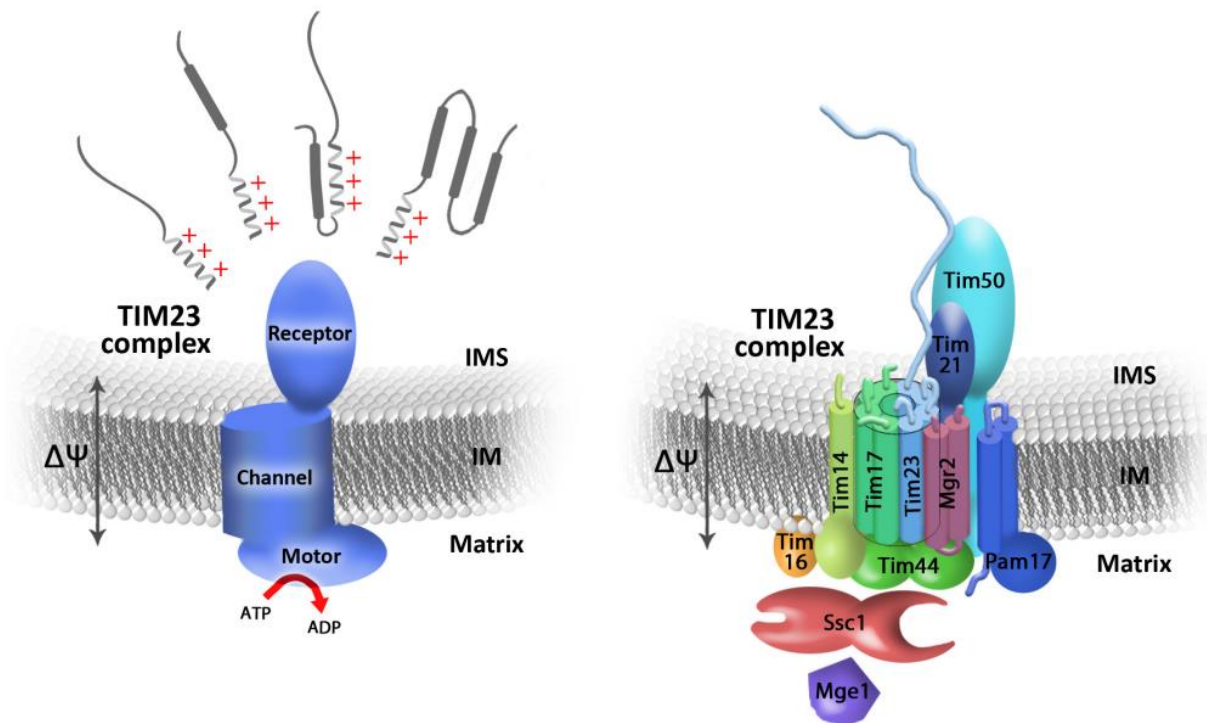


Figure 5 : Représentation schématique du complexe TIM23. Le complexe TIM23 peut être divisé fonctionnellement en trois unités : récepteur, canal de translocation et moteur d'import.

11 sous-unités du complexe TIM23 ont été identifiées à ce jour. Ce sont Tim23, Tim17, Tim50, Tim44, Tim14, Tim16, mtHsp70 (Ssc1), Mge1, Tim21, Mgr2 et Pam17 (panneau de droite) (Günsel and Mokranjac, 2019).

Tim50 est la protéine réceptrice principale du complexe TIM. Grâce à son domaine C-terminal présent dans l'espace intermembranaire, elle reconnaît les séquences d'adressage N-terminales portées par les protéines mitochondriales en cours d'import dès leur sortie du complexe TOM. Elle est d'ailleurs située à proximité de ce dernier. Lors de cette reconnaissance, Tim50 coopère avec Tim23. Les mutations qui perturbent l'interaction entre ces deux sous-unités altèrent la fonction du récepteur Tim50 ce qui a pour conséquence de réduire l'efficacité d'import et donc la croissance des levures. Les protéines formant le canal de translocation des protéines mitochondriales sont Tim23 et Tim17, qui appartiennent à la même famille. Elles possèdent toutes deux quatre segments membranaires qui permettent leur insertion dans la membrane interne. Tim44 est la protéine qui fait le lien entre le canal de translocation et le moteur d'import. En effet son domaine C-terminal recrute Tim17 et Tim23 tandis que son domaine N-terminal recrute les composants du moteur d'import. C'est également une protéine reconnaissant les séquences d'adressage N-terminales.

PAM (« Translocase-Associated Motor ») constitue le moteur d'import. Il est composé de mtHsp70 (ou Ssc1 chez la levure) et ses co-chaperonnes Tim44, Tim14 (Pam18), Tim16 (Pam16) et Mge1 (Chacinska et al., 2009). mtHSP70 est le composant central du moteur, qui fournit la force mécanique nécessaire pour assurer le transport unidirectionnel des protéines mitochondriales via l'hydrolyse d'ATP. Tim14 (Pam18), Tim16 (Pam16) et Mge1 jouent un rôle régulateur de mtHSP70.

Suite à leur passage dans le canal d'import de la membrane interne, les protéines sont libérées dans la matrice ou sont insérées dans la membrane interne. Le processus de décision entre ces deux destinées possibles n'est pas encore totalement connu mais trois protéines non essentielles présentes dans le complexe semblent y jouer un rôle : Tim21, Mgr2 et Pam17. Mgr2 joue un rôle dans le couplage entre la protéine Tim21 et le complexe TIM23. En son absence on observe une altération du couplage TOM-TIM23 et l'import des précurseurs est perturbé (Gebert et al., 2012). De plus elle contrôle la libération latérale des protéines dans la membrane interne. En effet, une souche de levure délétée de Mgr2 présente une libération excessive des protéines dans la membrane interne et l'inverse est observé dans une souche dans laquelle Mgr2 est sur-exprimée.

A l'heure actuelle le rôle de Tim21 et Pam17 reste mal défini. Une publication de 2008 indique que ces deux protéines jouent un rôle antagoniste (Popov-Čeleketić et al., 2008). Pam17 est indirectement responsable de la stabilité de l'interaction Tim23-Tim21. En effet, la délétion de

Pam17 entraîne une réduction de l'expression de Mgr2 laquelle est importante pour la stabilité du complexe Tim23-Tim21 (Moulin et al., 2019).

b. Le complexe TIM22

Il existe un autre complexe d'import présent dans la membrane interne, spécialisé dans l'insertion de protéines en cours d'import dans la membrane interne. Il ne concerne que les protéines portant des signaux internes d'adressage. L'import via ce complexe n'utilise que le potentiel de membrane. Il est composé de 7 sous-unités : Tim22, Tim18, Tim54, Sdh3, Tim9, Tim10 et Tim12. La protéine Tim22 forme le canal de ce complexe. Tim54 est la protéine réceptrice recrutant le complexe Tim9-Tim10-Tim12. Tim9-Tim10 présents dans l'espace intermembranaire vont jouer le rôle de protéines chaperonnes pour les précurseurs afin qu'ils ne s'agrègent pas. Ils vont ensuite transférer les protéines au complexe Tim9-Tim10-Tim12 appartenant lui-même au complexe TIM22 et Tim18-Sdh3 formant le complexe d'amarrage. Tim18 jouerait également un rôle d'ancrage de Tim22 et Sdh3 semble être nécessaire à la biogenèse de Tim18 et Tim22 (D'après les revues (Ferramosca and Zara, 2013; Wiedemann and Pfanner, 2017)). Je détaillerai l'import de ces protéines via le complexe TIM22 dans la partie I.4.a.

3. Les mitochondrial targeting sequences (MTS)

Les MTS sont les séquences N-terminales portées par un grand nombre de protéines mitochondriales et qui permettent leur import vers les mitochondries via les complexes TOM et TIM. Suite à l'import des précurseurs, les MTS sont clivés et dégradés. Les MTS ont été découverts en 1979 par l'observation chez la levure, d'une différence de taille des 3 sous-unités composant le complexe F1-ATPase avant et après import dans la mitochondrie (Maccacchini et al., 1979). Dans les années 80 de nombreuses études ont montré que ce signal d'adressage est situé au niveau N-terminal des protéines. Des expériences ont par exemple permis l'import de protéines non mitochondriales dans la mitochondrie en plaçant en amont de ces dernières seulement les quelques premiers acides aminés d'un précurseur mitochondrial (Emr et al., 1986; Hurt et al., 1984; van Loon et al., 1986; van Loon and Young, 1986). Ainsi la fusion des 27 premiers acides aminés d'Atp1 avec la GFP permet son import dans la mitochondrie (Emr et al., 1986).

Une étude publiée en 2009 a estimé qu'environ 70% des protéines mitochondriales possèdent un MTS. La taille des MTS peut être très variable, mais elle est généralement comprise entre 15 et 55 acides aminés (Vögtle et al., 2009). Ce signal d'adressage permet surtout l'import des protéines dans la matrice, mais en combinaison avec d'autres signaux, il peut également permettre un adressage dans tous les autres sous-compartiments mitochondriaux (Chapitre I.3.b et figure 3).

a. Caractéristiques biochimiques et prédiction des MTS

Il n'existe pas de séquence consensus simple pour définir les MTS (Allison and Schatz, 1986). Plusieurs publications importantes en 1986 ont fait émerger l'idée que la présence d'hélices alpha amphiphiles en N-terminal est une caractéristique fondamentale des MTS. Ainsi, l'étude des préséquences de 23 protéines mitochondriales de cellules eucaryotes a montré que quasiment toutes ces protéines sont capables de former des hélices alpha amphiphiles N-terminales (von Heijne, 1986). La même année, (Roise et al., 1986) ont étudié de manière in vitro la structure du MTS de Cox4 et ont montré que les 18 premiers résidus du MTS de cette protéine sont capables de former une hélice alpha amphiphile. Par ailleurs, le laboratoire de Gottfried Schatz a créé une préséquence synthétique de 14 acides aminés permettant l'adressage efficace vers les mitochondries du rapporteur Cox4 tronqué de son propre MTS, en n'utilisant que 3 types d'acides aminés : Leucine, Sérine et Arginine (Allison and Schatz, 1986). Cette séquence synthétique avait les caractéristiques requises pour former une hélice alpha amphiphile. En 2009, l'étude de 105 protéines mitochondriales de levure a montré que 71% d'entre elles sont capables de former une hélice alpha amphiphile sur les 10 premiers acides aminés (Huang et al., 2009). Ce résultat a été obtenu grâce à l'utilisation par les auteurs d'un algorithme de prédiction de la structure secondaire, Jpred3. Dans la même étude les auteurs ont analysé des protéines mitochondriales d'*Arabidopsis thaliana* et d'*Oryza sativa* et ont remarqué qu'elles ont également une forte propension à former des hélices alpha sur les 10 premiers acides aminés, pour 83% et 90,4% d'entre elles respectivement. L'autre élément qui montre que les MTS ont pour caractéristique la formation d'hélices alpha amphiphiles est leur enrichissement en acides aminés hydrophobes et chargés positivement comme l'Arginine et au contraire leur sous-enrichissement en acides aminés chargés négativement, comme l'acide aspartique et l'acide glutamique (Huang et al., 2009; von Heijne, 1986). Des études ont montré que des mutations consistant à remplacer certains acides aminés chargés positivement en N-

terminal par d'autres acides aminés chargés négativement perturbent l'import des protéines (Horwich et al., 1986; Kumamoto et al., 1987).

En résumé, les MTS forment une hélice alpha amphiphile qui se caractérise par un enrichissement en acides aminés chargés positivement et en acides aminés hydrophobes et un sous-enrichissement en acides aminés chargés négativement.

Il a été démontré de nombreuses fois qu'il existe une conservation fonctionnelle inter-espèce des MTS. Cela est illustré par le fait que des préséquences d'adressage N-terminal de protéines humaines sont fonctionnelles dans des cellules végétales et inversement (Baysal et al., 2020; Lee et al., 2020). En effet, les MTS présents dans ces espèces sont enrichis en acides aminés chargés positivement et ont une forte propension à former des hélices alpha amphiphiles (Huang et al., 2009).

b. Algorithmes de prédiction

Des algorithmes de prédiction ont été mis en place depuis plus de 25 ans afin de prédire la présence d'un MTS pour une protéine donnée. Les plus utilisés sont MitoProtII (Claros, 1995), TargetP (Emanuelsson et al., 2000), Predatr (Small et al., 2004), TPpred3.0 (Savojardo et al., 2015) et MitoFates (Fukasawa et al., 2015). MitoProtII et MitoFates sont les seuls algorithmes spécifiques aux MTS, tandis que les autres permettent aussi la recherche de diverses séquences telles que la séquence d'adressage aux chloroplastes. Ces différents algorithmes n'utilisent pas les mêmes méthodes d'analyse, ce qui explique le fait qu'ils n'ont pas le même taux d'efficacité de prédiction du MTS et/ou du site de clivage. D'après la littérature, MitoFates, et TargetP2.0 sont les algorithmes les plus performants. Une étude récente réalisée sur des protéines mitochondriales de souris a montré que Mitofates est capable de prédire la préséquence MTS à 75,9% et TargetP2.0 à 79,9% (Imai and Nakai, 2020).

La prédiction par MitoFates se base, notamment, sur la combinaison entre la composition des premiers acides aminés et leurs propriétés physicochimiques (Fukasawa et al., 2015). L'algorithme de prédiction est capable d'attribuer à un résidu un score en fonction de son hydrophobicité et de sa capacité à former une hélice alpha amphiphile avec le résidu situé sur la face opposée. De plus, nous savons que la majorité des protéines mitochondriales à MTS ont une arginine en position -2 du site de clivage par la MPP (voir chapitre I.3.d.). C'est un

paramètre qui est également pris en compte par MitoFates. D'autres algorithmes de prédiction, comme par exemple TargetP 2.0, utilisent des approches de type « machine learning » ou « deep learning » pour prédire les signaux d'imports ou alors pour mettre en évidence de nouvelles caractéristiques. Par exemple, à partir de protéines mitochondriales avérées, l'IA est capable de trouver des points communs entre ces protéines et d'utiliser ces paramètres pour prédire la présence de MTS dans de nouvelles protéines (Almagro Armenteros et al., 2019).

c. L'import des protéines avec MTS

La première association entre le complexe TOM et le MTS d'un précurseur mitochondrial se fait via une interaction hydrophobe entre ce dernier et la région cytosolique hydrophobe de Tom20 (Mukhopadhyay et al., 2006). La séquence signal est ensuite reconnue par Tom22 (Moczko et al., 1997). Cette interaction va permettre de stabiliser et de favoriser le passage du précurseur dans le pore que forme le tonneau β de Tom40. De plus, la surface de ce pore est chargée négativement ce qui permet d'attirer les MTS chargés positivement pour initier la translocation (Palmer et al., 2021) (Figure 6).

Les protéines avec MTS qui ont pour destination la matrice sont prises en charge par le complexe TIM23. C'est la protéine réceptrice Tim50 du complexe qui va prendre en charge le précurseur et le transmettre au complexe TIM23. La dissociation de la région C-terminale de Tom22 et du précurseur est favorisée par Tim21 (Chacinska et al., 2005). La présence du précurseur active le canal que forme Tim23 (Bauer et al., 1996; Truscott et al., 2001). Le potentiel de membrane va alors conduire la protéine portant le MTS vers la membrane interne. C'est Tim44 et le complexe PAM qui vont permettre de libérer le précurseur dans la matrice (Frazier et al., 2004; Krimmer et al., 2000) (Figure 6).

Même si la majorité des protéines possédant un MTS ont pour destination la matrice, certaines d'entre elles doivent être adressées à la membrane interne ou à l'espace intermembranaire. Elles ont alors des signaux hydrophobes situés directement en aval de leur MTS. Cette séquence va jouer le rôle d'arrêt de la voie d'import au niveau de l'espace intermembranaire. La protéine va commencer son processus d'import dans la matrice via le complexe TIM23. Ce processus d'import va s'arrêter quand la séquence d'arrêt se trouve dans la membrane interne. Pour les protéines de la membrane interne, cette séquence est conservée et fait partie intégrante de la

protéine mature. En revanche pour les protéines destinées à l'espace intermembranaire le MTS présent dans la matrice est clivé par la MPP et une autre protéase va cliver la séquence hydrophobe toujours fichée dans la membrane interne. La protéine mature va ainsi être libérée dans l'espace intermembranaire.

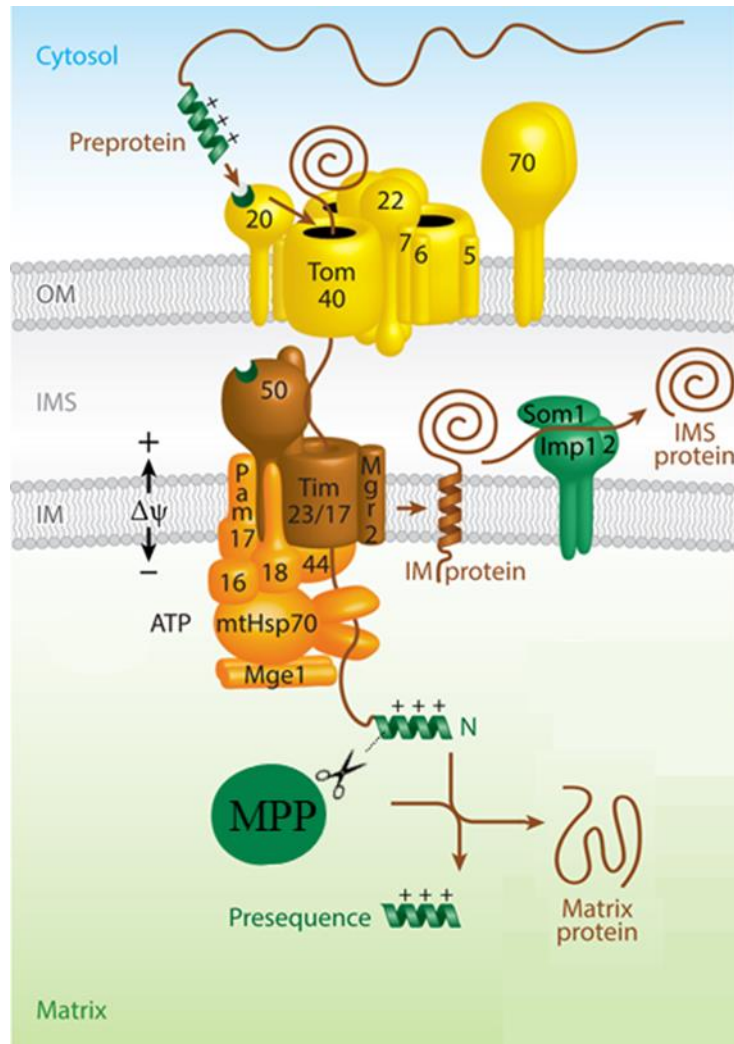


Figure 6 : : La voie d'adressage des protéines avec MTS vers la membrane interne et la matrice mitochondriale. Les précurseurs portant une préséquence d'adressage N-terminale (MTS) sont reconnus par les récepteurs Tom20 et Tom22 et sont transloqués à travers la membrane externe par Tom40. Le récepteur Tim50 se lie aux précurseurs et les transfère au canal TIM23, qui est étroitement associé à Tim17. Le potentiel membranaire ($\Delta\psi$) active le canal TIM23 et entraîne la translocation des préséquences chargées positivement. La translocation des protéines dans la matrice est conduite par le moteur associé à la translocase des préséquences (PAM) avec la protéine mitochondriale mtHsp70 dépendante de l'ATP. Tim44 couple mtHsp70 au canal TIM23. La peptidase MPP clive les préséquences. Les précurseurs présentant un signal de tri hydrophobe peuvent être libérés latéralement dans la membrane interne, sous le contrôle du portier latéral Mgr2. Pour certains précurseurs, le signal de tri est éliminé par la peptidase de la membrane interne (IMP, avec les sous-unités Imp1 et Imp2), et les protéines sont libérées dans l'espace intermembranaire (IMS). (Modifié d'après (Wiedemann and Pfanner, 2017))

d. Devenir du MTS après import

Lorsqu'un précurseur est importé dans la matrice, son MTS est clivé très rapidement par des aminopeptidases, grâce au site de clivage présent dans le MTS. La MPP est l'aminopeptidase impliquée dans ce phénomène qui est la mieux décrite à ce jour, et le site de clivage qu'elle reconnaît est un acide aminé situé 2 positions après une arginine (R-2) (Vögtle et al., 2009).

D'autres enzymes sont impliquées dans le passage de la forme précurseure à la forme mature des protéines mitochondriales. Une analyse globale du N-protéome de mitochondries de levure a montré que dans des conditions sauvages, on retrouve une forme mature suite au clivage du MTS des précurseurs par la MPP et une forme dite intermédiaire qui est la forme mature +1 acide aminé (Vögtle et al., 2009) (Figure 7). On trouve exclusivement la forme intermédiaire dans une souche délétée pour le gène de la peptidase *ICP55*. Ainsi, un sous-ensemble de protéines portant des séquences clivables subissent deux clivages successifs : un premier par la MPP qui va cliver les protéines au niveau d'un site situé en R-2, et un deuxième clivage de l'acide aminé N-terminal par Icp55 (Figure 7).

Une troisième peptidase a été identifiée. Il s'agit d'Oct1, qui reconnaît un site de clivage se situant généralement 10 acides aminés après une Arginine, ce qui correspond à 8 acides aminés après un premier clivage par la MPP (Isaya et al., 1992; Vögtle et al., 2011) (Figure 7).

En résumé, la MPP est une enzyme qui va cliver le MTS lorsque la protéine est importée dans la matrice. Certaines protéines peuvent subir un deuxième clivage par Icp55 ou Oct1 (Vögtle et al., 2011).

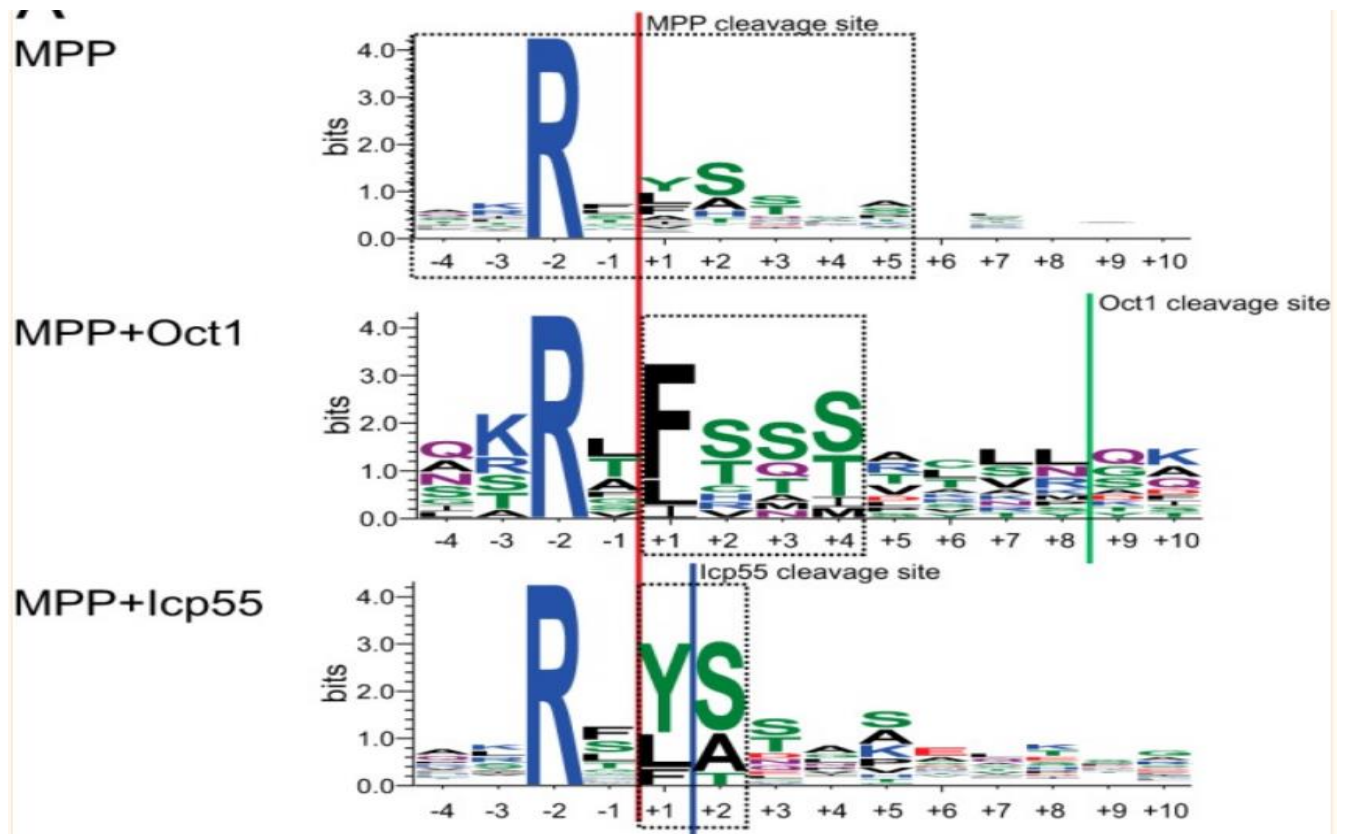


Figure 7 : Les différents sites de clivages des MTS en fonction de l'enzyme MPP seule ou avec Icp55 ou Oct1, Icp55 et Oct1. Ces sites de coupure ont été obtenus à partir de données de protéomique (Fukasawa et al., 2015).

Chez la levure, la protéase Cym1 est responsable de la dégradation des MTS libres via un processus ATP dépendant. A l'heure actuelle, le mécanisme moléculaire par lequel agit cette enzyme n'est pas connu. L'élimination de ces MTS clivés est critique pour la cellule, car une accumulation de MTS dans la mitochondrie conduit à une inhibition de l'action de la MPP par rétrocontrôle positif, ce qui sera à l'origine d'une déstabilisation de la mitochondrie (Mossmann et al., 2014). Par ailleurs, il a été démontré que la protéine Prep, l'homologue humaine de Cym1, est impliquée dans la maladie d'Alzheimer. En effet cette enzyme est capable de dégrader les peptides bêta-amyloïdes à l'intérieur des mitochondries, peptides dont l'accumulation est caractéristique de cette maladie (Falkevall et al., 2006).

4. Les autres signaux d'import

Je vais maintenant présenter les autres signaux portés par les protéines qui n'ont pas de MTS. Cela concerne toutes les protéines de la membrane externe, de nombreuses protéines de la membrane interne et quelques protéines matricielles (D'après la revue (Chacinska et al., 2009)).

a. Protéines de la membrane interne

L'import des protéines qui n'ont pas de MTS dans la membrane interne fait intervenir des signaux internes (i-MTS) (D'après les revues (Horten et al., 2020; Manganas et al., 2017)).

Ces signaux peuvent être nombreux et répartis tout le long de la protéine. De plus ils ont les mêmes caractéristiques que les MTS (Backes et al., 2018). Une étude a prédit les iMTS de protéines mitochondriales à l'aide de l'algorithme TargetP permettant la prédiction des MTS N-terminaux. D'ailleurs, si ces signaux internes sont situés en N-terminal ils peuvent permettre l'import des protéines au même titre que les MTS (Backes et al., 2018). Ces signaux hydrophobes internes sont reconnus par Tom70. De plus, les protéines portant ces signaux peuvent aussi être ciblées par HSP70 (Figure 8).

Ces protéines mitochondriales traversent ensuite le pore du complexe TOM et sont prises en charge par les protéines chaperonnes Tim9 et Tim10 (Figure 8). La formation de ce complexe empêche l'agrégation des précurseurs dans l'espace intermembranaire, un phénomène qui peut arriver d'autant plus probablement que les protéines ont des séquences hydrophobes comme c'est le cas pour les protéines membranaires. Elles sont ensuite transférées vers le complexe protéique TIM22 qui insère les protéines dans la membrane interne via un potentiel de membrane.

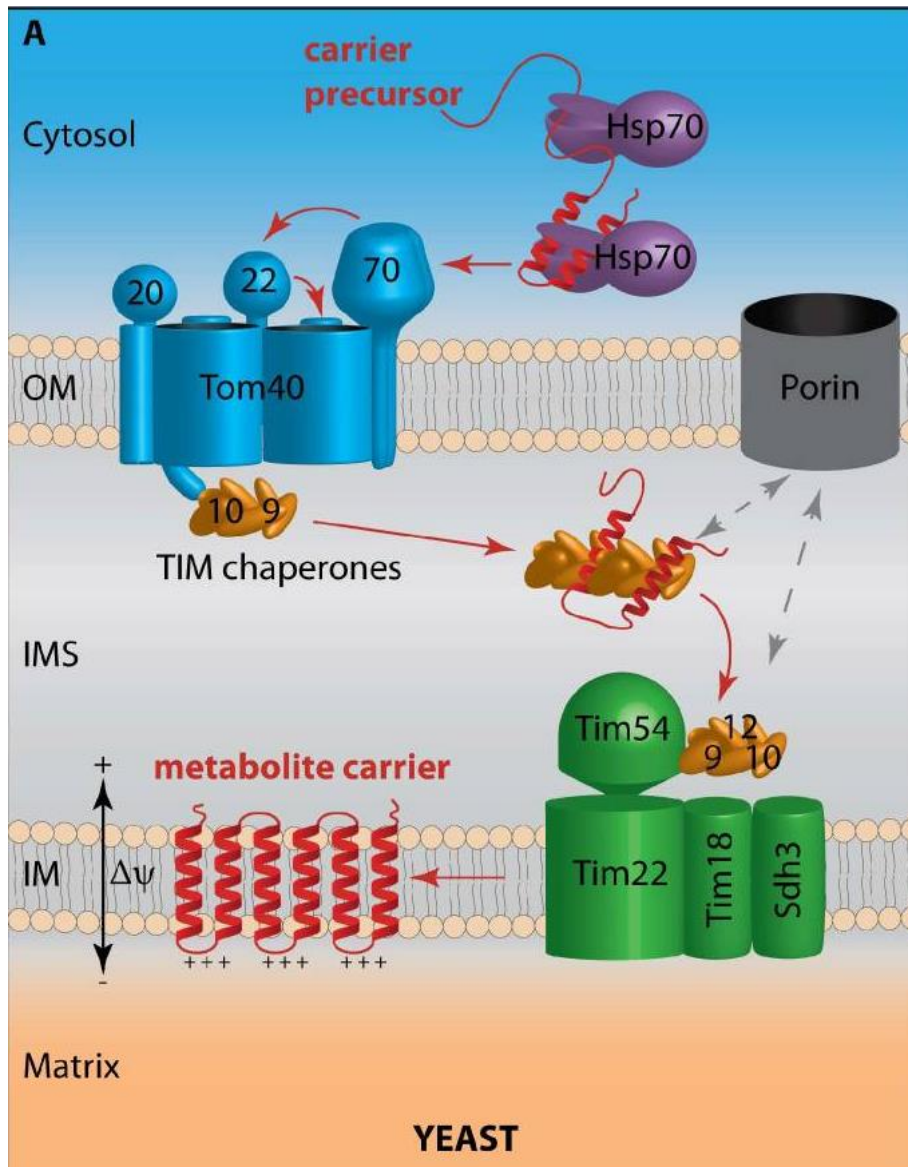


Figure 8 : Les différentes étapes de l'import des protéines mitochondriales qui ont pour destination la membrane interne. Les précurseurs sont liés par des chaperons dans le cytosol et reconnus au niveau de la translocase de la membrane externe (TOM) par le récepteur Tom70. Après leur transfert à travers la membrane externe, ils sont liés dans l'espace intermembranaire par les chaperons TIM hexamériques, Tim9-Tim10. Les chaperons TIM guident le précurseur à travers le compartiment aqueux jusqu'au complexe chaperon TIM lié à la membrane composé de Tim9-Tim10-Tim12 dans la levure. Le transfert du substrat vers la translocase porteuse de la membrane interne (TIM22) est facilité par les interactions avec les protéines de la membrane externe (flèches en pointillés) impliquant le canal métabolite porine/VDAC chez la levure. La translocase porteuse TIM22 intègre les précurseurs dans la membrane interne d'une manière dépendante du potentiel de membrane. (Horten et al., 2020)

b. Protéines de la Membrane Externe

Il existe deux types de protéines membranaires intégrales destinées à la membrane externe : celles qui forment des barillets β et celles qui ont un ou plusieurs domaines α hélicoïdaux. En fonction du type de protéine, une voie d'import différente est privilégiée.

Les protéines en barillet β ne sont associées qu'aux mitochondries et aux chloroplastes. Elles sont importées via le complexe TOM et le complexe SAM. Ce dernier est composé des trois protéines Sam50, Sam35 et Sam37, Sam50 étant la protéine principale du complexe. Seule la délétion de Sam37 n'empêche pas le bon fonctionnement du complexe (Chan and Lithgow, 2008). Comme pour les protéines de la membrane interne, les précurseurs sont transférés vers le complexe TOM à l'aide de la protéine chaperonne HSP70 avant d'être importés dans l'espace intermembranaire. À l'aide de protéines chaperonnes Tim9 et Tim10, les précurseurs sont guidés vers le complexe SAM qui va se charger de les insérer dans la membrane externe (Höhr et al., 2015) (Figure 9). La reconnaissance des protéines en barillet β se fait grâce à un signal présent en C-terminal de ces protéines (Imai et al., 2011; Kutik et al., 2008). Ce signal suit un motif en particulier : PoXGXXHyXHy. Po étant un acide aminé polaire, X un acide aminé quelconque, G une glycine et Hy un acide aminé hydrophobe de grande taille (Imai et al., 2011).

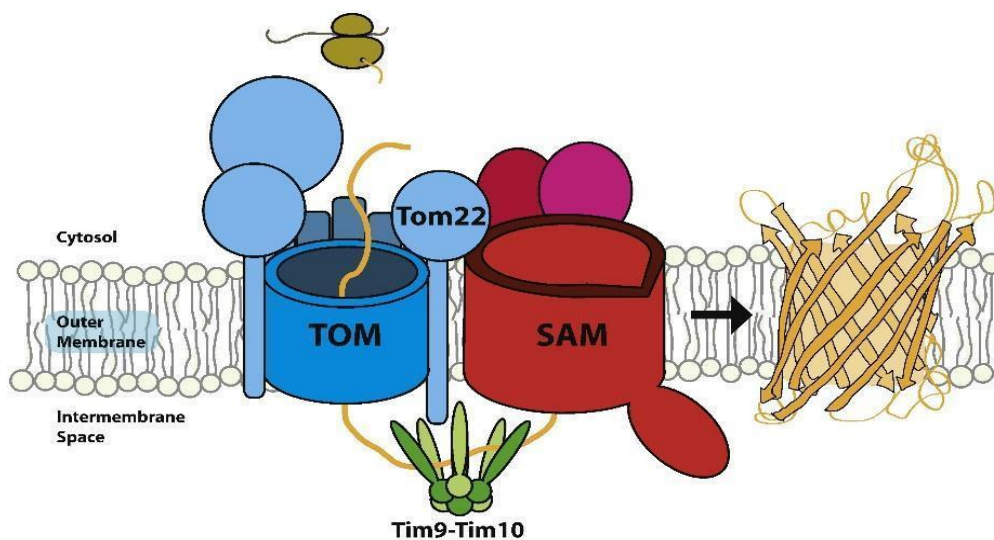


Figure 9 : Schéma simplifié de la machine d'importation des protéines mitochondriales dans la membrane externe. Les protéines sont synthétisées dans le cytosol avant d'être guidées par le complexe TOM. Dans l'espace intermembranaire, elles sont ensuite prises en charge par des protéines chaperonnes. Enfin, le complexe SAM insère les protéines dans la membrane externe. (Höhr et al., 2015)

Il existe une autre voie d'import des protéines membranaires qui ont un ou plusieurs segments transmembranaires hydrophobes disposés en hélices alpha. Cette voie implique la reconnaissance d'un signal pouvant être localisé en C-terminal, comme c'est le cas pour Tom5, Tom6 et Tom7. Il peut également s'agir de la reconnaissance d'un signal situé en N-terminal, comme pour Tom20 et Tom70 (Doan et al., 2020) (Figure 3). L'adressage de ces protéines s'appuie majoritairement sur le complexe MIM. Ce complexe est composé des deux protéines Mim1 et Mim2. La délétion d'une de ces sous-unités induit un phénotype de croissance très ralentie (Dimmer et al., 2012). Encore une fois, la protéine chaperone HSP70 va jouer un rôle d'adressage des précurseurs situés dans le cytosol vers la sous-unité Tom70, qui transfère les précurseurs vers le complexe MIM, lequel assure leur insertion dans la membrane externe (Becker et al., 2011; Papić et al., 2011). Cependant, certaines protéines peuvent être importées indépendamment du complexe Tom, via la coopération des complexes MIM et SAM (Doan et al., 2020).

c. Protéines de l'espace intermembranaire

L'espace intermembranaire mitochondrial n'abrite que 5% du protéome mitochondrial. Pourtant deux systèmes d'import assurent l'adressage des protéines vers ce sous-compartiment.

Le premier système d'adressage des protéines de l'espace intermembranaire utilise un « signal d'arrêt » (d'après la revue (Edwards et al., 2021)). Ce système concerne les protéines portant un MTS décrit précédemment et une séquence hydrophobe de type transmembranaire située en aval du MTS (voir chapitre I.3.c.). Cette séquence joue le rôle d'arrêt de la voie d'import au niveau de l'espace intermembranaire. La protéine commence son processus d'import dans la matrice via le complexe TIM23. Ce processus s'arrête quand la séquence d'arrêt se trouve dans la membrane interne. Le MTS présent dans la matrice est alors clivé par la MPP et des protéases IMP1/2 clivent la protéine juste avant la séquence hydrophobe toujours présente dans la membrane interne. La protéine mature est ainsi libérée dans l'espace intermembranaire (Figure 10).

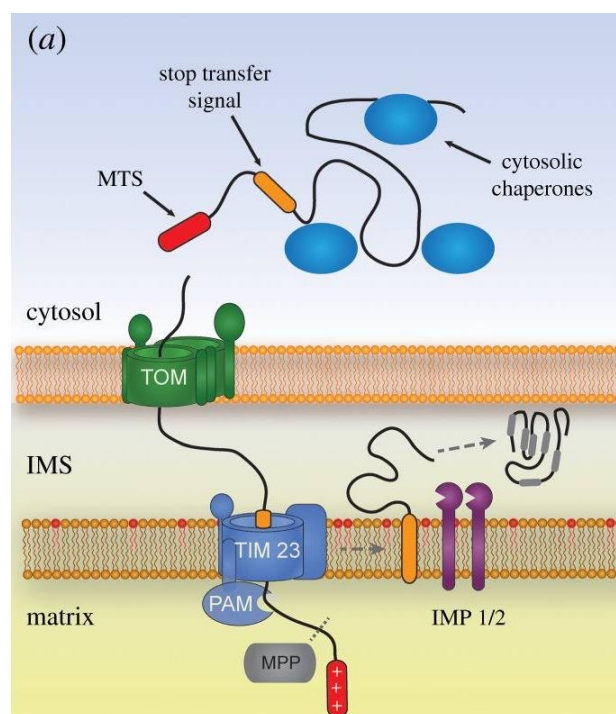


Figure 10 : Certaines protéines ont deux signaux d’adressage leur permettant d’être adressées dans la membrane interne de la mitochondrie. Elles ont un signal à l’extrémité N-terminale (MTS) qui va permettre à la protéine d’être reconnue et importée dans l’espace intermembranaire via le complexe TOM et de commencer son adressage dans la matrice à l’aide du complexe TIM23. Cette translocation dans la matrice va être bloquée lorsqu’un deuxième signal hydrophobe, situé en aval du MTS, entre dans le complexe TIM23. Ce blocage va provoquer la translocation du signal hydrophobe dans la membrane interne. Cela va libérer la protéine mature dans l’espace intermembranaire après que son MTS ait été clivé par la MPP et son extrémité C terminale clivée par des protéases Imp1 et Imp2 (D’après (Edwards et al., 2021)).

La deuxième voie concerne la majorité des protéines qui ont des domaines riches en cystéines. La protéine Mia40 est responsable de leur adressage dans l’espace intermembranaire et de leur repliement. Il s’agit d’une oxydoréductase qui agit comme une protéine donneuse de disulfure pour les précurseurs importés (Banci et al., 2009). Elle est ancrée dans la membrane interne et entraîne le repliement des précurseurs grâce à l’introduction de liaisons disulfures. Le repliement de ces précurseurs les piège dans l’espace intermembranaire. L’interaction entre Mia40 et ses substrats est réalisée via l’association entre des motifs comprenant 3 ou 9 acides aminés entre deux cystéines portées par ces substrats (CX3C ou CX9C), et une fente

hydrophobe présente dans Mia40 (Sideris et al., 2009). Afin que la protéine Mia40 puisse de nouveau jouer son rôle auprès d'autres substrats, elle doit être réoxydée par la sulphydrile oxydase Erv1 (Figure 11).

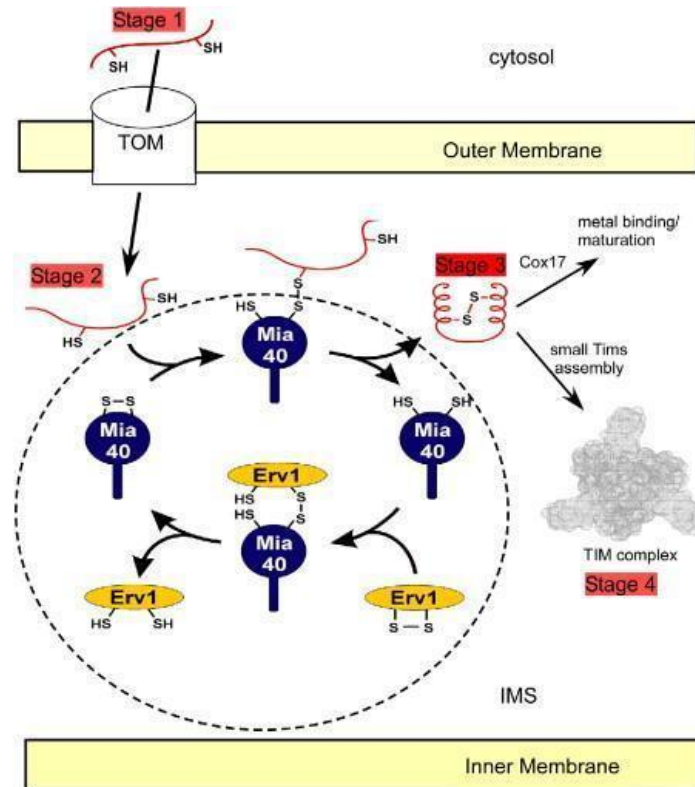


Figure 11 : Au centre de l'import et du repliement des protéines riches en cystéines dans l'espace intermembranaire se trouve le cycle redox Mia40/Erv1 (cercle) impliquant un système de relais de liaisons disulfures. La protéine précurseur à importer dans l'espace intermembranaire est d'abord transloquée par le canal TOM dans les mitochondries. La protéine réduite forme un pont disulfure intermoléculaire transitoire avec la forme oxydée de Mia40. Il en résulte une oxydation de la protéine précurseur qui contient alors une liaison disulfure intramoléculaire et est capable de se replier dans l'espace intermembranaire (Tokatlidis, 2005).

5. Import mitochondrial et cytotoxicité

L'import des protéines dans la mitochondrie est critique pour le bon fonctionnement de ces dernières, mais il permet également d'éviter une accumulation protéique dans la cellule. Un dysfonctionnement dans l'import des protéines mitochondriales peut avoir des conséquences sur la santé humaine en étant lié à de nombreuses maladies comme des troubles neurodégénératifs ou des cancers (Vyas et al., 2016; Wang et al., 2019). Ainsi, de nombreuses recherches ont été effectuées pour étudier l'impact de défauts d'import mitochondrial sur l'homéostasie du protéome cellulaire. Deux publications de 2015 dans Nature ont montré que la perturbation de l'import mitochondrial peut entraîner une surcharge de la machinerie d'import qui se caractérise par une accumulation des précurseurs dans le cytosol. Cette accumulation déclenche une voie de signalisation appelée « mPOS » pour « mitochondrial Precursor Over-accumulation Stress ». L'activation de ce signal a pour conséquence une réorganisation du programme transcriptionnel afin de réguler négativement la machinerie de traduction et ainsi désencombrer la machinerie d'import mitochondrial (Wang and Chen, 2015; Wrobel et al., 2015). Des gènes peuvent inhiber ce signal mPOS pour éviter qu'il ne conduise à la mort de la cellule. C'est le cas notamment de Gis2 et Nog2 qui sont régulés positivement dans la réponse au stress mitochondrial (Figure 12). Nog2 inhibe l'exportation nucléaire de la sous-unité ribosomale 60S et Gis2 favorise la traduction indépendante de la coiffe (Wang and Chen, 2015). Ces éléments font partie d'une boucle de rétroaction afin de promouvoir la survie de la cellule en inhibant le signal mPOS (Figure 12). Un autre mécanisme a été révélé pour contrôler le signal mPOS. Il a été observé chez l'homme, que la surexpression de ANT1 (l'orthologue mammifère de la Aac2 de levure), une protéine impliquée dans le transfert de l'ATP de la mitochondrie vers le cytosol, induit la formation d'agrégats, qui est surtout liée à une surcharge de la machinerie de dégradation (Liu et al., 2019). Cette induction est accompagnée par la régulation à la hausse de protéines comme, EGR1, eEF1a et l'ubiquitine qui favorisent le traitement de ces agrégats.

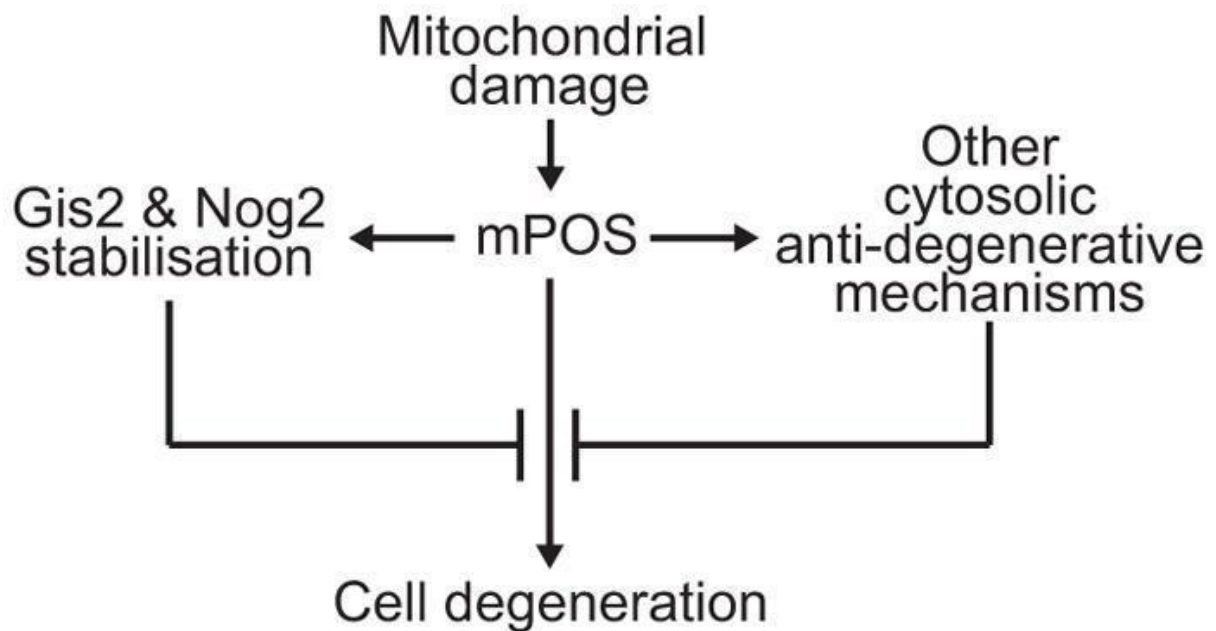


Figure 12 : Une boucle de rétroaction pour inhiber le signal mPOS et favoriser la survie cellulaire en surexprimant Gis2 et Nog2 (Wang and Chen, 2015)

L'un des deux articles de 2015 présente un système dans lequel le mutant Aac2 A128P de la protéine Aac2 codant l'isoforme majeure de la translocase qui est impliquée dans l'échange ATP/ADP. Cette mutation provoque un mauvais repliement et une accumulation des précurseurs (Wang and Chen, 2015). D'autres perturbations peuvent également déclencher le signal mPOS. D'après une revue de 2018, il s'agit de mutations dans la machinerie centrale d'import des protéines mitochondriales, d'un mauvais repliement des protéines intermembranaires, d'un contrôle qualité réduit des protéines de l'espace intermembranaire, des mutations de l'ADN mitochondrial, de la dissipation du potentiel de membrane, des mutations affectant l'adressage de précurseurs, et des défauts de stabilisation des protéines mitochondriales pré-importées ou de dégradation des protéines mitochondriales non importées (Coyne and Chen, 2018). Il a été montré que des protéines mitochondriales sont dégradées par le protéasome lorsqu'elles sont accumulées dans le cytoplasme (Kowalski et al., 2018; Wrobel et al., 2015). Cela a été observé plus particulièrement pour la protéine Cox12, qui est un substrat du complexe MIA et qui possède donc des résidus cystéines. Il a été constaté que la dégradation de cette protéine est plus rapide lorsqu'elle s'accumule dans le cytosol. De plus, l'ubiquitination d'une protéine empêche sa translocation par le complexe TOM, l'ubiquitine étant d'une taille plus importante que le canal que forme le complexe TOM (Kowalski et al., 2018). Ainsi, il

existe des signaux de dégradation portés par les protéines mitochondriales afin de réguler la biogénèse de la mitochondrie.

De nombreuses publications suggèrent l'existence d'un lien entre mPOS et maladies humaines. Tout d'abord, les deux caractéristiques principales des maladies neurodégénératives sont l'accumulation de protéines cytosoliques et un dysfonctionnement des mitochondries. Or il a été démontré que chez la souris l'expression de protéines qui ont une forte propension à s'agréger peut être associée à des mitochondries dysfonctionnelles (Mattiuzzi et al., 2002; Tabrizi et al., 2000; Xu et al., 2011). Il a été démontré que la mutation de protéines mitochondriales peut provoquer la maladie de Parkinson (Bose and Beal, 2016). De plus, l'inhibition de la traduction et la diminution de la synthèse protéique qui sont des phénomènes inhibant le mPOS ont été mis en évidence chez 12 patients atteints de maladies mitochondriales (Zhang et al., 2013). Une des protéines mutantes connues pour être responsables de la maladie d'Alzheimer est Htra. Or la délétion de Htra2 chez la souris conduit à une accumulation de protéines mal repliées de la membrane interne mitochondriale, phénomène connu pour activer la voie mPOS (Moiso et al., 2009). Une étude chez la souris a montré que l'inhibition de la biosynthèse des protéines induite indirectement par un traitement à la rapamycine atténue le syndrome de Leigh (Johnson et al., 2013). On peut également imaginer que le mPOS n'est pas forcément à l'origine de maladies mais peut être responsable de la progression de la maladie. En effet, une accumulation des précurseurs dans le cytosol pourrait s'ajouter à la charge protéostatique causée par une maladie sous-jacente. Ces éléments montrent que l'adressage efficace des protéines précurseurs vers les mitochondries est essentiel pour prévenir l'activation de la voie mPOS et donc pour la survie de la cellule.

6. Import mitochondrial et dynamique traductionnelle

Comme je l'ai expliqué dans le chapitre précédent, la perturbation de l'import mitochondrial peut avoir de lourdes conséquences sur le fonctionnement de la cellule et peut même être à l'origine de maladies. L'import efficace des protéines mitochondriales est donc crucial. Une optimisation de cette efficacité d'import est rendue possible par une traduction localisée au niveau des mitochondries et par l'implication de systèmes d'adressage faisant intervenir des chaperonnes dédiées.

a. Localisation des ARNm des protéines mitochondriales

Plusieurs études ont montré que l'import des protéines à la mitochondrie peut se faire de façon co-translationnelle ou très rapidement après la fin de la traduction.

En 1974 déjà, des chercheurs ont observé à l'aide de la microscopie électronique des sous-unités du ribosome autour des mitochondries (Kellems et al., 1974). Cette découverte a été depuis confirmée par de nombreuses études. Il a été observé très tôt que la traduction de certaines protéines mitochondriales est localisée dans l'environnement proche des mitochondries (Corral-Debrinski et al., 2000; Suissa and Schatz, 1982). En 2002, une étude globale basée sur l'utilisation de puces à ADN a montré que plus de la moitié des ARN codant des protéines mitochondriales ont une localisation péri-mitochondriale (Marc et al., 2002). De plus d'autres études ont montré par imagerie fluorescente que de nombreux ARNm d'origine nucléaire codant des protéines mitochondriales sont localisés à la surface des mitochondries (Gadir et al., 2011; Garcia et al., 2007). Ces observations ont été confirmées par des techniques plus avancées comme la technique de « proximity ligation » et la « cryo-microscopie » (Gold et al., 2017; Williams et al., 2014). Le premier mécanisme mis en lumière pour expliquer cette localisation mitochondriale est le rôle des séquences 3'UTR. Des études ont en effet montré que la modification ou la délétion des séquences 3'UTR de certains ARNm a pour conséquence une délocalisation des protéines mitochondriales et que, réciproquement, le 3'UTR de certains ARNm codant des protéines mitochondriales était suffisant pour localiser à la mitochondrie un ARNm GFP non traductible (Gadir et al., 2011; Garcia et al., 2007; Marc et al., 2002; Margeot et al., 2005). Par ailleurs, la délétion du gène codant la protéine Puf3, connue pour cibler les 3'UTR d'ARNm de protéines mitochondriales afin de promouvoir leur dégradation (figure 13) (Gadir et al., 2011; Zhu et al., 2009) affecte la localisation péri-mitochondriale de nombreux ARNm (Saint-Georges et al., 2008).

Cependant deux études globales, qui s'appuient sur l'utilisation de bibliothèques de protéines fusion en C ou N-terminal chez la levure, ont montré que certaines protéines ont une localisation mitochondriale indépendamment de leur UTR (Huh et al., 2003; Weill et al., 2018).

De plus, l'utilisation de la puromycine pour dissocier les ribosomes des mitochondries entraîne une délocalisation mitochondriale pour 70% des ARNm (Saint-Georges et al., 2008). Inversement, un traitement au cycloheximide, qui stabilise les ribosomes en les bloquant en cours de traduction, augmente la localisation mitochondriale de certains ARNm (Saint-Georges et al., 2008). Ces résultats indiquent qu'il existe un lien entre la localisation des ARNm et leur traduction. Ils ont été confortés par l'étude de la protéine Atp2, dans laquelle il a été démontré

qu'il existe des régions de la protéine responsables de la localisation de l'ARNm correspondant (Garcia et al., 2010). En effet, *ATP2* possède deux régions (R1 et R2) présentes en aval du MTS et lorsqu'elles ne sont pas traduites, la localisation péri-mitochondriale de l'ARNm est perdue. Cette localisation est également perdue lorsque le MTS est clivé. Ainsi, le MTS est nécessaire mais pas suffisante pour avoir une bonne localisation de l'ARNm d'*ATP2*. Une des explications que l'on peut émettre est que le MTS va interagir avec les complexes TOM/TIM dès sa sortie du tunnel de traduction. Si cette interaction a lieu avant la fin de la traduction de la protéine alors le précurseur et l'ARNm qui lui est associé via le ribosome vont avoir une localisation péri-mitochondriale. Par ailleurs, une étude a montré à l'aide de la technique de cryotomographie électronique que les ribosomes sont localisés au niveau de la mitochondrie de manière dépendante des chaînes naissantes (Gold et al., 2017). On parle alors de complexe ribosome-chaîne naissant (RNC) et il a été montré que la chaîne naissante interagit avec le complexe d'import TOM tout en étant liée au ribosome. De par sa position N-terminale et sa capacité à interagir avec la machinerie d'import, le MTS est déterminant pour l'import co-traductionnel et pour la localisation de certains ARNm à la mitochondrie. De plus une publication a montré que de nombreux ARNm, dont ceux codant des protéines avec un signal d'import N-terminal, ont une localisation mitochondriale dépendante de Tom20 (Eliyahu et al., 2010). Cette interaction est dépendante du MTS car la modification de certaines caractéristiques du MTS, comme sa charge positive et son hydrophobicité, ont un impact sur la localisation des ARNm. La présence de MTS permet donc l'interaction du complexe ARNm/ Ribosome au complexe TOM. Cependant, la bonne localisation mitochondriale de l'ARNm de la GFP couplée à une séquence MTS dans un contexte dans lequel Tom20 est délété semble indiquer que le MTS interagit avec d'autres protéines que Tom20. En effet, l'ARNm de *ATP2*, qui code une protéine avec MTS, est délocalisé dans des mutants Tom70 et Tom7 (Gadir et al., 2011). L'intérêt d'un import contradictionnelle pourrait être multiple pour la cellule. On peut imaginer que cela limiterait le risque d'agrégation de protéines dans le cytosol et réduirait l'utilisation de protéine chaperonne et donc économiserait à la cellule de l'ATP. Enfin, une traduction locale permettrait d'augmenter la concentration de protéines et favoriserait la formation de complexes protéiques.

b. Protéines chaperonnes capable d'interagir avec les protéines mitochondriales

Une autre façon de trier et d'assurer le bon adressage des protéines en cours de synthèse est l'implication de protéines chaperonnes. Il existe de nombreuses protéines chaperonnes capables d'interagir précocement (co-traductionnellement) avec les protéines mitochondriales afin de faciliter leur translocation dans les mitochondries.

Une des protéines chaperonnes impliquées dans ce processus est le complexe protéique NAC (Nascent chain Associated Complex) (figure 13). Ce complexe protéique est composé de deux sous-unités : les sous-unités α et β' ou β . Il est capable de reconnaître et d'interagir avec les chaînes naissantes à l'intérieur du tunnel du ribosome (Gamerding et al., 2019). En effet, la sous-unité β peut s'insérer dans le tunnel du ribosome alors que la sous-unité α reste à la surface de ce dernier et n'interagit avec la chaîne naissante que lorsqu'elle émerge du tunnel ribosomique (Gamerding et al., 2019). Cette interaction précoce du complexe NAC avec les chaînes naissantes pourrait prévenir un mauvais adressage des protéines en empêchant d'autres chaperonnes de les adresser dans la mauvaise organelle. (George et al., 2002; Lesnik et al., 2014). Il a été démontré qu'en absence de NAC, la SRP (Signal Recognition Particle), un complexe responsable de l'adressage de protéines en cours de traduction vers le réticulum endoplasmique, interagit moins avec ses substrats (del Alamo et al., 2011). Chez la levure plusieurs études ont montré que le complexe protéique $\alpha\beta'$ -NAC est capable de reconnaître très précocement les acides aminés chargés positivement des MTS et permet un adressage vers le complexe TOM (Avendaño-Monsalve et al., 2022; Fünfschilling and Rospert, 1999a; George et al., 1998; Wiedmann et al., 1994). De plus, d'après des analyses *in silico* la NAC ciblerait environ 200 protéines mitochondriales (Avendaño-Monsalve et al., 2022). La déplétion de la NAC induit une diminution de la quantité de protéines dans la mitochondrie et un appauvrissement en NAC conduit à une baisse de l'efficacité d'import des protéines mitochondriales (Fünfschilling and Rospert, 1999b; George et al., 2002, 1998). Cette reconnaissance par NAC semble être dépendante des acides aminés chargés positivement des MTS (Avendaño-Monsalve et al., 2022). Par exemple il a été démontré que les 10 premiers acides aminés de Oxa1 sont importants pour son interaction avec la NAC et que cette interaction a un effet positif sur son import. Outre son rôle dans l'adressage des protéines, la NAC peut également assister le repliement co-traductionnel des protéines. De plus, il a été démontré que la déplétion de la NAC augmente l'ubiquitination de certaines protéines conduisant à leur dégradation (Duttler et al., 2013; Kirstein-Miles et al., 2013; Wang et al., 2013). De plus, il a

été démontré que le complexe NAC empêche l'agrégation de protéines (Shen et al., 2019). La délétion d'un des gènes codant une des sous-unités du complexe NAC conduit à un phénotype létal chez la souris et les mouches alors qu'elle n'affecte pas la viabilité d'une cellule de levure (Deng and Behringer, 1995; Markesich et al., 2000; Ponce-Rojas et al., 2017).

D'autres protéines sont connues pour jouer le rôle de chaperonne pour les protéines mitochondriales. Les chaperonnes principales sont celles de la famille HSP70 et HSP90 (figure 13). Elles sont connues pour pouvoir interagir avec la protéine réceptrice du complexe TOM à savoir TOM70 (Young et al., 2003; Zanphorlin et al., 2016). Des études de profilage des ribosomes ont montré chez la levure que les isoformes *ssb1* et *ssb2* de HSP70 s'associaient au niveau du ribosome à la sortie du tunnel. Cela va ensuite leur permettre d'interagir avec les chaînes naissantes et d'éviter leur agrégation mais également de les maintenir dans un état déplié pour qu'elles puissent être transloquées dans le canal TOM40 (Döring et al., 2017; Stein et al., 2019). De plus grâce à cette même technique il a été démontré qu'une autre protéine chaperonne en forme d'anneau, TRiC/CCT, peut interagir avec les protéines à la sortie de tunnel sans rentrer en compétition avec les protéines *ssb*. Son rôle serait d'améliorer l'adressage de ces protéines dans les mitochondries (Stein et al., 2019). D'autres protéines chaperonnes sont impliquées dans l'import de protéines mitochondriales. C'est le cas des protéines de la famille de HSP40 : *Sti1*, *Ydj1*, *Xdj1* et *Djp1* (Hoseini et al., 2016) Il a été démontré que la délétion de *Sti1* conduit à une altération morphologique des mitochondries. *Xdj1* favorise l'import de précurseurs mitochondriaux des complexes TOM en les délivrant à la protéine réceptrice TOM22. *Djp1*, quant à lui est capable d'interagir avec les protéines mitochondriales et TOM70 (Opaliński et al., 2018). L'ensemble de ces protéines de la famille de HSP40 sont considérées comme étant des co-chaperonnes. Elles interagissent avec les protéines en collaboration avec HSP70. Le rôle de ces protéines est de faciliter l'adressage des protéines mitochondriales mais également d'éviter leur agrégation dans le cytosol.

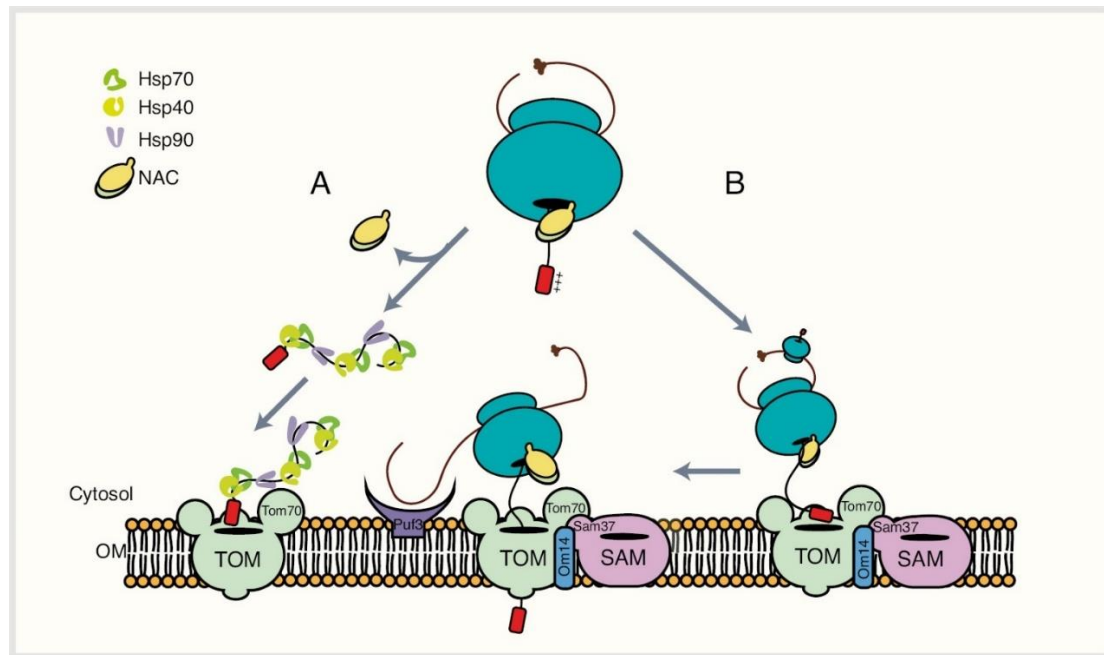


Figure 13 : La plupart des protéines mitochondriales sont synthétisées par des ribosomes cytosoliques (turquoise) et ciblées vers l'organite guidé par différents chaperons dont la reconnaissance et la liaison sont essentielles pour éviter les erreurs de ciblage et une éventuelle agrégation. (A) Pour les protéines précurseurs mitochondriales dont la traduction est terminée dans le cytosol, les chaperons solubles des familles Hsp70, Hsp40 et Hsp90 (vert, vert citron et violet) les reconnaissent une fois leur synthèse terminée et guident leur chemin vers les récepteurs de le complexe TOM, permettant ainsi l'import de protéines. (B) Les ribosomes traduisant les protéines mitochondriales peuvent être dirigés vers le voisinage de la surface mitochondriale par l'action de chaperons ribosomiques tels que NAC (jaune/vert) et/ou par des protéines fonctionnant comme des récepteurs pour les ARNm codant pour les protéines mitochondriales (mMP) telles que Puf3 (violet) (Avendaño-Monsalve et al., 2020)

L'ensemble des protéines chaperonnet que je viens de citer ne vont pas toutes interagir avec les protéines mitochondriales, étant donné qu'une partie de ces protéines sont importées de manière co-traductionnelle. Pour les autres, la prise en charge rapide des chaînes naissantes doit être importante, notamment pour les protéines mitochondriales avec MTS qui sont susceptibles de former des agrégats à cause de la face hydrophobe de leur hélice alpha amphiphile (Endo et al., 1995). La chaîne naissante semble donc jouer un rôle important dans l'import des protéines mitochondriales. Il est alors pertinent de s'intéresser à l'ensemble des enzymes et des modifications que peuvent subir ces chaînes naissantes pour mieux comprendre comment leur reconnaissance peut, très précocement dans la vie des protéines, influencer leur devenir cellulaire.

II. Les modifications N-terminales co-traductionnelles jouent un rôle important dans le devenir des protéines

Certaines modifications co-traductionnelles N-terminales sont dépendantes de la nature de l'acide aminé situé en position 2, c'est-à-dire suivant la méthionine initiatrice. C'est ce type de modifications qui m'a plus particulièrement intéressé pendant ma thèse. Dans ce chapitre, je présenterai d'abord les principales enzymes impliquées dans ces modifications, puis je parlerai de leur rôle et enfin je montrerai qu'il existe des biais de composition significatifs et spécifiques à la position 2 des précurseurs mitochondriaux suggérant une pression de sélection sur cette position liée aux spécificités de ces enzymes.

1. Les enzymes de modifications N-terminales co-traductionnelles

Lors de ma thèse, j'ai été amené à m'intéresser à certaines enzymes qui modifient les chaînes naissantes de manière co-traductionnelle et dont le recrutement dépend de la nature du résidu en position 2 de ces chaînes naissantes. Je vais donc vous présenter certaines de ces enzymes qui seront nécessaires à la compréhension de mes travaux de thèse et des hypothèses qui en découlent.

a. La Méthionine aminopeptidase

L'acide aminé en deuxième position des chaînes naissantes a un impact très important sur la conservation ou non de la méthionine initiatrice. Une protéase appelée Méthionine aminopeptidase clive en effet la méthionine initiatrice lorsque celle-ci est suivie, en position 2, par un résidu de petite taille : alanine, glycine, proline, sérine, thréonine, cystéine et valine (Sherman et al., 1985). La spécificité de cette enzyme a été démontrée grâce à l'utilisation de protéines rapportrices exprimées chez la levure et variant l'une de l'autre par leur résidu en position 2 (Moerschell et al., 1990). Cependant ce clivage n'est pas systématique. En effet, lorsqu'il y a une valine en position 2, la méthionine initiatrice n'est pas toujours clivée. Ce clivage est complètement inhibé quand une valine en position 2 est suivie d'une proline en position 3 (Prchal et al., 1986). Des modifications N-terminales peuvent également influencer le clivage de la méthionine initiatrice. Cela a été démontré lorsque la méthionine initiatrice subit une acétylation ou une formylation (Solbiati et al., 1999; Van Damme et al., 2015).

Contrairement aux procaryotes et aux archées, chez la levure, il existe deux protéines de la famille des Méthionine aminopeptidases : MetAP1 et MetAP2. La délétion d'une de ces MetAPs n'est pas létale pour la cellule, contrairement à la double délétion des MetAPs. Cela montre que la fonction des MetAPs est essentielle pour la cellule et que MetAP1 et MetAP2 ont des fonctions redondantes (Chang et al., 1992; Li and Chang, 1995). Cependant, ces protéines n'ont pas la même efficacité de clivage. MetAP1 joue un rôle dominant, ce qui explique le défaut de croissance plus important dans le mutant MetAP1 par rapport au mutant MetAP2 (Chen et al., 2002). Cette modification d'activité de la MetAP2 est peut-être liée à l'insertion dans le site catalytique de la MetAP2 d'environ 60 acides aminés (Tahirov et al., 1998).

L'action de ces enzymes peut avoir un effet sur le temps de demi-vie des protéines. Dans la littérature, une protéine conservant à son extrémité N-terminale une méthionine est considérée comme stable. Cependant, il a été montré que la conservation de la méthionine pour la protéine recombinante Met-Ala-GST induit une baisse de son temps de demi-vie (Chen et al., 2002). C'est un résultat à prendre toutefois avec précaution, car le maintien de la méthionine n'était que partiel dans ce cas. Seulement 54% de la version recombinante Met-Ala-GST a conservé sa méthionine initiatrice lorsque la MetAP1 est délétée. Le maintien partiel de méthionine peut donc expliquer ce résultat contradictoire. Outre son potentiel rôle dans la stabilité des protéines, le clivage de la méthionine initiatrice peut être nécessaire pour la bonne fonction d'une protéine. Par exemple, la protéine Gfa1 de la levure, qui a pour rôle de catalyser la synthèse de chitine est fonctionnelle uniquement si la méthionine initiatrice est clivée (Dummitt et al., 2005). De plus, le clivage peut avoir un effet sur l'adressage de protéines. C'est le cas pour la carboxypeptidase Y de levure qui perd sa localisation dans le réticulum endoplasmique lorsque sa méthionine initiatrice est clivée (Forte et al., 2011). La MetAP peut également avoir un rôle important dans le devenir des protéines de manière indirect. En effet le clivage de la méthionine initiatrice peut également être une condition nécessaire pour d'autres modifications N-terminales, comme la N-acétylation par une N-acétyltransférase. Je vais d'ailleurs vous présenter cette famille d'enzymes.

b. Les N-Acétyltransférases (NATs)

L'acétylation est l'une des modifications les plus courantes des protéines eucaryotes (Polevoda and Sherman, 2003). Elle consiste en l'ajout d'un groupe acétyl (COCH₃) à un groupe amine libre (NH₃⁺) (Figure 14).

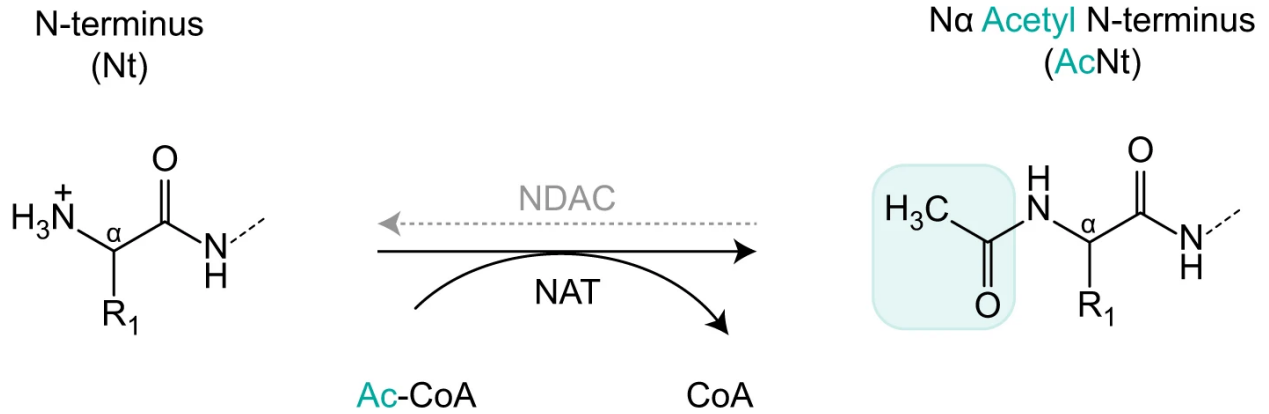


Figure 14 : Les N acétyltransférases N terminales (NATs) catalysent le transfert du groupe acétyle (CH₃CO) de l'acétyl -CoA vers le groupe amine libre (NH₃⁺) de la protéine N terminale (Varland et al., 2018).

À partir de l'étude du N-acétylome, de la fréquence des 20 acides aminés à la position 2 et de la spécificités des NATs, il a été montré qu'environ 60% des protéines de levures et 80 à 90% des protéines chez l'homme sont acétylées à leur extrémité N-terminale (Arnesen et al., 2009). L'acétylation N-terminale est considérée comme étant une modification irréversible car aucune protéine jouant le rôle de N-désacétylase n'a été caractérisée à ce jour. Cette acétylation N-terminale est assurée par une famille d'enzymes appelées N-acétyltransférases (NATs). À ce jour, 8 NATs ont été identifiées dans les cellules eucaryotes : NatA, NatB, NatC, NatD, NatE, NatF, NatG et NatH (Aksnes et al., 2019). Certaines NATs ne sont retrouvées que dans certains clades. C'est le cas notamment de NatG qui n'est présente que chez les plantes, dans les chloroplastes (Dinh et al., 2015). NatH ne cible que les molécules d'actine chez les métazoaires (Drazic et al., 2016). Quant à NatF, elle n'est présente que chez les métazoaires et les plantes et assure l'acétylation des protéines membranaires (Van Damme et al., 2011).

La levure a été le modèle le plus utilisé pour identifier et étudier les NATs, certainement car la délétion d'une des NATs n'est pas létale dans cette espèce. Chez la levure *S. cerevisiae*, il existe 4 NATs fonctionnelles : NatA à NatD. Elles agissent de manière co-traductionnelle et chacune d'elles présente une spécificité enzymatique différente en fonction de l'acide aminé en position 2 de ses cibles (Gautschi et al., 2003; Polevoda et al., 2008). Presque toutes les acétylations N-

terminales sont réalisées par NatA, NatB et NatC chez la levure. En effet, seules deux protéines cibles ont été identifiées pour NatD. Il s'agit des deux histones H2A et H4 (Polevoda et al., 2009; Song et al., 2003). NatE est présente chez la levure, mais ne semble pas être fonctionnelle (Knorr et al., 2019).

Toutes les NATs à l'exception de NatD sont composées d'au moins deux sous-unités. Les différentes sous-unités ont plusieurs noms dans la littérature, mais une nomenclature universelle a été mise en place avec comme référence les noms utilisés chez l'homme par exemple la sous unité catalytique de NatC est Mak3 et la nomenclature universelle voudrait qu'on la nomme Naa30.

Afin de comprendre la spécificité des NATs, les laboratoires ont longtemps utilisé la méthode de l'électrophorèse sur gel bidimensionnel. Cette technique consiste à identifier les protéines qui sont acétylées via les profils de migration sur des gels 2D dans des souches déplétées d'une NAT en comparaison avec une souche contrôle. L'ajout d'un groupe acétyl conduit à un déplacement du point isoélectrique de la protéine ce qui modifie son profil de migration par rapport à sa version non acétylée. Cette technique a permis d'étudier des mutants de la position 2 d'une protéine rapportrice comme l'iso cytochrome C, dans des souches déplétées d'une des NATs avant de permettre l'étude de protéines naturelles (Arnold et al., 1999; Mullen et al., 1989; Polevoda, 1999). Afin d'étudier les substrats des différentes NATs in vivo, différentes techniques ont été mises en place. La technique de référence est appelée COmbined FRActional DIagonal Chromatography (COFRADIC) (Arnesen et al., 2009) (figure 15). Elle consiste à enrichir des échantillons protéiques en peptides N-terminaux qui sont ensuite étudiés par spectrométrie de masse. A ce jour, elle a permis l'étude d'environ 10% du protéome de levure. Elle permet de mesurer l'acétylation N-terminale dans des protéines extraites de souches sauvages ou délétées d'une des NATs. De nombreuses cibles des NATs ont ainsi été identifiées, et leur spécificité vis-à-vis de leurs substrats a pu être établie chez l'Homme et chez la levure *S. cerevisiae* (Arnesen et al., 2009; Van Damme et al., 2012).

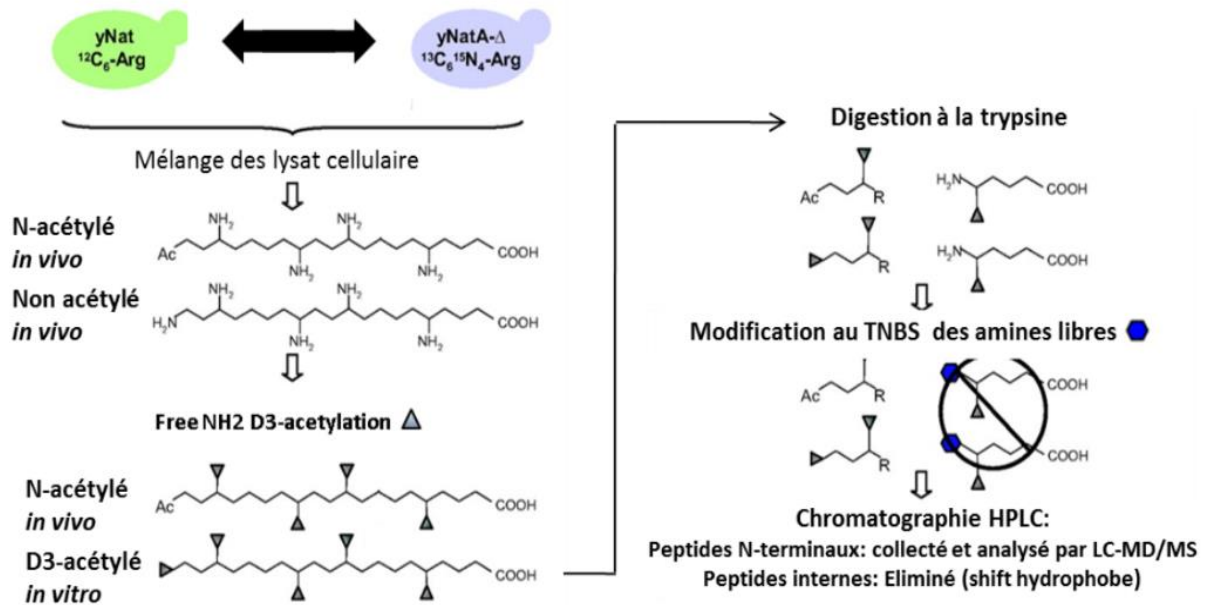


Figure 15 : La technique du COFRADIC : Les protéines sont extraites et leurs groupements amines libres sont bloqués par des groupements acétyl marqués au deutérium, un isotope lourd de l'hydrogène. Après extraction, elles sont digérées par la trypsine. Par la suite, seuls les peptides qui ne sont pas des peptides N-terminaux présentent un groupement amine libre.

Ces groupements amines libres sont bloqués par de l'acide Trinitrobenzenesulfonique (TNBS), un groupement chimique très hydrophobe. Les peptides N-terminaux sont alors purifiés par chromatographie en phase liquide (HPLC) : les peptides internes et C-terminaux portant le TNBS sont retenus par la colonne, et seuls les peptides N-terminaux sont élués et analysés par spectrométrie de masse (LC-MD/MS) (Arnesen et al., 2009)

NatA peut acétyler l'alanine, la cystéine, la glycine, la sérine, la thréonine ou la valine qui se retrouvent à l'extrémité N-terminale des protéines après clivage de leur méthionine initiatrice par les MetAPs. NatA est un complexe composé de deux sous-unités : la sous-unité catalytique Ard1 (Naa10) et la sous-unité auxiliaire Nat1 (Naa15) (Mullen et al., 1989) (Figure 16). D'après la littérature, NatA est définie seulement par le complexe composé de Ard1 et Nat1 (Grunwald et al., 2020). Cependant elle a également un partenaire permanent qui est Nat5 (Naa50) qui est considéré par certains auteurs comme faisant partie de NatA (Gautschi et al., 2003; Knorr et al., 2019). Nat1 est connue pour permettre l'ancrage de Ard1 au ribosome afin d'assurer une reconnaissance co-translationnelle par NatA de ses cibles (Gautschi et al., 2003). Il a été montré que la délétion de l'une ou l'autre des sous-unités de NatA conduit au même phénotype de croissance ralentie sur milieu riche. De plus, une sur-expression d'une des sous-unités ne

permet pas d'augmenter l'activité de NatA, ce qui met en évidence l'importance de la stoechiométrie du complexe pour son activité (Mullen et al., 1989).

NatB acétyle la méthionine initiatrice lorsqu'elle est suivie en position 2 par une asparagine, un acide aspartique, une glutamine ou un acide glutamique. NatB est composée de la sous-unité catalytique Nat3 (Naa20) et de son auxiliaire Mdm20 (Naa25) (Polevoda, 1999; Polevoda and Sherman, 2003; Van Damme et al., 2012) (Figure 16). La délétion de NatB conduit à un phénotype de croissance ralentie, qui est d'ailleurs plus important que pour les mutants NatA ou NatC.

NatC se compose de 3 sous-unités : Mak30 (Naa35), Mak31 (Naa38) et Mak3 (Naa30) (Figure 16). Cette dernière est sa sous-unité catalytique (Polevoda and Sherman, 2001). Elle a été identifiée pour la première fois dans une étude réalisée en 1992 qui a montré que sa délétion empêche la propagation du virus à ARN de levure L A (Tercero and Wickner, 1992). NatC acétyle la principale protéine d'enveloppe nucléaire GAG de ce virus et permet la propagation et l'assemblage des particules virales (Tercero and Wickner, 1992). NatC acétyle la méthionine initiatrice lorsqu'elle précède un large résidu hydrophobe tel qu'une isoleucine, une leucine, une phénylalanine ou un tryptophane. En plus de sa sélectivité pour certains résidus hydrophobes en position 2, une étude ancienne avait tout de même démontré que la substitution des résidus en position 3 et 4 par l'acide glutamique perturbait l'acétylation de la protéine Gag par NatC (Tercero et al., 1993). Une étude plus récente a confirmé le rôle de la position 3 des cibles de NatC dans le recrutement de l'enzyme (Grunwald et al., 2020). Dans cette publication l'étude de la structure de NatC par cristallographie aux rayons X a permis de montrer que la présence d'une arginine en position 3 augmente l'efficacité catalytique de NatC en se fixant dans une poche catalytique de Mak3.

De plus, la structure de NatC est différente de celles de NatA et NatB. Tout d'abord NatA et NatB n'ont qu'une seule sous-unité auxiliaire alors que NatC en a deux (Hong et al., 2017; Liszczak et al., 2013; Tercero et al., 1993). En outre, la disposition des sous-unités de NatC est différente de celle de NatA et NatB. Les sous-unités catalytiques de NatA et NatB n'englobent que la région N-terminale de leur sous-unité catalytique respective alors que la sous-unité auxiliaire de NatC, Mak30, englobe également une partie C-terminale de Mak3 (Grunwald et al., 2020). Cette disposition particulière à NatC lui permet de former une poche hydrophobe ce qui explique sa spécificité pour des substrats hydrophobes.

La délétion d'une des sous-unités rend NatC inefficace. Chez la levure *S. cerevisiae*, la délétion d'une des trois sous-unités entraîne un phénotype de croissance ralentie sur milieu riche à 37°C

qui est aggravé sur un milieu avec une source de carbone non fermentable comme du glycérol (Dihanich et al., 1989; Lee and Wickner, 1992) (Figure 17). Ce dernier phénotype est partagé avec les autres NATs mais étant donné que les souches déplétées de NatC n'ont pas de phénotype de croissance ralentie sur d'autres milieux, on peut imaginer que le phénotype sur glycérol provient de mécanismes différents. En effet le glucose est métabolisé par une voie différente de celle du glycérol qui met obligatoirement en jeu la respiration (Grauslund and Rønnow, 2000). On peut donc penser que NatC a un lien important avec les mitochondries et cette hypothèse est soutenue par certaines études. Tout d'abord il a été démontré que l'expression de Mak30 est inhibée par la présence de glucose et augmente en présence de glycérol (Lee and Wickner, 1992). Il a été rapporté très tôt que des protéines mitochondriales, comme Fum1 et Mrp1 ont les mêmes 4 premiers résidus que la protéine GAG, une cible avérée de l'enzyme (Tercero et al., 1993). Enfin il a été démontré que la délétion de Naa30 chez l'homme conduit à une expression réduite des protéines mitochondriales, une perte de potentiel de la membrane mitochondriale et une fragmentation mitochondriale (Van Damme et al., 2016). La fonction de NatC est conservée au cours de l'évolution de la levure à l'homme car il est possible de sauver le phénotype de croissance ralentie dû à la délétion de *MAK3* avec son orthologue humain Naa30 (Drazic and Varland, 2021).

NatC est impliquée dans de nombreuses maladies et notamment dans le cancer. En effet, dans le cancer du cerveau (glioblastome) il a été observé une sur-expression de Naa30 (Mughal et al., 2015). La déplétion de NatC provoque l'apoptose de manière dépendante de p53 (Starheim et al., 2009). Aussi, dans une étude de 2015, une mutation de Naa35 a été détectée chez des patients atteints de paralysie cérébrale (McMichael et al., 2015).

L'activité enzymatique du complexe appelé NatE est associée à l'activité catalytique de Naa50. En effet, la version métazoaire de cette sous-unité accessoire de NatA a été montrée comme pouvant aussi catalyser la N-acétylation, soit seule, soit en association avec Naa10 et Naa15 (Deng et al., 2020; Hou et al., 2007; Van Damme et al., 2015). Bien qu'ayant une composition proche de NatA, l'enzyme NatE humaine a en fait la même spécificité de substrat que NatC (Van Damme et al., 2015). Cependant l'activité NatE ne semble pas exister chez la levure, du fait que dans cette espèce Nat5 (Naa50) a une structure apparemment incompatible avec la fixation de l'acétyl CoA (Knorr et al., 2019).

Type	NatA	NatB	NatC	NatD	NatE
Catalytic subunit	Ard1p ^a	Nat3p ^a	Mak3p ^a	Nat4p ^a	Nat5p ^a
Auxiliary subunit	Nat1p ^{a,b}	Mdm20p ^a	Mak10p ^a Mak31p		Nat1p ^{a,b}
Substrates ^c	Ser- Ala- Gly- Thr-	Met-Glu- Met-Asp- Met-Asn- Met-Met-	Met-Ile- Met-Leu- Met-Trp- Met-Phe	Ser-Gly-Gly-Lys-etc- Ser-Gly-Arg-Gly-etc-	Unknown
No. of substrates	>2,000	>600	Low	At least 2	Unknown

a Subunits shown to be associated with ribosomes.

b We suggest that the auxiliary subunit Nat1p is common to both NatA and NatE, and that *nat1*-Δ mutants are deficient in both acetyltransferases.

c Acetylation occurs only on subclasses of proteins containing the indicated termini, except for Met-Glu- and Met-Asp- termini, which are apparently always acetylated.

Figure 16 : Tableau récapitulatif des différentes NATs chez la levure. Pour chaque NAT sont indiqués les substrats qu’elles ciblent, leur sous-unité catalytique (Polevoda et al., 2008).

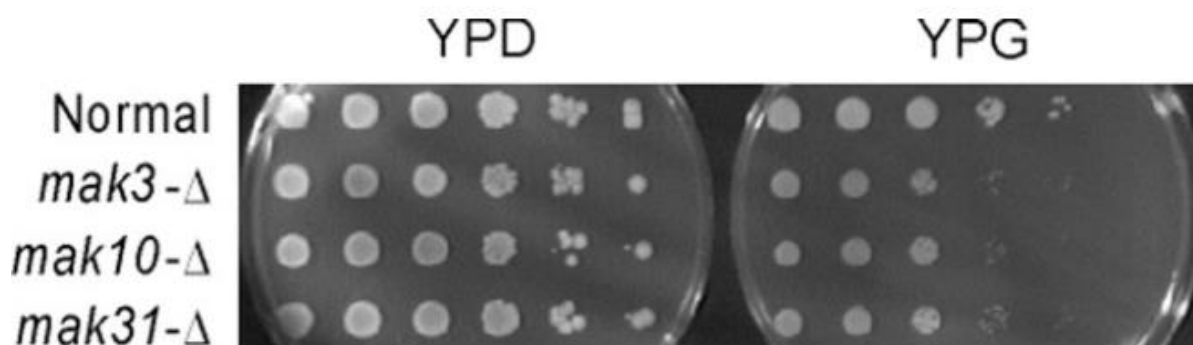


Figure 17 : Test de croissance en gouttes de différents mutants sur milieu riche (YPD) et sur milieu avec une source de carbone non fermentable (YPG). Le phénotype de croissance des souches délétées d’une des sous unités de Mak3 est comparé à celui d’une souche contrôle « Normal ». (Polevoda and Sherman, 2001)

L’existence de différentes N-acétyltransférases qui ont la même activité enzymatiques mais dont la spécificité est différente pourrait permettre d’acétyleur un maximum de protéines. Ceci expliquerait cette large proportion de protéines acétylées. Cependant, nous pouvons également imaginer que cette sélectivité différentielle des NATs permettrait à chacune de cibler des

ensembles de protéines avec des fonctions biologiques différentes. Cette hypothèse peut expliquer le fait que la délétion d'une des NATs conduit à un phénotype cellulaire spécifiquement associé à cette NAT. C'est une hypothèse qui a été étayée par des études *in silico* réalisées dans le laboratoire et que je vous présenterai dans ce manuscrit. Ce résultat très intéressant m'a conduit à étudier plus spécifiquement la N-acétyltransférase et son rôle potentiel sur le devenir des protéines mitochondriales avec MTS.

2. Impact de ces modifications sur le devenir des protéines

a. Dégradation (Voie N-End Rule)

Il existe une voie biologique appelée *N-End Rule* découverte en 1986, qui met en jeu le système UPS (Ubiquitin-Proteasome System) pour déterminer la demi-vie d'une protéine en fonction de son deuxième acide aminé. Elle a été découverte par Varshavsky et ses collègues (Bachmair et al., 1986). Ils ont exprimé chez la levure des mutants de la β -galactosidase portant à leur extrémité N-terminale chacun des 20 résidus possibles et ont remarqué que le temps de demi-vie des mutants pouvait varier entre moins de 3 min et plus de 20 h en fonction de ce résidu N-terminal (Figure 18.A). Cette voie biologique existe chez les procaryotes et les eucaryotes et implique un système d'ubiquitine. Elle assure le renouvellement spécifique des protéines dans la cellule et comprend 3 branches distinctes : Ac/N-end rule, Arg/N-end rule et Pro/N-end Rule.

A

Residue X in ub-X-βgal	Radius of gyration of X (Å)	Deubiquitination of ub-X-βgal	$t_{1/2}$ of X-βgal
Met	1.80	+	} >20 hours
Ser	1.08	+	
Ala	0.77	+	
Thr	1.24	+	
Val	1.29	+	
Gly	0	+	} ~30 minutes
Ile	1.56	+	
Glu	1.77	+	} ~10 minutes
Tyr	2.13	+	
Gln	1.75	+	} ~3 minutes
Phe	1.90	+	
Leu	1.54	+	
Asp	1.43	+	
Lys	2.08	+	} ~2 minutes
Arg	2.38	+	
Pro	1.25	-*	~7 minutes

*The rate of in vivo deubiquitination of ub-Pro-βgal is extremely low. The $t_{1/2}$ shown is that of the initial ub-Pro-βgal fusion protein (Fig. 4, lanes j to p).

B

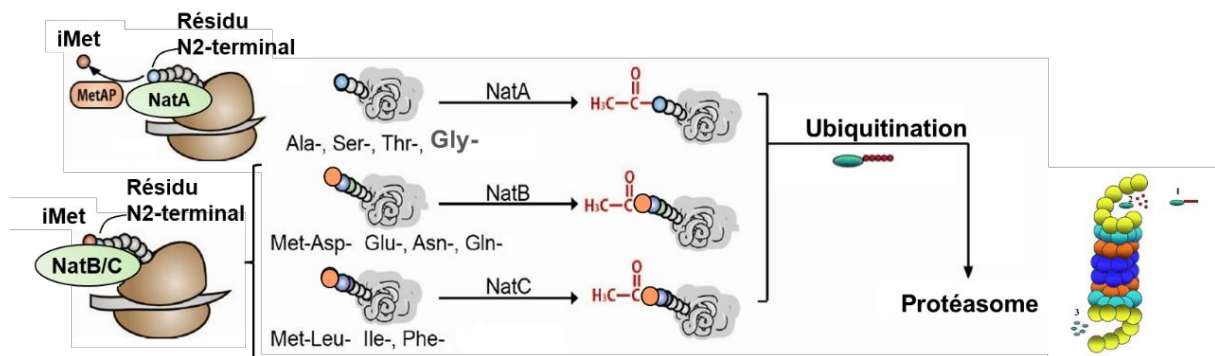


Figure 18 : (A) Tableau représentant le temps de demi-vie estimé pour chaque mutant en position 2 de la β-galactosidase. Ces données ont été obtenues *in vivo* à partir de culture d'*E. coli* et de levure *S. cerevisiae* (D'après (Varshavsky, 1996)) (B) La N-end rule repose sur l'activité de 3 N-acétyltransférases (NatA, NatB, NatC) présentant des spécificités enzymatiques distinctes en fonction de la nature du résidu en position 2. L'acétylation cotraductionnelle de leurs protéines cibles aboutit à leur adressage au protéasome. (Modifié d'après (Lee et al., 2016))

L'Ac/ N End Rule repose sur l'activité des trois N acétyltransférases : NatA, NatB et NatC. L'acétylation N-terminale a d'abord été vue comme un moyen de protéger la protéine de la dégradation (Ben-Saadon et al., 2004; Ciechanover and Ben-Saadon, 2004; Hershko et al., 1984; Jörnvall, 1975; Kuo et al., 2004; Persson et al., 1985). Cette hypothèse est venue du fait que les exopeptidases ont une efficacité enzymatique plus importante quand l'extrémité protéique est libre. C'est seulement en 2010 que l'acétylation N-Terminale est décrite comme un signal de dégradation (Hwang et al., 2010). En effet grâce à un système de protéine rapportrice il a été démontré que la délétion d'une ubiquitine ligase Doa10, impliquée dans la N end Rule, stabilisait les protéines NT-acétylées ayant en position 2 une Alanine, Valine, Serine, Thréonine et Cystéine (Figure 18.B).

Cependant, la stabilité des protéines Nt-acétylées semble très dépendante du contexte et de la protéine. Dans certains cas l'acétylation induit une dégradation mais ce signal de dégradation peut être masqué via un repliement intramoléculaire de la protéine ou via l'interaction de l'extrémité N-terminale de la protéine avec d'autres partenaires rendant impossible l'entrée de la protéine dans la voie de dégradation par le protéasome (Figure 19). C'est le cas par exemple pour la protéine Cog1 de levure qui est stabilisée lorsque son extrémité N-terminale acétylée est protégée grâce à son interaction avec d'autres partenaires du complexe Cog (Shemorry et al., 2013).

De plus, une étude récente a permis de montrer que la N acétylation semble rarement reconnue comme un signal de dégradation (Kats et al., 2018). De façon intéressante, les protéines pour lesquelles l'acétylation N terminale fonctionne comme un signal de dégradation sont plus particulièrement celles avec un signal hydrophobe en N terminal. Il semblerait donc que le niveau d'hydrophobicité N terminale soit une caractéristique importante pour la stabilité des protéines. Ce résultat est cohérent avec des publications encore plus récentes. Dans l'hypothèse où la N acétylation jouerait un rôle de dégradation, alors on s'attendrait à ce que la délétion des NATs augmente la stabilité des protéines et inhibe les processus de dégradation, or ce n'est pas le cas (Friedrich et al., 2021). NatA est impliquée dans le maintien de l'intégrité chromosomique et NatB agit principalement sur le repliement des protéines. Enfin, la délétion de NatA, induit une augmentation de l'activité des ubiquitine ligase (Kats et al., 2021). L'ensemble de ces publications montrent que la fonction de la N acétylation et des NATs, ne peut être généralisée, et n'est pas encore totalement comprise.

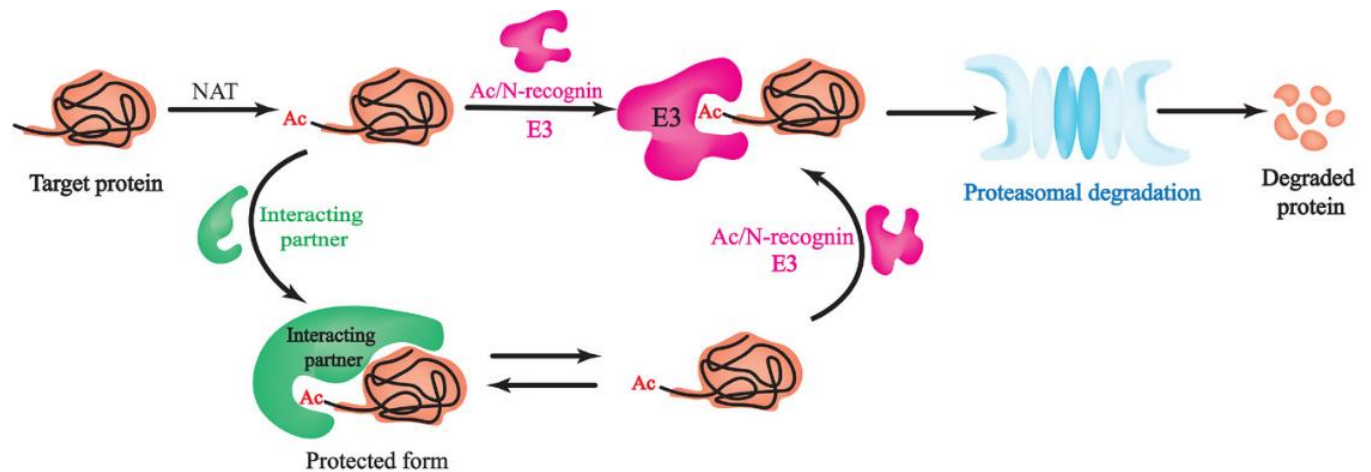


Figure 19 : Les protéines sont ciblées par les Nats en fonction de leur résidu N-terminal. Après acétylation, la protéine peut être ciblée par une ubiquitine ligase E3 qui va l’ubiquitiner afin qu’elle soit dégradée par le protéasome. Dès lors que la région N-terminale interagit avec une autre protéine, cela empêche la liaison de l’ubiquitine ligase E3 et donc la dégradation de la protéine.

Même si la très grande majorité des protéines est acétylée en N-terminal, une autre voie de dégradation permet la prise en charge de protéines conservant leur méthionine initiatrice et n’étant pas acétylées. Chez la levure, cette voie est appelée Arg/ N end Rule et implique l’ubiquitine ligase E3 Ubr1 (Xie and Varshavsky, 1999). Ubr1 fait partie de la famille des Ubiquitines ligases UBR. Cette enzyme cible les protéines ayant à leur extrémité des acides aminés déstabilisateurs (aussi appelés N-degrons) de type 1 et de type 2. Le site de type 1 est spécifique des substrats protéiques portant des résidus N-terminaux basiques (Arg, Lys et His), alors que le site de type 2 est spécifique des protéines ayant à leur extrémité N-terminale des résidus hydrophobes volumineux (Trp, Phe, Tyr, Leu et Ile) (Tasaki et al., 2009). Le troisième site de liaison d'Ubr1 cible les protéines par leur degron interne (non N-terminal). C’est le cas notamment de la protéine Cup9 qui est un substrat de Ubr1 grâce à un signal interne (Byrd et al., 1998). Il aura fallu attendre 2014 et le travail de Kim et al pour mettre en évidence que Ubr1 est aussi capable de reconnaître certaines protéines conservant leur méthionine initiatrice si celles-ci portent en position 2 un grand résidu hydrophobe (Kim et al., 2014).

La troisième voie de la N-end Rule est la Pro/N-End Rule. Comme son nom l’indique elle concerne les protéines qui ont à leur extrémité N-terminale une proline. Il s’agit notamment des enzymes gluconéogéniques (Fbp1, Icl1, Mdh2, ou Pck1). Ces enzymes sont dégradées via la Pro/N-end rule lorsque la concentration de glucose est élevée (D’après la revue (Nguyen et al., 2018)). Des études

ont démontré que l'ubiquitine ligase E3 Gid4, qui est impliquée dans cette voie reconnaît aussi des protéines avec une proline en deuxième position, c'est le cas par exemple pour la protéine Pck1. Une étude combinant un test de double hybride sur différents mutants d'une protéine connue pour être ciblée par la Gid4, Fbp1, montre que le sillon de liaison de substrat Gid4 permet une interaction avec le 6ème résidu d'une protéine (Chen et al., 2017). Tout comme l'Ac et l'Arg/N-end Rule, les protéines ubiquitinées vont être dégradées par le protéasome. Cependant les protéines ciblées par la Pro/N-end rule ne peuvent pas être ciblées par les deux autres voies (Chen et al., 2017).

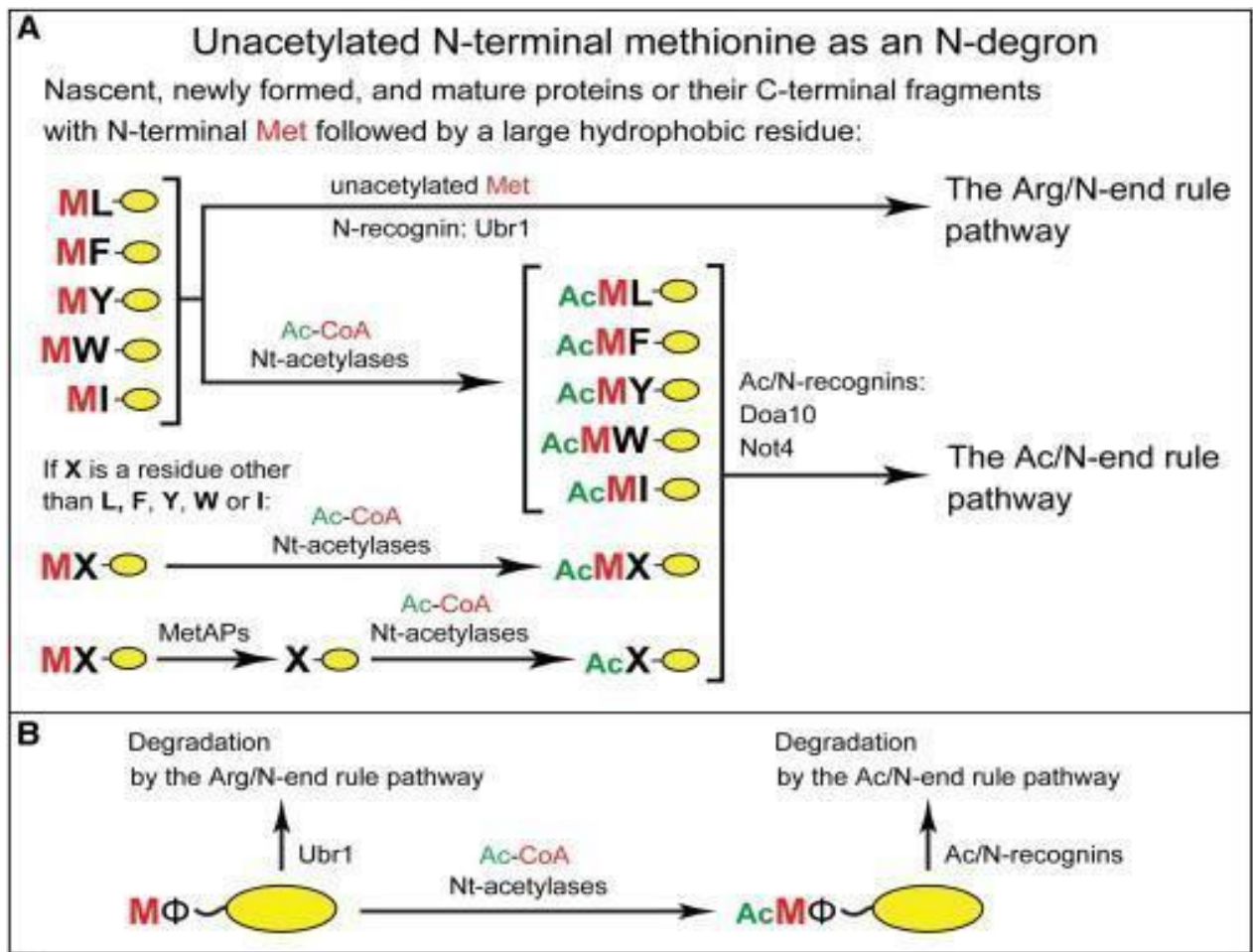


Figure 20 : Ce diagramme représente la complémentarité fonctionnelle entre les voies de la règle Arg/N-end et de la règle Ac/N-end. Cette complémentarité découle de la reconnaissance des Met Φ /N-degrons pour les protéines avec un résidu hydrophobe suivant une méthionine (Met- Φ) par rapport à la reconnaissance des AcMet Φ /N-degrons précédemment caractérisés (Kim et al., 2014).

b. Adressage et interaction protéines/protéines

L'acétyl est un groupement hydrophobe. Ainsi l'acétylation N-terminale a pour conséquence de rendre l'extrémité des protéines concernées plus volumineuse et plus hydrophobe. Cette augmentation de l'hydrophobicité N-terminale peut jouer un rôle sur les interactions protéines/protéines. C'est ce qui a été démontré par une étude biochimique et biophysique (Monda et al., 2013). Il a été démontré que la N-acétylation peut optimiser l'adressage des protéines vers leur compartiment cellulaire cible. C'est le cas notamment pour l' α -synucléine qui est une cible de NatB et qui, lorsqu'elle est acétylée en N-terminal, est plus stable, s'agrège moins et a une meilleure affinité pour les vésicules lipidiques (Bartels et al., 2011; Dikiy and Eliezer, 2014; Trexler and Rhoades, 2012). Or l'une des caractéristiques de la maladie de Parkinson est l'agrégation de cette protéine pour former ce que l'on appelle « corps de Lewy ». Il a également été vu que l'acétylation, notamment via NatC, permettait le bon adressage de Arl3 et Grh1 vers l'appareil de golgi (Behnia et al., 2004). Cependant, ce n'est pas un effet généralisable à l'ensemble des substrats de NATs. Une publication a montré que sur 15 protéines cibles putatives de NatC, 13 n'ont pas perdu leurs localisation dans un mutant de NatC (Aksnes et al., 2013). Une étude similaire a été réalisé dans le mutant de NatB et n'a pas vu de délocalisation chez 13 cibles putatives de NatB étudiées (Caesar et al., 2006).

c. Repliement des protéines

Un autre rôle que peut avoir l'acétylation N-terminale des protéines est de favoriser leur repliement. Plusieurs protéines présentent un défaut de repliement lorsqu'une des NATs est déplétée. Ce défaut peut être indirectement lié à la perte de N-acétylation de protéines chaperonnes dont le rôle dans le repliement est indispensable. Dans une souche déplétée de NatA, on observe des problèmes de repliement de nombreuses protéines (Holmes et al., 2014). En absence de NatA, il a en particulier été démontré que la protéine chaperonne HSP90 perd certaines de ses fonctions, notamment sa capacité à interagir avec Chk1 (Oh et al., 2017). De plus, le repliement des protéines est dépendant de la stabilité des structures secondaires qu'elles forment. Or il a également été démontré que l'acétylation peut jouer un rôle stabilisateur d'hélices alpha amphiphiles (Greenfield et al., 1994; Jarvis et al., 1995). Une étude récente de 2021 a montré par protéomique quantitative que dans le mutant NatB on retrouve des agrégats de protéines mal repliés (Friedrich et al., 2021). Ce phénomène a également été observé dans le mutant NatA mais dans une moindre mesure. On aurait pu penser que cette agrégation est liée principalement à une activité réduite de protéines chaperonnes à cause du manque

d'acétylation. Pourtant, NatA acétyle les principales chaperonnes et on retrouve des agrégats principalement dans le mutant NatB. De plus, il a été montré qu'il existait un chevauchement de substrat entre NatB et les chaperonnes ssb. Ces éléments combinés au fait que NatB agisse de manière co traductionnelle poussent à penser que l'acétylation joue un rôle direct sur le repliement des protéines (Friedrich et al., 2021).

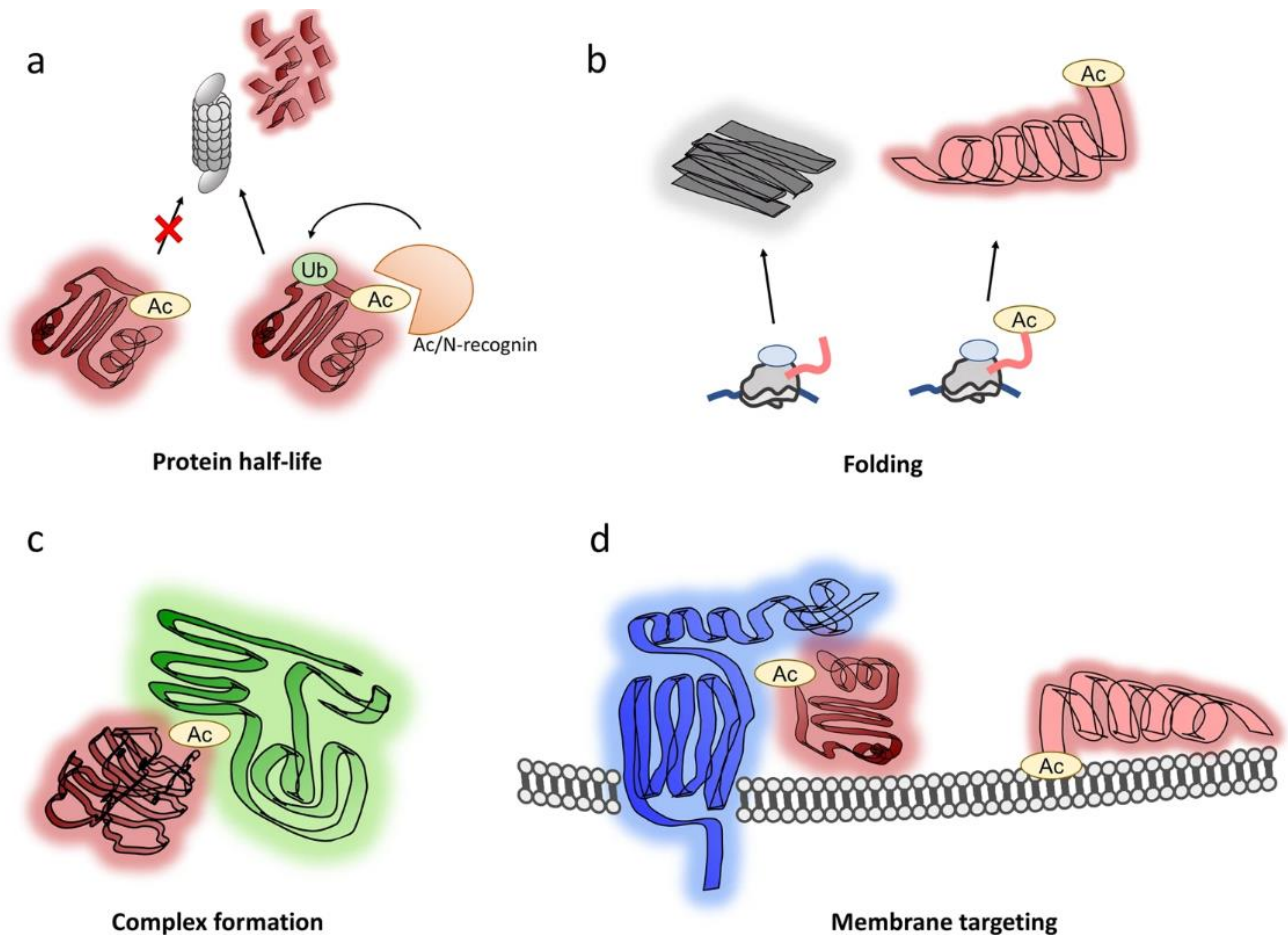


Figure 21 : Résumé schématique des différentes conséquences d'une acétylation N-terminale des protéines : la stabilité (A), le repliement (B), les interactions protéines-protéines (C) et l'adressage des protéines vers les membranes (D) (Ree et al., 2018)

3. Biais de composition à la position 2

La très grande majorité des protéines commence par une méthionine initiatrice. Le résidu suivant cette méthionine est particulier car c'est le premier résidu de la séquence qui peut différer d'une protéine à l'autre. Comme je l'ai expliqué dans les paragraphes précédents, la nature du résidu à cette position détermine le recrutement de diverses enzymes de modification qui jouent un rôle important sur le devenir des protéines. En 1987 déjà, l'analyse de 96 ARNm de la levure *Saccharomyces cerevisiae* a montré une sur-représentation du codon UCU codant la sérine à la position 2 de la phase codante de ces ARNm (Hamilton et al., 1987). Cette surreprésentation a de nouveau été observée quelques années plus tard, dans le cadre de l'étude de 211 séquences d'ARNm (Miyasaka, 1999). Elle a été mise en évidence en particulier parmi les ARNm codant des protéines fortement exprimées, suggérant que ce biais en position 2 de la phase codante pourrait être lié à un contexte nucléotidique permettant de favoriser la traduction (Hamilton et al., 1987). D'ailleurs une autre publication allant dans ce sens indique que parmi les 80 gènes les plus exprimés dans la cellule, il existe une sur-représentation d'alanine en position 2 de la protéine produite. Cette observation a été faite chez neuf bactéries, trois archées et trois eucaryotes unicellulaires dont *S. cerevisiae*. Elle concerne un plus grand nombre d'ARNm que celui utilisé dans l'étude de 1987, ce qui pourrait expliquer cette observation d'un enrichissement en alanine plutôt qu'en sérine en position 2 de ces protéines abondantes (Tats et al., 2006). Une étude publiée en 2001 a également montré chez la bactérie *E. coli* que la présence de codons riches en adénine en position 2 d'une séquence codante permet d'optimiser l'initiation de la traduction, rendant cette dernière jusque 20 fois plus efficace que pour d'autres codons (Stenström et al., 2001).

La reconnaissance de la position 2 des protéines par des enzymes de modification et son lien avec l'efficacité de traduction pourraient impliquer que cette position est soumise à de fortes contraintes évolutives et que l'utilisation des acides aminés à cette position diffère de celle observée en moyenne aux autres positions. Une première étude globale de la fréquence des acides aminés en position 2 dans différents taxons a été réalisée en 2010 (Shemesh et al., 2010). Ce travail a mis en lumière des différences de distribution des acides aminés en position 2 par rapport à leur utilisation moyenne chez les bactéries, les eucaryotes unicellulaires et les mammifères. Chez *Escherichia coli* le résidu le plus abondant en position 2 est la lysine. Chez *S. cerevisiae* il s'agit de la sérine, chez *Homo sapiens* et chez *Arabidopsis thaliana*, c'est l'alanine qui est sur-représentée à cette position (Figure 22). Il a également été démontré que le biais en position 2 des protéines peut être lié à la fonction et au lieu d'adressage des protéines.

Une publication récente a notamment démontré qu’il existait un lien entre les peptides de ciblage N terminaux et l’acide aminé situé en deuxième position (Almagro Armenteros et al., 2019). C’est l’utilisation de l’algorithme de prédiction TargetP 2.0 qui utilise la méthode du Deep Learning, qui a montré ce lien. Ainsi, chez les champignons et les organismes unicellulaires moins de 30% des protéines avec un peptide d’adressage vers la mitochondrie et le réticulum endoplasmique ont un acide aminé qui conduit au clivage de la méthionine initiatrice contre 60% pour les protéines qui n’ont pas ce peptide (Almagro Armenteros et al., 2019).

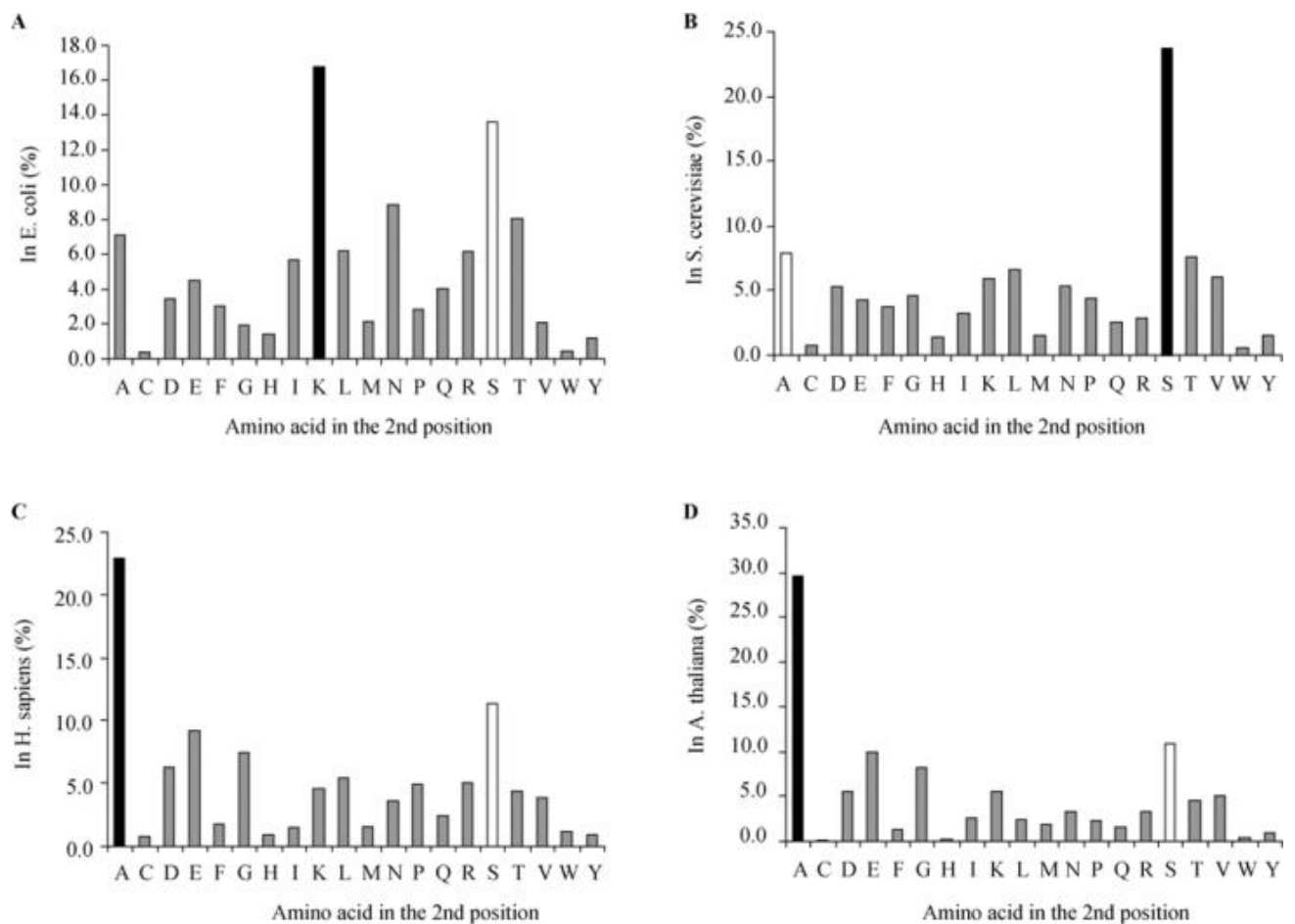


Figure 22 : Fréquences des 20 acides aminés en position 2 chez différentes espèces. (A) chez *E. coli* (n=4 474) (B) Chez *S. cerevisiae* (n=4,819) (C) Chez *H. sapiens* (n=10 790) (D) Chez *A. thaliana* (n=3 056). Les colonnes noires représentent les acides aminés de deuxième position les plus abondants. Les colonnes blanches représentent les acides aminés suivants les plus abondants (Shemesh et al., 2010)

Au laboratoire, une étude globale et fonctionnelle de la position 2 des protéines a été réalisée. Dans un premier temps, les biais d'acides aminés observés dans la littérature en position 2 chez la levure ont été confirmés par des analyses *in silico*. Les fréquences d'acides aminés en position 2 des protéines ont été comparées à leur fréquence moyenne d'utilisation aux autres positions. De manière frappante, on peut distinguer une distribution asymétrique des acides aminés. On peut voir que ce biais est flagrant pour la sérine dont la fréquence passe de 9 % dans le protéome à 24 % en position 2. Inversement, certains acides aminés sont sous-représentés à cette position. C'est le cas notamment de R, Q, E, L, Y, I (Figure 23). Il n'y a pas de corrélation entre l'abondance des acides aminés dans le protéome et ceux en position 2. Ce résultat suggère qu'il doit exister une sélection fonctionnelle pour ce résidu (Shemesh et al., 2010).

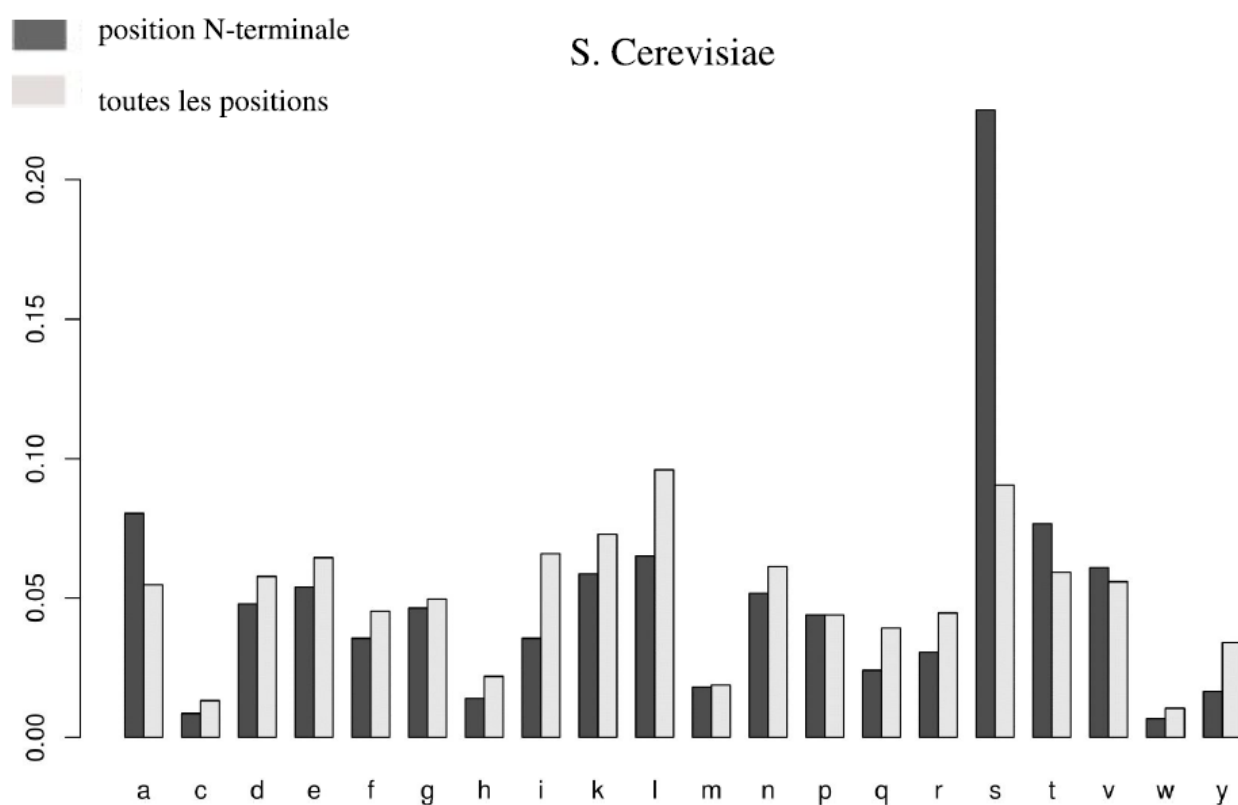


Figure 23 : Histogramme obtenu à partir d'analyse *in silico* qui représente la fréquence des 20 acides aminés en position 2 par rapport à leurs fréquences dans le protéome (D'après le rapport de M2 de Lucie Gaulier)

III. Problématique de la thèse : exploration du lien entre les biais de composition en position 2, NatC et l'import des protéines mitochondriales

1. Identification d'un biais en position 2 des protéines mitochondriales

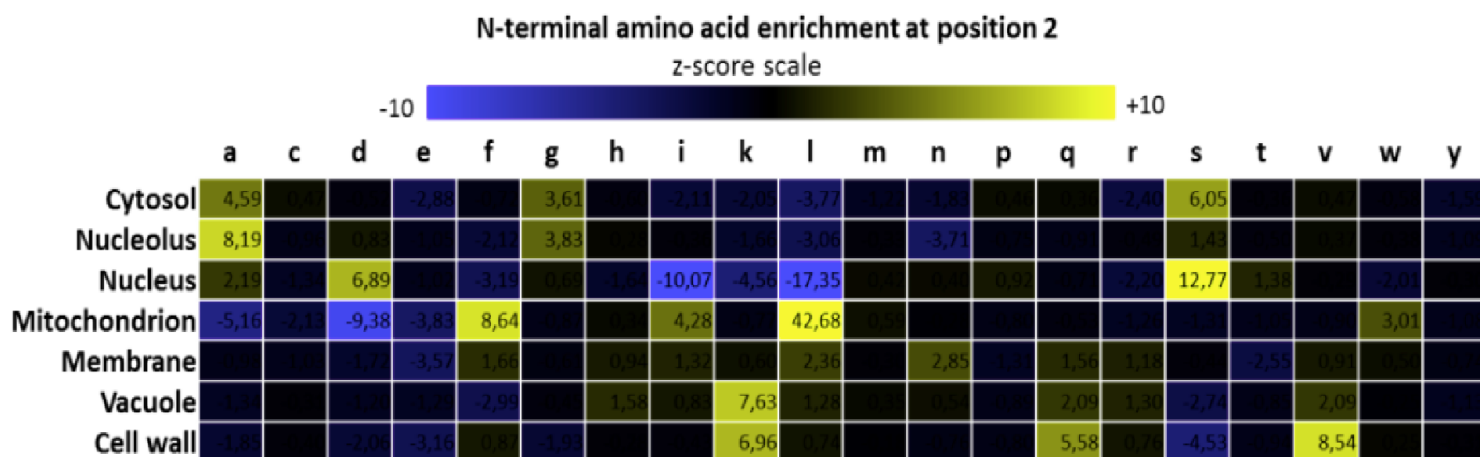


Figure 24 : Analyse de l'enrichissement des différents acides aminés en position 2 selon le compartiment cellulaire dans le protéome de *S. cerevisiae*. L'enrichissement est mesuré en utilisant un Z-score du log10 des p-values calculées par un test hypergéométrique. Une valeur positive indique un enrichissement, une valeur négative une sous-représentation (Données de l'équipe non publiées)

Grâce à une exploration globale in silico, l'équipe a mis en évidence qu'il existe un lien entre le résidu en position 2 des protéines et leur appartenance aux catégories fonctionnelles définies par la Gene Ontology (GO). L'un des résultats les plus forts obtenu lors de cette analyse est une sur-représentation d'acides aminés hydrophobes comme la Leucine, la phénylalanine et l'isoleucine en position 2 des protéines mitochondriales (Figure 24). Alors que ces résidus en position 2 ne représentent que 15% des protéines cellulaires, ils concernent plus de 50 % des protéines destinées aux mitochondries.

Réciproquement, une autre analyse a été réalisée afin de déterminer comment sont représentées les catégories GO au sein de sous-groupes de protéines classées en fonction du résidu qu'elles portent à leur position 2 (Figure 25). Le résultat de cette étude est représenté sous forme de diagrammes. On peut voir que la moitié des protéines ayant une leucine en position 2, 38 % des protéines ayant une phénylalanine et 20 % ayant un tryptophane à cette position sont des protéines mitochondriales.

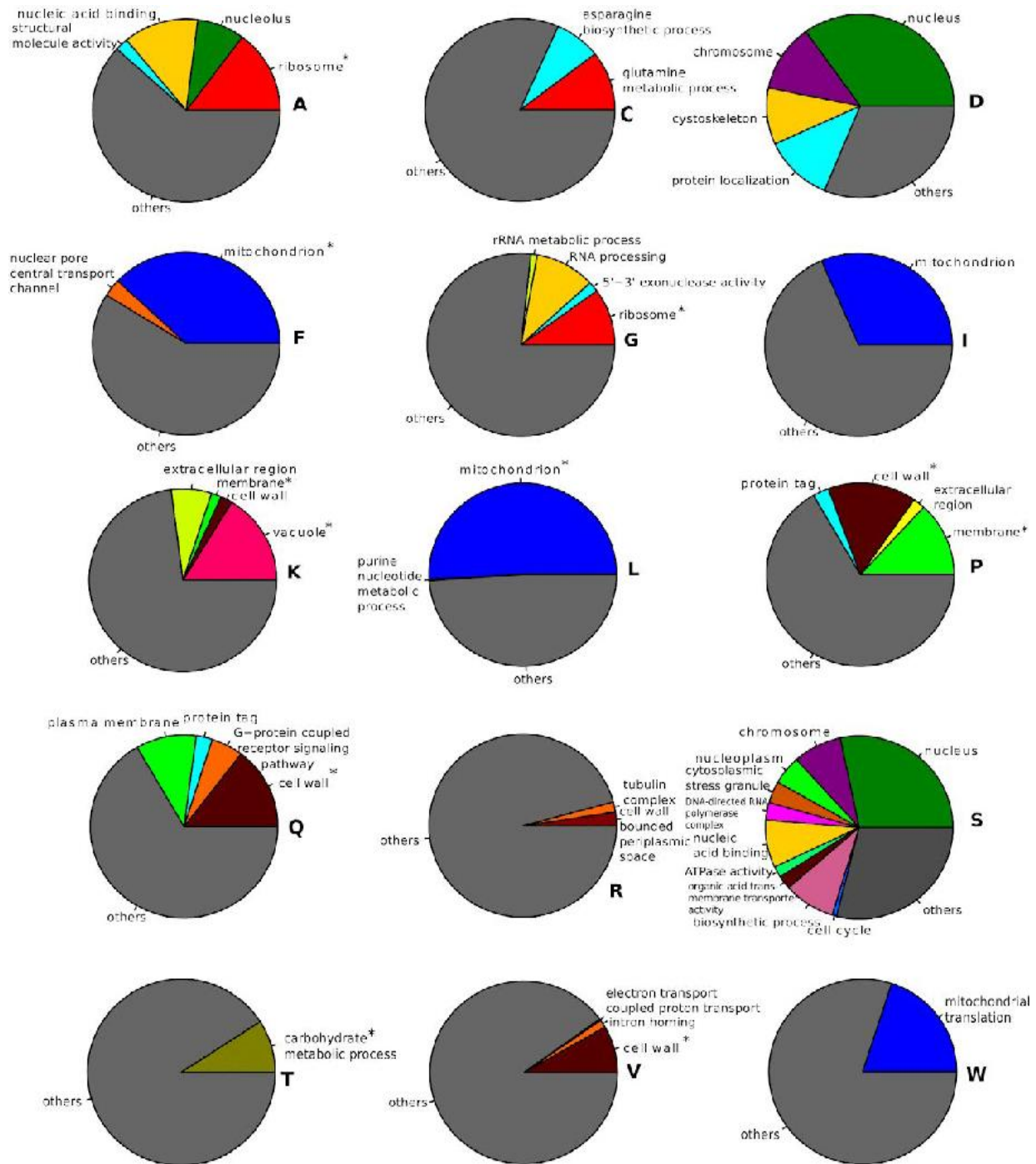


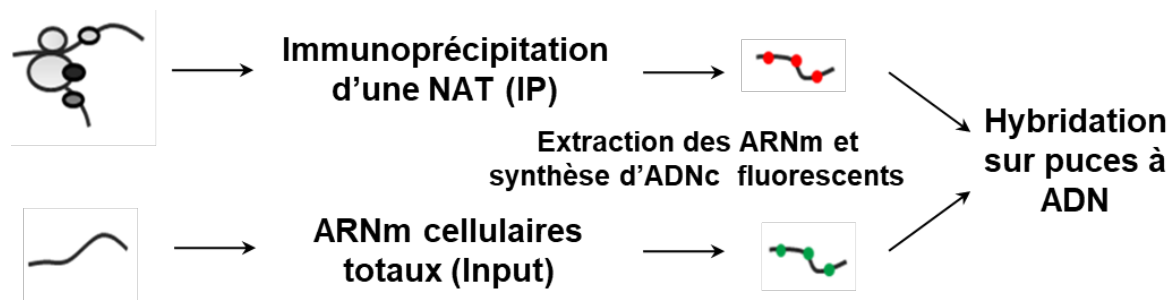
Figure 25 : Diagramme circulaire représentant pour chaque acide aminé la part des termes GO sélectionné par l'analyse de biais au niveau du RNt. Les termes GO sélectionnés correspondent aux termes enrichis pour chaque acide aminé ($p \text{ value} < 10^{-4}$). (D'après le rapport de M2 de Lucie Gaulier)

2. Confirmation de l'interaction entre l'enzyme NatC et les protéines mitochondriales

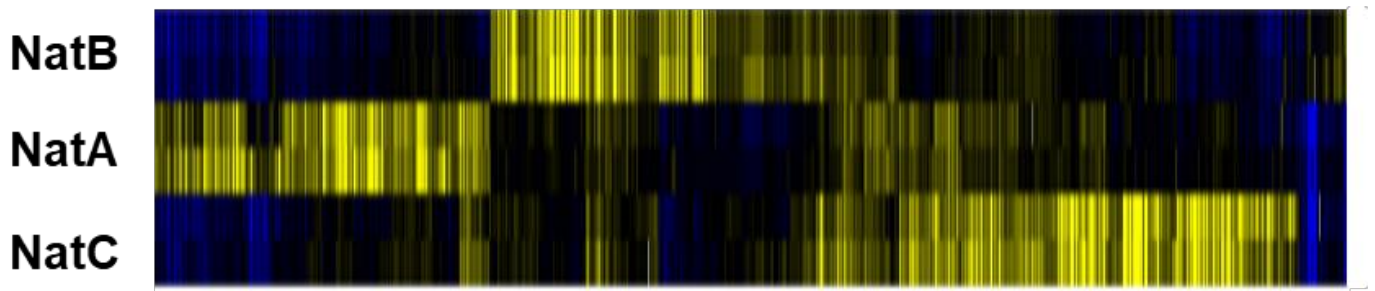
Les acides aminés enrichis en position 2 des protéines mitochondriales (Leucine, Phénylalanine, Isoleucine et Tryptophane) sont ceux connus comme étant ciblés par NatC (voir chapitre précédent). Cependant, le nombre de cibles de NatC validées expérimentalement publiées dans la littérature est faible. En effet, la sélectivité de NatC a été déduite via l'utilisation de protéines rapportrices mais aucune étude publiée au début de ma thèse n'avait permis de décrire globalement les substrats de NatC *in vivo*. L'équipe a voulu vérifier l'hypothèse selon laquelle les protéines mitochondriales sont des substrats de NatC. Or, les précurseurs mitochondriaux sont extrêmement instables car suite à leur traduction, ils sont rapidement importés dans les mitochondries où leur extrémité N-terminale est clivée et dégradée. Il est donc difficile de les étudier par des approches N-terminal protéomiques type COFRADIC. Pour contourner ce problème, l'équipe a mis en place une technologie basée sur l'identification co-traductionnelle des cibles protéiques de NatC à partir des ARNm correspondants : le Sel-TRAP.

Le "Selective Affinity Purification of Translating Ribosomes" (sel-TRAP) est une technologie qui consiste en l'immunoprécipitation des Nats dans un contexte où l'on maintient les complexes de traduction (Figure 26.A). Cela permet d'identifier indirectement les protéines cibles de chacune des NATs via les ARNm contenus dans les immunoprécipitats (IP) comparés aux ARNm totaux (input) et utilisant comme référence l'enrichissement de chaque ARNm dans une expérience de Sel-TRAP immunoprécipitant la protéine Rpl16, présente dans tous les ribosomes. L'équipe a immunoprécipité NatC mais également NatA et NatB. Il a alors été observé que chaque NAT cible bien un sous-ensemble d'ARNm qui lui est propre ce qui est cohérent avec la littérature, à savoir le fait que chaque NAT présente une sélectivité distincte (Figure 26.B).

A



B



C



Figure 26 : (A) Stratégie expérimentale du Sel-Trap afin de capturer de manière cotraductionnelle les NATs (B) Heatmap après clustering hiérarchique, généré avec le logiciel TMeV montrant les \log_2 de ratio IP/Input mesurés dans les 2 réplicats des sel-TRAP de NatA, NatB et NatC. Chaque ligne représente une expérience, chaque colonne un ARNm. (C) Score d'un test hypergéométrique d'enrichissement des cibles de chaque NAT dans chacun des grands compartiments cellulaires. Une valeur positive indique un enrichissement, une valeur négative une déplétion. Le score limite correspondant à une p-value de 5% est de 1.3 (Données de l'équipe non publiées)

Cette analyse a révélé que les différentes NATs ciblent de manière différentielle les protéines appartenant aux divers compartiments cellulaires. En particulier, cette analyse a démontré que le complexe NatC cible majoritairement des protéines destinées au compartiment mitochondrial. (Figure 26.C)

3. Problématique de la thèse

Dans ce contexte, ma thèse avait pour objectif d'élucider le rôle de NatC et l'importance de la signature N-terminale des précurseurs mitochondriaux dans leur import mitochondrial. Compte tenu de l'implication de NatC dans la N-end rule et de la faible demi-vie observée des protéines présentant une signature de recrutement de NatC (Kats et al., 2018), une hypothèse était que la signature N-terminale des précurseurs mitochondriaux permettrait leur dégradation rapide en cas de problème d'import, protégeant ainsi les cellules du mPOS. Une autre hypothèse était que l'acétylation par NatC participerait au processus d'import mitochondrial lui-même. J'ai donc évalué le(s) rôle(s) de la position 2 des précurseurs mitochondriaux et de NatC par des approches de mutagenèse couplées avec des analyses des profils d'import de ces précurseurs et des analyses phénotypiques. Deux approches complémentaires ont été mises en place au cours de ma thèse : (1) une mutagenèse dirigée du résidu en position 2 du précurseur d'une protéine mitochondriale modèle, Hsp60, (2) la création de souches délétées de NatC dans le but d'observer l'impact global de NatC sur les précurseurs mitochondriaux.

La partie « Résultats » sera scindée en deux parties distinctes. Dans un premier temps, vous découvrirez un article dans lequel je suis co-premier auteur. Celui-ci contient le travail que j'ai réalisé pour analyser le rôle du résidu en position 2 de Hsp60 chez la levure *S. cerevisiae*. La seconde partie, qui n'a pas encore donné lieu à publication, sera consacrée à l'étude de NatC.

Résultats

I. L'article : Functional mapping of N-terminal residues in the yeast proteome uncovers novel determinants for mitochondrial protein import

Mon travail de thèse a permis la rédaction d'un article qui est actuellement en révision dans la revue « Plos Genetics ». Vous trouverez une version de cet article ci-après, ainsi que sur BioRxiv : <https://www.biorxiv.org/content/10.1101/2022.08.19.504527v2>

Ce travail a été réalisé en étroite collaboration avec Salomé Nashed, une doctorante de l'équipe, co-première auteure de l'article qui décrit nos résultats. La thèse de Salomé avait pour objectif de réaliser une analyse globale et fonctionnelle, *in silico*, de la position 2 des protéines, d'apporter une dimension évolutive et multi-espèces à ces analyses et d'analyser en détail les résultats de Sel-TRAP et de COFRADIC concernant les NATs obtenus par le laboratoire et disponibles dans la littérature, respectivement. C'est Salomé qui a réalisé toutes les analyses bio-informatiques présentes dans l'article et celles que je vous présenterai dans cette première partie.

Je me suis particulièrement intéressé au résidu en position 2 des protéines mitochondriales possédant une séquence d'adressage N-terminale (MTS) car des analyses *in silico* réalisées par Salomé Nashed ont montré que ces protéines présentent, à cette position, une forte sur-représentation de larges résidus hydrophobes à la fois par rapport à leur utilisation en position 2 dans le protéome entier et par rapport aux protéines mitochondriales sans MTS. Il s'agit de la leucine, la phénylalanine, l'isoleucine, le tryptophane et dans une moindre mesure, la méthionine. Elle a de plus montré que ce biais est spécifique de la position 2, c'est-à-dire qu'il est distinct de la composition globale du MTS, et qu'il est très conservé chez les *Saccharomyces cerevisiae*, une lignée de levures dont l'évolution s'étend sur près de 400 millions d'années.

Mon but a été d'étudier le rôle de ce résidu par des approches expérimentales. Pour cela j'ai travaillé sur la levure *S. cerevisiae* et j'ai utilisé comme modèle d'étude la protéine Hsp60. Mon choix s'est porté sur cette protéine pour trois raisons : (1) il s'agit d'une protéine mitochondriale

possédant un MTS, (2) cette protéine présente une leucine en position 2, qui est le résidu le plus sur-représenté dans les protéines mitochondriales à MTS d'après nos analyses *in silico* et enfin (3) l'équipe possède des anticorps dirigés contre cette protéine ce qui permet de la visualiser par la technique de Western Blot.

L'étude des précurseurs mitochondriaux est complexe car dans des conditions physiologiques, ils sont présents en très faible quantité dans la cellule. En effet, suite à leur synthèse dans le cytoplasme, ces précurseurs sont très rapidement importés dans les mitochondries dans lesquels ils sont maturés. Ainsi, j'ai dû travailler dans un système expérimental qui me permette d'optimiser la visualisation de la forme précurseure de Hsp60. Pour cela j'ai utilisé une souche dotée d'un fond génétique particulier : *YPH-pam16Δ-hpam16(N76D)* dont la souche de référence correspondante est *YPH499*. Cette souche est mutée dans la protéine Pam16 impliquée dans le système d'import mitochondrial, qui est remplacée par son homologue humain avec une mutation ponctuelle (N76D) qui altère sa fonctionnalité. Cette souche a été construite par Agnès Delahodde qui nous l'a gracieusement fournie (Mehawej et al., 2014). Cette mutation N76D confère à la souche un phénotype thermosensible, de sorte qu'à 37°C, l'import mitochondrial est très perturbé.

Mon premier objectif a donc été de tester ce système expérimental et de faire un criblage phénotypique de cette souche. Comme attendu j'ai été capable de détecter par Western Blot la forme précurseure de Hsp60 à 37°C alors qu'elle n'est pas toujours détectable à la température permissive. Cependant, j'ai constaté que cette souche présente un phénotype de croissance ralentie dès 30°C sur milieu riche par rapport à sa souche de référence (Figure 27). Cela signifie que la souche *YPH-pam16Δ-hpam16(N76D)* a un import mitochondrial perturbé même à température permissive et expliquerait pourquoi il est possible de détecter la bande précurseur de cette souche à 30 degrés.

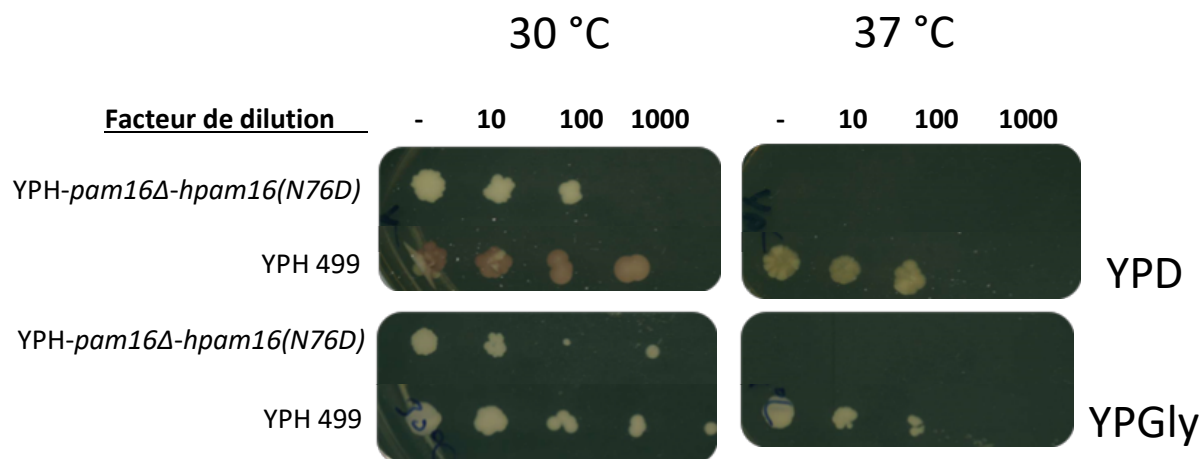


Figure 27 : Test de croissance en gouttes des mutants souches YPH499 et YPH-*pam16Δ-hpam16(N76D)* sur milieu riche (YPD) et sur milieu avec une source de carbone non fermentable (YPGly). Le test de croissance est réalisé sur 3 jours d'incubation

Après avoir validé mon système expérimental, j'ai généré des mutants ponctuels de la position 2 de Hsp60 (L2X) en utilisant la technologie CRISPR Cas9. Le premier mutant que j'ai analysé par Western Blot est le mutant L2S, la sérine n'étant pas un acide aminé sur-représenté en position 2 dans nos analyses *in silico* contrairement à la leucine. De façon intéressante, j'ai constaté un profil perturbé pour ce mutant, qui se traduit par une baisse de la quantité de la forme mature et une accumulation visible de la forme précurseure dès 30°C. C'était un résultat encourageant, car il démontre l'importance de ce résidu en position 2, et je l'ai obtenu pour d'autres résidus : L2V, L2P, L2K et L2D. Comme attendu, lorsque j'ai remplacé la leucine par la phénylalanine F, un autre acide aminé sur-représenté en position 2 des protéines à MTS, dans le but de générer un contrôle positif, j'ai observé un phénotype moléculaire comparable à celui de notre souche WT (Figure 5.a de l'article). Parce que je travaillais sur une souche N76D déjà malade à 30°C, j'ai décidé de mettre en place un contrôle interne afin de m'assurer que le résultat obtenu était bien lié à la mutation en position 2. J'ai donc exprimé deux versions de *HSP60* dans la cellule : une version chromosomique non mutée et une version plasmidique mutée en position 2 et portant en C-terminal une étiquette composée de 13Myc. Cette étiquette devait me permettre de distinguer la version mutée de la version sauvage par Western Blot. J'ai commencé par faire un premier essai avec un plasmide permettant l'expression d'une version WT de Hsp60 et un autre permettant celle d'une version présentant une sérine en position 2. Lors de la transformation des levures par l'un ou l'autre de ces deux plasmides, j'ai obtenu un résultat surprenant. J'ai en effet obtenu de nombreux transformants avec le plasmide L2S mais

très peu voire pas de clones avec le plasmide WT. J'ai donc essayé de comprendre ce résultat. Dans la littérature il a été montré que l'ajout d'étiquettes en C-terminal de Hsp60 perturbe la croissance de la cellule et peut même être létal si ces étiquettes sont trop longues (Fang and Cheng, 2002). Hsp60 formant un homotétradécamère, l'absence de clone portant le plasmide WT peut être expliquée par un effet dominant négatif de la version *HSP60* étiquetée Myc. Ce phénotype létal est sauvé par la mutation leucine vers sérine en position 2, qui pourrait perturber l'import de la protéine dans la mitochondrie. En effet, la perturbation de l'import de la forme étiquetée permettrait d'éviter qu'elle ne vienne interférer avec la formation dans les mitochondries de complexes fonctionnels, qui seraient alors exclusivement constitués de la forme chromosomique sauvage.

J'ai mis à profit cet effet dominant négatif qui me permettait d'avoir un nouveau système expérimental permettant de tester rapidement et indirectement l'efficacité d'import du MTS de Hsp60. J'ai alors décidé de tester l'ensemble des 20 acides aminés à la position 2 de Hsp60.

J'ai obtenu deux types de résultats. Les larges résidus hydrophobes enrichis en position 2 des MTS dans l'analyse *in silico* (L, F, I, W et M) conduisent à un phénotype dominant négatif. Au contraire, les autres résidus sont associés à un sauvetage phénotypique causé par une perturbation de l'import mitochondrial.

Mon travail a ainsi permis de confirmer la pertinence des prédictions découlant des analyses *in silico* réalisées au laboratoire. Il a également montré que le résidu en position 2 des protéines portant un MTS est critique pour leur devenir cellulaire, au moins dans le cas de Hsp60. Ces résultats sont exposés en détail dans l'article qui suit.

En complément des expériences dans le fond génétique N76D incluses dans l'article, j'ai testé l'impact sur Hsp60 endogène des mutations L2F et L2S dans la souche YPH499 (figure 28). Je n'ai pas observé d'effet de la mutation L2S en western blot, contrairement à ce que j'avais vu chez la souche N76D. Ce résultat est discuté dans la partie discussion de ce manuscrit de thèse.

Functional mapping of N-terminal residues in the yeast proteome uncovers novel determinants for mitochondrial protein import.

Nashed S^{1*}, El Barbry H^{1*}, Benchouaia M¹, Dijoux-Maréchal A¹, Ruiz Gutierrez N¹, Gaulier L¹, Chevreux G², Le Crom S¹, Palancade B², Devaux F¹, Laine E¹, Garcia M^{1**}

¹ Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative, Paris, France

² Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France.

* Equally contributing authors

**To whom correspondence should be addressed. Tel: +33 1 44278140; Fax: +33 1 44277336; Email: mathilde.garcia@sorbonne-universite.fr

ABSTRACT

N-terminal ends of polypeptides are critical for the selective co-translational recruitment of N-terminal modification enzymes. However, it is unknown whether specific N-terminal signatures differentially regulate protein fate according to their cellular functions. In this work, we developed an in-silico approach to detect functional preferences in cellular N-terminomes, and identified in *S. cerevisiae* more than 200 Gene Ontology terms with specific N-terminal signatures. In particular, we discovered that Mitochondrial Targeting Sequences (MTS) show a strong and specific over-representation at position 2 of hydrophobic residues known to define potential substrates of the N-terminal acetyltransferase NatC. We validated mitochondrial precursors as co-translational targets of NatC by selective purification of translating ribosomes, and found that their N-terminal signature is conserved in Saccharomycotina yeasts. Finally, systematic mutagenesis of the position 2 in a prototypal yeast mitochondrial protein confirmed its critical role in mitochondrial protein import. Our work highlights the hydrophobicity of MTS N-terminal residues and their modification by NatC as critical features for the definition of the mitochondrial proteome, providing a molecular explanation for mitochondrial defects observed in yeast or human NatC-depleted cells. Functional mapping of N-terminal residues thus has the potential to support the discovery of novel mechanisms of protein regulation or targeting.

INTRODUCTION

As soon as they emerge from the ribosomal tunnel, the protein nascent chains recruit factors that will play key roles in their life cycle (1–5). Such recruitment is dependent on the nature of the protein N-terminal amino acid residues. In particular, the specific recognition and co-translational action of several N-terminal modification enzymes is determined by the type of the amino acid residue directly following the initiator methionine, namely the residue at position 2.

For instance, methionine aminopeptidases (MetAPs) will remove the initiator methionine (iMet) in a large fraction of the nascent chains displaying an amino acid with a small radius of gyration (Ala, Cys, Gly, Pro, Ser, Thr, or Val) at position 2 (6). In addition, N-terminal acetyltransferases (NATs) catalyze the irreversible covalent attachment of an acetyl group (CH₃CO) to the free α -amino group (NH₃⁺) at the protein N-terminus. Three major NATs, namely NatA, NatB and NatC, that acetylate nascent chains in a co-translational manner, are conserved from yeast to human (7). The actual impact of the inactivation of NATs on the N-terminal acetylation of subsets of cellular proteins was measured by

various proteomics techniques, eventually providing experimental data on the N-terminal acetylation status for up to 10% of the cellular proteome in yeast and in human (8–11). All these studies led to the conclusion that NatA can acetylate the N-termini beginning with Ala, Cys, Gly, Ser, Thr, or Val after removal of iMet, whereas the other two can acetylate the uncleaved iMet if it is followed by a second specific residue, namely Asn, Asp, Gln, or Glu for NatB, and Ile, Leu, Phe, or Trp for NatC. Based on these substrate specificities together with the observed frequencies of the 20 amino acids at position 2 in proteomes and the partial N-terminal acetylome data, it was estimated that 80-90% of human and 50-70% of yeast proteins could potentially be acetylated by one of these three enzymes (7, 12, 13). Finally, Methyl-, myristoyl-, and palmitoyltransferases can also selectively modify, during or after translation, the N-terminus of proteins after iMet cleavage.

The biological importance of these N-terminal modifications is underlined by the strong defects observed when the corresponding enzymes are inactivated. Complete inactivation of MetAPs activity is lethal in *Escherichia coli*, *Salmonella typhimurium*, and *Saccharomyces cerevisiae*, and the identification of human MetAP1 and MetAP2 as targets for putative anticancer drugs further confirmed the importance of this enzyme family (14, 15). In addition, N-terminal acetylation has been implicated in several diseases, including cancers, developmental disorders, as well as Parkinson's disease (5). At the molecular level, the importance of the nature of the second residue and the associated N-terminal modifications have been implicated in all major steps of proteins life cycle including folding and aggregation, protein interactions and complex formation, protein subcellular targeting, and protein turnover through proteasomal degradation pathways (reviewed in [6]). In particular, during the last decades, the exploration of the N-end rule pathway, linking the nature of the N-terminus to protein *in vivo* stability has emphasized the importance of protein N-end signatures and demonstrated that iMet cleavage and N-terminal acetylation are important players in the early control of proteins life cycle (16–19). Finally, preferences for the use of amino acids at position 2, different from those observed globally in the proteome, have been described in various species, suggesting that selection pressures act on this particular protein position. In addition, these biases are taxon-specific, indicating progressive changes during evolution (20).

This extensive knowledge on the general importance of protein N-terminal signatures sharply contrasts with the lack of data describing their potential implication in specific functional pathways. To address this issue, we have developed an *in-silico* approach at the proteome scale allowing for the systematic detection of functional biases in the use of the 20 amino acids at position 2 of proteins. The rationale of our approach is that, given the importance of the residue at position 2 in the life cycle of proteins, evolutionary constraints could have led to the selection of specific amino acids at this position in groups of proteins from the same cellular compartment or involved in the same biological pathway. As a proof of concept, we analyzed the *S. cerevisiae* proteome to statistically assess all significant and specific overrepresentations of one or more amino acids at position 2 in protein subsets sharing the same Gene Ontology (GO) annotations. Using this approach, we were able to identify various groups of proteins with common GO annotations and characterized by particular amino acid preferences at position 2. We hypothesized that these newly identified N-terminal signatures are likely associated with important functional roles. We further characterized, with a combination of *in-silico* analyses and experimental assays, the function of the strong position-specific bias detected at position 2 for mitochondrial proteins.

Most mitochondrial proteins are translated by cytoplasmic ribosomes and addressed to the mitochondrial compartment via a cleavable presequence, the so-called Mitochondrial Targeting Sequence (MTS). We found that, throughout the Saccharomycotina yeast lineage, MTS-bearing

mitochondrial precursors present, just after the iMet, predominantly hydrophobic amino acids known to define potential NatC substrates (i.e., Leu, Phe, Ile, and Trp). Affinity-selective purification of the translating ribosomes then confirmed their co-translational recognition by this N-terminal acetyltransferase. Finally, we demonstrated the functional significance of the bias detected at position 2 of MTSs by site-directed mutagenesis of this position in a dominant negative allele of the essential mitochondrial protein Hsp60p. Only the amino-acids found to be overrepresented at position 2 of MTSs in our *in-silico* analyses allowed the efficient mitochondrial import of the toxic Hsp60p, as revealed by the associated loss of cell viability.

Our work has revealed an unknown and critical feature of MTSs. Defects in mitochondrial biogenesis are implicated in a number of serious pathologies, including neuropathies, cardiovascular disorders, myopathies, neurodegenerative diseases and cancers (21). In a long-term perspective, understanding the role of this signature for mitochondrial protein import is therefore crucial for human health. For example, this knowledge could be very useful for the design of mitochondrial gene therapies based on the targeting of specific proteins to the mitochondrial compartment (22). Beyond the results obtained for the proteins of the mitochondrial compartment, we provide the community with the first functional mapping of N-terminal residues, whose future exploration will undoubtedly contribute to the discovery of novel mechanisms of protein regulation or targeting.

RESULTS

Detection of functional biases in amino acid usage at position 2 in *S. cerevisiae* proteome

The amino acid usage at position 2 of *S. cerevisiae* proteins (Fig. 1a) is clearly different from the average amino acid distribution in its proteome. Strikingly, we observed a serine at this position in nearly 25% of the yeast proteins. This overrepresentation is both highly significant (HGT score equal to 268 corresponding to a p-value after hypergeometric test equal to 10^{-268} , see Methods for a definition of HGT score) and specific to position 2 (aspecificity score of 0%, see Methods). By contrast, several amino acids are underrepresented (HGT score <-3 and aspecificity score $<3\%$), including the hydrophobic residues leucine, isoleucine, and tyrosine as well as charged or polar residues such as arginine, glutamate and glutamine. Such asymmetric distribution, with the over-representation of serine, is conserved in budding yeasts (Fig. Sup1), as indicated by proteome analysis of 17 yeasts spanning nearly 400 million years of evolution of the Saccharomycotina lineage (23, 24).

We further developed a computational approach to explore the functional relevance of amino acid biases at the proteome scale and assess their statistical significance. We took advantage of the robust Gene Ontology annotations of the *S. cerevisiae* proteome to detect all significant and specific overrepresentations of one or more amino acids at position 2 in protein subsets defined by common GO annotations.

In a first step, we scanned the diverse GO terms describing the components, the cellular pathways or the protein molecular functions and we calculated the frequencies of the 20 amino acids at position 2 in the associated protein subsets. We then identified the subsets displaying significantly higher frequencies than expected by the overall distribution of amino acids at position 2. We assessed the significance of the enrichments by calculating the HGT scores derived from p-values obtained in hypergeometric tests. With this strategy, we identified 232 GO terms with a significantly increased frequency (HGT >4) for at least one amino acid. These GO terms were distributed as follows in the different GO categories: 72 GO terms corresponding to components, 138 to cellular pathways and 22

to molecular functions (Table S1). We designed a dedicated algorithm to reduce this initial list and eliminate information redundancy in each category. Rather than relying on the hierarchical relationships between the GO terms, as usually done by existing solutions (25), our algorithm directly analyzes the characteristics of the protein subsets defined by these GO terms, such as the number of proteins, the enrichment factors, and the overlaps between the subsets. In a first filtering step, we eliminate GO terms with the lowest frequency bias (<1.8 fold change) and corresponding to very generic components or processes, which typically include a very high number of proteins. Our algorithm then retains the GO terms encompassing the largest number of proteins and maximizing the coverage of the original dataset (called BestN GO terms, see Methods). When possible, it complements this set of bestN GO terms with the one or several smaller GO terms displaying the highest position 2 frequency biases (called BestF GO terms, see Methods). Each BestF is ultimately linked to the nearest BestN based on the overlap of their respective set of proteins.

Figure 1b illustrates the results of this filtering for the GO category "components" and represents the amino acid usage biases observed in the 17 GO terms that were finally retained after processing the initial list of 72 GO terms. Results for the other two GO categories, "cellular pathways" and "molecular functions", are available as supplementary data (see Table S1 and Fig. sup2 in which the original lists were reduced from 138 to 21 and from 22 to 10 GO terms, respectively). The final list of GO terms "components" includes the 6 BestN GO terms "Ribonucleoprotein complex", "Chromosome", "Mitochondrion", "Nuclear pore central transport channel", "Extracellular Region" and "Storage vacuole" (fig. 1b1). These GO terms covered 85% of the proteins with the amino acid preferences associated with the 55 GO terms retained after the first filtering step that removed low bias generic GO terms. The 11 GO BestF terms (Fig. 1b2) provided a more detailed picture of protein subsets with specific N-terminal amino acid usage. For example, "90S preribosome" (3.4-fold increase in glycine utilization at position 2) and "large cytosolic ribosomal subunit" (5.1-fold increase in alanine utilization at position 2) led to a much better characterization of the proteins involved in the bias observed in the more general GO term "ribonucleoprotein complex". Similarly, the "Kinetochore" proteins (4.4-fold increase in aspartate utilization at position 2) explain a large part of the aspartate bias detected in the GO term "Chromosome" since it accounts for 12 of the 21 proteins responsible for this bias detection. The BestF GO terms also pointed to several mitochondrial sub compartments with high preferences for leucine at position 2 (4.0- to 13.8-fold increase in leucine usage), suggesting that this bias involved only a specific subset of the mitochondrial proteins localized in the "Mitochondrial inner membrane" and the "Mitochondrial matrix".

The physiological significance of the amino acid utilization biases detected remains to be elucidated, and some of them might be related to the activity of N-terminal modifying enzymes such as MetAPs or NATs. One of the strongest biases revealed by our approach is the dramatic overrepresentation of leucine at position 2 in GO terms related to the mitochondrial compartment. In this work, we sought to clarify, in *S. cerevisiae*, the functional significance of usage bias at position 2 of mitochondrial precursors and its potential relationship with NatC.

The N-terminal Mitochondrial Targeting Sequence has a specific conserved signature at position 2 typical of NatC potential substrates.

We first investigated whether the high overrepresentation of leucine at position 2 of the mitochondrial precursors could correlate with other specific features in the amino acid composition of their N-terminal region. Especially, precursors of matrix, inner membrane and intermembrane space proteins are dependent for their mitochondrial import on an N-terminal sequence forming an amphiphilic alpha

helix, 25-35 residues-long, enriched in hydrophobic and positively charged residues (26, 27). Such a mitochondrial targeting sequence, called MTS, is observed in 361 of the 726 yeast proteins annotated as mitochondrial with high confidence (Table S2, see Methods). Consistently, the 361 mitochondrial precursors harboring a MTS showed a dramatic overrepresentation of arginine and an underrepresentation of negatively charged residues in the positions 3 to 20 of their N-terminal region. This profile contrasted strongly with that of mitochondrial precursors lacking MTS (Fig2a). Strikingly, we found that the bias at position 2 described above for the proteins associated with mitochondrial GO terms was specific of mitochondrial proteins with a MTS (Fig2a). More precisely, four hydrophobic residues (Leu, Phe, Trp and Ile) are significantly overrepresented (HGT value ranging from 3.9 for isoleucine to 71 for leucine) at position 2 of proteins with MTS (Fig.2a). Furthermore, it should be noted that these biases in favor of Leu, Phe, Trp and Ile are strictly restricted to position 2 of the MTS.

Quantitative analysis of the distribution of amino acids at position 2 of the MTSs (Fig. 2b) shows the importance of this previously unknown N-terminal signature of the mitochondrial addressing sequence: nearly 60% of them have an N-end signature with a Leu, Phe, Trp and Ile, which contrasts strongly with the low representation of the latter at position 2 in the proteome (see L, F, I, W residues frequencies in Fig. 1a). For instance, leucine accounts for more than 35% of the amino acids observed at position 2 of the MTSs, whereas it is poorly used in the *S. cerevisiae* proteome at this position (Fig 1a). Interestingly, we also found an overall significant under-representation at position 2 of the MTSs of amino acids with small radii of gyration (Gly, Ala, Ser, Cys, Thr, Pro and Val), known to induce the cleavage of iMet by methionine aminopeptidase (6). Finally, the analysis of amino acid biases in the MTSs of 17 budding yeasts of the Saccharomycotina lineage (23, 24), demonstrated that this specific signature at position 2 is a conserved characteristic of mitochondrial addressing sequences (Fig.2c). Indeed, in all the studied species, we observed: (1) a strong over-representation of the same hydrophobic amino acids at position 2 and in particular of leucine whose HGT score exceeds the value of 20 in all species, and (2) a bias restricted to position 2 since the HGT scores of leucine, phenylalanine, isoleucine and tryptophan were systematically lower than 3.9 for all the other positions of the MTSs (Fig. Sup1).

Interestingly, the conserved amino-acid bias described above at position 2 of MTSs perfectly correlated with the known specificities of N-acetyl transferases in yeast (Fig. 2b). Indeed, leucine, phenylalanine, isoleucine and tryptophan, which are overrepresented at this position, are known to define potential substrates of NatC in yeast when positioned after iMet (9). Conversely, serine, alanine and threonine, which are the most under-represented at this position, are known to induce iMet cleavage and subsequent targeting by NatA (8).

Together, these observations strongly suggest that position 2 of N-terminal mitochondrial targeting sequences is under selective pressure and that the nature of the residue at position 2 is crucial for the fate of mitochondrial precursors imported through the MTS-dependent pathway. The correspondence between the amino acid bias at position 2 of MTSs and the preferences of NatC suggests a link between mitochondrial precursors fate and NatC targeting. They raised two important questions addressed in the following sections. First, are mitochondrial precursors indeed targeted by the N-terminal acetyl transferase NatC? And second, what are the functional consequences of changing the residue located at position 2 of a typical MTS?

Selective Translating Ribosome Affinity Purification reveals co-translational targeting by NatC of mitochondrial precursors harboring the identified N-terminal signature

Our current knowledge of the substrate specificities of NatC in yeast identifies proteins with ML-, MF-, MI- and MW- N-termini as its potential substrates. It was deduced from the observation of the loss of N-terminal acetylation of various iso-1 cytochrome c reporter proteins mutated at position 2 in yeast strains lacking NatC, which was further confirmed for a small subset of cellular proteins (9, 28–30). Partial N-terminal acetylome of *S. cerevisiae* were more recently obtained (8, 10, 11) that covered up to 10% of the proteome, corresponding to the most abundant cellular proteins. Unfortunately, these data included very few NatC potential substrates (between 0 and 20 for each NatC typical N-termini) and, to date, there are no published data describing globally the *in vivo* substrates of NatC by studying the impact of NatC depletion on the yeast N-terminal acetylome.

Mitochondrial precursors steady state levels are extremely low, since, after their synthesis, they are quickly imported in mitochondria where their N-terminal addressing sequence is cleaved (26). Hence, detecting them with protein-based approaches is challenging. To overcome this limitation, we decided to use the Selective Affinity Purification of Translating Ribosomes (sel-TRAP) method, to capture by immunoprecipitation the various NatC-targeted nascent chains, as well as their associated mRNAs, which we characterized using transcriptomic analyses (Fig. 3a, see also microarray data in Table S3). This approach had already proven sufficient sensitivity to study the selectivity of various nascent chain associated factors (31–34). To validate our strategy, we also conducted sel-TRAP experiments on NatA and NatB, which preferences were previously confirmed by the impact of their deletion on N-terminal acetylome. Hence, we immunoprecipitated protein A-tagged versions of the catalytic subunits of each of the three NATs, Ard1p (NatA), Nat3p (NatB), and Mak3p (NatC) (Fig. 3b, top and middle panel). The detection of rRNAs in the immunoprecipitated fractions (Fig. 3b, bottom panel, IP lane) and the analysis of the protein fractions of the three immunoprecipitations compared with the mock experiment confirmed that we had purified translating ribosomes that were mostly distinct by the specific presence of the catalytic and accessory subunits of the immunoprecipitated N-terminal acetyltransferase complexes (Fig. 3c, see also proteomic data in Table S4). Transcriptomic analyses of the immunoprecipitated samples compared to the input RNAs identified non-overlapping sets of enriched mRNAs likely encoding the respective substrates of each NAT. As a control, we also analyzed mRNA co-immunoprecipitated with the canonical Rpl16a ribosomal protein, which confirmed the specificity of the mRNAs identified in the NATs sel-TRAP experiments (Fig. 3c).

To verify that the NATs sel-TRAPs have specifically purified their respective substrates, we performed a threshold-independent analysis of the proportion of amino acids encoded at ORFs position 2 in the transcriptomic data. We detected a specific and significant increase in serine in the NatA immunoprecipitation experiment, aspartate in that of NatB, and leucine and phenylalanine in that of NatC (Fig. Sup3). These results are globally consistent with the known substrate specificity of the three NATs in *S. cerevisiae* (see Introduction). However, all the expected N-termini were not detected as enriched, which could reflect differences in the efficiency of recognition of their potential substrates by the NATs. Consistent with this idea, N-terminal acetylome data has previously revealed that, in many cases, only a fraction of NATs expected substrates is actually acetylated (8, 10, 11). For example, among the proteins potentially targeted by NatA, almost all of those with a Ser- N-termini were acetylated, whereas only about 50% of those with Ala- or Thr- N-termini were, and almost none of the other potential NatA targets (Pro-, Gly-, and Val- N-termini). Similarly, most of the few NatC potential substrates detected in N-terminal acetylome data were only partially acetylated. It should also

be kept in mind that sel-TRAP technology captures co-translational targets that co-purify with the immunoprecipitated NATs and that the transient nature of enzyme-substrate connections could limit the detection of the weakest interactions.

As N-terminal acetylome data in absence of NatC are not available, sel-TRAP allowed for the first time a global study of NatC targets *in vivo* in *S. cerevisiae*. After immunoprecipitation of Mak3p, we observed in our threshold-independent analysis a significant increase specific to leucine and phenylalanine at position 2 of the proteins encoded by the most enriched mRNAs (green and red curves in Fig.4a1, see also Fig. Sup3). The significance of these increases was confirmed by the observation that the HGT scores for the ML- and MF- N-termini reached a maximum value near 15 and 25, respectively (green and red curves in Fig.4a2). This is consistent with the known substrate specificity of NatC and suggests more efficient co-translational acetylation of the initiator methionine when followed by these residues rather than by isoleucine or tryptophan. Furthermore, in agreement with our prediction, nearly 50% of the enriched mRNAs (enrichment threshold value equal to 0.8) encoded mitochondrial precursors (Fig 4b) and only those imported through the MTS-dependent pathway were observed enriched in the sel-TRAP data (Fig 4c). Among these precursors, the most significant enrichments were observed for those with a leucine (HGT=24), phenylalanine (HGT=23), or isoleucine (HGT=5) at position 2 of their mitochondrial addressing sequence (Fig 4d).

Intriguingly, we also observed in the sel-TRAP data a slight enrichment of MTSs without a typical signature of NatC targets at position 2 (amino acid other than F, L, I, W in Fig. 4d). With the exception of glutamine at position 2 (HGT=4), these enrichments were not significant, but this could be due to the low number of precursors with a given amino acid at position 2 of their MTSs sequence (between 3 and 42). Indeed, a threshold-independent analysis showed that the enrichment of mitochondrial precursors with MTS was mainly related to their N-termini that define them as potential NatC substrates (F, L, I, W amino acid at position 2, green curve in Fig. 4e1) but that MTSs lacking this signature at position 2 contributed significantly to MTSs enrichment in the sel-TRAP data (blue curve in Fig. 4e1). This enrichment for proteins that are not putative NatC substrates was not observed for mitochondrial precursors lacking MTS (blue curve in Fig.4e2), as well as in the overall analysis of NatC sel-TRAP results (Fig. Sup3).

Finally, sel-TRAP analysis clearly demonstrated that mitochondrial proteins with an N-terminal MTS sequence are co-translationally recognized by NatC mainly when they have a leucine, a phenylalanine or to a lesser extent an isoleucine at position 2. Furthermore, unexpectedly, the observation that MTSs with other amino acids at position 2 were selectively enriched after NatC immunoprecipitation strongly suggests that other features of MTSs might be involved in their recognition by NatC.

Substitution of leucine 2 by amino acids not overrepresented at position 2 of MTSs result in precursor accumulation and rescue of the dominant negative effect of an HSP60-13myc transgene

Finally, we explored the functional significance of selecting during evolution specific amino acids at the N-terminus of the mitochondrial addressing sequences. To this aim, we set up a mutagenesis experiment at position 2 of a typical MTS. We chose Hsp60p, an essential chaperonin of the mitochondrial matrix (35) as a model of the MTS-dependent import pathway because of (1) the presence of a highly conserved leucine at position 2 of its MTS (leucine 2 is observed in 13 of 17 Saccharomycotina yeasts studied) and (2) the availability of an antibody allowing sufficiently sensitive detection of its precursor when its mitochondrial import is impaired. In addition, to improve detection of

Hsp60p precursors, we used the thermosensitive strain YPH499 pam16Δ-MAGN76D (36), which, even at the permissive temperature, exhibits a slight defect in mitochondrial import as evidenced by the detection of a low-intensity band corresponding to the precursor form of Hsp60p (Fig. 5a, Hsp60p WT lane). We used CRISPR/CAS9 technology to mutagenize this strain and create L2X genomic mutants of Hsp60p in which the leucine at position 2 was replaced by another amino acid (Fig. 5a, lanes corresponding to L2X mutants). Replacing leucine 2 in Hsp60p with aspartate, lysine, serine, valine or proline (Fig.5a L2D, L2K, L2S, L2V, L2P lanes) resulted in a dramatic increase in the intensity of the band corresponding to the Hsp60p precursor compared to that corresponding to its mature form, strongly suggesting an altered mitochondrial import of Hsp60p. In contrast, substitution of leucine with a phenylalanine, another hydrophobic amino acid over-represented at position 2 of the MTS-containing protein group, had no impact on the Hsp60p profile observed by western blot (Fig. 5a, L2P lane).

In a second step, we developed a screen to easily test the effect of all possible amino acid substitutions at position 2 of Hsp60p. For this purpose, we took advantage of an HSP60-13myc allele that we found to have a dominant negative effect. Indeed, the transformation of *S. cerevisiae* with the pHSPmyc(L) plasmid, which allows the expression of the HSP60-13myc allele, led to no colony (plate L in Fig.5b1). To interpret these results, we hypothesized that the 13-myc C-terminal extension abolished the activity of Hsp60p C-terminal domain, previously shown to play a crucial role for the function of the protein (37, 38). Thus, since Hsp60p is essential for cell survival, we attributed the near absence of clones after transformation with pHSPmyc(L) to a dominant negative effect of the myc-tagged version of Hsp60 on the endogenous wild-type form. After being imported in mitochondria, Hsp60p forms a tetradecameric chaperonin complex in the mitochondrial matrix required for its own assembly (37). Thus, the association of the deficient protein Hsp60p-13myc with endogenous wild-type Hsp60p would lead to the inactivation of the Hsp60p complexes and the observed drastic loss of viability (Fig. 5b2). We further observed that an altered mitochondrial import of Hsp60p, as induced by the cis mutations L2D, L2K, L2S, L2V, L2P (Fig. 5a), resolves the dominant negative effect of the 13-myc tag. More than 1,000 transformants were obtained with the pHSPmyc(X) plasmids corresponding to these mutations, where X represents the amino acid that replaces leucine at position 2 of Hsp60 in the Hsp60p-13myc protein (Fig. 5b1).

Strikingly, after testing all possible amino acids at position 2 of Hsp60p-13myc, we found that the two groups of amino acids defined on the basis of their ability to rescue C-terminal tag toxicity correlated strongly with their enrichment at position 2 of *S. cerevisiae* mitochondrial addressing sequences, as assessed by our statistical analyses (Fig. 2 and Fig.5b). Indeed, the amino acids that did not result in more than 1-3 clones after transformation were those we showed to be enriched at position 2 of MTSs, i.e., leucine (HGT=71), phenylalanine (HGT=15), isoleucine (HGT=4), tryptophan (HGT=6), and methionine (HGT=1.6). Western blot analysis revealed that the few clones obtained for these 4 amino acid substitutions did not express Hsp60p-13myc protein and likely escaped toxicity through plasmid mutations, which was confirmed by sequencing some of these plasmids (Fig. Sup4b). For all other amino acids (except tyrosine which showed a lower number of transformants), we observed more than 1,000 colonies after transformation by pHSPmyc(X) plasmids. Western blot analysis of the resulting clones confirmed that they all express Hsp60p-13myc protein at a level roughly comparable to that of the endogenous protein (Fig. Sup. 4a). The only exceptions are the clones transformed with pHSPmyc(Y) in which the level of Hsp60-13myc protein was systematically reduced (Fig. Sup4b). In addition, we consistently observed, in all clones with a rescue phenotype, an Hsp60p-myc band corresponding predominantly to the non-imported precursor form (see for example the profile of the

L2Q mutant for which both bands are visualized with the anti-myc antibody) (Fig. Sup4a). More surprisingly, we also observed a defect in import and/or processing of the endogenous, wild type, Hsp60p, reflected by the clearly visible band corresponding to the precursor form of the protein (Fig. Sup4a). This observation may indicate that a small fraction of the Hsp60p-13myc proteins may have been imported, resulting in a significant, although not lethal, loss of Hsp60p function that could be associated with impaired processing of mitochondria-targeted endogenous Hsp60p precursors, as previously observed upon inactivation of mitochondrial chaperonin (35, 39).

The results obtained in our mutagenesis experiments allowed us to demonstrate the dramatic functional consequences of modifying the residue located at position 2 of the mitochondrial addressing sequence. These results, which fully validated our *in-silico* predictions, also supported the existence of strong evolutionary constraints that shaped the current amino acid biases at position 2 of MTSs, thus ensuring an optimized mitochondrial import.

DISCUSSION

In this work, we developed an *in-silico* approach to identify functional groups of proteins with N-terminal amino acid usage biases. The resulting functional mapping of N-terminal residues reveals possible evolutionary constraints at position 2 of nascent chains and highlights potential early mechanisms of protein regulation or targeting. In particular, we discovered the potential critical role of the residue located at position 2 of mitochondrial N-terminal targeting sequences in *S. cerevisiae*. This prediction was experimentally validated using an original system based on a dominant-negative allele of HSP60, that we set up to test the efficiency of mitochondrial import after mutagenesis at position 2 of the MTS. Strikingly, we observed that only the 5 hydrophobic amino acids (Leu, Phe, Ile, Trp and Met) over-represented at this position allowed efficient mitochondrial import of Hsp60p protein. This finding raises the question of the nature of selective pressures acting specifically on position 2 of the MTSs in the context of their other well-established characteristics. It also strongly suggests the involvement in optimizing mitochondrial precursor import of N-terminal modification enzymes, including NatC, which we found to co-translationally interact with MTS-bearing precursors.

The N-terminal signature of MTSs in light of the current knowledge about mitochondrial import mechanisms

MTSs are typically 15-50 amino acid residues in length and form positively charged amphipathic α -helices. It has been suggested that these sequences possess several biochemical features involved in their interaction with binding sites of the mitochondrial translocation machinery (26). First, the amphipathic alpha helix interacts with two outer membrane translocase (TOM) receptors: the hydrophobic surface is recognized by Tom20 and the positively charged one by Tom22 (40–42). In addition, translocation of the precursor into mitochondria can occur via an acidic binding chain involving the interactions between MTSs basic residues and some residues from Tom7p, Tom22p, and Tim23p (43). Finally, the mitochondrial membrane potential ($\Delta\psi$) generates an electrophoretic force on the MTSs and the net positive charge of the MTSs is the critical determinant of this effect (44, 45). Taken together, these data indicate that the hydrophobicity of MTSs is mainly determinant for their primary interaction with the outer mitochondrial membrane receptor Tom20p.

In addition, TargetP2.0, a new N-terminal sorting signal detection tool based on a machine learning approach, has recently further improved the prediction of mitochondrial presequences (46). By

analyzing the attention layer of their final network, these authors found that the residue at position 2 of the presequences had a strong influence on their classification. More specifically, they observed that, in fungi, peptides targeted to the endoplasmic reticulum (ER) and to mitochondria showed at position 2 an underrepresentation of residues that allow the removal of the initiator methionine, i.e. Ala, Cys, Gly, Ser, Thr, or Val (46). This observation is consistent with the results we obtained in *S. cerevisiae* (Fig2b). Indeed, we found that only 20% of MTSs have amino acids at position 2 allowing the removal of their iMet compared to 56% within the proteome and 69% among cytosolic proteins. Of note, Forte et al. also previously observed that less than 25% of ER signal sequences would be predicted to be MetAPs substrates and further demonstrated that removal of iMet from carboxypeptidase Y precursor inhibits its ER translocation (47). Similarly, we can speculate based on our results and on the observation made by Almagro Armenteros et al. that removal of iMet may be deleterious to mitochondrial import.

In favor of this hypothesis, we observed that the optimal residues at position 2 of MTSs, i.e. Leu, Phe, Ile, Trp and Met, have high hydrophobicity indices, but also that none of them allow iMet removal. Conversely, the small residues Ala and Cys, which are also highly hydrophobic but induce iMet cleavage when positioned at position 2 of the nascent chains, are underrepresented at this position of the yeast MTSs. Therefore, the N-terminus of most MTSs will be characterized by the sequence M(L/F/I/W/M) with two consecutive hydrophobic residues in positions 1 and 2 and this hydrophobic N terminus may be important in the early binding of mitochondrial precursors to the mitochondrial surface. Indeed, several analyses indicate that consecutive hydrophobic residues could be involved in the binding of MTSs to Tom20p. In particular, by measuring the binding of a peptide library to Tom20, a consensus sequence $\theta\phi\chi\beta\phi\phi$ was identified (where θ , β , ϕ , and χ represent a hydrophilic, basic, hydrophobic, and any residue, respectively) and proposed as a Tom20 binding motif (48).

Bioinformatic analysis extended this consensus by identifying statistically significantly enriched hexamers in the 30 first residues of MTSs, and their detection was implemented in MitoFates, one of the most efficient tools for MTSs prediction to date (49). Most of these hexamers showed strong potential to form amphiphilic helices and includes several consecutive hydrophobic residues.

By integrating these data from the literature and our results on MTS position 2, we propose a model explaining the critical role of hydrophobic residues at the N-terminus of mitochondrial addressing sequences (Fig. 6). We can hypothesize that the selection of large hydrophobic amino acids at position 2 of MTSs was a successful evolutionary solution, the consequences of which were high N-terminal hydrophobicity due to (1) conservation of the hydrophobic iMet, and (2) possible N-terminal neutralization of the N-terminal amine by acetylation by NatC. Positioning the hydrophobic site directly at the N-terminus of MTS would then allow for early interaction with Tom20p and the mitochondrial membrane. Consequently, the arginines that provide the positive charge necessary for translocation of the precursors to mitochondria could be highly enriched from residue 3 of MTS, without compromising the interaction with Tom20p.

A conserved role of the N-terminal signature in the mitochondrial import of MTS-bearing proteins

In support to our observation of significant biases at position 2 of the MTS in yeast, our mutagenesis experiments using Hsp60p as a model revealed the importance for proper mitochondrial import of having at this position large hydrophobic amino acids that do not lead to cleavage of the initiator

methionine. Only leucine, phenylalanine, isoleucine, tryptophan and methionine at position 2 did not affect the mitochondrial import of our dominant negative Hsp60p-13myc allele.

Strikingly, mutagenesis of position 2 of the rat aldehyde dehydrogenase (Aldh) MTS, whose first residues are MLRAAL, yielded similar results to those we obtained by mutating Hsp60p in yeast (50). Replacement of leucine at position 2 with a valine, which is also a hydrophobic residue but which drives the removal of iMet, did not allow efficient *in vitro* import of Aldh, whereas isoleucine, tryptophan, or phenylalanine at position 2 had no impact on mitochondrial import. Interestingly, the correct import of the L2V mutant was restored when it was translated using a system that did not remove the hydrophobic iMet. Similarly, a mutation that increased the N-terminal hydrophobic surface area compensated for the L2V mutation. These data underscore the general importance during eukaryotic evolution of the N-terminal hydrophobicity of MTSs and the necessary balance with positive charges that are also critical for precursor translocation into mitochondria. In this context, it would be crucial to further study the evolution of MTSs in order to better understand the characteristics that allow their optimized targeting to mitochondria.

The characteristics of MTSs allow their efficient recognition by the N-terminal acetyltransferase NatC

In this work, we discovered the specific and optimal N-terminal signature of mitochondrial targeting sequences, defined by the presence of a hydrophobic amino acid directly after the uncleaved initiator methionine. This led us to hypothesize that, in *S. cerevisiae*, mitochondrial precursors with a MTS could be cotranslationally targeted by NatC.

Interestingly, previous N-terminal acetylome analyses in *S. cerevisiae* detected three mitochondrial precursors: two of them had an ML-terminus, Pda1p and Glr1p, and were fully acetylated (51). However, no global experimental identification of NatC targets in yeast had been achieved before our work. Therefore, we used the Selective Affinity Purification of Translating Ribosomes (sel-TRAP) technology to identify nascent chains co-immunoprecipitated with NatC. In this experiment, 30% of the captured nascent chains identified by their corresponding mRNA were mitochondrial precursors with a MTS, most of them having a leucine or phenylalanine at position 2. However, the nature of the residue at position 2 may not be the sole determinant of the affinity of NatC for these precursors.

The structure of NatC bound to a *S. cerevisiae* ligand highlighted the importance of residue at position 3 of the substrates that is fully involved in the binding in the catalytic pocket. More specifically, the analysis of the ligand bound structure revealed that when an arginine is present at position 3 of the peptide, it binds to a specific site in the catalytic pocket and strongly increases the catalytic efficiency of NatC towards the peptide measured in a catalytic assay (52). Hence, many mitochondrial precursors have features for optimal NatC binding (retention of iMet, hydrophobic residue at position 2 and arginine at position 3), as depicted in figure 6. As a matter of fact, one-third of MTSs that have a leucine at position 2 also have an arginine at position 3.

In summary, it appears that many MTSs possess N-terminal features that are optimal for efficient binding and acetylation by NatC. Using sel-TRAP technology, we have indeed demonstrated that MTSs are recognized co-translationally by NatC, suggesting their early N-terminal acetylation. More surprisingly, we also observed that some MTSs without features compatible with binding to the active site of NatC, i.e., those for which the i-Met is cleaved, were enriched in our sel-TRAP data. It is then

tempting to speculate that other features of these MTSs, besides their N-terminal composition, could allow their recognition by NatC.

A role for NatC in mitochondrial biogenesis?

From yeast to human, phenotypic analyses of cells lacking NatC have highlighted its critical importance for mitochondrial function. In *S. cerevisiae*, deletion of any of the NatC subunits has been shown to induce growth defects on non-fermentable carbon sources, such as glycerol and ethanol (52–56).

The importance of NatC for mitochondrial function was also underscored by the observation that depletion of NatC catalytic subunit, Naa30, in a human cell line resulted in reduced levels of mature mitochondrial matrix membrane proteins, many of whose precursors were substrates of NatC (57). This phenotype was associated with loss of mitochondrial membrane potential and fragmentation of mitochondria.

Furthermore, in glioblastoma-initiating cells (GICs), that drive tumorigenesis of the most common primary brain malignancy, a strong overexpression of the protein Naa30 was detected. shRNA-mediated knockdown of *NAA30* gene in GICs demonstrated that NatC may serve as a therapeutic target in these types of cancer, as intracranial transplantation of such NatC depleted cells resulted in prolonged animal survival compared to control mice transplanted with unmodified GICs (58). At the cellular level *NAA30* knockdown in GICs cells results in reduced viability and decreased mitochondrial hypoxia tolerance, with a more abrupt and severe mitochondrial membrane depolarization compared to control cells in such condition. These observations suggested that NatC contribute to the stability of the mitochondrial potential, which was supported by the observation that several genes involved in hypoxia response, including HIF1 α , a central regulator of transcriptional response to hypoxia, were upregulated in absence of NatC even in normal culture conditions.

We have shown that, in the yeast *S. cerevisiae*, a large fraction of proteins harboring MTSs is susceptible to acetylation by NatC during their translation. At the molecular level, N-terminal acetylation by NatC was shown to modulate proteins ability to interact with other proteins or membranes allowing protein complexes formation or targeting to specific subcellular subsites. Examples of such NatC-dependent interactions include the formation of the Gag capsid of the yeast L-A helper virus (28, 59), or of the Ubc12-Dcn1 complex (60, 61), the insertion of the tRNA-specific methyltransferase Trm1-II into the inner nuclear membrane (62), the targeting of the human ADP-ribosylation factor-like protein 8b to lysosomes (63), or the localization of Arf-related protein 1 (hARFRP1) and its yeast orthologue yArl3 to the trans-Golgi via an interaction of their acetylated N-terminus with the hSys1/Sys1p protein (64, 65).

Similarly, it is tempting to speculate that N-terminal acetylation of MTSs further increases their N-terminal hydrophobicity, which may enhance their binding to the mitochondrial receptor Tom20p and their subsequent mitochondrial addressing. A previous study of the GFP localization pattern of potential NatC substrates, including three mitochondrial proteins, failed to identify altered localization in yeast lacking the catalytic subunit of NatC (51). However, the fluorescent imaging approach used in this study may not be sensitive enough to detect subtle mitochondrial import defects that could be induced by the loss of N-terminal acetylation of MTSs in the absence of NatC. Moreover, N-terminal acetylation could have various effects on mitochondrial addressing efficiency, depending on other

characteristics of the MTS, including its overall N-terminal hydrophobicity. Thus, the mitochondrial phenotype of cells lacking NatC activity could be due either to the cumulative effect of mild defects in the import of many mitochondrial proteins or to strong effects on a limited number of mitochondrial precursors for which N-terminal acetylation by NatC would be critical. A systematic exploration of the impact of NatC depletion on the various MTS-bearing precursors is still needed to link the strong mitochondrial phenotypes of NatC-depleted cells to the specific role of N-terminal MTS acetylation on mitochondrial precursors fate.

MATERIAL AND METHODS

***S. cerevisiae* strains, plasmids, and growth conditions**

All *Saccharomyces cerevisiae* strains used in this study are described in Supplementary Methods (Table S5). The plasmids used for yeast transformation are also described in the Supplementary Methods (Table S6) as well as the oligonucleotides used for the diverse genetic constructs (Table S7 and S8).

The strains used for the sel-TRAP experiment were derived from BY4741 by homologous recombination with the ProtA-His5 cassette amplified from the pBXA plasmid (66). The temperature-sensitive and respiratory-deficient strain YPH499 pam16Δ-MAGN76D was kindly provided by Agnes Delahodde with the corresponding reference strain YPH499 (36). The substitutions of the leucine at position 2 of Hsp60p were obtained by site-directed mutagenesis using CRISPR/CAS9 technology. YPH499 strains were co-transformed with the plasmid pAEF-HSP60 (67) encoding the Cas9 endonuclease and expressing a guide RNA targeting the HSP60 sequence, and a HSP60 repair template including the desired mutation. The resulting strains were used for PCR amplification of the mutant HSP60 sequences to obtain the diverse pHSPmyc(X) plasmids (X representing the amino acid at position 2 of Hsp60p sequence) by cloning into pZMYA7 (32), a pRS416-derived plasmid allowing C-terminal tagging with the 13Myc epitope.

Yeast were grown on rich medium (1% bactopectone, 1% yeast extract). When required for plasmid selection, cells were grown on CSM medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% dextrose, 0.07% CSM mixture) depleted of the selective compound. *Hygromycin* B (200 µg/ml) was used in rich medium to select mutants containing pAEF5-derived plasmid after CRISPR/CAS9 procedure. After genome sequencing verification, plasmid was removed from HSP60 mutant strains by overnight culture in rich medium without antibiotic selection.

Antibodies

The following antibodies were used for western blotting or immunoprecipitation: rabbit IgG-HRP polyclonal antibody (PAP; code Z0113; Dako), rabbit Hsp60 polyclonal antibody (gift from J.P. Di Rago), anti-PGK1 mouse monoclonal antibody (ab113687, abcam) and mouse monoclonal anti-myc antibody (11667149001, Roche).

Data for the *in-silico* analyses

The 6,050 sequences of *S. cerevisiae* proteome (proteome ID: UP000002311) were downloaded from the UniProtKB database (68)

(<http://www.uniprot.org/uniprot/?query=proteome:up000002311&format=fasta>). Since proteins encoded by transposons have redundant sequences and their distribution vary depending on yeast

strains (69), the 91 proteins annotated as transposons according to the Saccharomyces Genome Database (<https://www.yeastgenome.org/>) were removed from the proteome.

Gene Ontology annotations were downloaded from the GO consortium database (70, 71) (<http://current.geneontology.org/annotations/sgd.gaf.gz>). For extensive analysis of mitochondrial proteins, the 1,190 proteins belonging to the mitochondrion Gene Ontology term (GO:0005739) were manually curated: proteins that were associated to mitochondria based on only one high-throughput experiment were removed, leaving 726 proteins for which association with mitochondria was strongly evidenced. Curated mitochondrial proteins were then sorted depending on whether they carry an N-terminal mitochondrial targeting sequence, according to sequence annotations provided by the UniProtKB database (68). 361 over 726 mitochondrial proteins were annotated as carrying a MTS (Table S2).

For the comparative analysis, 17 species were selected from the Saccharomycotina yeast lineage (23) (see Table S9 for the list of selected species). Orthology groups recently established for 322 budding yeast species (24) were used to retrieve the 311 groups containing a mitochondrial protein from *S. cerevisiae*, among which 161 had an N-terminal MTS. More than 95% of these groups (298) contained orthologs for at least 13 of the 17 selected species, and all recovered orthologs for at least 9 of these species.

HGT scores derived from hypergeometric test

Hypergeometric tests (HGT) have been classically used to identify significant differences between the observed frequencies (F_{obs}) in various subsets of interest and the reference frequency (F_{ref}) of the corresponding input set. The p-values provided by these tests were then used to calculate hypergeometric scores (HGT scores) as follows: $|HGT\ score| = \log_{10}(P\text{-value})$, with a negative sign assigned to the score in case of underrepresentation ($F_{obs} < F_{ref}$).

Analysis of amino acids usage bias at position 2 in the proteome

For each of the 20 amino acids, the frequency of use at position 2 (F_2) was calculated and compared with the respective average use of the amino acid in the proteome (F_{mean}). Differences were statistically evaluated using hypergeometric tests and HGT scores were calculated. The observed frequency value at position 2 was also used to define a reference value called F_{2ref} calculated as follows: if $F_2 > F_{mean}$, $F_{2ref} = 0.9 * F_2$, otherwise $F_{2ref} = 1.1 * F_2$. The F_{2ref} value was then used to test the observed frequencies at any other position and calculate an aspecificity score, representing the percentage of positions with a frequency greater than F_{2ref} when $F_2 > F_{ref}$, or less than or equal to F_{2ref} when $F_2 \leq F_{ref}$. To calculate this score, the first 500 positions of the proteome were analyzed by a 100-position long sliding window. At each step of the analysis, the window slides by one position and the number of positions within the window with a frequency value higher (when $F_2 > F_{ref}$) or lower (when $F_2 \leq F_{ref}$) than the reference value F_{2ref} was evaluated, to calculate a local aspecificity score. The final retained aspecificity score corresponds to the maximum value of the measured local aspecificity scores.

Analysis of the amino acid usage bias at position 2 in subsets of proteins of the Gene Ontology

We have implemented a set of functions in R to detect gene ontology terms showing, among their associated proteins, a significant and specific amino acid usage bias at position 2. The gene

annotations were downloaded from the GO consortium and stored in MGSA objects (R package MGSA (72)). The second residue associated with each protein was extracted from the downloaded proteome (UniProtKB database). For each GO domain (cellular component, molecular function, biological process), contingency tables were generated that count, for all GO terms, the number of proteins with a given amino acid at position 2. To avoid detecting amino acid bias due to proteins with the same N-terminal sequence, proteins sharing the same first 10 residues were grouped under the same identifier, which was counted only once. These contingency tables were used to calculate HGT scores and detect GO terms with significant over-representation relative to the proteome (HGT score > 4) of one or more amino acids in position 2 of their associated proteins. GO terms that showed the lower over-representation (fold change < 1.8) and mainly correspond to GO term regrouping very large number of proteins were first removed from this initial list. Then, to reduce the redundancy of the detected GO terms, we developed a two-step algorithm to (1) select the minimum number of low-redundant GO terms (overlap of GO terms with larger ones < 40%) maximizing the number of proteins associated with the detected amino acid usage biases (BestN GO), and (2) identify smaller low-redundant GO terms (overlap with larger GO term < 30%) that are partially or fully included in one BestN GO (overlap ≥ 40%) terms but have at least 30% higher frequency bias (BestF GO). Finally, to verify that the amino acid usage bias detected in a given GO term was not due to a specific N-terminal composition of the associated proteins, the search for position-specific amino acid bias was extended by calculating the HGT enrichment scores relative to the proteome on the first 100 positions of these proteins. Details of the GO reduction algorithm are available in the supplementary methods.

Selective Translating Ribosome Affinity purification (sel-TRAP)

For analysis of canonical ribosomes and NatA, NatB and NatC-associated ribosomes, ribonucleoparticle purifications from cells expressing proteinA-tagged baits (respectively Rpl16a, Ard1, Nat3 and Mak3) were performed essentially as described in (32). Briefly, frozen grindates obtained from cells grown in rich medium were homogenized in nine volumes of extraction buffer (20 mM Hepes pH 7.5, 110 mM KOAc, 2 mM MgCl₂, 0.1% Tween-20, 0.5% Triton X-100, 1 mM DTT, 1× protease inhibitors cocktail, complete EDTA-free, Roche and antifoam B, Sigma, 1:5000). The obtained extract was clarified by filtration through 1.6 µm GD/X Glass Microfiber syringe filters (25 mm, Whatman) and input samples were collected for analysis of total cellular RNA and protein content. The immunoprecipitation of proteinA-tagged baits was performed by further incubation of the extract for 30 min at 4°C with IgG-conjugated magnetic beads. Beads were then washed three times with extraction buffer, once with washing buffer (0.1 M NH₄OAc, 0.1 mM MgCl₂) supplemented with 0.02% Tween-20 and four times with washing buffer without Tween-20. Beads were then split into two samples for protein and RNA analysis respectively. Immunoprecipitated proteins were eluted with 0.5 M NH₄OH, 0.5 mM EDTA, lyophilized and resuspended either in SDS-sample buffer for SDS-PAGE or in 25 mM ammonium carbonate for mass spectrometry analysis. Immunoprecipitated RNA was eluted by using the RLT buffer from Quiagen RNeasy Mini Kit[®] that was used to purified total and immunoprecipitated RNAs.

Mass spectrometry analysis of sel-TRAP experiment

Mass spectrometry analysis procedures, performed at Proteomics Core Facility of Institut Jacques Monod, are described in Supplementary methods, and mass spectrometry proteomic data were deposited on the ProteomeXchange Consortium (73) via the PRIDE partner repository with dataset identifier PXD034922.

Mascot score distributions of immunoprecipitations from control (unlabeled BY4741) or protein A-labeled strains were compared to distinguish specifically interacting proteins from background. Based on the distribution of Mascot scores in the control, a cutoff score of 250 was chosen to retain less than 5% false positives. Only proteins with a Mascot score greater than 250 in the immunoprecipitation data of protein A-labeled strains were considered for further analysis. Proteins with a score greater than 250 in the control or with a ratio of protein A-labeled strain to control less than 2 were excluded in the final list of partners detected in sel-TRAP experiments.

Venn diagrams obtained from two independent experimental sets were used to compare the protein partners of Ard1, Nat3 and Mak3 and determine the shared and specific partners observed in a reproducible manner. In this representation, for a more accurate definition of partners, all proteins with at least a 5-fold higher Mascot score in one of the sel-TRAP experiments compared with the others was assigned as specific partner of the corresponding immunoprecipitated protein.

Microrray analysis of sel-TRAP experiment

The microarray data and the related protocols are available at ArrayExpress website (<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11772/>) with the dataset identifiers E-MTAB-11772.

After sel-TRAP experiment, input and immunoprecipitated (IP) RNAs were purified using the RNeasy Mini Kit (Qiagen). One μg of RNA from each sample was then reverse-transcribed and labeled with Cy3 or Cy5 dye using the indirect labeling procedure. cDNAs of input and IP samples were then hybridized on Agilent's ab 8 \times 60K *S. cerevisiae* custom DNA chip (AMADID: 027945). Arrays were read using an Agilent scanner at 2 μm resolution and the signal segmentation was done using the feature extraction software (Agilent). The data was normalized without background subtraction using the global Lowess method (74).

To access the specifically enriched mRNAs in the Ard1, Nat3, or Mak3 sel-TRAP experiments, the measured $\log_2(\text{IP}/\text{input})$ ratios were corrected by subtracting the corresponding value obtained from the Rpl16a immunoprecipitation. Average specific enrichment values were then calculated from two independent experiments. The obtained data were filtered according to the mRNA enrichment values using a threshold value increasing in 0.1 steps from -3 to +4. From the lists of mRNAs detected above these thresholds, the evolution of the proportions of the 20 amino acids at position 2 of the corresponding proteins was analyzed to identify any possible enrichment related to this position. HGT scores were also calculated for each threshold value to validate the statistical significance of the detected enrichments. After this analysis, the maxima of the HGT scores were observed for an enrichment threshold value around 0.8 (see Figure 4a) and this value was then used for subsequent analyses of Mak3-associated mRNAs. The LIMMA algorithm (75, 76) statistically validated the reproducibility of the list of enriched mRNA obtained with this threshold: all mRNAs with a mean \log_2 ratio greater than 0.8 in the two independent sel-TRAP experiments had an adjusted p-value lower than 0.05.

ACKNOWLEDGEMENT

We are grateful to Jean-Paul Di Rago and Deborah Tribouillard for the fruitful discussions on mitochondria and suggestions of experimental systems. Our thoughts go to the late Agnes Delahodde who provided us with the heat-sensitive yeast strain that proved crucial for the validation of our *in-silico* prediction.

FUNDING

This research was funded by the Emergence program of the Sorbonne University, the Systems Biology Network of the Paris Seine Biology Institute, and Fondation ARC pour la recherche sur le Cancer, <https://www.fondation-arc.org/recherche-cancer> (PJA 20171206624). The PhD of El Barbry H. was funded by the doctoral school Complexité Du Vivant of Sorbonne Université, and the PhD of Nashed S. by the program for disabled students of the French National Center for Scientific Research.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Balchin,D., Hayer-Hartl,M. and Hartl,F.U. (2016) In vivo aspects of protein folding and quality control. *Science*, **353**, aac4354.
2. Giglione,C., Fioulaine,S. and Meinnel,T. (2015) N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie*, **114**, 134–146.
3. Kramer,G., Boehringer,D., Ban,N. and Bukau,B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol*, **16**, 589–597.
4. Pechmann,S., Willmund,F. and Frydman,J. (2013) The ribosome as a hub for protein quality control. *Mol Cell*, **49**, 411–421.
5. Varland,S., Osberg,C. and Arnesen,T. (2015) N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics*, **15**, 2385–2401.
6. Sherman,F., Stewart,J.W. and Tsunasawa,S. (1985) Methionine or not methionine at the beginning of a protein. *BioEssays*, **3**, 27–31.
7. Aksnes,H., Ree,R. and Arnesen,T. (2019) Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Mol. Cell*, **73**, 1097–1114.
8. Arnesen,T., Van Damme,P., Polevoda,B., Helsens,K., Evjenth,R., Colaert,N., Varhaug,J.E., Vandekerckhove,J., Lillehaug,J.R., Sherman,F., *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**, 8157–8162.
9. Polevoda,B. and Sherman,F. (2000) N-terminal acetylation of eukaryotic proteins. *J Biol Chem*, **275**, 36479–36482.
10. Van Damme,P., Lasa,M., Polevoda,B., Gazquez,C., Elosegui-Artola,A., Kim,D.S., De Juan-Pardo,E., Demeyer,K., Hole,K., Larrea,E., *et al.* (2012) N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12449–12454.
11. Van Damme,P., Støve,S.I., Glomnes,N., Gevaert,K. and Arnesen,T. (2014) A *Saccharomyces cerevisiae* model reveals in vivo functional impairment of the Ogden syndrome N-terminal acetyltransferase NAA10 Ser37Pro mutant. *Mol. Cell Proteomics*, **13**, 2031–2041.
12. Deng,S. and Marmorstein,R. (2021) Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation. *Trends Biochem Sci*, **46**, 15–27.
13. Van Damme,P., Hole,K., Pimenta-Marques,A., Helsens,K., Vandekerckhove,J., Martinho,R.G., Gevaert,K. and Arnesen,T. (2011) NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet.*, **7**, e1002169.

14. Hu,X., Addlagatta,A., Lu,J., Matthews,B.W. and Liu,J.O. (2006) Elucidation of the function of type 1 human methionine aminopeptidase during cell cycle progression. *Proc Natl Acad Sci U S A*, **103**, 18148–18153.
15. Lowther,W.T. and Matthews,B.W. (2000) Structure and function of the methionine aminopeptidases. *Biochim Biophys Acta*, **1477**, 157–167.
16. Hwang,C.-S., Shemorry,A. and Varshavsky,A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science*, **327**, 973–977.
17. Kim,H.-K., Kim,R.-R., Oh,J.-H., Cho,H., Varshavsky,A. and Hwang,C.-S. (2014) The N-terminal methionine of cellular proteins as a degradation signal. *Cell*, **156**, 158–169.
18. Nguyen,K.T., Mun,S.-H., Lee,C.-S. and Hwang,C.-S. (2018) Control of protein degradation by N-terminal acetylation and the N-end rule pathway. *Exp. Mol. Med.*, **50**, 91.
19. Varshavsky,A. (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci.*, **20**, 1298–1345.
20. Shemesh,R., Novik,A. and Cohen,Y. (2010) Follow the leader: preference for specific amino acids directly following the initial methionine in proteins of different organisms. *Genomics Proteomics Bioinformatics*, **8**, 180–189.
21. Palmer,C.S., Anderson,A.J. and Stojanovski,D. (2021) Mitochondrial protein import dysfunction: mitochondrial disease, neurodegenerative disease and cancer. *FEBS Lett*, **595**, 1107–1131.
22. Chin,R.M., Panavas,T., Brown,J.M. and Johnson,K.K. (2018) Optimized Mitochondrial Targeting of Proteins Encoded by Modified mRNAs Rescues Cells Harboring Mutations in mtATP6. *Cell Reports*, **22**, 2818–2826.
23. Morel,G., Sterck,L., Swennen,D., Marcet-Houben,M., Onesime,D., Levasseur,A., Jacques,N., Mallet,S., Couloux,A., Labadie,K., *et al.* (2015) Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. *Sci Rep*, **5**, 11571.
24. Shen,X.-X., Oplente,D.A., Kominek,J., Zhou,X., Steenwyk,J.L., Buh,K.V., Haase,M.A.B., Wisecaver,J.H., Wang,M., Doering,D.T., *et al.* (2018) Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, **175**, 1533-1545.e20.
25. Supek,F. and Škunca,N. (2017) Visualizing GO Annotations. *Methods Mol Biol*, **1446**, 207–220.
26. Chacinska,A., Koehler,C.M., Milenkovic,D., Lithgow,T. and Pfanner,N. (2009) Importing Mitochondrial Proteins: Machineries and Mechanisms. *Cell*, **138**, 628–644.
27. von Heijne,G. (1986) Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J*, **5**, 1335–1342.
28. Tercero,J.C., Dinman,J.D. and Wickner,R.B. (1993) Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus. *J Bacteriol*, **175**, 3192–3194.
29. Plevoda,B., Norbeck,J., Takakura,H., Blomberg,A. and Sherman,F. (1999) Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J*, **18**, 6155–6168.
30. Kimura,Y., Takaoka,M., Tanaka,S., Sassa,H., Tanaka,K., Plevoda,B., Sherman,F. and Hirano,H. (2000) N(alpha)-acetylation and proteolytic activity of the yeast 20 S proteasome. *J Biol Chem*, **275**, 4635–4639.
31. del Alamo,M., Hogan,D.J., Pechmann,S., Albanese,V., Brown,P.O. and Frydman,J. (2011) Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol.*, **9**, e1001100.
32. Delaveau,T., Davoine,D., Jolly,A., Vallot,A., Rouvière,J.O., Gerber,A., Brochet,S., Plessis,M., Roquigny,R., Merhej,J., *et al.* (2016) Tma108, a putative M1 aminopeptidase, is a specific nascent chain-associated protein in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, **44**, 8826–8841.

33. Stein, K.C., Kriel, A. and Frydman, J. (2019) Nascent Polypeptide Domain Topology and Elongation Rate Direct the Cotranslational Hierarchy of Hsp70 and TRiC/CCT. *Mol. Cell*, **75**, 1117–1130.e5.
34. Willmund, F., del Alamo, M., Pechmann, S., Chen, T., Albanèse, V., Dammer, E.B., Peng, J. and Frydman, J. (2013) The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell*, **152**, 196–209.
35. Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L. and Horwich, A.L. (1989) Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature*, **337**, 620–625.
36. Mehawej, C., Delahodde, A., Legeai-Mallet, L., Delague, V., Kaci, N., Desvignes, J.-P., Kibar, Z., Capo-Chichi, J.-M., Chouery, E., Munnich, A., *et al.* (2014) The Impairment of MAGMAS Function in Human Is Responsible for a Severe Skeletal Dysplasia. *PLOS Genetics*, **10**, e1004311.
37. Cheng, M.Y., Hartl, F.U. and Horwich, A.L. (1990) The mitochondrial chaperonin hsp60 is required for its own assembly. *Nature*, **348**, 455–458.
38. Fang, Y.-C. and Cheng, M. (2002) The effect of C-terminal mutations of HSP60 on protein folding. *J Biomed Sci*, **9**, 223–233.
39. Hallberg, E.M., Shu, Y. and Hallberg, R.L. (1993) Loss of mitochondrial hsp60 function: nonequivalent effects on matrix-targeted and intermembrane-targeted proteins. *Mol Cell Biol*, **13**, 3050–3057.
40. Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T. and Kohda, D. (2000) Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell*, **100**, 551–560.
41. Saitoh, T., Igura, M., Obita, T., Ose, T., Kojima, R., Maenaka, K., Endo, T. and Kohda, D. (2007) Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J*, **26**, 4777–4787.
42. Yamano, K., Yatsukawa, Y.-I., Esaki, M., Hobbs, A.E.A., Jensen, R.E. and Endo, T. (2008) Tom20 and Tom22 share the common signal recognition pathway in mitochondrial protein import. *J Biol Chem*, **283**, 3799–3807.
43. Komiya, T., Rospert, S., Koehler, C., Looser, R., Schatz, G. and Mihara, K. (1998) Interaction of mitochondrial targeting signals with acidic receptor domains along the protein import pathway: evidence for the ‘acid chain’ hypothesis. *EMBO J*, **17**, 3886–3898.
44. Krayl, M., Lim, J.H., Martin, F., Guiard, B. and Voos, W. (2007) A cooperative action of the ATP-dependent import motor complex and the inner membrane potential drives mitochondrial preprotein import. *Mol Cell Biol*, **27**, 411–425.
45. van der Laan, M., Meinecke, M., Dudek, J., Hutu, D.P., Lind, M., Perschil, I., Guiard, B., Wagner, R., Pfanner, N. and Rehling, P. (2007) Motor-free mitochondrial presequence translocase drives membrane integration of preproteins. *Nat Cell Biol*, **9**, 1152–1159.
46. Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A. and Nielsen, H. (2019) Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance*, **2**, e201900429.
47. Forte, G.M.A., Pool, M.R. and Stirling, C.J. (2011) N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. *PLoS Biol.*, **9**, e1001073.
48. Obita, T., Muto, T., Endo, T. and Kohda, D. (2003) Peptide library approach with a disulfide tether to refine the Tom20 recognition motif in mitochondrial presequences. *J Mol Biol*, **328**, 495–504.
49. Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P. and Imai, K. (2015) MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol Cell Proteomics*, **14**, 1113–1126.

50. Hammen,P.K., Heard,T.S., Waltner,M. and Weiner,H. (1999) The loss in hydrophobic surface area resulting from a Leu to Val mutation at the N-terminus of the aldehyde dehydrogenase presequence prevents import of the protein into mitochondria. *Protein Sci*, **8**, 890–896.
51. Aksnes,H., Osberg,C. and Arnesen,T. (2013) N-terminal acetylation by NatC is not a general determinant for substrate subcellular localization in *Saccharomyces cerevisiae*. *PLoS ONE*, **8**, e61012.
52. Grunwald,S., Hopf,L.V.M., Bock-Bierbaum,T., Lally,C.C.M., Spahn,C.M.T. and Daumke,O. (2020) Divergent architecture of the heterotrimeric NatC complex explains N-terminal acetylation of cognate substrates. *Nat Commun*, **11**, 5506.
53. Dihanich,M., van Tuinen,E., Lambris,J.D. and Marshallsay,B. (1989) Accumulation of viruslike particles in a yeast mutant lacking a mitochondrial pore protein. *Mol Cell Biol*, **9**, 1100–1108.
54. Lee,Y.J. and Wickner,R.B. (1992) MAK10, a glucose-repressible gene necessary for replication of a dsRNA virus of *Saccharomyces cerevisiae*, has T cell receptor alpha-subunit motifs. *Genetics*, **132**, 87–96.
55. Polevoda,B. and Sherman,F. (2001) NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. *J Biol Chem*, **276**, 20154–20159.
56. Drazic,A. and Varland,S. (2021) Human NAA30 can rescue yeast mak3Δ mutant growth phenotypes. *Biosci Rep*, **41**, BSR20202828.
57. Van Damme,P., Kalvik,T.V., Starheim,K.K., Jonckheere,V., Myklebust,L.M., Menschaert,G., Varhaug,J.E., Gevaert,K. and Arnesen,T. (2016) A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity. *Mol. Cell Proteomics*, **15**, 3361–3372.
58. Mughal,A.A., Grieg,Z., Skjellegrind,H., Fayzullin,A., Lamkhannat,M., Joel,M., Ahmed,M.S., Murrell,W., Vik-Mo,E.O., Langmoen,I.A., *et al.* (2015) Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells. *Mol Cancer*, **14**, 160.
59. Tercero,J.C. and Wickner,R.B. (1992) MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly. *J Biol Chem*, **267**, 20277–20281.
60. Monda,J.K., Scott,D.C., Miller,D.J., Lydeard,J., King,D., Harper,J.W., Bennett,E.J. and Schulman,B.A. (2013) Structural Conservation of Distinctive N-terminal Acetylation-Dependent Interactions across a Family of Mammalian NEDD8 Ligation Enzymes. *Structure*, **21**, 42–53.
61. Scott,D.C., Monda,J.K., Bennett,E.J., Harper,J.W. and Schulman,B.A. (2011) N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science*, **334**, 674–678.
62. Murthi,A. and Hopper,A.K. (2005) Genome-wide screen for inner nuclear membrane protein targeting in *Saccharomyces cerevisiae*: roles for N-acetylation and an integral membrane protein. *Genetics*, **170**, 1553–1560.
63. Starheim,K.K., Gromyko,D., Evjenth,R., Rynningen,A., Varhaug,J.E., Lillehaug,J.R. and Arnesen,T. (2009) Knockdown of human N alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. *Mol. Cell. Biol.*, **29**, 3569–3581.
64. Behnia,R., Panic,B., Whyte,J.R.C. and Munro,S. (2004) Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat Cell Biol*, **6**, 405–413.
65. Setty,S.R.G., Strohlic,T.I., Tong,A.H.Y., Boone,C. and Burd,C.G. (2004) Golgi targeting of ARF-like GTPase Arl3p requires its Nalpha-acetylation and the integral membrane protein Sys1p. *Nat Cell Biol*, **6**, 414–419.
66. Rout,M.P., Aitchison,J.D., Suprpto,A., Hjertaas,K., Zhao,Y. and Chait,B.T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol*, **148**, 635–651.
67. Agier,N., Fleiss,A., Delmas,S. and Fischer,G. (2021) A Versatile Protocol to Generate Translocations in Yeast Genomes Using CRISPR/Cas9. *Methods Mol Biol*, **2196**, 181–198.

68. The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**, D480–D489.
69. Bleykasten-Grosshans,C., Fabrizio,R., Friedrich,A. and Schacherer,J. (2021) Species-Wide Transposable Element Repertoires Retrace the Evolutionary History of the *Saccharomyces cerevisiae* Host. *Mol Biol Evol*, **38**, 4334–4345.
70. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**, 25–29.
71. Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*, **49**, D325–D334.
72. Bauer,S., Robinson,P.N. and Gagneur,J. (2011) Model-based gene set analysis for Bioconductor. *Bioinformatics*, **27**, 1882–1883.
73. Vizcaíno,J.A., Deutsch,E.W., Wang,R., Csordas,A., Reisinger,F., Ríos,D., Dianes,J.A., Sun,Z., Farrah,T., Bandeira,N., *et al.* (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol*, **32**, 223–226.
74. Lemoine,S., Combes,F., Servant,N. and Le Crom,S. (2006) Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics*, **7**, 467.
75. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, **3**, Article3.
76. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47.
77. Wagih,O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645–3647.

Functional mapping of N-terminal residues in the yeast proteome uncovers novel determinants for mitochondrial protein import.

Nashed S^{1*}, El Barbry H^{1*}, Benchouaia M¹, Dijoux-Maréchal A¹, Ruiz Gutierrez N¹, Gaulier L¹, Chevreux G², Le Crom S¹, Palancade B², Devaux F¹, Laine E¹, Garcia M^{1**}

¹ Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative, Paris, France

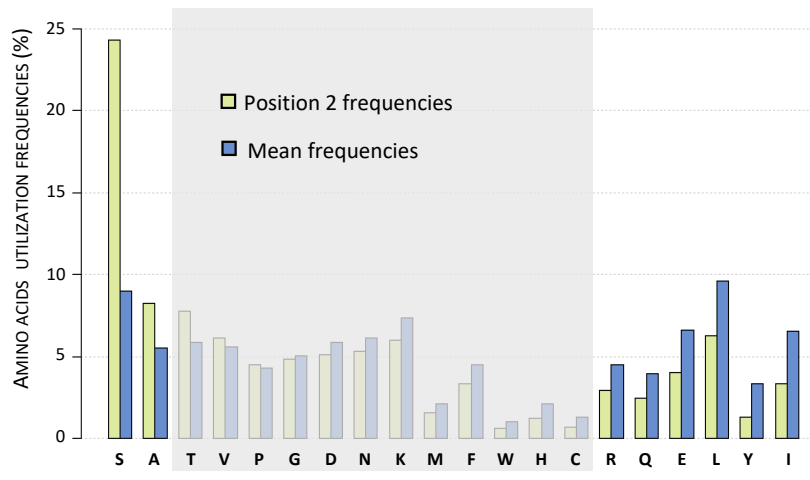
² Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France.

* Equally contributing authors

**To whom correspondence should be addressed. Tel: +33 1 44278140; Fax: +33 1 44277336; Email: mathilde.garcia@sorbonne-universite.fr

FIGURES

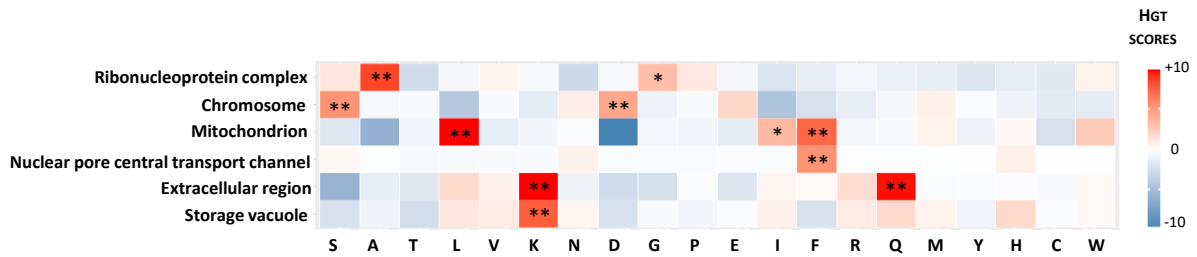
a



AMINO ACID	LOG2 RATIO	HGT SCORE	ASPECIFICITY SCORE (%)
S	1,4	268	0
A	0,6	17	0
T	0,4	9,2	5
V	0,1	1,5	79
P	0,1	0,7	94
G	-0,1	-0,5	83
D	-0,2	-2,1	46
N	-0,2	-2,3	43
K	-0,3	-4,9	9
M	-0,4	-2,6	72
F	-0,4	-5,6	2
W	-0,7	-3,3	7
H	-0,8	-6,7	2
C	-0,9	-5,0	1
R	-0,6	-9,2	0
Q	-0,7	-9,6	0
E	-0,7	-17	0
L	-0,6	-20	0
Y	-1,3	-22	0
I	-1,0	-28	0

b

b1



b2

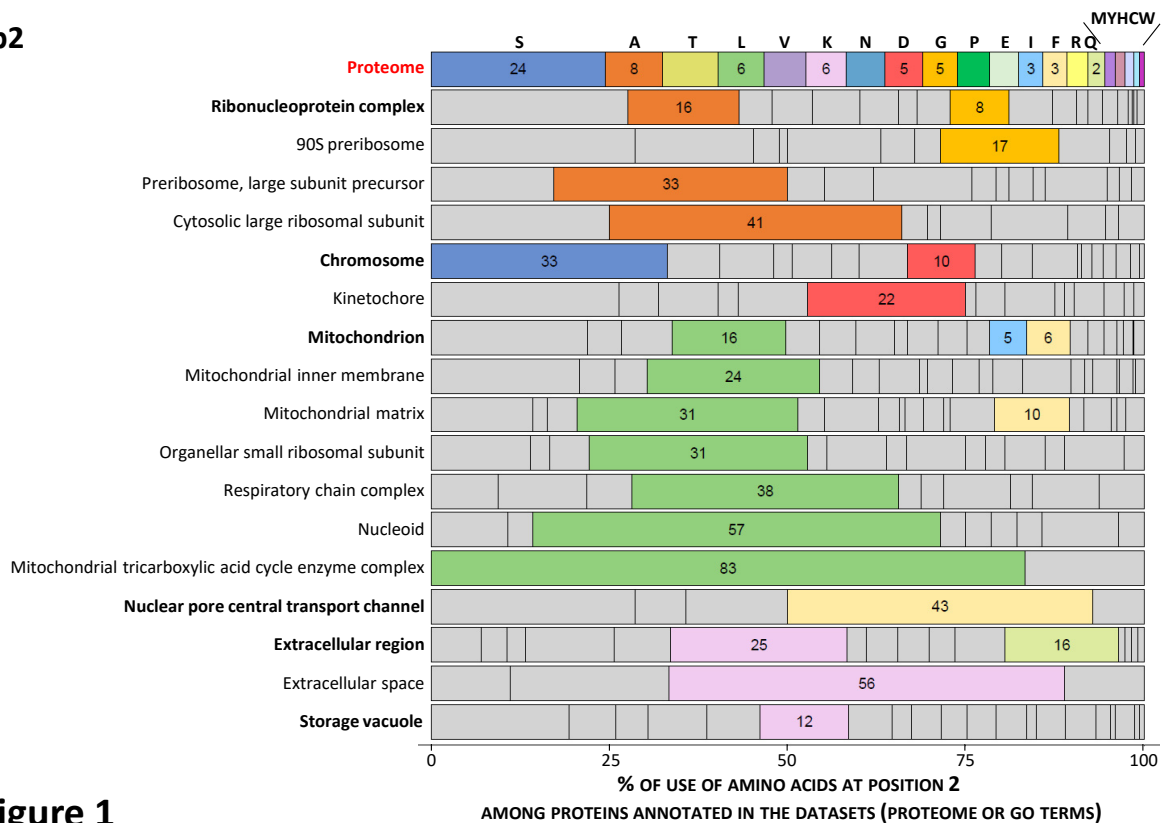


Figure 1

Figure 1: Amino acid usage biases at position 2 of *S. cerevisiae* proteins

(a) Amino acid usage at position 2 compared to their average usage at any position in *S. cerevisiae* proteome. Frequencies (Barplot, left) of the 20 amino acids measured at position 2 among the 5,959 sequences of *S. cerevisiae* proteins retrieved from UniProtKB database (68). Table (right) recapitulates quantitative values calculated to identify significant amino acid usage biases at position 2: position2/mean log2Ratio, HGT score obtained from hypergeometric tests, and aspecificity scores indicating the percentage of positions in sequences with usage bias as high as that measured at position 2 (see Methods). Amino acids for which the absolute value of the HGT score is greater than 3 ($p\text{-value} < 10^{-3}$) and the aspecificity score is lower than 5% are highlighted in yellow. Such values indicate a significant and specific bias of usage at position 2 revealing a possible selection pressure.

(b) Detection of biases in the use of amino acids at position 2 in subsets of proteins of the various cellular compartments of *S. cerevisiae*. We performed a systematic detection of the significant over-representation of amino acids at position 2 among groups of proteins sharing the same GO annotations (see Methods). (b1) Heatmap of HGT scores for the 20 amino acids in GO terms in the "component" category identified with this strategy. Seventy-two GO terms were first identified with HGT score higher than 4 (i.e. $p\text{-value} < 10^{-4}$) for at least one amino acid. Our reduction algorithm finally retained 6 bestN GO terms corresponding to the minimal number of terms maximizing the number of proteins with the detected usage biases (see Methods). Stars indicate significant HGT scores (*HGT>3, **HGT>4). (b2) Barplot showing the percentages of use of the 20 amino acids at position 2 in retained GO terms compared to their use at the same position in the proteome. For a better characterization of position 2 functional biases, bestF GO terms were retrieved by our algorithm (see Methods) and added to bestN GO terms in bold. Each of them contains proteins included in one BestN GO term and has higher position 2 frequency biases. For each GO term, highlighted amino acids are those for which the percentage of use at position 2 is significantly higher than that observed at the same position in the proteome (HGT score > 3). In figures b1 and b2, the amino acids are sorted in decreasing order of use at position 2 in the proteome.

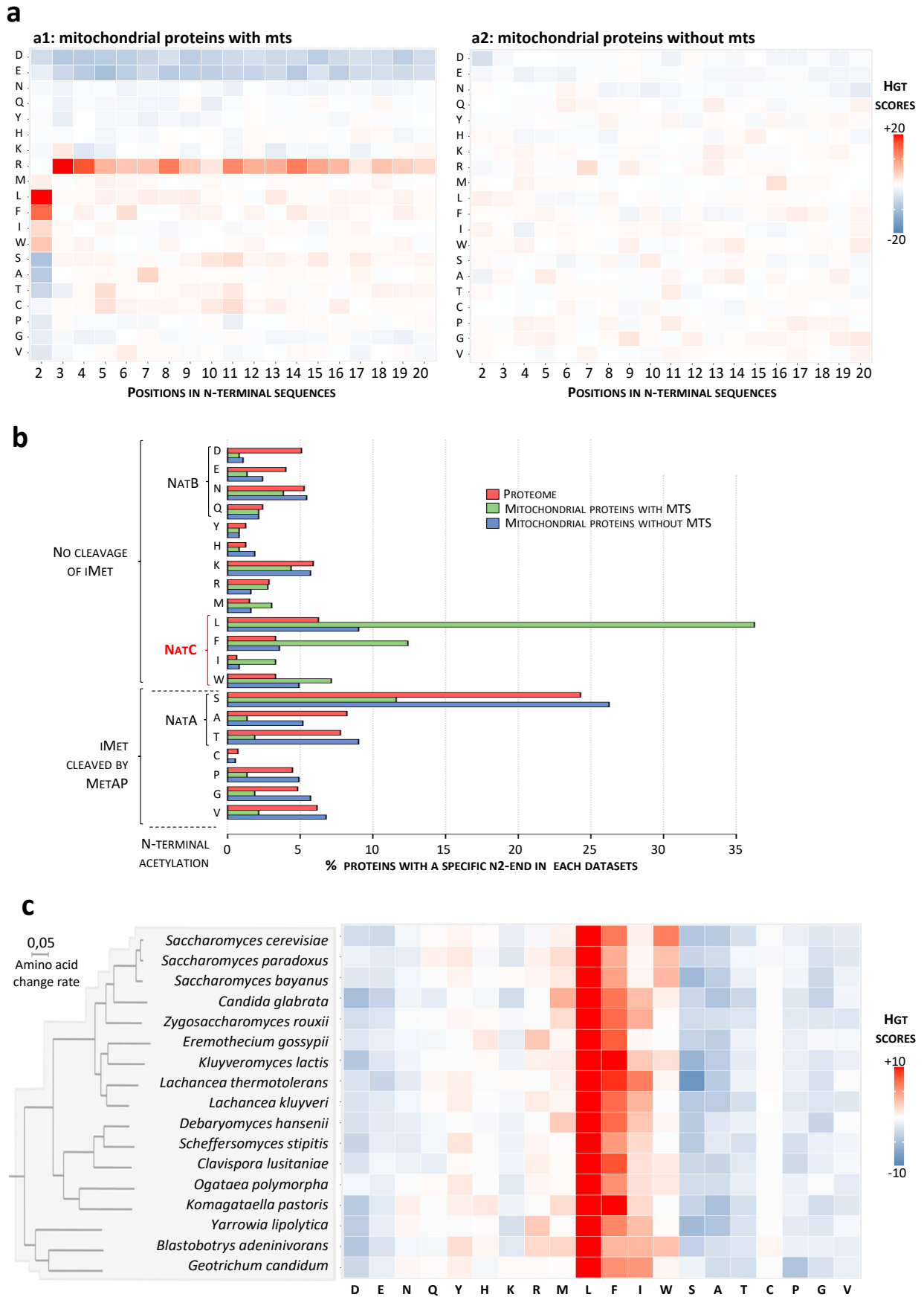


Figure 2

Figure 2: Mitochondrial precursors harboring a MTS are putative substrates of NatC

To characterize the bias detected at position 2 of mitochondrial precursors, the list of mitochondrial proteins extracted from the Gene Ontology database was manually curated (see Methods), and mitochondrial precursors were sorted according to whether they carry a MTS based on annotations retrieved from the Uniprot database (68). The two classes of mitochondrial proteins were then analyzed separately.

(a) Heatmap showing the hypergeometric enrichment scores (HGT score) obtained by comparing the amino acid usage at the first 20 positions of the mitochondrial precursors harboring (a1) or not (a2) a MTS with their average usage in the proteome at these same positions. Mitochondrial precursors with a MTS display the expected MTS characteristics (positive bias toward R and negative bias toward D and E) and exhibit specific biases at position 2 (positive bias toward F, L, I and W). This position 2 specific biases were not observed among precursors without N-terminal addressing sequence.

(b) Distribution of the 20 amino acids at position 2 in the proteome and among the two classes of mitochondrial precursors. Several amino acids are overrepresented at position 2 of mitochondrial precursors harboring a MTS, which is not the case for the mitochondrial precursors lacking a MTS. Amino acids showing significant increase compared to the proteome are those that define NatC putative substrates according to the literature: Leucine (6 fold increase: 36% vs 6%, p-value = 10^{-71}), Phenylalanine (4 fold increase: 13% vs 3%, p-value = 10^{-15}), Tryptophan (5 fold increase : 3% vs 0.6%, p-value = 10^{-6}) and Isoleucine (2 fold increase: 7% vs 3%, p-value = 10^{-4}).

(c) Heatmap showing, for 17 budding yeasts of the Saccharomycotina lineage, the HGT score obtained by comparing amino acids usage at position 2 of MTS-harboring mitochondrial precursors and usage at the same position in the proteome. The positive bias, particularly for leucine, at position 2 of mitochondrial precursors is conserved in all analyzed species. The phylogenetic tree was retrieved and modified from (23).

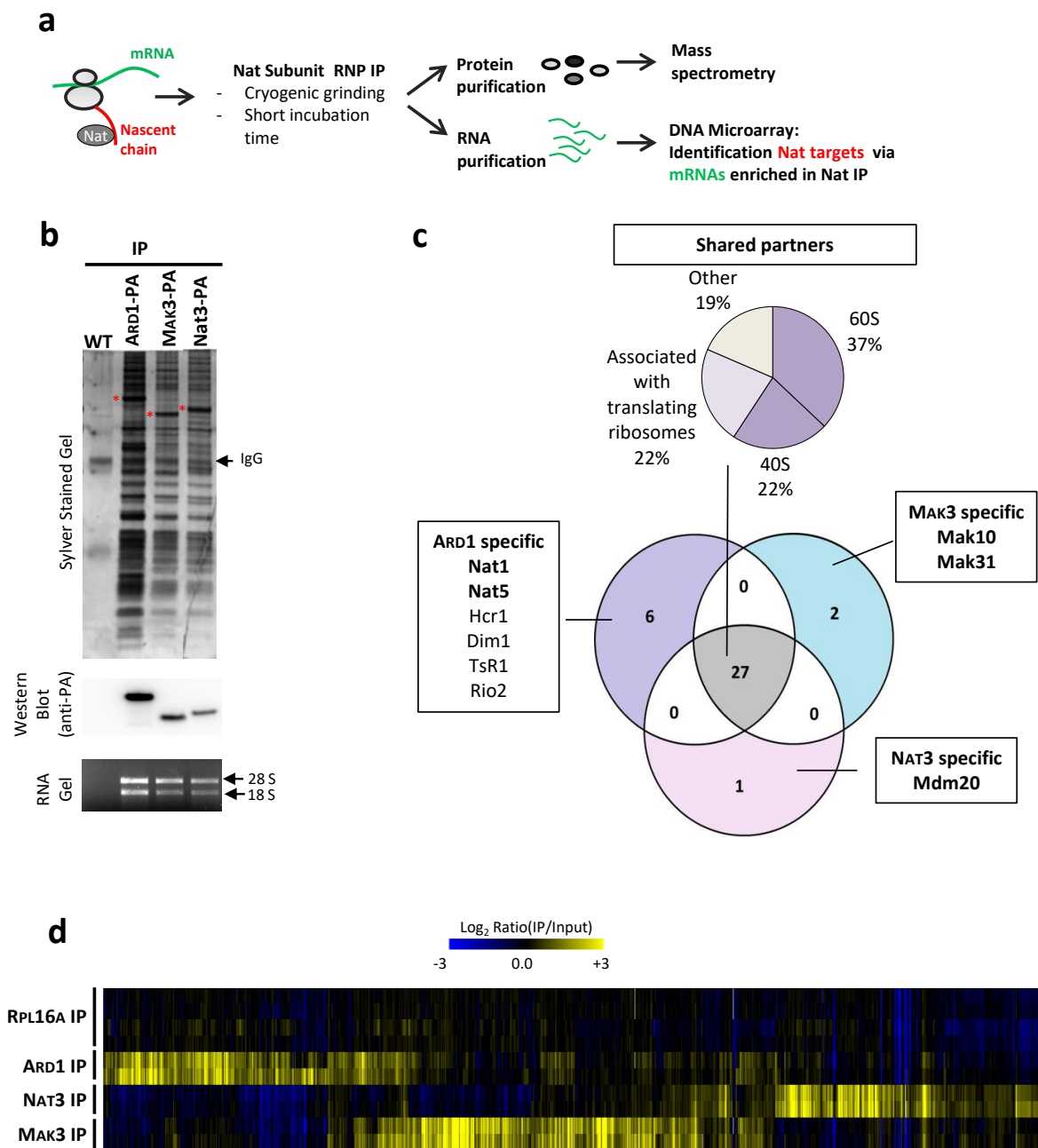


Figure 3

Figure 3: Selective Translating Ribosome Affinity Purification (sel-TRAP): a comprehensive approach to identify translational targets of N-acetyltransferases.

(a) Experimental approach. Immunoprecipitations of NatA, NatB and NatC were performed on cells expressing respectively Ard1-PA, Nat3-PA and Mak3-PA. A control experiment was also performed with an untagged strain (WT). Protein (b and c) and RNA (b and d) contents were analyzed from immunoprecipitated samples (IP).

(b) Top panel: Immunoprecipitated fractions obtained from untagged or PA fusion protein expressing-cells were analyzed by SDS-PAGE followed by silver staining. The position of the Ard1-PA, Nat3-PA and Mak3-PA fusion and IgG are indicated. Middle panel: The same samples were analyzed by SDS-PAGE followed by western-blots using antibodies recognizing the PA epitope. Bottom panel: Following RNA isolation, immunoprecipitated fractions obtained were analyzed by agarose gel electrophoresis with ethidium bromide staining. The bands corresponding to 28S and 18S rRNA are indicated.

(c) Comparison of the composition of NatA (Ard1), NatB (Nat3) and NatC (Mak3)-associated ribosomes. Immunoprecipitations of proteinA-tagged versions of the catalytic subunit of the three Nats were followed by mass spectrometry analysis of their protein partners. Venn Diagram represents the overlap of the three different sets of co-immunoprecipitated proteins: for each sector the number and the name of interactants reproducibly detected in two independent experiments are indicated (Table S4). Partners indicated in bold correspond to the known subunits of the N-terminal acetyltransferase complexes. Chart pie shows the functional distribution of proteins for shared partners in the three datasets.

(d) To identify mRNAs specifically enriched after sel-TRAP with cells expressing protein A tagged versions of Ard1, Nat3 or Mak3, RNA contents from total extract (Input) and immunoprecipitated samples (IP) were analyzed using microarray technology. As a control, the same approach was applied with the whole set of ribosomes immunoprecipitated using cells expressing a proteinA-tagged version of Rpl16a. Hierarchically clustered heat map represents IP/input log₂(ratios) for mRNAs enriched, in two independent experiments, in at least one of the NAT immunoprecipitations compared to Rpl16a (Table S3). Each line represents an experiment, each column a gene.

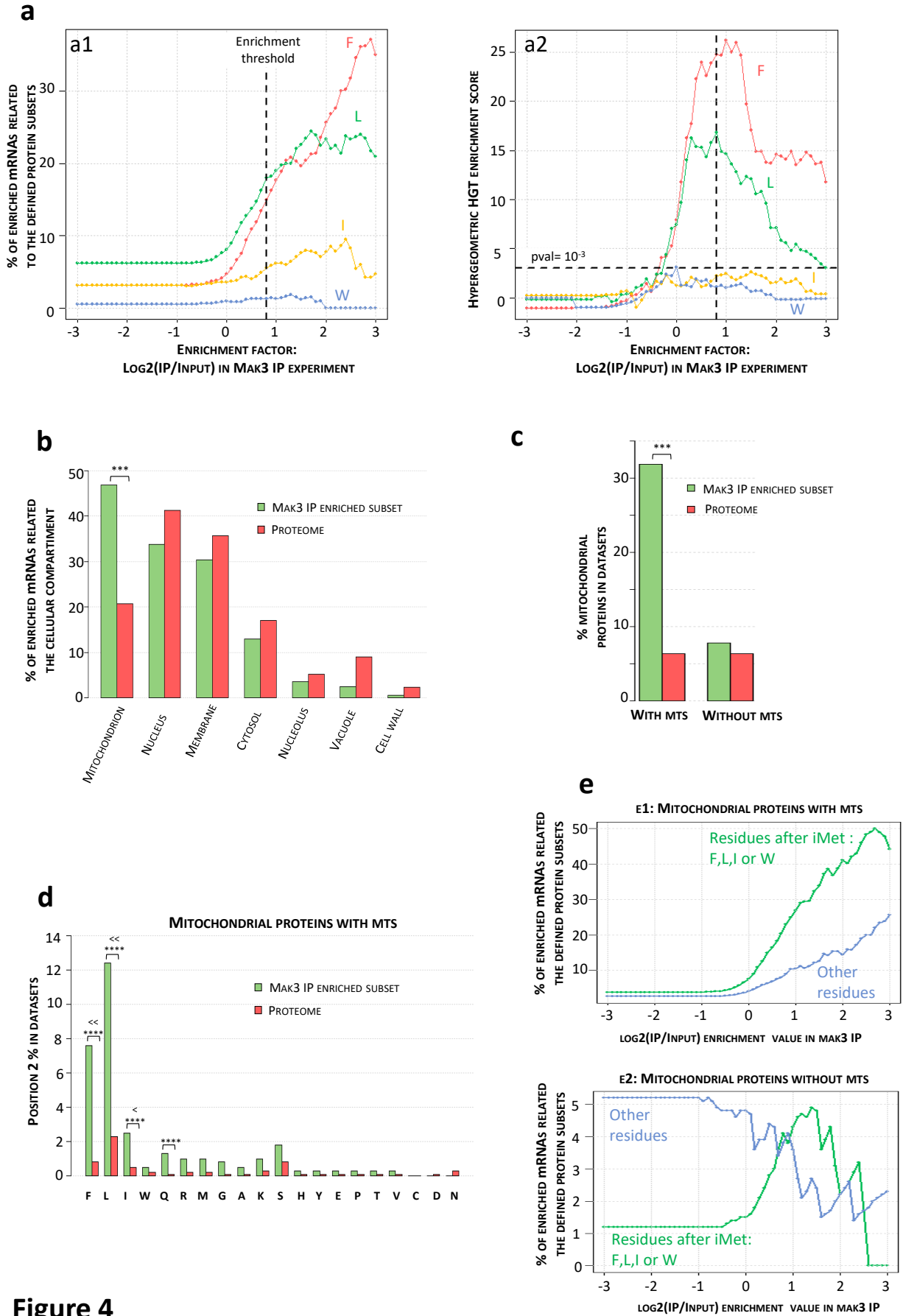


Figure 4

Figure 4: Analysis of NatC selectivity with sel-TRAP technology confirms mitochondrial precursors harboring a MTS as its main targets.

Global analysis of mRNAs copurified after immunoprecipitations (IP) from cells expressing Mak3-PA allowed the indirect characterization of nascent protein chains translated by ribosomes associated with NatC. For each cellular protein, the corresponding mRNA enrichment value, defined as $\text{IP}/\text{input} \log_2(\text{ratio})$, was calculated after microarray hybridizations of cDNA reverse-transcribed from immunoprecipitation (IP) and input samples. Since NatC, as well as other NATs, is known to target proteins according to the nature of their N-terminus, the evolution of the proportion of the 20 possible residues at position 2 associated with the subsets of mRNAs detected with increasing enrichment values in IP experiments was studied (Fig. sup3).

(a) Curves obtained for N-terminal residues defining the putative substrates of NatC (i.e. F, L, I and W), are shown in a1. To assess the statistical significance of the observed increases, the evolution of HGT scores as a function of the enrichment value is also plotted (a2), which allowed the detection of a significant enrichment of the MF- and ML- N-termini in the most enriched protein set (horizontal dotted line indicates the HGT score cutoff of 3 corresponding to a p-value equal to 10^{-3}). Because highest HGT scores are reached for an enrichment value around 0.8, this value was chosen as a threshold (vertical dotted lines in a1 and a2) to determine enriched mRNAs identified with sel-TRAP experiments of NatC.

(b) The percentages of each residue at position 2 among the identified targets (enrichment value ≥ 0.8) were compared to those observed in the proteome. MF- and ML- N-termini represent respectively 15 and 18% of the enriched dataset versus 3 and 6% in the proteome.

(c) The percentages, among the identified targets, of proteins with GO annotation corresponding to the main cellular compartments were compared to their respective percentage in the proteome. This analysis confirmed the strong selectivity of NatC for mitochondrial precursors, which represent 47% of the identified NatC targets compared to 21% in overall cellular proteins (p-value = 10^{-33}).

(d) The percentages of the two classes of mitochondrial precursors among the identified targets (enrichment value ≥ 0.8) were compared to those observed in the proteome. Mitochondrial precursors with a MTS represent 32% of the identified NatC targets versus 6% in the proteome (p-value= 10^{-54}). No significant enrichment of the mitochondrial precursors without MTS was detected in the sel-TRAP data.

(e) The percentages of the different N-termini among the MTS identified as NatC targets were compared with those observed globally among MTSs. Significant enrichment was detected for the MF-, ML-, and MI- N-termini (p-values equal to 10^{-24} , 10^{-23} , and 10^{-25} , respectively) and to a lesser extent for the MQ- N-termini (p-value equal to 10^{-4}). Stars indicate significant p values (p value= $<10^{-4}$).

(f) Evolution of the proportion of specific N-termini in the two classes of mitochondrial precursors in the mRNA subsets defined with increasing enrichment values. Precursors with (e1) or without mts (e2) were sorted depending on whether they carry at position 2 a putative NatC-targeted residue (L, F, W or I). As expected, we observed a significant increase in the proportion of precursors with NatC-targeted N-termini, regardless of whether they carry a MTS. This analysis also revealed an increase in the proportion of precursors with MTSs that lack an N-terminus recognized by NatC, suggesting that NatC may recognize other MTS features.

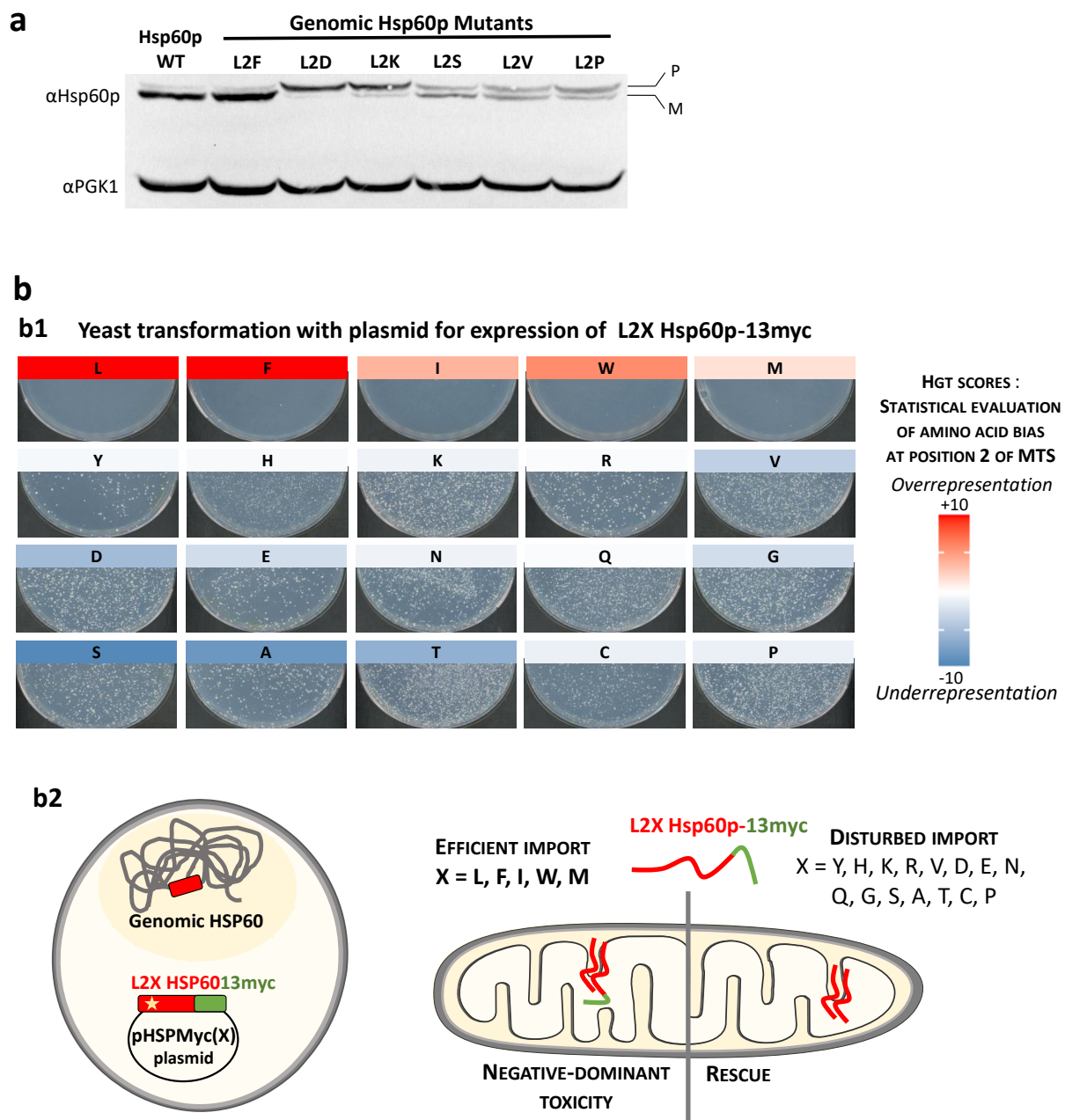


Figure 5

Figure 5: Mutagenesis of position 2 of Hsp60p MTS confirmed our predictions of the optimal residues for efficient mitochondrial import.

(a) L2X genomic mutants of Hsp60p were generated using CRISPR/CAS9 technology (see Methods) and proteins were analyzed by SDS-PAGE followed by western blots using antibodies against Hsp60p and Pgk1p, used as loading control. The position of the precursor and mature Hsp60p are indicated. In L2D, L2K, L2S, L2V, L2P mutants, the increase in precursor bands relative to the mature band suggests a defect in mitochondrial import of the mutated Hsp60p protein.

(b) The pHSPmyc(X) plasmids expressing different versions of Hsp60p-13myc mutated at position 2 were transformed into the thermosensitive strain YPH499 pam16Δ-MAGN76D (see b2 for experimental strategy). Photographs of the transformation plates (b1) showed that only a few clones were obtained after transformation with the plasmid encoding the wild-type leucine at position 2 of Hsp60p, a result probably due to the dominant-negative effect of the HSP60-13myc allele (see b2 for a schematic of the interpretation of results). Similar results were obtained after transformation of the strain with plasmids expressing the L2F, L2I, L2W and L2M mutants of Hsp60p-13myc. In contrast, transformations with plasmids expressing all other L2X mutants of Hsp60p-13myc resulted in more than 1000 transformants, which is likely due to impaired mitochondrial import of the mutated Hsp60p in these cases (see b2 for interpretation scheme). Interestingly, the two sets of residues at position 2 defined according to the results of the transformation experiment correlate perfectly with the amino acid usage biases at position 2 of the MTSs detected in our *in-silico* analysis, as indicated by the color scale representing the previously calculated HGT scores (Fig.4a1). The experimental strategy and interpretation scheme are presented in b2 (see main text for details).

S. cerevisiae amino acid usages of MTSs with optimal residues at position 2

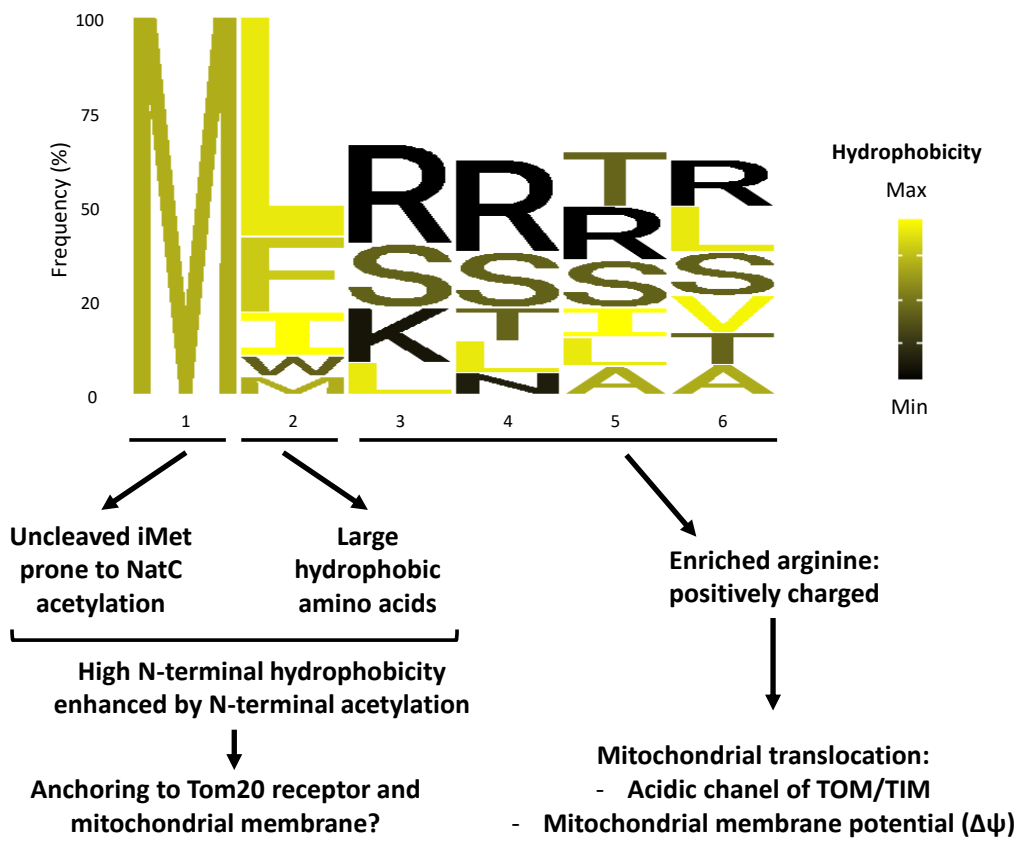


Figure 6

Figure 6: A model to explain the bias at position 2 of MTSs and the importance of their N-terminal hydrophobicity

Sequence logo (77) of the first 6 residues of *S. cerevisiae* MTSs with optimal residues at position 2 (Leu, Phe, Ile, Trp, Met) according to our amino acid usage bias analysis and our results of Hsp60p MTS mutagenesis. For residues 3 to 6, only the most frequent residues that represent at least 60% of the observed residues are represented. The N-terminus of this class of MTS exhibits high hydrophobicity due to (1) the lack of iMet cleavage, (2) the intrinsic hydrophobicity of the residue at position 2, and (3) potential acetylation of the N-terminal amine by NatC. We propose that this N-terminal feature is favorable for an efficient interaction of MTSs with the mitochondrial membrane and the Tom20p receptor. In addition, we observe, from the residue at position 3 of MTSs, a high enrichment of positively charged arginine involved in the translocation of MTSs into mitochondria. We propose that this residue configuration in MTSs has been selected during evolution because of its ability to promote efficient mitochondrial import with limited interference between the two driving forces, hydrophobicity vs. positive charge, required in this mechanism.

Functional mapping of N-terminal residues in the yeast proteome uncovers novel determinants for mitochondrial protein import.

Nashed S^{1*}, El Barbry H^{1*}, Benchouaia M¹, Dijoux-Maréchal A¹, Ruiz Gutierrez N¹, Gaulier L¹, Chevreux G², Le Crom S¹, Palancade B², Devaux F¹, Laine E¹, Garcia M^{1**}

¹ Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative, Paris, France

² Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France.

* Equally contributing authors

**To whom correspondence should be addressed. Tel: +33 1 44278140; Fax: +33 1 44277336; Email: mathilde.garcia@sorbonne-universite.fr

SUPPLEMENTARY INFORMATION

TABLE OF CONTENT

The supplementary information refers to additional data on methods and results, and includes several tables.

Some of the additional results are available as supplementary figures S1 to S4.

Tables S1-S4 are available as independent excel files. Table S5-S9 are grouped in Supplemental Methods.

Table S1: Complete table of GO terms identified with amino acids biases at position 2

Table S2: Annotation of mitochondrial proteins of *S. cerevisiae*

Table S3: Microarray data from sel-TRAP experiments

Table S4: Mass spectrometry data from sel-TRAP experiments

Table S5: Yeast strains

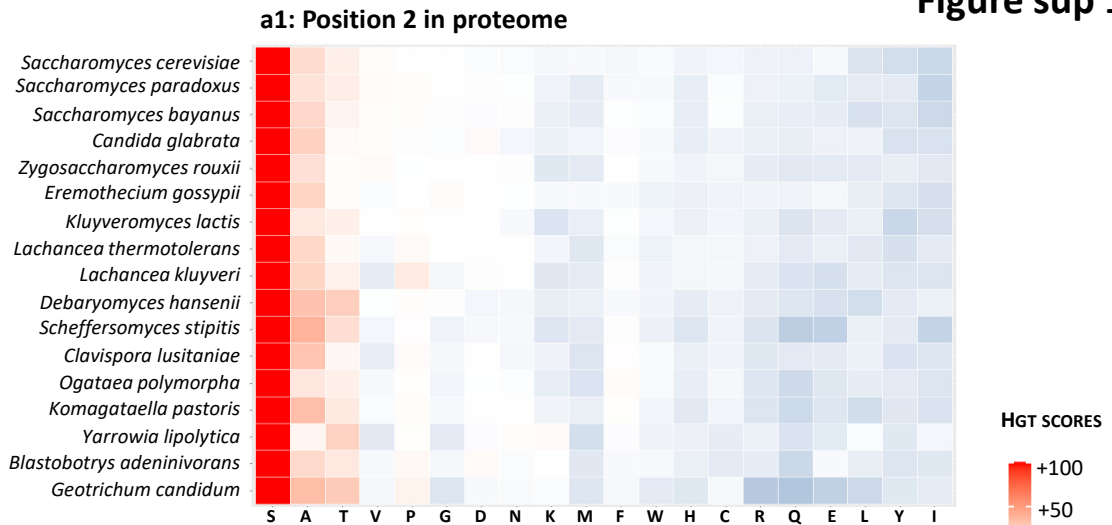
Table S6: Plasmids and derived strains

Table S7: Oligonucleotides

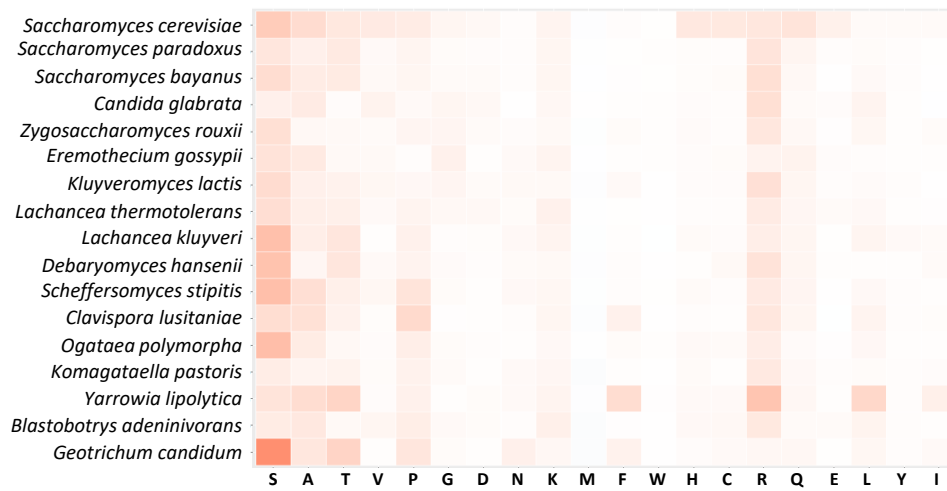
Table S8: Codon used in the reparation cassette to encode the desired X mutation.

Table S9: List of the yeast species used for the genomic comparative studies

a



a2: Max Positions 3-30 in proteomes



b

Max positions 3-30 in mitochondrial targeting sequences

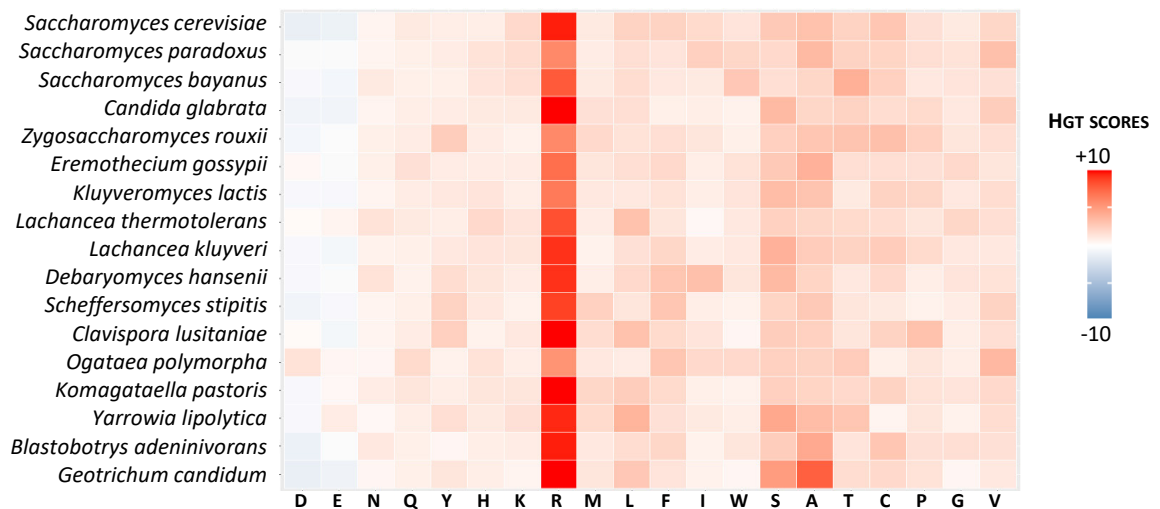


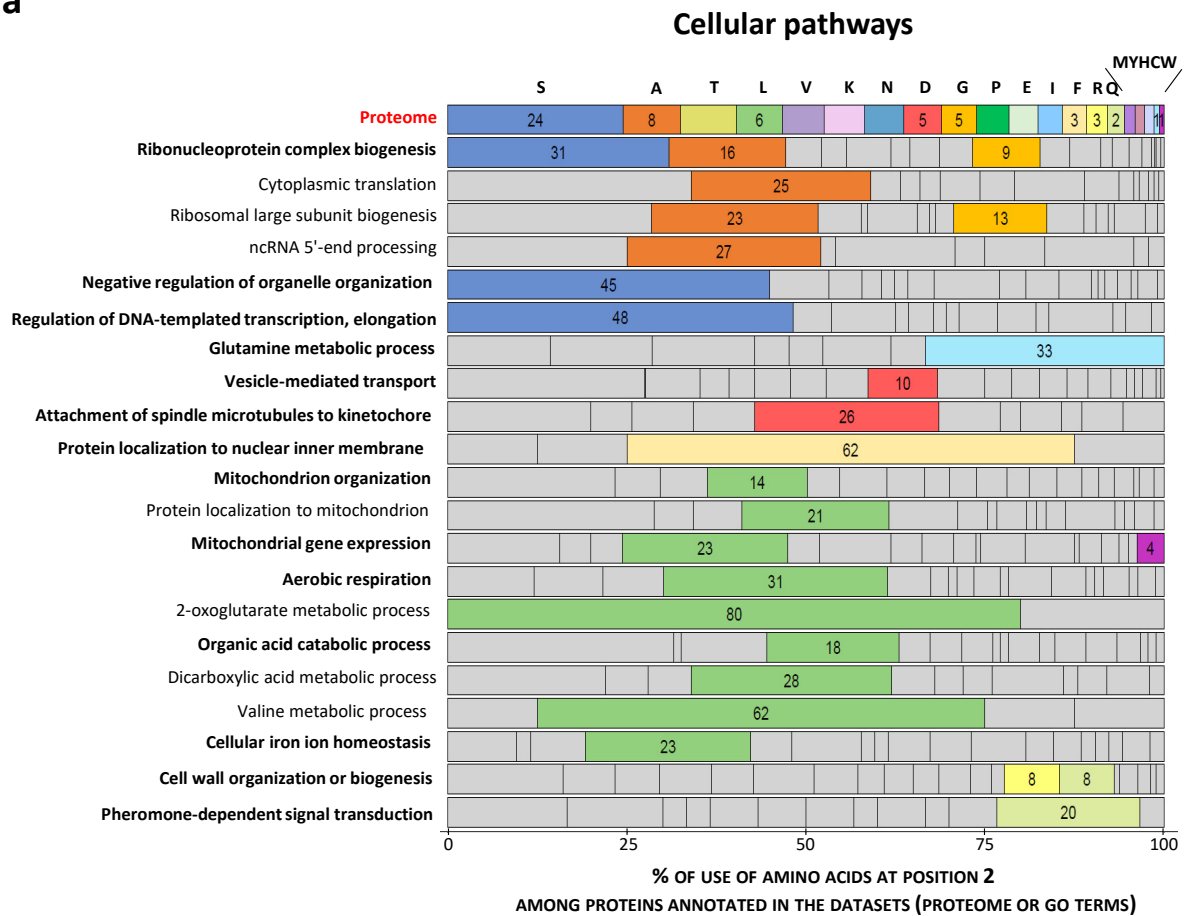
Fig Sup 1: Conservation and specificity analysis of positional bias in amino acid utilization in the Saccharomycotina lineage.

(a) Heatmap showing, for 17 budding yeasts of the Saccharomycotina lineage, the hypergeometric enrichment scores (HGT scores) obtained by comparing the utilization of amino acids at position 2 of the proteome to their respective average usage at any position (a1). Amino acid biases at position 2, particularly the high enrichment of serine, are similar in all species analyzed. The maximum values of the HGT scores obtained by comparing the amino acid usage at each position between 3 and 30 in the proteome to the average usage (a2) do not reach the values observed at position 2, as observed in the heatmap presented in a2 which confirmed the specificity of the amino acid bias observed at position 2.

(b) Heatmap showing, for 17 budding yeasts of the Saccharomycotina lineage, the maximum values of the HGT scores obtained by comparing the amino acid utilization at each position between 3 and 30 of the MTSs with the amino acid usage in the proteome at these same positions. Comparison with the heatmap analyzing the amino acid bias at position 2 of the MTSs (Fig. 2c) clearly validated the specificity of the biases detected at position 2 of MTSs (i.e. enrichment in L, F and to a lesser extent I and W).

Figure sup 2

a



b

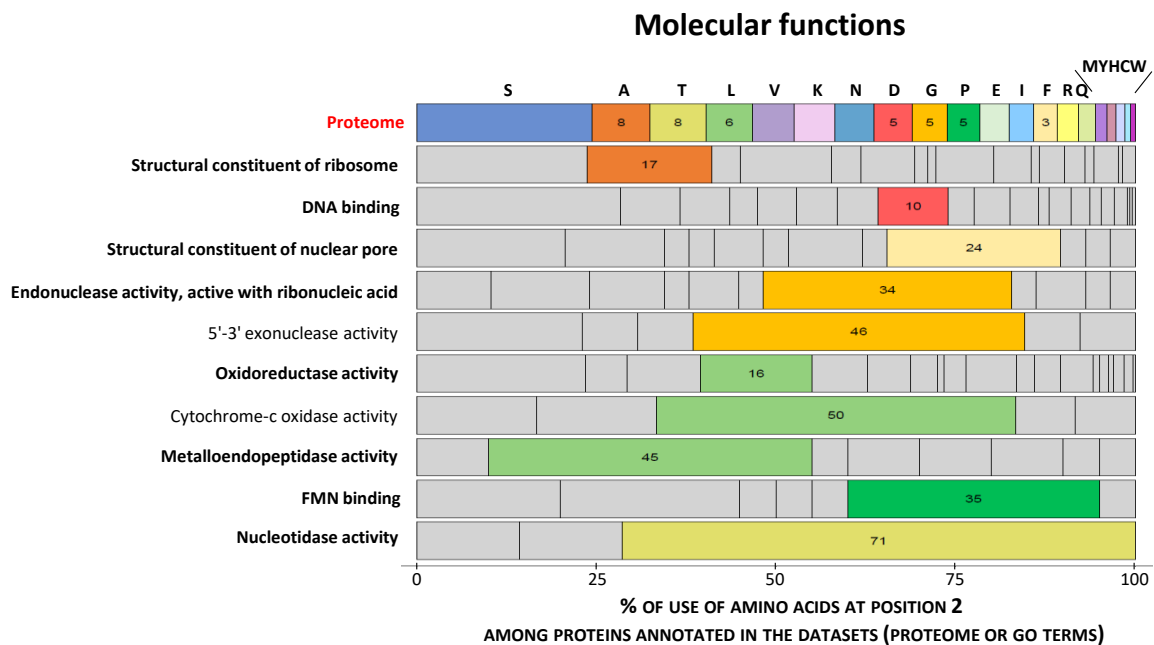


Fig Sup 2: Amino acid usage biases at position 2 in GO terms related to “Cellular pathways” and “Molecular functions” GO categories

Barplot showing the percentages of use of the 20 amino acids at position 2 in retained GO terms related to cellular pathways (a) and molecular function (b) compared to their use at the same position in the proteome. bestN GO terms corresponding to the minimal number of terms maximizing the number of proteins with the detected usage biases (see Methods) are indicated in bold. For a better characterization of position 2 functional biases, bestF GO terms were retrieved also by our algorithm (regular typography). Each of them contains proteins included in one BestN GO term and has higher position 2 frequency biases. For each GO term, highlighted amino acids are those for which the percentage of use at position 2 is significantly higher than that observed at the same position in the proteome (HGT score > 3). Amino acids are sorted in decreasing order of use at position 2 in the

Figure Sup 3

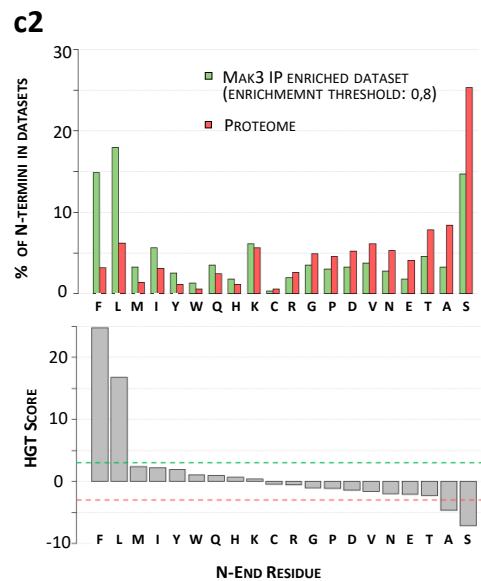
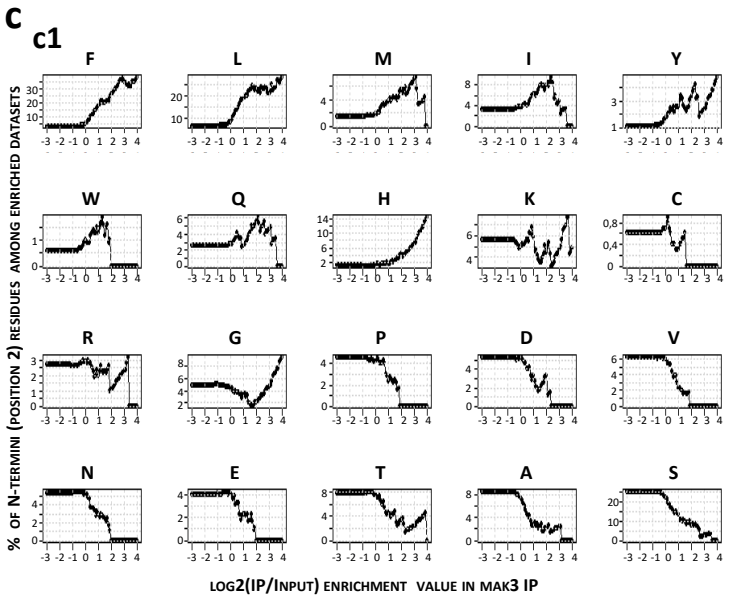
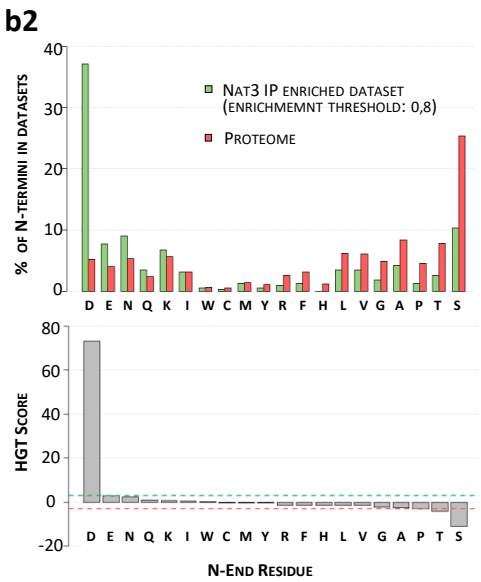
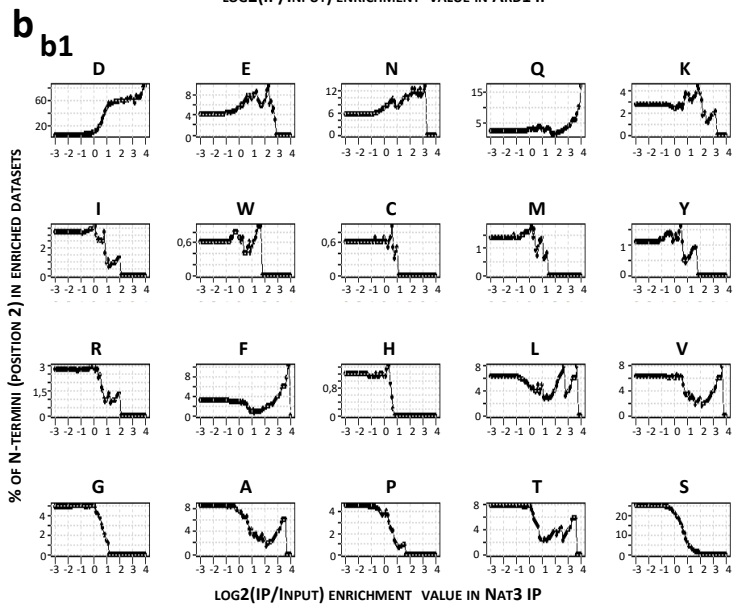
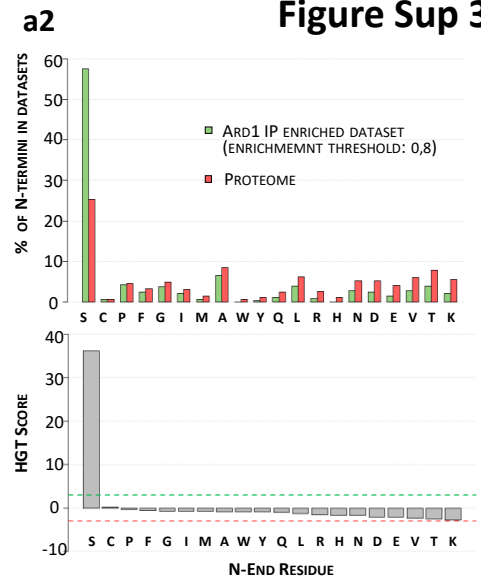
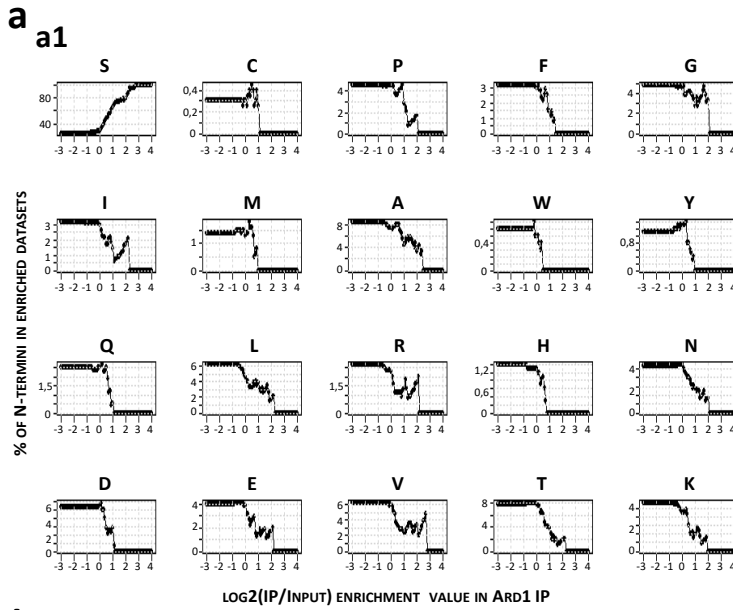
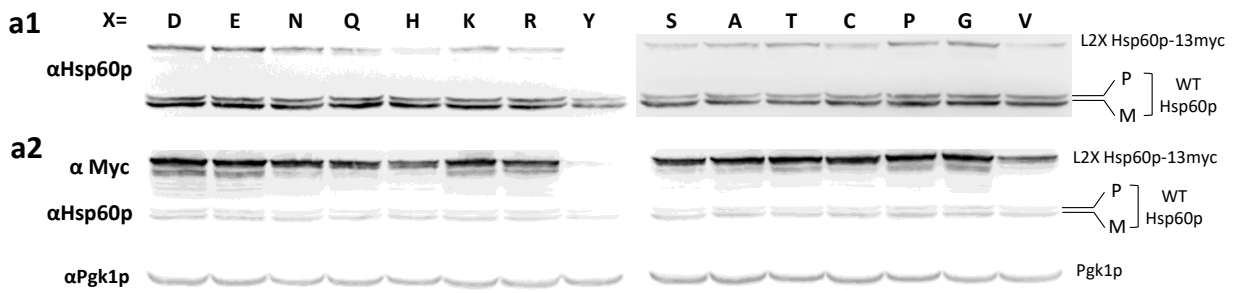


Fig Sup 3: Analysis of the selectivity of NatA, NatB and NatC toward proteins N-termini using the sel-TRAP technology.

Global analysis of mRNAs copurified after immunoprecipitations (IP) from cells expressing Ard1-PA (a), Nat3-PA (b) and Mak3-PA (c) allowed the indirect characterization of nascent protein chains translated by ribosomes associated respectively with NatA, NatB and NatC. For each cellular protein, the corresponding mRNA enrichment value, defined as IP/input $\log_2(\text{ratio})$, was calculated after microarray hybridizations of cDNAs reverse-transcribed from immunoprecipitation and input samples. Since NATs are known to target proteins according to the nature of their N-terminus, the left panels (a1, b1, c1) show the 20 curves representing, for all possible residues at position 2, the evolution of their proportion associated with the subsets of mRNAs detected with increasing enrichment values in IP experiments. After this analysis, an enrichment value of 0.8 was used as a threshold to define enriched mRNAs identified with sel-TRAP experiments of NatA, NatB and NatC. The percentages of each residue at position 2 among the identified targets were compared to those observed in the proteome (a2, b2, c2: upper panels) and the corresponding HGT scores were calculated (a2, b2, c2: lower panels) for statistical evaluation of significant differences. Dotted lines indicate the HGT score cutoff of 3 corresponding to a p-value obtained by a hypergeometric test equal to 10^{-3} . This analysis of the sel-TRAP data revealed the specific co-translational selectivity of NatA for serine (S), NatB for aspartate (D), and NatC for phenylalanine (F) and leucine (L).

Figure Sup 4

a



b

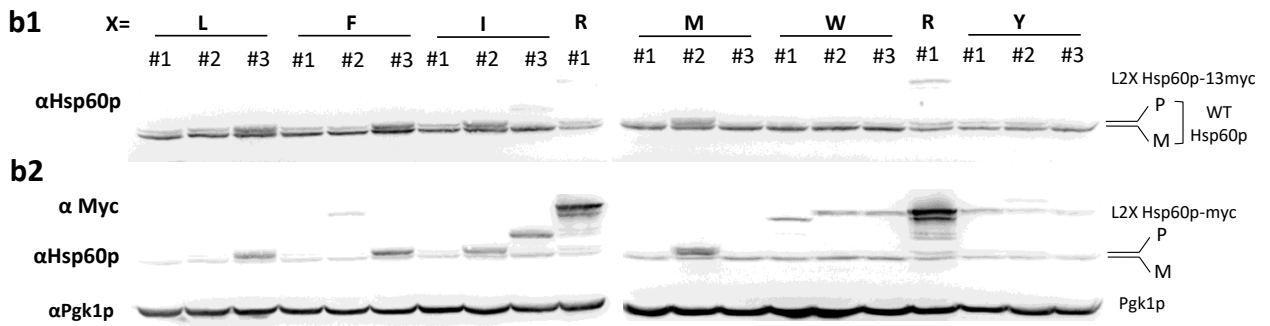


Fig Sup 4: Western blot analysis of protein content in clones obtained after transformation with the diverse pHSPmyc(X) plasmids.

Protein extracts were analyzed by SDS-PAGE followed by western blots using antibodies against myc epitope, Hsp60p, and Pgk1p. The position of the precursor and mature Hsp60p are indicated.

(a) Analysis of clones transformed with plasmids that allow the rescue of the dominant negative toxicity of the Hsp60p-13myc protein (X=D, E, N, Q, H, K, R, Y, S, A, T, C, P, G, V, see Fig 5).

(b) Analysis of clones transformed with plasmid expressing a toxic Hsp60p-13myc protein (X= L, F, I, M, W, see Fig 5) and of those obtained with the pHSP60(Y) for which a longer exposure time was necessary to detect the Hsp60p-13myc protein (see Fig. Sup3a).

SUPPLEMENTARY METHODS

Table S5: Yeast strains

Strain	Genotype	Origin
BY4741	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0	Euroscarf (Brachmann et al,1998)
Ard1-PA	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0, ARD1-ProtA-His5	This Study* ¹
Nat3-PA	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0, NAT3-ProtA-His5	
Mak3-PA	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0, MAK3-ProtA-His5	
YPH499	MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1	Mehawej et al, PLoS Genet, 2014
YPH499 pam16Δ-MAGN76D	MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ2 pma16Δ + pmagMAT N76D	

*¹Strains tagged at their C terminus were obtained from BY4741 strain, after classical procedure of lithium acetate transformation, by homologous recombination with ProtA-His5 cassette amplified from pBXA (Rout et al., 2000, provided by M. Rout, The Rockefeller, University, New York, NY)

Table S6: Plasmids and derived strains

YPH99 or YPH499 pam16Δ-MAGN76D strain were transformed with the following plasmids.

Plasmids (Origin)	Inserted sequence	Experiment
pAEF-HSP60 (This Study * ²)	Cas9 sequence HSP60 Guide Sequence	CRISPR/CAS9 mutagenesis of HSP60 genomic locus
pZymA7 (32)	ZWF1 promotor ASN1(1-1717)-13myc PGK1 terminator	Study of the impact of L2X mutations on the dominant negative toxicity of the Hsp60p-13myc protein
pHSPMyc(X)(This study* ³)	HSP60 Promotor HSP60 L2X-13myc PGK1 terminator	

*² The plasmid pAEF-HSP60 contains the guide RNA sequence targeting HSP60 and the sequence coding for the Cas9 endonuclease. It allows to cut the genomic sequence of HSP60 to obtain the desired mutation of the second residue of HSP60 by homologous recombination with an ad hoc repair cassette (see table S3).

*³X represents the amino acid that replaces leucine at position 2 of Hsp60p. This modification was first introduced into the *S. cerevisiae* genome with CRISPR/CAS9 technology using HSP60 repair cassettes that include the desired mutations (see Table S3 and S4). The mutated HSP60 sequences, including its promoter (680 bp upstream of the initiator codon), were then amplified by PCR (see oligonucleotides used in Table S3) and inserted between the SacI and PacI restriction sites into the plasmid pZMYA7 in fusion with the 13myc sequence.

Table S7: Oligonucleotides

Oligonucleotide Name	Sequence 5' → 3'	
HSP60 guide for	<u>ATC</u> AGTAGCGCGACTACGAACAA	CRISPR/CAS9 mutagenesis of HSP60* ⁴
HSP60 guide rev	<u>AAC</u> TTGTTTCGTAGTCGCGCTACT	
K7HSP60(X)_fw	ACATCATAAGCAAAAAAGTTTTCAAAA <u>ATG</u> YYYAGATCATCA <u>A</u> GTTG TTCGTAGTCGCGCTACTTTAAGGCCTTATTGCGTCGTGCTTACT	
K7HSP60(X)_rev	AGTAAGCACGACGCAATAAAGGCCTTAAAGTAGCGCGACTACGA ACAAC <u>T</u> GATGATCT <u>ZZZ</u> CATTTTGAAAACTTTTTGCTTATGATGT	
SacI_HSP60(-800)_for	AGCCACAT <u>GAGCTC</u> GCTCAGCAGCTTACGTTCC _{Sac}	PCR amplification of HSP60 sequence for insertion in pZMYA7
PacI_HSP60(stop)_rev	AGCCACATTT <u>TAATTA</u> ACATCATACCTGGCATTCTGGCATAAC _{Pac}	

*4: The forward and reverse HSP60 guide oligonucleotides were hybridized and the resulting dsDNA was inserted in Lg_{ul} site (Lg_{ul} extensions underlined in sequences) of the pAEF5 plasmid (67). The resulting plasmid (pAEF HSP60) was co-transformed in yeast (YPH499 or YPH499 pam16Δ-MAGN76D strain) with the repair cassette previously obtained by hybridization of the forward and reverse oligonucleotides K7HSP60. In these oligonucleotides, the codon following the initiator methionine (underlined in the sequence) was replaced with codons encoding the desired X mutation (indicated by YYY/ZZZ in the sequence, see Table S4 for the codon chosen for each mutation). In the repair cassette, a synonymous mutation was introduced into the HSP60 sequence to eliminate the Pam site targeted by the CAS9 endonuclease with the chosen guide RNA sequence.

Table S8: Codon used in the reparation cassette to encode the desired X mutation.

The selected codons are not rare codons in *S. cerevisiae*.

X	Amino Acid	YYY	ZZZ
L	Leucine	TTA	TAA
N	Asparagine	AAT	ATT
K	Lysine	AAG	CTT
T	Threonine	ACT	AGT
S	Serine	TCC	GGA
R	Arginine	AGA	TCT
I	Isoleucine	ATT	AAT
M	Methionine	ATG	CAT
H	Histidine	CAT	ATG
Q	Glutamine	CAA	TTG
P	Proline	CCT	AGG
D	Aspartic Acid	GAT	ATC
E	Glutamic Acid	GAA	TTC
A	Alanine	GCT	AGC
G	Glycine	GGT	ACC
V	Valine	GTC	GAC
Y	Tyrosine	TAT	ATA
C	Cysteine	TGT	ACA
W	Tryptophan	TGG	CCA
F	Phenylalanine	TTT	AAA

Table S9: List of the yeast species used for the genomic comparative studies

<i>Saccharomyces cerevisiae</i>
<i>Saccharomyces paradoxus</i>
<i>Saccharomyces bayanus</i>
<i>Candida glabrata</i>
<i>Zygosaccharomyces rouxii</i>
<i>Eremothecium gossypii</i>
<i>Kuyveromyces lactis</i>
<i>Lachancea thermotolerans</i>
<i>Lachancea kluyveri</i>
<i>Debaryomyces hansenii</i>
<i>Scheffersomyces stipitis</i>
<i>Clavispora lusitaniae</i>
<i>Ogataea polymorpha</i>
<i>Komagataella pastoris</i>
<i>Yarrowia lipolytica</i>
<i>Blastobotrys adenivorans</i>
<i>Geotrichum candidum</i>

GO reduction algorithm

Definitions

Let G_i be the ensemble of genes from the protein-coding fraction \mathcal{P} associated with the GO term GO_i , such that $G_i = \{g \in \mathcal{P} \mid g \sim GO_i\}$, where g is a gene and the symbol \sim denotes the association between a gene and a GO term. We define G_i^a as the subset of genes from G_i whose resulting proteins display the amino acid a in position 2. We evaluate the representativity of a GO term with respect to a according to the following score,

$$s_i^a(w) = w_N \frac{|G_i^a|}{\max_j(|G_j^a|)} + w_F \frac{f_i^a}{\max_j(f_j^a)}, \quad (1)$$

where $f_i^a = \frac{|G_i^a|}{|G_i|}$ is the frequency of occurrence of a in the second position of the proteins resulting from the genes included in G_i , and w_N and w_F are weights controlling the relative contributions of the two terms such that $w_N + w_F = 1$. The first term reflects how well the GO term GO_i covers the proteins displaying a in position 2. The GO term associated with the maximum number of such proteins has the maximum value of 1. The second term reflects how specific the GO annotation is for the amino acid a . The lower the variability of amino acids in position 2 among the proteins associated with a given GO term, the higher this value.

Initialization: Pre-selection of GO terms

We start from 20 sets of pre-selected GO terms corresponding to the 20 amino acids. For each amino acid a , the N_a selected GO terms are those displaying an enrichment for a with a p-value $p - val \leq 10^{-4}$, and also for which $f_i^a > 1.8f^a$, where f^a is the frequency of occurrence of a at position 2 in the proteome. Notice that some sets may be empty.

First step: identifying the most inclusive and selective GO terms for each amino acid

The goal of this first step is to define, for each amino acid, a minimal set of GO terms representative of the full pre-selected set, both in terms of coverage and of specificity. We first order the GO terms according to the score expressed in Eq. 1, from the highest to the lowest score. We set the weight $w_N = w_F = 0.5$ so that coverage and specificity contribute equally to the score. Then, we apply the following Algorithm 1, independently for each amino acid a . The algorithm considers the N_a GO terms pre-selected for a , from the highest-scored one to the lowest-scored one, and decides whether each term should be retained or discarded based on its gene overlap with the previously retained terms. At each iteration of the algorithm, we require that the GO term considered shares less than $\alpha = 40\%$ of its associated genes with the current GO term list. If this criterion is not met, then the GO term is disregarded.

Second step: reducing redundancy at the global level while prioritizing coverage

While the first step treats each amino acid independently from the others, this second step aims at reducing redundancies across all amino acids. More precisely, for each amino acid a , we look for the GO terms not included in its *bestNGO* list but present in the *bestNGO* list of at least one other amino acid b and displaying an enrichment for a with a p-value $p - val \leq 10^{-3}$. We extend the *bestNGO* list of a with these GO terms. They are of interest because they display high enrichments for several amino acids and they potentially cover more genes than the terms selected in the previous step. To allow for them to be retained, and potentially replace GO terms with lower coverage, we again apply Algorithm 1 for each amino acid setting $w_N = 1$ and $w_F = 0$.

Algorithm 1 Redundancy reduction applied to a list of pre-selected GO terms

Input: an ordered list of GO terms $[GO_1, GO_2, \dots, GO_{N_a}]$ and the value of α **Output:** a reduced list of GO terms

```
1:  $bestNGO \leftarrow [GO_1]$ 
2:  $bestN \leftarrow G_1^a$ 
3: for  $i \leftarrow 2$  to  $N_a$  do
4:    $overlap \leftarrow |bestN \cap G_i^a|$ 
5:   if  $overlap \leq \alpha|G_i^a|$  then
6:     Add  $GO_i$  to  $bestNGO$ 
7:      $bestN \leftarrow bestN \cup G_i^a$ 
8:   end if
9: end for
10: Output the list  $bestNGO$ 
```

Algorithm 2 Rescue of GO terms from a pre-filtered list

Input: an ordered list of GO terms $[GO_1, GO_2, \dots, GO_{n_a}]$ and the value of β **Output:** a reduced list of GO terms

```
1:  $bestFGO \leftarrow [GO_1]$ 
2: for  $i \leftarrow 1$  to  $n_a$  do
3:    $c \leftarrow 0$ 
4:   for  $j \leftarrow 1$  to  $length(bestFGO)$  do
5:      $overlap \leftarrow |G_j^a \cap G_i^a|$  #  $j$  and  $i$  are the indices of GO terms from  $bestFGO$ 
6:                                     and from the input list, respectively
7:     if  $|G_j^a| \leq |G_i^a|$  then
8:        $overlapRel \leftarrow overlap/|G_i^a|$ 
9:     else
10:       $overlapRel \leftarrow overlap/|G_j^a|$ 
11:    end if
12:    if  $overlapRel \leq \beta$  then
13:       $c \leftarrow c + 1$ 
14:    end if
15:  end for
16:  if  $c = length(bestFGO)$  then
17:    Add  $GO_i$  to  $bestFGO$ 
18:  end if
19: end for
20: Output the list  $bestFGO$ 
```

Third step: rescuing highly selective GO terms

This goal of this third step is to rescue the GO terms displaying a very high selectivity for a particular amino acid. Here, for each amino acid a , we start from the n_a GO terms pre-selected in the initialization phase but not retained in the $bestNGO$ list. We order them according to the score expressed in Eq. 1, from the highest to the lowest score, with the weights set to $w_N = 0$ and $w_F = 1$. The rationale is to prioritize the most selective GO terms. Then, we apply the following Algorithm 2. At each iteration, we require for a GO term to be retained that it shares less than $\beta = 30\%$ of its

associated genes with each one of the GO terms in the current list.

Fourth step: reducing redundancy at the global level while prioritizing specificity

Similarly to what is done in the second step, the goal of this fourth step is to reduce redundancies across all amino acids. We consider the nF GO terms appearing in at least one of the *bestFGO* lists defined in the previous step. For each of these GO terms, we identify the amino acids for which it displays an enrichment with a $p - val \leq 10^{-4}$. This operation leads to nF potentially overlapping subsets of amino acids. The subsets with only one member are discarded. We encode this information in a graph where the nodes are the 20 amino acids and the set of edges is defined from the subsets.

Algorithm 3 Build a graph representing the overlaps between GO terms enriched for a set of amino acids

Input: a list $[GO_1, GO_2, \dots, GO_m]$, a list of amino acids $[a_1, a_2, \dots, a_l]$, and the value of β

Output: a graph

```

1: Initialize the graph  $\mathcal{G} = (\mathcal{V}_G, \mathcal{E}_G)$  with  $\mathcal{V}_G = \{GO_1, GO_2, \dots, GO_m\}$ 
2: for  $i \leftarrow 1$  to  $m - 1$  do
3:   for  $j \leftarrow i + 1$  to  $m$  do
4:      $c \leftarrow c + 1$ 
5:     for  $k \leftarrow 1$  to  $l$  do
6:        $overlap \leftarrow |G_i^{a_k} \cap G_j^{a_k}|$ 
7:       if  $|G_i^{a_k}| \leq |G_j^{a_k}|$  then
8:          $overlapRel \leftarrow overlap / |G_j^{a_k}|$ 
9:       else
10:         $overlapRel \leftarrow overlap / |G_i^{a_k}|$ 
11:      end if
12:      if  $overlapRel \geq \beta$  then
13:         $c \leftarrow c + 1$ 
14:      end if
15:    end for
16:    if  $c = l$  then
17:      Add an edge between  $GO_i$  and  $GO_j$ 
18:    end if
19:  end for
20: end for
21: Output the graph  $\mathcal{G}$ 

```

More specifically, any pair of nodes is linked by an edge if the corresponding amino acids were found in the same subset at least once. We further define groups of amino acids as the connected components of the graph. Each group $\{a_1, a_2, \dots, a_l\}$ is associated with a set of GO terms $\{GO_1, GO_2, \dots, GO_m\}$ coming from the *bestFGO* lists. We then build a graph encoding the gene overlaps between these GO terms by applying Algorithm 3. For each connected component in the graph, we determine a representative GO term GO_{imax} as the most selective one for the amino acid group. Formally, we determine $\arg \max_i (f_i^{a_1} + f_i^{a_2} + \dots + f_i^{a_l})$. The final operation consists in replacing all the GO terms appearing in the *bestFGO* lists defined for the amino acids a_1, a_2, \dots, a_l by their representative GO_{imax} .

Mass spectrometry analysis

Digestion was performed overnight at 37°C in the presence of 12.5 µg/ml of sequencing grade trypsin (Promega, Madison, WI, USA). Peptides mixtures were analyzed by a Q-Exactive Plus coupled to a Nano-LC Proxeon 1000 (both from Thermo Scientific). Peptides were separated by chromatography using the following conditions: Acclaim PepMap100 C18 pre-column (2 cm, 75 µm i.d., 3 µm, 100 Å), Pepmap-RSLC Proxeon C18 column (50 cm, 75 µm i.d., 2 µm, 100 Å), 300 nl/min flow rate, a 98 min gradient from 95 % solvent A (water, 0.1 % formic acid) to 35 % solvent B (100 % acetonitrile, 0.1% formic acid) followed by column regeneration, giving a total time of 120 minutes. Precursor peptides were analyzed in the Orbitrap cell in positive mode, at a resolution of 70,000, with a mass range of m/z 375-1500 and an AGC target of $3 \cdot 10^6$. MS/MS data were acquired in the Orbitrap cell in a Top20 data-dependent mode with a dynamic exclusion of 30 seconds. Fragments were obtained by Higher-energy C-trap Dissociation (HCD) activation with a collisional energy of 27% and a quadrupole isolation window of 1.4 Da. The Orbitrap cell was set at a resolution of 17,500, m/z 200-2000 and an AGC target of $2 \cdot 10^5$. Peptides with unassigned charge states or monocharged were excluded from the MS/MS acquisition. The maximum ion accumulation times were set to 50 ms for MS acquisition and 45 ms for MS/MS acquisition.

Data were processed with Proteome Discoverer 2.2 software (Thermo Fisher scientific, San Jose, CA) coupled to an in-house Mascot search server (Matrix Science, Boston, MA; version 2.5.1). MS/MS spectra were searched against the SwissProt protein database release 2017_09 with the *Saccharomyces Cerevisiae* (*baker's yeast*) taxonomy and a maximum of 2 missed cleavages. Precursor and fragment mass tolerances were set to 6 ppm and 0.02 Da respectively. The following post-translational modifications were included as variable: Acetyl (Protein N-term), Oxidation (M), Phosphorylation (STY). Spectra were filtered using a 1% FDR with the percolator node.

Résultat complémentaires non inclus dans l'article



Figure 28 : Western Blot des souches mutantes au niveau du deuxième acide aminé pour la protéine Hsp60. Détection de la protéine Hsp60 pour 2 clones avec comme deuxième acide aminé une phénylalanine (L2F) ou 4 clones avec comme deuxième acide aminé une Sérine (L2S). La souche N76D correspond à la souche contrôle. Je n'ai pas réalisé de témoin de charge.

II. Exploration fonctionnelle de l'impact de NatC sur les précurseurs mitochondriaux

1. Préambule

Comme vous avez pu le voir dans l'article, nos résultats suggèrent que l'acétylation N-terminale des MTS par l'enzyme NatC pourrait être une des raisons de la sélection des résidus hydrophobes en position 2 de ces MTS. En effet, leucine, phénylalanine, tryptophane et isoleucine en position 2 sont connus pour permettre le recrutement co-translationnel de cette enzyme, et nous avons par ailleurs confirmé par des expériences de purification sélective des ribosomes que les précurseurs mitochondriaux possédant un MTS sont les cibles principales de NatC. Il est alors tentant d'imaginer que l'acétylation N-terminale des MTS en neutralisant la charge positive de leur amine N-terminale augmente leur hydrophobicité et donc leur potentiel d'adressage mitochondrial.

Par ailleurs, d'après la littérature, l'enzyme NatC est impliquée dans une voie de dégradation des protéines dépendante du protéasome et appelée la AC/N-end rule (cf chapitre II.2.a de l'introduction). Dans cette voie, l'acétylation N-terminale par NatC de protéines ayant en position 2 un résidu hydrophobe de grande taille induit leur reconnaissance par les ubiquitines ligases Doa10 et Not4 (Lee et al., 2016). Ces dernières ubiquitinent les protéines acétylées qui sont adressées au protéasome pour dégradation. On peut donc émettre l'hypothèse que NatC joue un rôle sur la stabilité des précurseurs mitochondriaux. En particulier on peut imaginer que dans un contexte de perturbation de l'import mitochondrial, l'acétylation des précurseurs qui s'accumulent dans le cytoplasme permet leur dégradation rapide. Un tel système éviterait la toxicité causée par cette accumulation.

Ainsi, les expériences présentées ci-dessous ont été motivées par deux hypothèses non exclusives, à savoir que NatC pourrait jouer un rôle dans l'import mitochondrial des précurseurs possédant un MTS, ou un rôle dans la détermination de leur temps de demi-vie via la AC/N-end rule.

2. La perte du génome mitochondrial (*rho0*) est un facteur de confusion dans l'interprétation des mutants *NatC*

J'ai souhaité déterminer si la délétion de *NatC* induit une accumulation de précurseurs mitochondriaux, en particulier celle de *Hsp60*, dans un contexte où l'import mitochondrial est déficient. J'ai donc commencé par générer des souches déplétées de *NatC*. Pour cela, nous avons décidé de déléter sa sous-unité catalytique *Mak3*. Cette délétion a été réalisée dans le fond génétique *YPH-pam16Δ-hpam16(N76D)* et a permis l'obtention de trois clones : *N76D Mak3Δ #21*, *#22* et *#23*. J'ai étudié leur phénotype moléculaire par Western Blot en utilisant des anticorps dirigés contre *Hsp60* (figure 28).

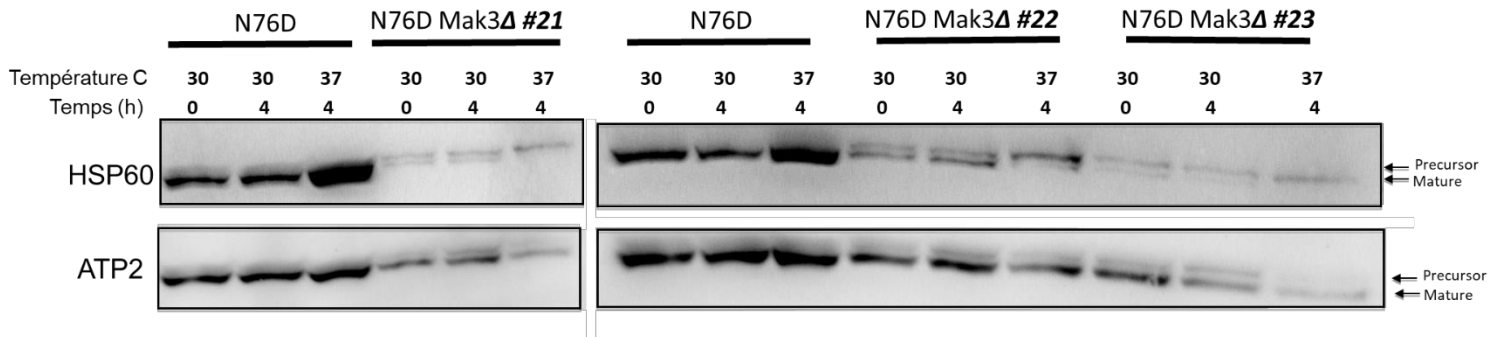


Figure 29 : Western blot des souches déléetées de *NatC* dans le fond génétique *YPH-pam16Δ-hpam16(N76D)* (*N76D Mak3Δ #21*, *N76D Mak3Δ #22* et *N76D Mak3Δ #23*). Détection des protéines *Hsp60* et *Atp2* endogènes avec des anticorps anti *Hsp60* et anti *Atp2*. La culture cellulaire se fait sur milieu riche (YPD) liquide et les protéines sont extraites après quatre heures d'incubation à 30 °C ou 37 °C.

La souche notée *N76D* dans la figure 29 correspond à *YPH-pam16Δ-hpam16(N76D)* qui est notre souche contrôle. Dans ce Western Blot, nous pouvons voir qu'à 30°C nous ne pouvons observer qu'une seule bande correspondant à la forme précurseure mature de la protéine *Hsp60* pour la souche contrôle. Nous obtenons le même résultat pour *Atp2*. Cependant, après 4 heures d'incubation à 37°C lorsque l'import mitochondrial est perturbé, nous nous attendions à ce qu'il y ait une accumulation visible des précurseurs de *Hsp60* dans la souche *N76D*, qui se caractériserait par l'apparition d'une deuxième bande plus haute. Ce n'est pas le cas ici. Ce résultat n'est pas cohérent avec ce que nous avons réussi à obtenir de nombreuses fois avec cette souche (voir les western blot présentés dans l'article au chapitre précédent ou encore la Figure 34 au chapitre 3b ci-dessous). Une raison pourrait en être la mauvaise résolution du

western blot couplée à la forte intensité de la bande Hsp60 à 37°C dans le contrôle N76D, qui pourrait masquer la distinction entre la bande mature et la bande précurseur.

On peut par contre observer dans les souches délétées de NatC une accumulation des précurseurs à 30 et 37°C. Cette accumulation est visible de par l'apparition d'une bande plus haute que la bande correspondant à la forme mature de Hsp60. Les deux bandes, correspondant à la forme précurseur et à la forme mature, sont visibles à 30°C et seule la bande correspondant à la forme précurseur est visible à 37°C. Ce résultat semblait confirmer notre hypothèse selon laquelle NatC a un impact sur le devenir des précurseurs mitochondriaux, soit en influençant leur stabilité, soit en augmentant leur efficacité d'import. Cependant, en plus de notre protéine modèle Hsp60, nous avons voulu voir comment se comportait une autre protéine mitochondriale, Atp2, qui possède un MTS mais qui présente une valine en position 2, un résidu qui n'est pas ciblé par NatC. Pour cette protéine, on peut identifier clairement la bande mature dans toutes les conditions et une deuxième bande qui pourrait correspondre à la forme précurseur d'Atp2, présente uniquement pour les souches mutées pour NatC à 30 et 37°C. C'était un résultat surprenant étant donné que Atp2 n'a pas un résidu en position 2 qui est connu pour être ciblé par NatC. Cette absence de reconnaissance par l'enzyme est d'ailleurs confirmée par le sel-TRAP. Ceci suggérait donc que le phénotype moléculaire observé dans les souches délétées pouvait être lié à d'autres raisons qu'à la simple inactivation de NatC. En manipulant ces trois souches, nous nous sommes très vite rendus compte que leurs DO n'atteignent pas une valeur de 1 après une culture de nuit en milieu riche glucose, ce qui n'est pas un phénotype attendu pour les souches *mak3*. J'ai alors décidé de caractériser le phénotype de croissance de ces souches sur milieu riche et sur milieu contenant du glycérol afin de tester leur capacité respiratoire (figure 30).

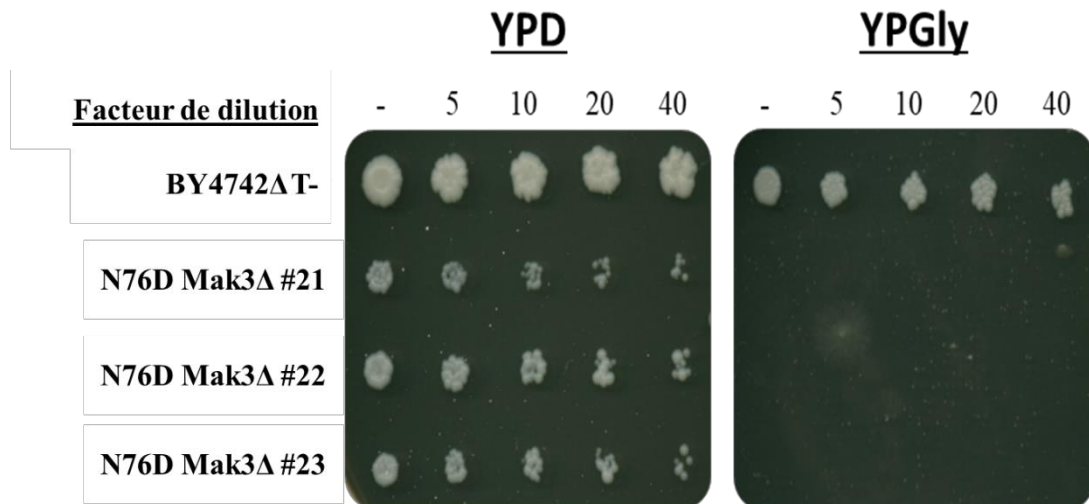


Figure 30 : Test de croissance en gouttes de souches mutantes de *NatC* sur milieu riche (YPD) et sur milieu avec une source de carbone non fermentable. La capacité respiratoire des souches est mesurée à 30°C par un test en gouttes après 3 jours d’incubation. Le phénotype de croissance des souches *N76D-mak3Δ* est comparé à celui d’une souche contrôle *BY4742 mak3Δ*.

On peut observer que les souches délétées de *MAK3* dans le fond N76D sont incapables de pousser sur un milieu contenant une source de carbone non fermentable. Il n’a pas été démontré dans la bibliographie et dans les expériences réalisées antérieurement par le laboratoire que la délétion de *MAK3* pouvait donner un phénotype aussi grave sur un milieu contenant une source de carbone non fermentable (voir le contrôle *BY4742 mak3Δ* sur la figure 30 par exemple). L’une des hypothèses que nous avons émises pour expliquer ce résultat est que la souche est une *Rho°*, c’est-à-dire qu’elle a perdu son génome mitochondrial. En effet, il est connu dans la littérature que les souches *Rho°* ont un défaut d’import mitochondrial, une croissance limitée sur milieu riche, forment de petites colonies et ne poussent pas sur un milieu avec une source de carbone non fermentable (Zubko and Zubko, 2014). Cela pouvait expliquer nos deux observations : accumulation de la forme précurseure de Hsp60 dès 30°C et pas de pousse sur YPGly. Pour tester cette hypothèse, nous avons décidé de faire un test de complémentation fonctionnelle (figure 31). Cela a été rendu possible grâce à des collaborateurs qui nous ont conseillé de réaliser ce test avec la souche *D160 Rho°*, ayant perdu le génome mitochondrial, qui nous a été gracieusement donnée par Cristina Panozzo. Nous avons croisé nos souches à tester avec cette souche *D160 Rho°*. Si les cellules diploïdes issues du croisement ne sont toujours pas capables de pousser sur du milieu avec du glycérol, alors cela montrerait que le

phénotype observé pour les souches testées est bien dû à un état Rho°. Nous avons aussi utilisé comme témoin positif une souche rho+ délétée de *ATP12* au niveau génomique qui possède également un phénotype d'arrêt de croissance sur un milieu avec une source de carbone non fermentable, mais dont le génome mitochondrial est intact. Avec cette souche, dans le cas d'un croisement avec un mutant Rho°, on devrait obtenir un sauvetage phénotypique chez les diploïdes puisque la souche *atp12Δ* devrait apporter des mitochondries sauvages et les souches testées un locus *ATP12* fonctionnel.

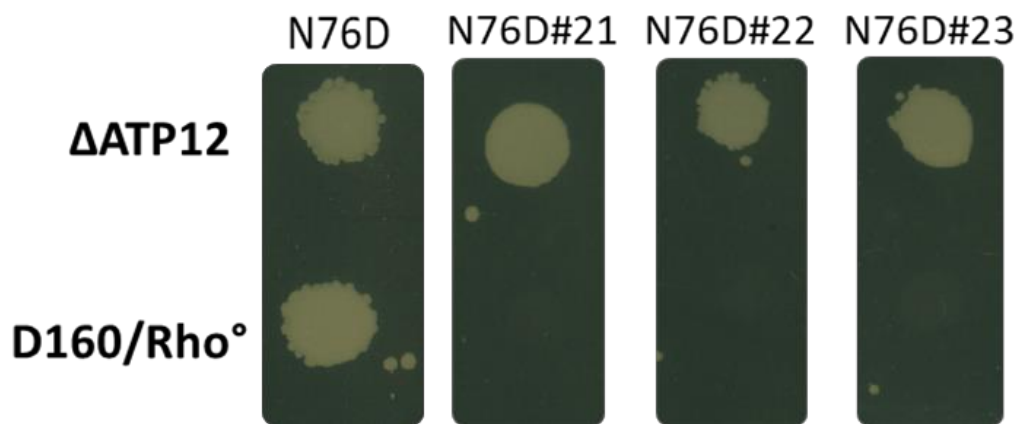


Figure 31 : Test de vérification des souches Rho°. Les souches suspectées d'avoir perdu leur génome mitochondrial sont croisées avec une souche contrôle *Δatp12* ou avec une souche Rho° avérée (*D160*) sur milieu riche pendant une nuit. Les souches sont ensuite repiquées sur milieu sélectif (YPGly) sur lequel est déposée une goutte de cellules dans 100 μ l d' H_2O (YPGly). Les photographies sont prises après un jour d'incubation.

Lors des croisements de nos mutants *N76D Mak3Δ #21, 22 et 23* avec la souche *D160* (Rho°, prototrophe) nous n'avons pas pu obtenir de pousse d'un diploïde sur milieu YPGly. Une expérience contrôle avec la souche *N76D* parentale a comme attendu permis d'obtenir des diploïdes capables de se développer sur YPGly (la pousse du parent testé ayant été préalablement éliminée sur la base de ses auxotrophies). Au contraire, le croisement des mutants *N76D Mak3Δ #21, 22 et 23* avec la souche *ATP12Δ* a conduit à un sauvetage fonctionnel de nos mutants. L'ensemble de ces résultats confirme que les souches *mak3Δ* qui présentaient des défauts de croissance et une accumulation massive de précurseurs mitochondriaux *Hsp60* et *Atp2* correspondent en définitive à des mutants Rho° ayant perdu leur génome mitochondrial.

Ces mutants ne permettent donc pas de tirer des conclusions sur l'impact de l'absence de NatC sur la stabilité ou l'import des précurseurs mitochondriaux puisque nous ne disposons pas des souches contrôles correspondantes (mutant Rho^o sans délétion de *MAK3* dans le même contexte génétique).

Cependant, ce résultat a été très important pour la suite de ma thèse. En effet, nous nous sommes rendus compte que notre souche *YPH-pam16Δ-hpam16(N76D)* présente une instabilité mitochondriale et est donc susceptible de générer des souches Rho0. Ainsi, à chaque fois que j'ai manipulé ou généré des souches ayant ce fond génétique, j'ai dû m'assurer qu'elles avaient conservé leur génome mitochondrial.

3. Construction et caractérisation de souches rho+ mutées pour NatC

J'ai ensuite généré deux autres clones déplétés de NatC, toujours en déléant sa sous-unité catalytique Mak3. J'ai généré ces souches sur deux fonds génétiques différents : le fond génétique *YPH499* et le fond génétique *YPH-pam16Δ-hpam16(N76D)*. Avant d'étudier ces souches, je me suis assuré qu'elles avaient bien conservé leur génome mitochondrial en réalisant un test de croissance sur YPGly.

a) Phénotype de croissance ralentie pour les souches *Δmak3*

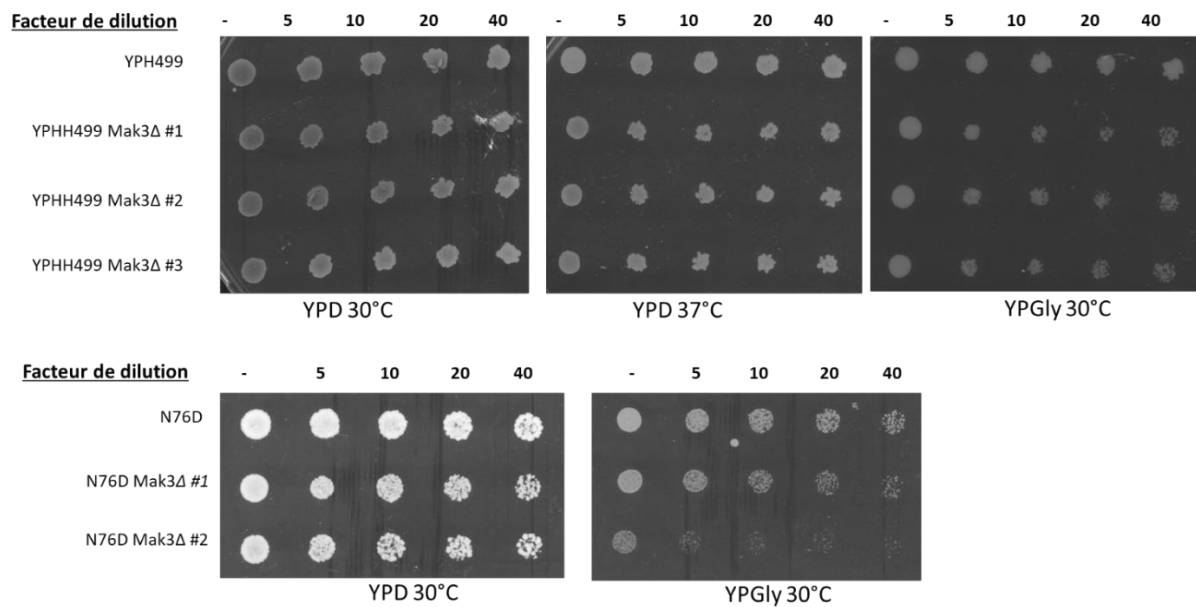
J'ai d'abord mesuré le temps de génération des différentes souches sur milieu riche glucose ou glycérol à 30 ou à 37°C (figure 32.A). Pour la souche *YPH499*, il semblerait que les mutants de NatC aient un phénotype de croissance ralentie à 37°C mais la différence n'est pas assez importante pour conclure avec certitude. En revanche, pour les deux souches déplétées de NatC dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*, la légère différence est visible dès 30°C. Je n'ai pas pu tester la souche *YPH-pam16Δ-hpam16(N76D)* à 37°C car son défaut d'import mitochondrial est trop marqué à cette température pour permettre la croissance.

Le léger phénotype de croissance observé en glucose est accentué quand le glycérol est la seule source de carbone, et ce dans les deux fonds génétiques (*YPH499* ou *YPH-pam16Δ-hpam16(N76D)*). Il est à noter que les deux clones mutants du fond N76D ne se comportent pas de la même manière sur YPGly. Le clone numéro 2 a un défaut de croissance plus important que le clone 1 (Voir figure 32.A).

Dans la littérature, des études de croissance des souches mutantes de NatC réalisées avec des tests en gouttes n'ont pas permis de détecter de phénotype de croissance sur milieu riche glucose (Polevoda, 1999; Polevoda and Sherman, 2001). Cependant, une étude récente a démontré que

dans le fond génétique *BY4741* un très léger phénotype de croissance ralentie pouvait être mesuré avec un test de croissance en liquide (Drazic and Varland, 2021). Ainsi, afin de mieux identifier le phénotype observé sur milieu glucose, j'ai décidé de mesurer le temps de génération des différentes souches en liquide sur milieu riche. Cette technique permet de distinguer plus facilement les petites différences de temps de génération entre les souches par rapport au test sur goutte (Figure 32.B) On retrouve bien un phénotype de croissance ralentie pour les mutants de NatC dans le fond *YPH-pam16Δ-hpam16(N76D)* mais on remarque également le même phénotype dans le fond *YPH499* (Figure 32.B). J'ai mesuré le temps de génération des souches mutantes à partir de deux réplicats. On peut voir que dans le fond *YPH499*, on observe une augmentation du temps de génération d'environ 15% des souches mutantes par rapport au contrôle. Cette augmentation est d'environ 24% dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*. Faute de ne pas avoir assez reproduit cette expérience, je n'ai pas pu faire un test statistique assez robuste pour déterminer si la différence observée est significative. On peut néanmoins noter que les résultats que j'ai obtenus pour l'instant vont dans le sens de ce qui est publié pour les mutants *mak3*, à savoir un léger défaut de croissance en milieu glucose et un défaut plus sévère en glycérol.

A



B

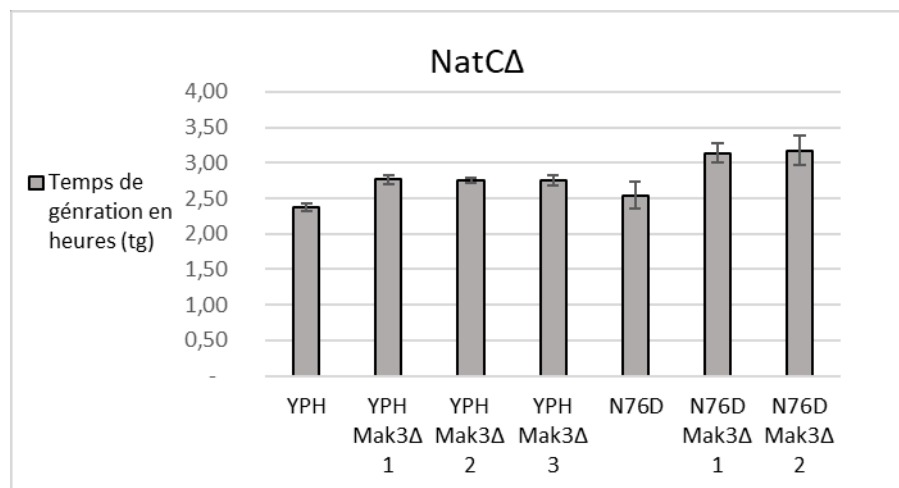


Figure 32 : (A) Test de croissance en gouttes de souches mutantes dans la souche *YP499* et *YPH-pam16Δ-hpam16(N76D)* de *NatC* sur milieu riche à 30 et 37°C et à 30°C avec une source de carbone non fermentable YPGly. (B) Histogramme représentant les temps de génération des souches mutantes de *NatC* dans les fond génétique *YPH499* et *YPH-pam16Δ-hpam16(N76D)*. Les temps de génération sont calculés à partir de deux répliquas d'un test de croissance en liquide sur milieu riche (YPD) à 30°C. La DO est mesurée toutes les heures pendant 7 heures.

b) Phénotype moléculaire des souches délétées de NatC

Nous avons ensuite voulu voir quel est le phénotype moléculaire de la protéine Hsp60 dans un contexte dans lequel NatC a été délétée. La première étude a été réalisée dans le fond YPH499, à 30 et à 37°C. En effet, nous avons montré précédemment que nos 3 clones délétés de NatC présentaient un phénotype de croissance ralentie à 37°C, même en milieu riche glucose (Figure 32.A).

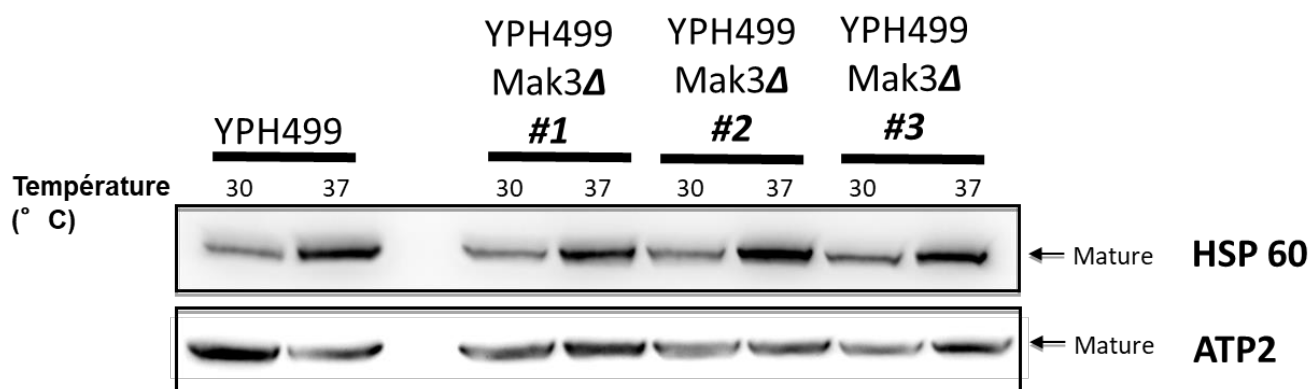


Figure 33 : Western blot des souches délétées de NATC dans le fond génétique YPH499.

Détection des protéines Hsp60 et Atp2 avec des anticorps anti Hsp60 et Atp2. La culture cellulaire se fait sur milieu riche (YPD) liquide et les protéines sont extraites après 4 heures d'incubation à 30 °C ou 37 °C.

Tout d'abord, nous pouvons voir que la souche contrôle est associée à une seule bande pour Hsp60, que ce soit à 30 ou 37°C (figure 33). Cependant, nous pouvons voir que l'intensité de la bande est plus importante à 37 qu'à 30°C. Le fait de n'observer qu'une bande était attendu compte tenu de l'absence de défaut d'import mitochondrial dans le fond génétique YPH499. L'augmentation de l'intensité de cette bande à 37°C n'est pas surprenante car l'expression de *HSP60* est plus importante après une augmentation de température (Yan et al., 2009). Nous n'observons pas d'effet de la délétion de NatC sur la protéine Hsp60 par Western Blot. En effet nous pouvons observer un phénotype similaire à celui de la souche contrôle pour les trois souches déplétées de NatC, c'est-à-dire une absence d'accumulation de précurseurs de Hsp60 qui se serait traduite par l'apparition d'une deuxième bande plus haute que la bande correspondant à la forme mature de Hsp60. Nous avons utilisé un anticorps dirigé contre la protéine Atp2 en tant que témoin de charge. On peut noter qu'il n'y a pas non plus

d'accumulation de précurseurs pour Atp2. Ainsi, il n'y a pas de perturbation visible du profil électrophorétique de Hsp60 dans des mutants NatC dans un fond génétique avec un import mitochondrial non perturbé.

Nous avons par la suite étudié la délétion de NatC dans un contexte dans lequel l'import mitochondrial est perturbé (figure 34). Dans ce contexte j'ai voulu voir comment se comportait la protéine Hsp60 en présence ou en absence de Mak3.

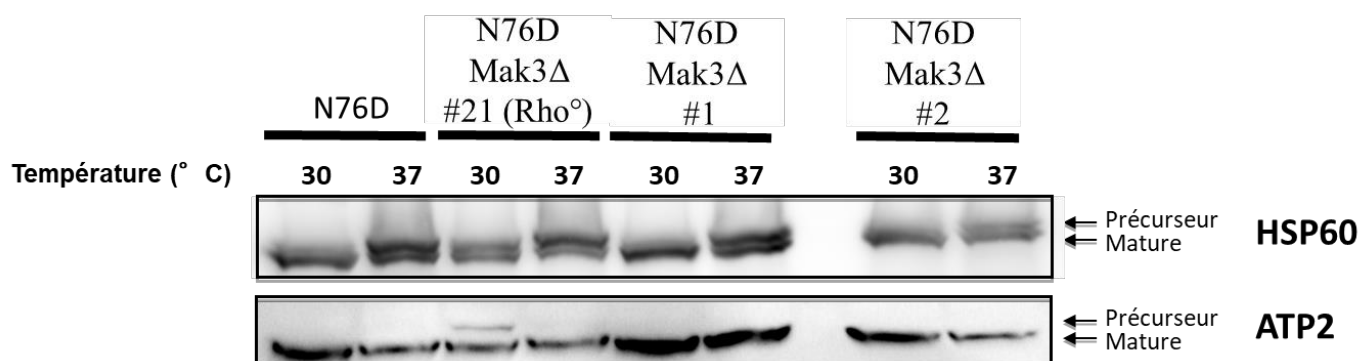


Figure 34 : Western blot des souches délétées de NATC dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*. Détection des protéines Hsp60 et ATP2p avec des anticorps anti Hsp60 et ATP2p. La culture cellulaire se fait sur milieu riche (YPD) liquide et les protéines sont extraites après 4 heures d'incubation à 30°C ou 37°C. A

En plus de la souche contrôle N76D, j'ai déposé également un extrait protéique d'une souche Rho° délétée de NatC (Clone #21) déjà étudiée au chapitre II.2. (figure 29) et qui présente un défaut d'import déjà bien visible à 30°C.

Pour 37°C, nous sommes capables d'identifier deux bandes clairement distinctes pour toutes les conditions. Il n'y a pas de différence d'accumulation particulière des précurseurs dans les deux clones rho+ mutants de NatC (Clones #1 et #2 sur la figure 34) par rapport à notre souche contrôle (N76D), que ce soit à 30 ou à 37°C (Figure 34).

Les résultats de la figure 33 et de la figure 34 suggèrent donc que la délétion de NatC n'a pas d'influence sur l'accumulation des précurseurs de Hsp60, que ce soit dans le fond génétique YPH499 ou le fond génétique *YPH-pam16Δ-hpam16(N76D)*.

4. L'arg/ N End rule empêche-t-elle la détection du rôle de NatC ?

Alors que le résidu en position 2 de Hsp60 a un rôle très important pour prévenir l'accumulation de précurseurs (voir chapitre I des résultats), nous n'avons pas pu montrer par western blot une accumulation des précurseurs de Hsp60 dans une souche délétée de NatC par rapport à la souche sauvage, que ce soit dans un fond génétique standard ou dans un fond dans lequel l'import mitochondrial est perturbé. Ceci pourrait être dû à l'existence d'une voie alternative de dégradation des précurseurs redondante à l'activité de NatC, hypothèse étayée par la bibliographie. En effet, en plus de la voie de l'Ac/N-end rule dans laquelle intervient NatC, il existe une autre branche de la N-end Rule, appelée la voie Arg/N-end rule (Voir chapitre 2 de l'introduction). Cette seconde voie consiste en l'ubiquitination et la dégradation des protéines après leur arginylation ou leur désamination N-terminale. La voie Arg/N-end rule cible notamment les protéines ayant en position 2 les acides aminés L, I et F, qui sont aussi les résidus reconnus par NatC. Ainsi, on peut penser que la délétion de NatC est compensée par une prise en charge des précurseurs mitochondriaux par la voie Arg/N-end rule, ce qui empêche l'accumulation cytoplasmique de ces précurseurs. Nous avons alors décidé d'inhiber la voie Arg/N-end rule pour forcer potentiellement l'accumulation des précurseurs de façon à étudier l'impact de la délétion de NatC sur leur stabilité en délétant l'ubiquitine ligase impliquée dans cette voie : Ubr1 (Kim et al., 2014).

a) Création de souches délétées d'Ubr1 et de doubles mutants Mak3/Ubr1

J'ai donc entrepris de créer une souche délétée non seulement de *MAK3* mais aussi de *UBR1*. Alors que l'obtention d'une telle souche s'est faite sans difficulté dans le contexte de référence YPH499 (voir tableau 1), je ne suis pas parvenu à obtenir le double mutant dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*, ce qui pourrait indiquer que NatC et Ubr1 jouent un rôle synergique important dans des conditions où l'import est perturbé. Des clones ont pu être obtenus pour le simple mutant *ubr1* dans le fond *YPH-pam16Δ-hpam16(N76D)*, mais ils se sont tous avérés être des Rho^o (tableau 1).

	Clones générés	Clones testés	Clones validés
YPH499 Ubr1Δ	12	12	7
YPH499 Mak3ΔUbr1Δ	Nombreux	20	3
hPam16 N76D Ubr1Δ	2	2	0
hPAM16 N76D Ubr1Δ Mak3Δ	1	1	0

Tableau 1 : Tableau récapitulatif de l'ensemble des souches obtenues dans les fonds génétiques *YPH499* et *YPH-pam16Δ-hpam16(N76D)*. J'ai utilisé la méthode de recombinaison homologue pour induire la délétion de *MAK3* et/ou *UBR1*. Clones générés : clones obtenus après la transformation. Clones testés : clones ayant subi un test de croissance sur glycérol (YPGly) afin de discriminer les clones qui ont conservé leur génome mitochondrial Clones validés : clones qui ont conservés leur génome mitochondrial (Rho⁺)

J'ai ensuite caractérisé le phénotype de croissance des clones rho⁺ obtenus dans le fond génétique YPH499. Pour cela, j'ai fait un test en gouttes pour les différentes souches mutantes sur milieu riche, sur milieu avec du glycérol et sur un milieu sans glucose, mais avec éthanol (figure 35).

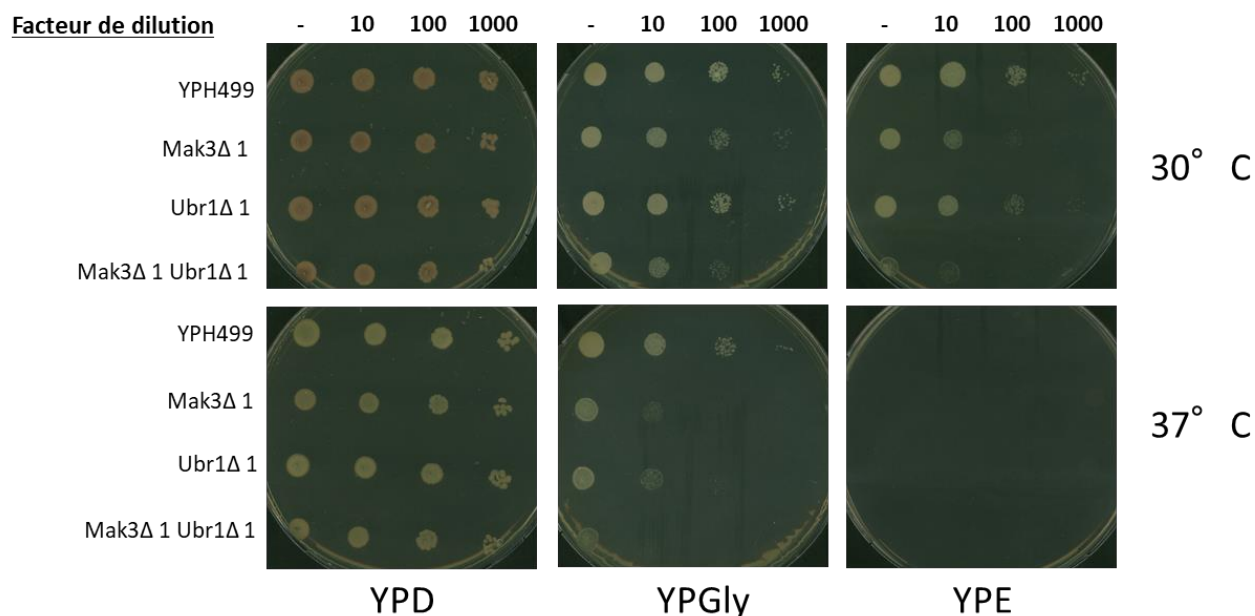


Figure 35 : Test de croissance en gouttes de souches mutantes de NatC et/ou de Ubr1 sur milieu riche (YPD), et sur milieu avec pour source de carbone non fermentable du glycérol (YPGly) ou de l'éthanol (YPE). Le test de croissance est réalisé après 3 jours d'incubation.

Sur ce test en gouttes que j'ai réalisé à 3 reprises, nous ne pouvons pas identifier un défaut de croissance sur milieu riche glucose à 30°C. Toutes les souches semblent pousser à la même vitesse. Cependant à 37°C, il y a un léger phénotype de croissance ralentie pour le mutant de NatC et pour le double mutant. On peut observer un défaut de croissance sur un milieu avec du glycérol comme source de carbone non fermentable pour les souches délétées de NatC, délétées d'Ubr1 et le double mutant. Cependant, le défaut de croissance n'est pas le même pour toutes les souches. Le clone *Δubr1* présente un léger défaut de croissance à 30°C, tandis que la souche *Δmak3* montre un défaut plus marqué. Le clone double mutant est celui qui a le défaut de croissance le plus fort. Le phénotype pour chacun des mutants est encore plus marqué à 37°C et avec de l'éthanol comme source de carbone (YPE). Aucune souche n'a poussé sur YPE à 37°C après 3 jours d'incubation.

À la vue du défaut de croissance très important du double mutant en milieu glycérol, nous étions optimistes à l'idée d'identifier cette fois une accumulation de précurseurs de Hsp60 en western blot.

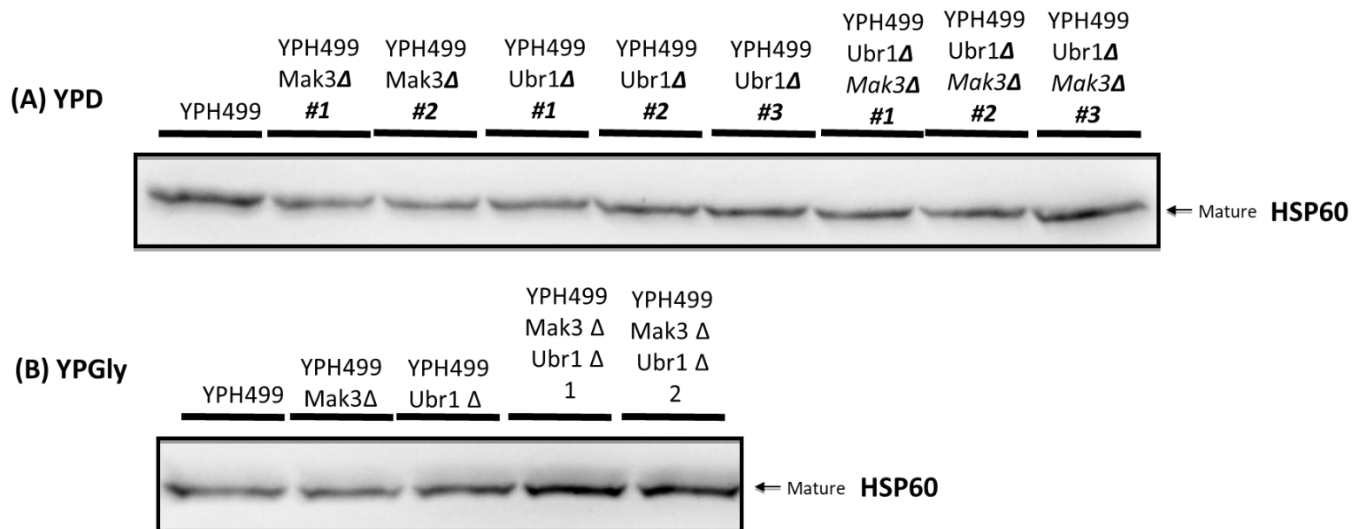


Figure 36 : Western blot des souches délétées de NATC et/ou d’UBR1 dans le fond génétique *YPH499*. Détection des protéines Hsp60 endogène par un anticorps anti Hsp60. La culture cellulaire se fait sur milieu riche (YPD) (A) et sur YPGly (B). Les protéines ont été extraites lorsque les cellules ont atteint une DO de 1.

Cependant, aucune accumulation n’a pu être observée (Figure 36).

On peut voir sur ce Western Blot 4 conditions : un clone contrôle YPH499, 2 clones déplétés de NatC, 3 clones délétés d’Ubr1 et enfin 3 clones délétés pour NatC et pour Ubr1.

De plus, étant donné qu’on a pu observer un phénotype de croissance très ralenti sur YPGly, j’ai alors voulu voir si cela pouvait se traduire par une accumulation de Hsp60 dans ce contexte. Ce n’est apparemment pas le cas (figure 36.B). J’ai déposé des extraits protéiques des différents mutants après 3 heures de culture dans un milieu avec une source de carbone non fermentable (YPGly). Quelle que soit la condition, nous pouvons identifier une seule bande correspondant à la forme mature de Hsp60. La délétion de Mak3 et d’Ubr1 ne permet pas de détecter une accumulation de précurseurs de Hsp60 dans le cytosol.

b) Étude des souches délétées d’Ubr1 dans le fond *YPH-pam16Δ-hpam16(N76D)*

Afin d’obtenir un double mutant NatC/Ubr1 dans un contexte dans lequel l’import mitochondrial serait perturbé, nous avons essayé de mettre en place une méthode alternative pour contourner l’utilisation de la souche *YPH-pam16Δ-hpam16(N76D)*. En effet, les seuls doubles mutants rho+ que j’ai pu créer ont été obtenus sur un fond génétique dans lequel l’import mitochondrial n’est pas perturbé. Dans la littérature, il a été montré que le CCCP est

une drogue perturbant l'import mitochondrial (Miyata et al., 2014; Reid and Schatz, 1982). J'ai alors réalisé un essai de traitement des mutants en fond *YPH499* avec une concentration de 10 μ M sur 5 heures. J'ai choisi cette concentration de CCCP car les auteurs de certaines études avaient réussi à observer une perturbation de l'import mitochondrial avec cette concentration (Itami et al., 2015; Zubko and Zubko, 2014). Cela n'a pas permis de détecter par Western Blot une accumulation des précurseurs de Hsp60 (Figure 37). En effet, quelle que soit la durée d'incubation, on ne peut identifier qu'une seule bande correspondant à la forme mature d'Hsp60 sur le Western Blot. L'utilisation de cette méthode pourrait impliquer de faire des mises au point pour identifier des paramètres permettant de perturber l'import mitochondrial de façon satisfaisante. Pour tester l'efficacité du CCCP, nous pourrions mettre en place une cinétique avec différentes concentrations de CCCP. Cela nous permettrait de savoir quelle est la concentration et le temps d'incubation optimaux pour visualiser un défaut d'import. Par Western Blot nous devrions voir une baisse de la quantité de la forme mature d'une protéine mitochondriale, Hsp60 par exemple, par rapport à une souche qui n'aurait pas subi de traitement au CCCP. Une souche Rho^o peut également être utilisée comme second contrôle. Dans cette souche, nous devrions pouvoir détecter le précurseur sans traitement au CCCP.

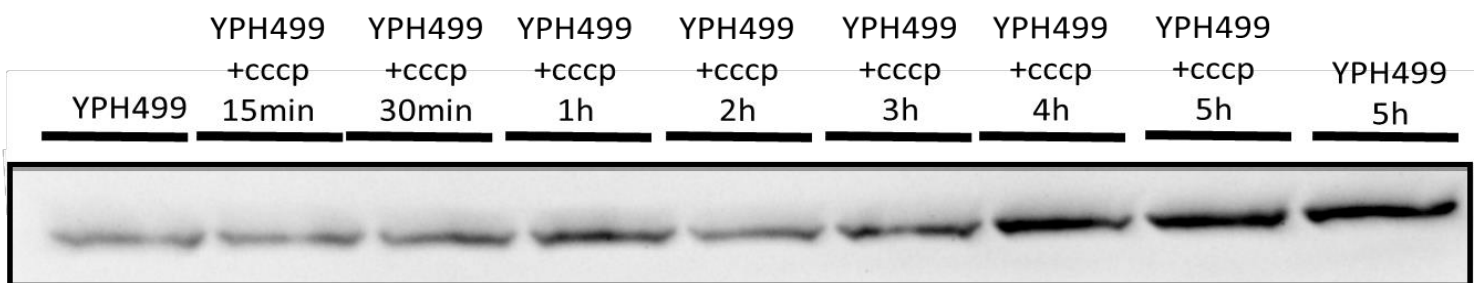


Figure 37 : Western blot des souches délétées dans le fond génétique YPH499 exposées à du CCCP. Détection des protéines Hsp60 grâce à des anticorps dirigés contre Hsp60. La culture cellulaire se fait sur milieu riche (YPD) avec du CCCP 10uM.

5. Utilisation du système de criblage Hsp60-myc pour étudier l'effet de la déplétion de l'enzyme NatC

Comme je vous l'ai montré précédemment (cf chapitre 1 des résultats), j'ai mis en place un système qui nous permet de cribler de manière efficace des mutants d'import. Pour cela, j'ai utilisé la létalité que provoque l'import d'une version de Hsp60 avec un épitope C-terminal

composé de 13 répétitions myc : lorsque cette version taguée Hsp60-myc est efficacement importée dans la mitochondrie, elle induit un effet dominant négatif létal pour la cellule. Ce système a été validé expérimentalement, car aucun clone n'est obtenu suite à une transformation des levures avec un plasmide contenant une version Hsp60-myc sans mutation au niveau du résidu en position 2 de Hsp60. Au contraire, la transformation par une version de *HSP60* présentant une mutation L2S, observée auparavant comme perturbant l'import de la protéine Hsp60, a permis l'obtention de nombreux clones, indiquant que la mutation en position 2 permet un sauvetage phénotypique. Il était donc tentant de combiner l'utilisation du système de criblage Hsp60-myc avec la déplétion de l'enzyme NatC. J'ai donc utilisé l'effet de la toxicité de la protéine Hsp60-myc pour voir si ce phénotype peut être sauvé par la délétion de *MAK3* et/ou *UBR1* dans le fond génétique *YPH499* (figure 38).

Lorsqu'une des enzymes ou les deux sont délétées, nous retrouvons très peu de colonies sur la boîte avec une version sauvage de *HSP60*. En revanche, quel que soit le contexte, on peut observer de nombreuses colonies sur les boîtes lorsque Hsp60 a une Sérine en position 2. La délétion de *Mak3* et/ou d'*Ubr1* ne permet pas de sauver l'effet dominant négatif de létalité. Cela suggère que la délétion de l'une ou l'autre de ces enzymes ne perturbe pas l'import de la protéine Hsp60-myc.

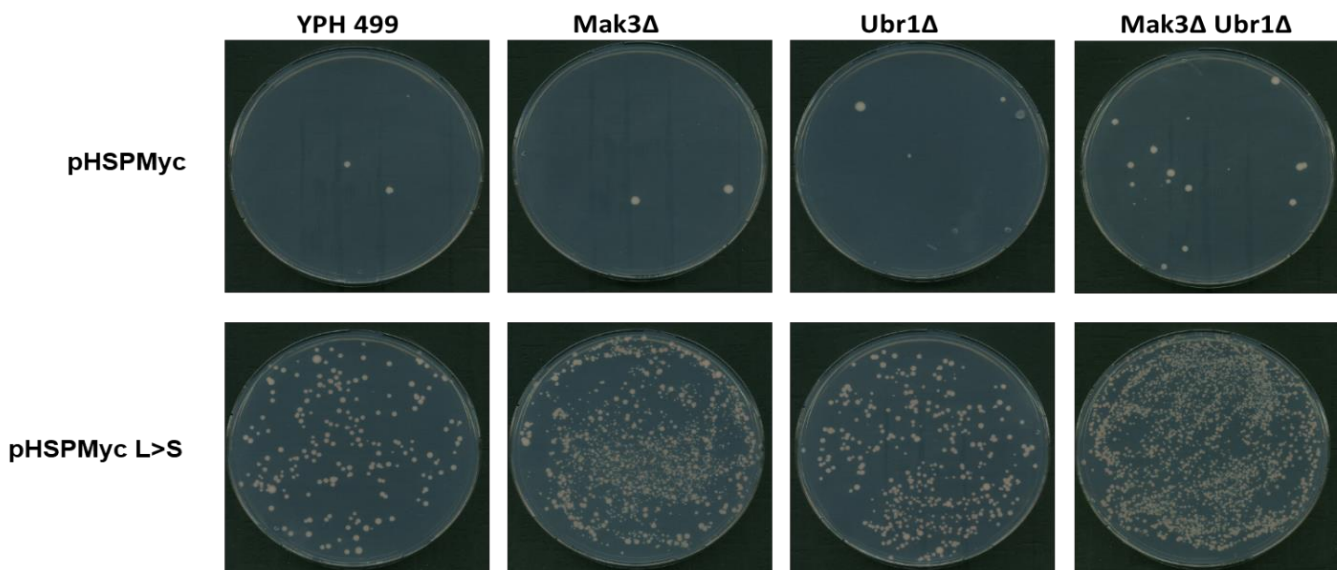


Figure 38 : Photographies des boîtes de pétri après transformation avec des plasmides pHSPmyc et pHSPmyc(L>S) dans les souches YPH499, YPH499 *Mak3Δ*, YPH499 *Ubr1Δ* et YPH499 *Mak3ΔUbr1Δ*. De nombreuses colonies ont été obtenues avec le plasmide pHSPmyc(L>S) et seulement quelques colonies avec le plasmide pHSPmyc.

6. De nouveaux résultats encourageants ?

Lors de la rédaction de l'article présenté dans le chapitre I des résultats, j'ai souhaité refaire certaines expériences pour obtenir de meilleures figures. En effet, au cours de ma thèse, j'ai pu optimiser certaines techniques de biologie moléculaire notamment les Western Blots qui ont progressivement acquis de bien meilleures résolutions. J'ai commencé par étudier de nouveau le phénotype moléculaire des deux clones rho+ délétés de Mak3 dans le fond *YPH-pam16Δ-hpam16(N76D)* à 30°C sur milieu riche (Figure 39)

Tout d'abord, on peut voir que la résolution est bien meilleure que celle de la figure 34. Pour le clone sauvage, on ne distingue qu'une seule bande claire correspondant à la forme mature de Hsp60 et on devine à peine une bande très faible de taille supérieure qui pourrait correspondre à une très légère accumulation de précurseurs. Ce résultat est cohérent avec le fait que nous sommes ici à température permissive et que le défaut d'import de la souche N76D est très tenu à cette température.

Nous observons le même profil pour le clone délété de *MAK3* #1. En revanche, pour le clone #2 nous observons un profil perturbé qui se caractérise par la présence de deux bandes : une bande pour la forme mature et une autre clairement visible pour la forme précurseure. Cette différence n'est pas due à une différence de quantité de protéines chargées dans les puits, comme le montre le témoin de charge. J'ai décidé d'utiliser ici comme témoin de charge la protéine Pgc1 plutôt qu'Atp2. En effet, Atp2 est une protéine mitochondriale, même si elle n'est pas censée être ciblée par NatC, on ne peut pas exclure que la délétion de NatC puisse avoir un effet indirect sur l'expression de la protéine. Nous avons donc deux profils différents pour les clones délétés de NatC. Après ce résultat surprenant, j'ai réalisé des vérifications par PCR afin de m'assurer que les deux clones étaient bien délétés de Mak3 et c'était bien le cas (données non montrées).

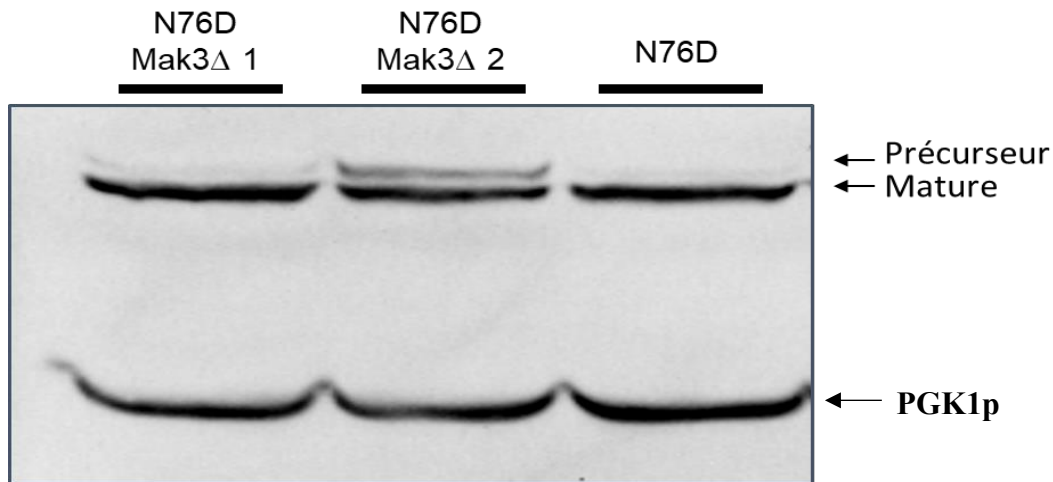


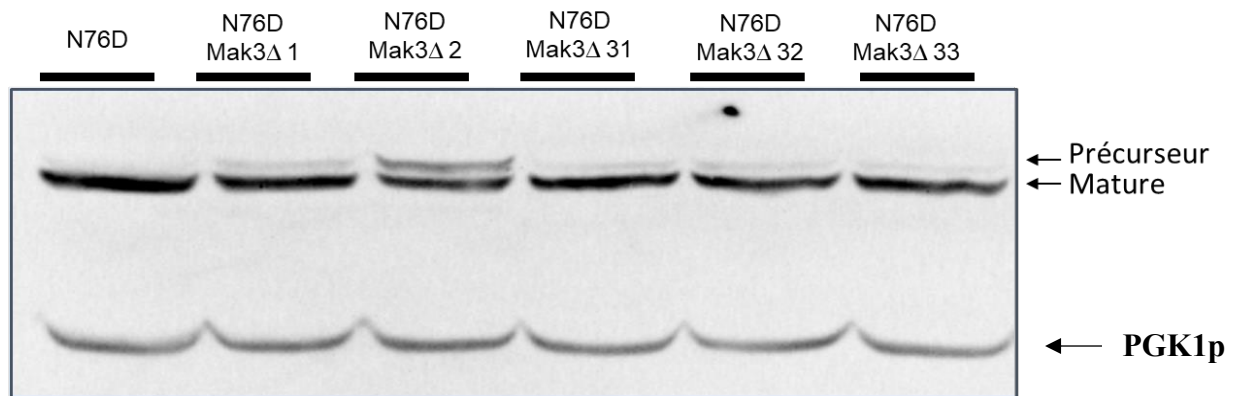
Figure 39 : Western blot des souches délétées de NATC dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*. Détection des protéines Hsp60 et Pgk1 avec des anticorps anti Hsp60 et Pgk1. La culture cellulaire se fait sur milieu riche (YPD) liquide à 30°C et les protéines sont extraites lorsque les cellules ont atteint une DO de 1.

J'ai donc entrepris de générer d'autres clones rho⁺ déplétés de NatC dans le fond génétique N76D pour mieux comprendre cette variabilité. J'ai alors obtenu 3 autres clones.

L'étude par Western Blot des cinq clones *mak3Δ* disponibles m'a permis de confirmer le profil perturbé pour le clone 2 (Figure 39). Pour ce clone, j'ai réalisé au total 2 répliquats techniques et 3 répliquats biologiques au cours de ma thèse et j'ai toujours pu détecter la bande précurseure. Pour les 3 nouveaux clones, nous pouvons distinguer dans le Western Blot de la figure 40 de fines bandes, peu intenses qui correspondraient à la forme précurseure de Hsp60. Cette bande supplémentaire est également visible pour le clone 1 mais pas du tout visible pour le contrôle N76D. D'après les résultats de la figure 40, la délétion de *MAK3* dans le fond génétique N76D entraînerait systématiquement une accumulation de précurseurs de Hsp60 à température permissive, ce phénotype étant plus marqué pour le clone 2 que pour les 4 autres. Ce qui rend ce résultat difficilement interprétable est qu'il existe un problème de reproductibilité. L'extrait protéique déposé pour le clone 1 dans le Western Blot de la figure 40 est le même que celui de la figure 39. Pourtant la bande correspondant à la forme précurseure n'est détectable pour ce clone que dans la figure 40. Pour mieux estimer la reproductibilité de ce phénotype, j'ai réalisé 3 répliquats biologiques des 3 nouveaux clones (clone 31, 32 et 33). Je n'ai pas pu détecter les précurseurs dans ces trois répliquats (Figure 40B). Le phénotype moléculaire des clones 1, 31,

32 et 33 observé en figure 40 semble donc ne pas être reproductible et il n'est pas possible de conclure sur l'effet de la délétion de NatC sur l'accumulation de précurseurs Hsp60 dans ces 4 clones.

A



B

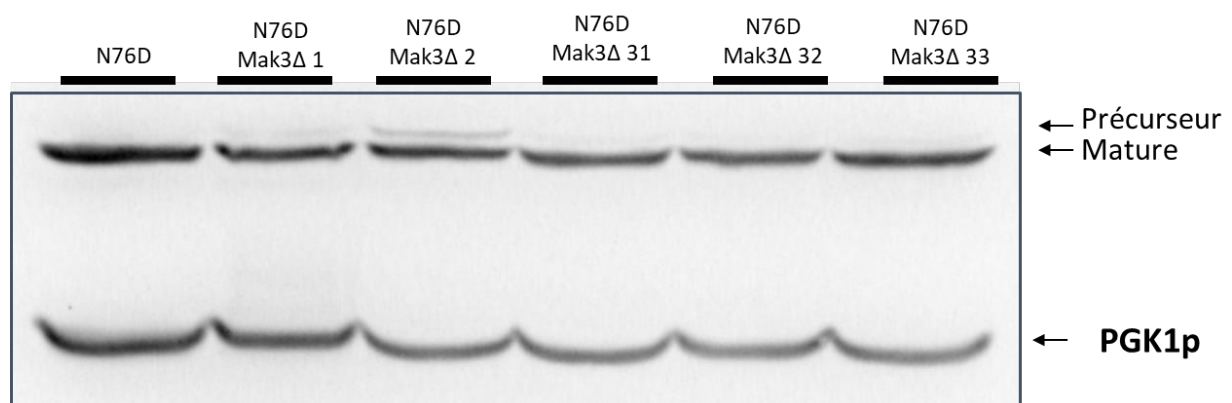


Figure 40 : Western blot de 5 souches délétées de NatC dans le fond génétique *YPH-pam16Δ-hpam16(N76D)*. Détection des protéines Hsp60 et Pgc1 avec des anticorps anti Hsp60 et Pgc1. La culture cellulaire se fait sur milieu riche (YPD) liquide à 30°C et les protéines sont extraites lorsque les cellules ont atteint une DO de 1. Les figures A et B correspondent à deux répliquats biologiques différents.

Étant donné que le clone 2 est le seul qui présente un phénotype clair et reproductible d'accumulation des précurseurs (figure 39 et 40), nous avons décidé d'approfondir notre étude de ce clone pour essayer de comprendre l'origine de ce phénotype original. Afin de savoir si ce phénotype moléculaire est bien lié à la déplétion de NatC, nous avons essayé de faire de la

complémentation fonctionnelle en réintroduisant le gène *MAK3* sous forme de plasmide centromérique avec le promoteur endogène. J’ai ainsi obtenu six clones transformés, à partir de deux plasmides *MAK3* obtenus de manière indépendante, que j’ai étudiés par western blot (Figure 41).

Nous observons comme attendu la seule bande correspondant à la forme mature de Hsp60 pour les contrôles, la souche *YPH-pam16Δ-hpam16(N76D)* et nous observons les deux bandes pour le clone 2 déplété de NatC. Pour les autres clones qui ont un plasmide exprimant Mak3, nous pouvons distinguer deux bandes également ce qui suggère que le phénotype du clone de départ n’est pas sauvé par l’expression de Mak3. Cependant, pour certains clones, l’intensité de la bande du dessus semble plus faible que pour le clone 2 déplété de NatC. Ce résultat n’est pas clair car il est difficile de quantifier ces résultats de western blot pour obtenir une vue statistique et quantitative des effets. J’ai réalisé cette expérience une deuxième fois avec de nouveaux extraits protéiques et j’ai obtenu les mêmes résultats. Ce premier essai de complémentation fonctionnelle ne permet donc pas de tirer de conclusions sur le rôle de NatC dans ce profil perturbé et des études complémentaires sont requises pour conclure. En effet, il se pourrait que l’expression induite de Mak3 par le plasmide ne soit pas suffisante pour sauver le phénotype. Pour tester cette hypothèse, nous pouvons mesurer l’expression de l’ARNm de *MAK3* par RT-qPCR en comparaison avec un clone contrôle exprimant Mak3 à partir de la version chromosomique du gène.

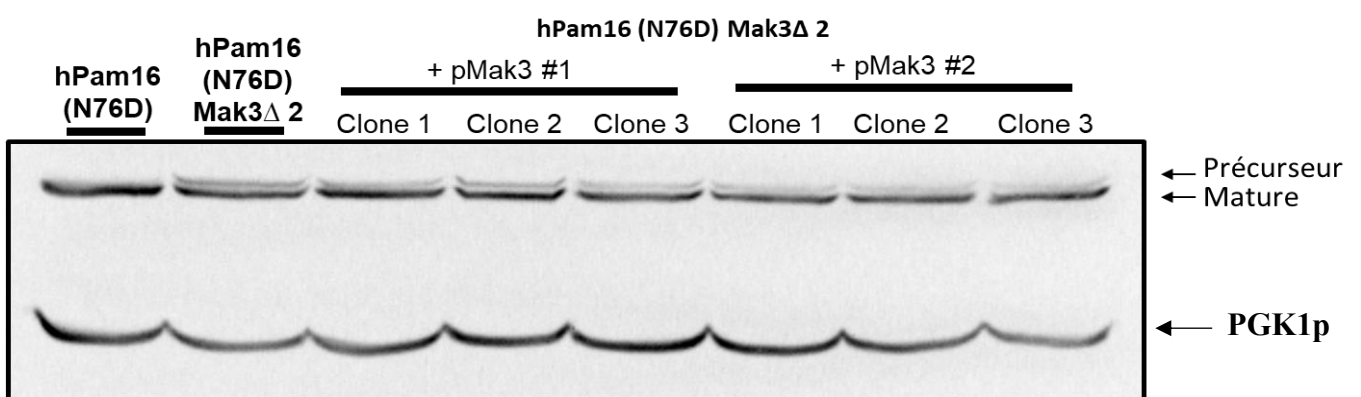


Figure 41 : Western blot des souches déléetées de NatC dans le fond génétique *YPH-pam16Δ-hpam16(N76D)* avec un plasmide pMak1 #1 ou 2. Détection des protéines Hsp60 et PGK1 avec des anticorps anti Hsp60 et Pgc1. La culture cellulaire se fait sur milieu riche (YPD) liquide à 30°C et les protéines sont extraites lorsque les cellules ont atteint une DO de 1.

Discussion

I. Le résidu en position 2 du MTS joue un rôle critique sur le devenir de Hsp60

1. Le résidu en position 2 de Hsp60 semble avoir un rôle sur son import mitochondrial...

Le principal résultat de ma thèse est d'avoir confirmé le rôle de l'acide aminé en position 2 de Hsp60 sur son devenir et d'avoir validé expérimentalement l'importance fonctionnelle de la présence d'un large résidu hydrophobe en position 2 de cette protéine possédant un MTS.

L'importance fonctionnelle de la nature de ce résidu avait été fortement suggérée par l'observation faite par Salomé Nashed d'une sur-représentation des acides aminés hydrophobes de grande taille que sont la leucine, la phénylalanine, l'isoleucine, le tryptophane et dans une moindre mesure, la méthionine en position 2 des protéines possédant un MTS. Notamment, plus de 35% des MTS ont une leucine en position 2, alors que ce résidu est faiblement utilisé en position 2 dans le protéome (6% des protéines cellulaires). La conservation de cette caractéristique des MTS nouvellement découverte sur l'ensemble de la lignée des Saccharomycotina qui couvre 400 millions d'années d'évolution, suggère qu'une forte pression de sélection s'exerce sur la position 2 des MTS. Par ailleurs, une signature spécifique en position 2 a aussi été observée lors de l'analyse des MTS humains, soulignant encore l'importance de ce biais en position 2 des MTS au cours de l'évolution (données non montrées et obtenues par Salomé Nashed). En formant une hélice alpha amphiphile capable d'interagir avec les protéines du complexe TOM, les MTS sont des séquences permettant l'import de nombreux précurseurs mitochondriaux. Nous avons donc émis l'hypothèse que le biais en position 2 des MTS pourrait jouer un rôle majeur dans le processus d'import mitochondrial.

Afin de tester cette hypothèse, j'ai dû travailler dans un système expérimental permettant de faciliter la visualisation des précurseurs mitochondriaux, lesquels sont très instables dans des conditions physiologiques, car rapidement importés dans les mitochondries après leur traduction cytoplasmique et maturés par clivage. J'ai donc travaillé dans la souche N76D, une souche thermosensible qui, même à la température permissive de 30°C, présente un léger défaut

d'import mitochondrial (Mehawej et al., 2014). J'ai utilisé comme modèle d'étude la protéine Hsp60. A l'aide de la technologie CRISPR CAS9, j'ai réalisé différentes mutations du résidu en position 2 de cette protéine, remplaçant la leucine par divers résidus (sérine, acide aspartique, valine, proline, phénylalanine et lysine). De manière frappante, nous avons observé, pour toutes les modifications introduites en position 2 de la protéine Hsp60, une perturbation drastique du profil observé en Western Blot à savoir une détection des précurseurs et une diminution de l'intensité de la bande correspondant à la forme mature de la protéine. Ce résultat, compatible avec un problème d'import, a confirmé le rôle crucial du résidu en position 2 du MTS pour le devenir des précurseurs mitochondriaux.

Après le résultat très prometteur obtenu par western blot, j'ai mis à profit un système qui nous permet de cribler de manière efficace des mutants d'import, que j'ai appliqué dans ce fond génétique N76D. Pour cela, j'ai utilisé la létalité que provoque l'import d'une version de Hsp60 avec un épitope C-terminal composé de 13 répétitions myc : lorsque cette version étiquetée Hsp60-myc est efficacement importée dans la mitochondrie, elle induit un effet dominant négatif létal pour la cellule (Fang and Cheng, 2002). Ce système a été validé expérimentalement car la transformation de levures avec un plasmide portant une version Hsp60-myc sans mutation au niveau du résidu en position 2 n'a permis d'obtenir que des revertants. En effet, après vérification par séquençage, j'ai constaté que les rares clones que j'avais obtenus avaient perdu des étiquettes Myc (données non montrées). Au contraire, une mutation L2S, observée précédemment comme perturbant l'import de la protéine Hsp60, a permis l'obtention de nombreux clones après transformation indiquant que cette mutation en position 2 permet un sauvetage phénotypique. J'ai ensuite effectué le criblage des 20 acides aminés en position 2 de la protéine Hsp60-myc. Nous avons alors observé que tous les acides aminés hydrophobes et qui sont sur-représentés en deuxième position des MTS sont associés à une toxicité de la protéine Hsp60-myc ; au contraire, les autres résidus sauvent l'effet dominant négatif. Plusieurs éléments indiquent que cette effet des mutations L2X est bien lié à l'import mitochondrial. Premièrement, sur les Western Blot on peut observer pour les mutants de la position 2 une baisse de l'intensité de la bande correspondant à la forme mature, ce qui est cohérent avec un défaut d'import réduisant la quantité de protéines Hsp60 clivées dans la mitochondrie. La mutation L2D se traduit même par la disparition presque totale de la bande mature au profit du précurseur. Ce résultat ne s'explique pas dans l'hypothèse où la mutation agirait seulement sur la stabilité du précurseur dans le cytoplasme, sauf à imaginer que l'accumulation de précurseurs stables cytoplasmiques bloque indirectement totalement l'import. Deuxièmement, la protéine

Hsp60 doit être importée dans la mitochondrie pour former un complexe tétradécamérique, et on peut supposer que l'effet dominant négatif est lié au fait que ce complexe est rendu non fonctionnel lorsqu'il se compose à la fois de la forme génomique et de la forme étiquetée myc. L'import mitochondrial serait donc nécessaire pour que l'effet dominant négatif puisse se produire. De plus une publication a montré qu'in vitro la mutation de la leucine en Valine en position 2 de la protéine mitochondriale ADH perturbe son import mitochondrial chez le rat (Hammen et al., 1999). Ce résultat suggère donc que la nature du résidu en position 2 du MTS de Hsp60 joue un rôle critique dans son import et confirme l'existence d'une forte pression de sélection exercée à cette position des MTS.

2. ... mais certains de mes résultats suggèrent une réalité plus complexe.

Je n'ai pas réalisé de quantification des Western Blot pour vérifier rigoureusement que la quantité totale de Hsp60 détecté ne varie pas dans mes différentes souches. Cette absence de quantification m'empêche d'exclure totalement un rôle du résidu en position 2 sur la stabilité des précurseurs. Nous pourrions également tester directement l'hypothèse de la stabilité en utilisant par exemple du cycloheximide, un inhibiteur de la traduction, et en calculant le temps de demi vie des précurseurs. Cependant cette expérience serait compliquée par les faibles quantités de précurseurs détectées.

En l'état, on peut toujours émettre l'hypothèse que l'apparition des bandes correspondant aux formes précurseuses des Hsp60-myc mutées est liée au moins en partie à un effet de la mutation en position 2 sur la stabilité de cette forme précurseure. Les protéines précurseures cytoplasmiques deviendraient plus stables, s'agrègeraient dans le cytosol et perturberaient indirectement les canaux d'import, sauvant ainsi le phénotype dominant négatif.

Cette hypothèse est soutenue par le western blot réalisé sur des cellules exprimant à la fois la version génomique sauvage et la version taguée dans le fond génétique N76D. On observe sur ce western blot à la fois la bande mature et la forme précurseure de la version endogène non taguée de Hsp60 (Figure Supp4.a de l'article). Or la forme endogène de Hsp60 n'a en théorie aucune raison de présenter un import perturbé puisqu'elle n'est pas mutée en position 2... Ce phénotype pourrait être expliqué par une perturbation globale de l'import induite par l'accumulation cytoplasmique de précurseurs stabilisés par la mutation de Hsp60-myc. Afin de

tester cette hypothèse, nous pourrions envisager de regarder le profil moléculaire d'une autre protéine mitochondriale, comme Atp2. Si les précurseurs d'Atp2 sont détectables dans ce système expérimental, alors cela signifierait que l'accumulation du précurseur de la version mutée L2S de Hsp60-myc a un impact sur l'import de l'ensemble des protéines mitochondriales. Le cas contraire suggérerait qu'il y a des interactions entre les protéines Hsp60 taguées et non taguées dans le cytoplasme, ce qui compliquerait potentiellement notre interprétation du sauvetage de l'effet dominant négatif. Nous ne verrions l'accumulation du précurseur L2S que dans le cadre du fond N76D car la perturbation d'import propre à cette souche induirait une charge protéostatique causant une saturation du système de dégradation à l'origine d'une accumulation des précurseurs.

D'autres résultats intrigants sont ceux obtenus dans le fond génétique YPH499, sans défaut d'import initial. Dans ce contexte, nous n'avons pas été en mesure d'observer de profil perturbé en western blot pour les mutations L2S. La forme précurseuse n'a pas pu être détectée par Western Blot (figure 28). J'ai également réalisé l'expérience de sauvetage phénotypique de la version taguée myc de Hsp60 L2S, et son expérience contrôle L2L dans le fond génétique YPH499 (figure 38). Une toxicité forte a été observée pour la forme L2L, tandis que la mutation L2S a permis un sauvetage phénotypique, exactement comme dans le fond génétique N76D (figure 38) Comment la mutation L2S peut-elle sauver l'effet dominant négatif dans la souche YPH499 sans perturber ni l'import ni la stabilité du précurseur (ou en tous cas pas de façon visible sur western blot quand la forme endogène d'Hsp60 est mutée) ?

Tout d'abord, on peut supposer que le système reposant sur l'utilisation de la version plasmidique Hsp60 taguée myc est plus sensible pour détecter un problème d'import que la technique du western blot. On peut également émettre l'hypothèse que les versions Myc perturbent l'import de façon suffisante pour permettre d'observer un phénotype qui n'était pas visible sur western blot quand seule la forme endogène non taguée de Hsp60 était exprimée. Une autre hypothèse que l'on peut émettre serait que l'effet dominant négatif n'est en réalité pas lié à l'import mitochondrial. Par exemple la mutation L2S pourrait perturber le repliement de la protéine Hsp60-myc et l'empêcher ainsi d'interagir avec la version sauvage endogène, sans pour autant perturber fortement son import mitochondrial. L'observation de l'accumulation du précurseur dans le fond N76D serait alors un effet synergique entre le défaut de repliement et le défaut d'import propre à cette souche.

Pour vérifier ces hypothèses, il faudrait réaliser le même type de western blot que celui de la Figure Supp4.a de l'article mais dans le fond YPH499, ce qui n'a pour l'instant pas été fait. Ceci permettrait de vérifier si le précurseur de la forme Hsp60-myc mutée L2S s'accumule ou non dans le fond YPH499.

II. Exploration globale du rôle de NatC

1. Un lien avéré entre NatC et les mitochondries

Durant ma thèse, l'un de mes objectifs était de caractériser le rôle de NatC sur le devenir des précurseurs mitochondriaux. En effet, dans la littérature il a été montré que chez la levure, la délétion de NatC est associée à un ralentissement de croissance sur un milieu contenant une source de carbone non fermentable, ce qui suggère un défaut respiratoire induit par cette délétion (Polevoda and Sherman, 2001). Chez l'homme, la délétion de NatC conduit à une fragmentation des mitochondries, soulignant le lien entre NatC et ces dernières (Van Damme et al., 2016). Au laboratoire, j'ai pu confirmer le phénotype de délétion de NatC observé chez la levure (Figure 32). De plus, le sel-TRAP a pu apporter pour la première fois une preuve expérimentale que les précurseurs mitochondriaux portant à leur extrémité N-terminale une séquence d'adressage MTS constituent une large proportion des cibles de NatC (figure 4.c de l'article). Afin de comprendre plus précisément l'impact de cette enzyme sur les précurseurs mitochondriaux, je me suis intéressé à son effet sur la protéine Hsp60. Nous avons deux hypothèses concernant le rôle de NatC. La première est qu'en augmentant l'hydrophobicité N-terminale des MTS via leur acétylation, NatC pourrait augmenter l'efficacité d'import des précurseurs portant des MTS. La seconde est liée au fait que d'après la littérature, NatC est impliquée dans une voie de dégradation protéasome dépendante appelée la AC/N-end rule (voir Chapitre II.2 de l'introduction). Dans le cadre de cette voie, les protéines N-acétylées par NatC sont reconnues par les ubiquitines ligases DOA10 et Not4. Ces protéines sont alors ubiquitinées et dégradées par le protéasome. NatC pourrait donc influencer le temps de demi-vie des précurseurs possédant un MTS. En particulier, elle pourrait induire leur dégradation s'ils s'accumulaient dans le cytoplasme à cause d'un problème d'import pour éviter la toxicité qui serait causée par une telle accumulation. Ces deux hypothèses ne sont pas exclusives, mais elles seraient associées au même phénotype, à savoir, en cas de délétion de NatC, une accumulation de la forme précurseure de Hsp60.

2. Les mutants de NatC ont un phénotype respiratoire mais leurs mitochondries respirent normalement

Nous avons démontré que la délétion de NatC dans le fond génétique YPH499 et dans le fond génétique N76D conduit à un phénotype de croissance ralentie sur un milieu contenant du glycérol. Afin de mieux comprendre l'origine de ce défaut respiratoire de la souche déplétée de NatC, une collaboration avec l'équipe de Jean-Paul Di Rago (Laboratoire « Molecular Genetics of Mitochondrial Systems », Bordeaux) a été initiée. Les résultats obtenus sont intrigants. Il n'a pas été observé d'altération de l'activité mitochondriale dans la souche déplétée. En comparaison avec la souche sauvage, aucune différence n'a été observée au niveau de la synthèse d'ATP, du potentiel de membrane et en termes de respiration à partir de mitochondries isolées. Cependant, quand ils ont testé la croissance de ces souches, ils ont retrouvé les mêmes résultats que moi, à savoir un phénotype de croissance ralentie sur YPD qui est exacerbé à 37°C et sur un milieu avec du glycérol comme source de carbone. De plus, sur milieu solide, la souche a une forte sensibilité à l'oligomycine ce qui suggère un défaut dans le fonctionnement de l'ATP synthase. L'une des hypothèses que nos collaborateurs ont émises pour expliquer ces observations, est que les cellules déplétées de NatC ont des mitochondries fonctionnelles mais en plus petit nombre. Afin de tester cette hypothèse nous pourrions utiliser l'imagerie électronique ou l'imagerie à fluorescence.

3. L'impact de la délétion de NatC sur Hsp60 est difficile à caractériser par Western Blot

Malgré le phénotype fort obtenu en mutant le résidu en position 2 ciblé par NatC chez Hsp60 et le phénotype de croissance ralentie sur glycérol lié à la délétion de l'enzyme, je n'ai pas pu observer de façon reproductible et claire un effet de la délétion de NatC sur le profil de Hsp60 en western blot, même dans le fond génétique N76D dans lequel l'import est perturbé.

En effet, il existe une hétérogénéité des phénotypes de délétion, pourtant obtenus via le même protocole expérimental et alors que cette délétion a bien été confirmée par PCR. En particulier, deux phénotypes mutants *mak3Δ* sont observés : celui du clone 2, qui présente dans chacune des 3 expériences de western blot réalisées en fin de thèse une accumulation très marquée du précurseur Hsp60 (figures 39, 40 et 41), et celui des 4 autres clones, pour lesquels une bande très fine pouvant correspondre à la forme précurseuse n'a été détectée qu'une fois sur 3 réplicats. Il serait donc intéressant de reproduire de nombreuses fois cette expérience afin

d'avoir assez de données pour pouvoir dire si statistiquement nous avons un phénotype de délétion. Bien qu'il soit difficile de justifier cette hétérogénéité de phénotypes de délétion de NatC, il faut noter que la souche YPH499 présente déjà une instabilité mitochondriale qui s'accroît dans le contexte N76D dans lequel cette délétion a été réalisée. Il est donc possible d'imaginer que l'absence de NatC est trop délétère dans ce contexte et est contrecarrée par des mutations compensatoires dans certains des clones obtenus. Cependant, cette hypothèse est peu probable car je n'ai pas eu de difficultés pour obtenir des mutants de NatC. Dans le clone 2, j'ai réalisé une réintroduction de *MAK3* sur un plasmide, mais le sauvetage phénotypique alors observé n'était, au mieux, que partiel. Cela pourrait s'expliquer soit par le fait que l'expression plasmidique de Mak3 n'était pas assez forte pour assurer un sauvetage phénotypique complet, soit par le fait que le phénotype de ce clone n'est pas lié uniquement à la délétion de NatC. Dans les autres clones obtenus suite à la délétion de NatC, une telle expérience de sauvetage phénotypique aurait été difficilement concluante car le phénotype n'est pas assez marqué et pas assez reproductible. Il est de plus à noter que la délétion de NatC a été obtenue par recombinaison homologue dans des cellules haploïdes. Une solution pour identifier le phénotype de délétion de NatC serait d'obtenir cette délétion à partir du croisement de cellules N76D et de cellules YPH499 delta NatC. Partant du principe que cette délétion est récessive, les cellules diploïdes maintiendraient une expression de *MAK3* ne nécessitant pas l'apparition de mutations compensatoires. Ces cellules pourraient alors être mises à sporuler ce qui permettrait d'obtenir des haploïdes N76D delta NatC ayant moins de probabilités que nos clones d'avoir développé des mutations secondaires. La comparaison du phénotype de Hsp60 dans ces haploïdes et de celui des clones déjà produits nous permettrait d'identifier le groupe de clones ayant un phénotype exclusivement lié à la délétion de NatC.

4. L'acétylation par NatC joue peut-être un rôle mineur sur le devenir de Hsp60 et n'est peut-être pas la raison de la conservation de la leucine en position 2

Nous faisons donc face à un paradoxe qui est que le résidu en position 2 de Hsp60, ciblé par NatC, a un rôle important dans le devenir de la protéine alors que la délétion de NatC ne semble avoir aucun impact. Ceci pourrait être lié à notre choix d'utiliser Hsp60 comme protéine modèle. Les enzymes de modifications tels que NatC peuvent présenter des effets différents sur leurs cibles. En particulier, deux études s'appuyant sur l'utilisation de protéines rapportrices différentes ont fourni des résultats contradictoires entre elles. Une des études se base sur

l'évaluation du temps de demi-vie de protéines rapportrices présentant toutes les combinaisons possibles de résidus aux deux premières positions et exprimées chez la levure. Elle montre en cas de double délétion de NatC et d'Ubr1 une absence de stabilisation de leurs cibles, à savoir des protéines ayant en position 2 un large résidu hydrophobe (Kats et al., 2018). Cependant une autre étude qui utilise des protéines rapportrices et des protéines naturelles a montré par Western Blot que la délétion d'une des enzymes induit une stabilisation des protéines cibles qui est accentuée quand les deux enzymes sont délétées. C'est le cas par exemple de la protéine rapportrice ML-Ura et de certaines protéines naturelles comme Msn4 (Kim et al., 2014). Ainsi, on ne peut exclure le fait que Hsp60 ne soit pas la protéine optimale pour observer l'effet d'une des enzymes.

De plus, nous avons vu que l'hydrophobicité N-terminale des MTS semble être un élément critique pour un import efficace à la mitochondrie. Or, Hsp60 présente cette hydrophobicité N-terminale forte du fait que sa méthionine initiateur hydrophobe est conservée et qu'elle a une leucine hydrophobe également en position 2. Dans ce cas il est possible qu'une acétylation par NatC ne contribue pas de façon cruciale à une hydrophobicité N-terminale de Hsp60 déjà grande. Au contraire, un remplacement de la leucine en position 2 par un acide aminé moins hydrophobe et/ou moins volumineux aurait un impact très fort sur cette hydrophobicité et donc sur l'efficacité du MTS. Dans ce modèle, la conservation de la leucine en position 2 de Hsp60 serait liée à son caractère hydrophobe et pas au fait que ce résidu permette le recrutement de NatC. L'acétylation de Hsp60 par NatC serait donc une conséquence fortuite et potentiellement neutre de la nécessité d'avoir un acide aminé hydrophobe tel que la leucine en position 2 de la protéine.

Néanmoins, il est avéré que les mutants de NatC présentent un phénotype respiratoire, ce qui suggère fortement que le recrutement de NatC doit avoir un rôle important au moins pour certaines protéines mitochondriales. Il pourrait ainsi être intéressant d'étudier l'impact de NatC sur une protéine possédant un MTS moins optimal pour l'import que celui de Hsp60. Il pourrait par exemple s'agir d'un MTS portant en position 2 une tyrosine, un résidu également ciblé par NatC mais dont l'hydrophobicité est moins forte que celle d'une leucine.

Dans l'hypothèse où nous confirmerions à l'avenir que l'accumulation du précurseur d'Hsp60 observée dans le clone 2 muté pour *MAK3* serait bien liée à NatC, le résultat serait tout de même différent de celui obtenu pour les mutants du résidu en position 2. En effet, contrairement aux mutants en position 2 pour lesquels on aperçoit une diminution de l'intensité de la bande

mature, pour les mutants de NatC, au mieux, la bande précurseure est de même intensité que la bande mature. Cela suggère que muter le 2^e résidu de la protéine n'est pas équivalent à l'effet d'une délétion de NatC. L'enzyme aurait plutôt un rôle sur la stabilité de la protéine. Cette hypothèse est appuyée par le fait que le phénotype de croissance ralentie causé par la délétion de NatC est aggravé lorsqu'on délète également l'ubiquitine ligase Ubr1, impliquée dans la dégradation des protéines via l'Arg/N-end Rule. D'ailleurs, nous n'avons pu obtenir ce double mutant que dans le contexte YPH499. Cependant, aucun clone n'a pu être obtenu dans le fond N76D. Il est possible que la double délétion dans ce fond génétique présentant un import déjà perturbé soit à l'origine d'une trop forte accumulation des précurseurs, létale d'après la littérature. Toutefois cette expérience n'a été réalisée qu'une fois dans un seul clone N76D (le clone N76D Mak3Δ1). Il serait donc intéressant de la reproduire pour s'assurer du phénotype de létalité du double mutant dans un contexte d'import perturbé.

Finalement, que l'effet de la délétion de NatC sur Hsp60 soit confirmé ou non par Western blot, nos résultats obtenus jusqu'à présent suggèrent qu'il n'y a pas d'équivalence entre muter le résidu ciblé par NatC et la délétion de NatC. Le phénotype moléculaire de Hsp60 dans les mutants de NatC n'est dans tous les cas pas similaire à ceux obtenus avec les mutants en position 2. Par ailleurs, je n'ai pas observé de sauvetage phénotypique pour la version HSP60-Myc dans le cas de la délétion de NatC (figure 38). Ces résultats vont dans le sens que NatC ne joue pas de rôle majeur dans l'import de Hsp60. Ceci nous ramène à l'idée exprimée ci-dessus que le MTS de Hsp60 permettrait un import optimisé de telle sorte que la délétion de NatC ne perturbe pas suffisamment son efficacité d'import pour que celle-ci puisse être visualisable. C'est la nature du résidu en position 2 qui semble être fonctionnellement importante. Le résidu doit être de grande taille, pour éviter le clivage de la méthionine initiatrice et être hydrophobe afin d'optimiser l'import dans la mitochondrie. Pour Hsp60, il semblerait que l'acétylation de NatC ne soit qu'un effet secondaire, lié au fait que l'enzyme reconnaît les protéines ayant en position 2 un large résidu hydrophobe. Cependant, cette acétylation pourrait jouer un rôle plus important pour d'autres protéines qui auraient un MTS considéré comme moins optimal.

Il serait alors intéressant d'étendre l'étude à d'autres protéines mitochondriales possédant un MTS. Cependant, même dans un contexte dans lequel l'import mitochondrial est perturbé il est très difficile de détecter les précurseurs par Western Blot. Nous avons par exemple essayé au cours de ma thèse de détecter la forme précurseure de la protéine Atp1, sans succès (données non montrées). Il serait donc intéressant d'effectuer une analyse plus globale, notamment via la

technique de protéomique quantitative des extrémités N-terminales des protéines. Pour étudier l'hypothèse de l'import nous pourrions purifier les mitochondries afin d'observer les protéines importées. Pour l'hypothèse de la stabilité, nous pourrions au contraire étudier des extraits protéiques enrichis en protéines cytosoliques.

Enfin, il faut rappeler que nous n'avons jamais vérifié au cours de ma thèse que les précurseurs mitochondriaux cibles de NatC étaient effectivement acétylés, ni que cette acétylation était perdue dans les mutants de NatC ou dans les mutants L2X de Hsp60. Notamment, une raison pour laquelle nous n'avons pas vu d'effet clair de la délétion de NatC sur l'accumulation de précurseurs de Hsp60 est peut-être que cette délétion est en partie compensée par une autre Nat. Même si NatE, qui a une spécificité similaire à NatC chez les métazoaires, est supposée ne pas être fonctionnelle chez la levure, il serait intéressant de faire du SelTrap sur Nat5 en présence ou en absence de NatC.

III. Conclusion générale

J'ai démontré durant ma thèse le rôle critique du résidu en position 2 des protéines mitochondriales portant un MTS sur leur devenir cellulaire. Or comprendre les caractéristiques optimales d'import à la mitochondrie est un enjeu majeur pour la santé publique. En effet d'après un rapport de l'institut Pasteur, les maladies mitochondriales, qui sont les maladies métaboliques les plus fréquentes, concernent plus de 200 nouveaux cas par an en France. Les seuls traitements disponibles actuellement permettent de mieux vivre avec la maladie sans pour autant en guérir. Une piste thérapeutique étudiée est la thérapie génique consistant à réintroduire au niveau mitochondrial la protéine défailante. De nombreuses recherches ont été faites récemment pour essayer de sélectionner des MTS efficaces pour conférer à des molécules médicamenteuses un bon adressage vers la mitochondrie. Une étude a montré que tous les MTS n'ont pas la même efficacité de ciblage, montrant qu'il est indispensable de déterminer les critères des MTS permettant l'adressage le plus efficace possible (Dong et al., 2021; Klimpel and Neundorf, 2018). En effet, plus la livraison est organite-spécifique et plus l'efficacité du médicament sera importante (Sakhrani and Padh, 2013). L'autre enjeu d'un bon adressage de la protéine est de limiter des effets secondaires causés notamment par l'administration aspécifique de médicaments dans des cellules ou organites sains (Daughton and Ruhoy, 2013). Or jusqu'à présent le rôle du résidu en position 2 des MTS n'a pas été pris en compte par les laboratoires qui ont pour but de synthétiser des MTS efficaces. Mes travaux contribuent donc

potentiellement à mettre en lumière de nouvelles voies de recherche thérapeutiques pour le traitement des maladies mitochondriales.

Matériels et Méthodes

1. Souches disponibles au laboratoire et milieux de croissance

Les souches de levure *S. cerevisiae* utilisées au cours de ma thèse sont présentées dans le tableau 2. Elles ont été cultivées à 30°C dans des milieux riches YP (1% bactopeptone, 1% yeastextract) contenant soit 2% de glucose (YPD) comme source de carbone, soit 2% de glycérol lorsque l'on souhaitait étudier leur capacité respiratoire. Les milieux solides ont été complétés avec 2% agar.

<u>Nom</u>	<u>Génotype</u>	<u>Sigle utilisé dans les figures</u>
YPH499	<i>MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	YPH
YPH499 pma16Δ + pmagMAT N76D	<i>MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ2 pma16Δ + phPAM16(N76D)</i>	N76D
BY4742 mak3Δ	<i>MATa his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0 mak3Δ</i>	BY4742mak3Δ
D160 Rho°	<i>Mata Rho°(absence de genome mitochondrial)</i>	D160/Rho°
ATP12 Δ	<i>Mata leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15</i>	ATP12Δ

Tableau 2 : Liste des souches utilisées lors de mon stage

2. Amplification par PCR de cassettes de délétion par recombinaison homologue

La délétion des gènes *MAK3* et *UBR1* a été effectuée par recombinaison homologue à l'aide d'une cassette de délétion contenant le gène de résistance à la kanamycine (KanMX) pour *MAK3* et le gène de résistance pour l'hygromycine pour *UBR1*.

Pour *MAK3*, la cassette a été obtenue par PCR à partir d'une souche *BY4742 mak3Δ* issue de la collection euroscarf. Cette souche contient une délétion génomique du gène *MAK3* lequel a été remplacé par le gène KanMX. J'ai amplifié cette cassette grâce à une PCR d'amplification en utilisant des amorces spécifiques au gène *MAK3* : 5'GACATACGTAACGAAGAGC 3' et 3' CTCAGTCTCACGTTTCTATGC 5'.

Pour *UBR1*, la cassette a été obtenue par PCR à partir d'un plasmide pAG32. Ce plasmide contient le gène de sélection hygromycine (Hygr^R). J'ai amplifié le gène de résistance à l'aide d'amorces spécifiques au gène de l'hygromycine flanquée par des séquences homologues du gène *ubr1* ce qui permettra leur intégration par recombinaison homologue : 5' TGAAGTCCCTAATCTTTACAGGTCACACAAATTACATAGAACATTCCAATCGTAC GCTGCAGGTCGAC 3' et 5' TTTATATACAAATATGTCAACTATAAAACATAGTA GAGGGCTTGAATCTAATCGATGAATTCGAGCTCG 3'

Les mélanges de PCR ont été obtenus en ajoutant 1 µl d'ADN génomique, 42.4 µl d'H₂O, 5 µl de Buffer 10x vert, 1 µl de dNTP (10 mM), 0.2 µl d'oligonucléotide et de 0.2 µl de la polymérase DreamTaq. L'ensemble subit 30 cycles successifs comportant une étape de dénaturation à 95°C pendant 30 secondes, une étape d'hybridation à 54°C pour la cassette de délétion d'*UBR1* et 58°C pour la cassette de délétion de *MAK3* pendant 30 secondes et enfin une étape d'élongation à 72°C pendant 1 minute la cassette de délétion de *MAK3* et 4 min pour la cassette de délétion d'*UBR1*.

Après vérification de l'amplification les produits de PCR ont été purifiés sur colonne (QiaQuick PCR purification, QIAGEN) puis une quantification du produit a été effectuée à l'aide d'un spectrophotomètre. Si nécessaire, de nouvelles PCR ont été effectuées de manière à obtenir environ 5µg de la cassette de délétion, pour chacune des transformations à effectuer.

3. Transformation et sélection phénotypique des levures en vue de l'obtention des mutants

La transformation des cellules a été réalisée en suivant la méthode classique d'Acétate de Lithium : les transformations se font à partir de plusieurs colonies de levure resuspendues dans 100 µL de solution A (100 mM lithium d'acétate, 100 mM Tris EDTA). La transformation est réalisée à partir de 50 µl de cellules, 5 µl d'ADN carrier qui est préalablement dénaturé à 100°C pendant 20 min, 300 µl de la solution B (100 mM lithium d'acétate, 100 mM Tris EDTA, 50% PEG). A ce mélange est ajouté l'ADN utilisé pour la transformation (environ 5 µg de cassette de délétion ou 5 µg de cassette d'ADN de réparation et 1,5 µg d'ADN plasmidique) avant d'effectuer un choc thermique pendant 25 min à 42°C. Le milieu de transformation est soigneusement éliminé par un lavage dans 1mL de H₂O avant re-suspension du culot dans 100

μL de H_2O . Un temps de génération est nécessaire : les cellules sont alors incubées 2h à température ambiante. Les cellules ayant potentiellement été génétiquement modifiées sont sélectionnées à l'aide d'un milieu Ad Hoc (contenant du G418 pour les mutants de délétion de *NatC* ou de l'hygromycine pour les mutants de délétion d'*UBR1*).

4. Vérification des mutants : PCR et séquençage

Pour les souches mutantes délétées de *MAK3* et d'*UBR1*, la vérification a été réalisée par PCR en utilisant des amorces spécifiques.

Deux types de PCR ont été effectuée (table 3) :

Pour *MAK3* : une première (notée Mak3 Locus dans la table 3) à l'aide d'amorces bordant le locus *MAK3*, une deuxième (notés Mak3+Kan dans la table 3) avec un oligonucléotide dans le gène KanMX de la cassette de délétion et l'autre localisé dans la région située en aval de la séquence codant Mak3 après la zone d'insertion de la cassette.

Pour *UBR1* : une première (noté Ubr1 Locus dans la table 3) à l'aide d'amorces bordant le locus *UBR1*, une deuxième (notés Ubr1 + Hygr dans la table 3) avec un oligonucléotide dans le promoteur d'*UBR1* et l'autre localisé dans la région située dans le gène de sélection de l'hygromycine.

<u>PCR</u>	<u>Amorce up (5'-3')</u>	<u>Amorce Down (3'-5')</u>
Kan + mak3	CTTGCCATCCTATGGAAGTGC	GAGAAGTGGAAGGCTCTAACG
Mak3 locus	GACATACGTAACGAAGAGC	CTCAGTCTCACGTTTCTATGC
Ubr1 + Hygro	GCAGATAGTGACCATAAGGC	GATGTGAGAACTGTATCCTAGC
Ubr1 locus	GCACTATGTCAAGATTCCAGTTCG	GCGCTTCGAACAATATATTGGAAG

Tableau 3 : Séquences des oligonucléotides utilisés pour les PCR

5. Tests de croissance des souches de levures en milieux solide et liquide

La croissance cellulaire a été étudiée sur milieu YPD, YPGly et YPE en déposant sur des boîtes de Petri des gouttes de 5µl de dilutions sériées réalisées à partir d'une culture cellulaire à DO= 5. Les boîtes ont été incubées pendant 24 à 72h à 37°C ou 30°C selon les conditions testées.

La croissance des souches a également été testée grâce à une cinétique de croissance en milieu liquide. Les concentrations cellulaires en milieu liquide ont été estimées par mesure de densité optique à 600 nm (DO). La croissance sur YPD a été mesurée toutes les heures dans des Falcons de 50 mlensemencés à une DO de 0.1 et placés à l'étuve avec agitation à 30°C. Les courbes de DO ont été analysées avec le logiciel Excel de manière à déterminer l'intervalle de temps pendant lequel la croissance est exponentielle et suit la relation $DO(t) = DO_i \cdot e^{\mu t}$ avec, $DO(t)$ DO mesurée à l'instant t, DO_i la DO initiale, μ le taux de croissance. Une régression de type exponentiel a permis de déterminer le coefficient μ puis le temps de génération T_g (temps nécessaire pour le doublement de la population cellulaire) a été calculé en utilisant la relation $T_g = \ln(2)/\mu$.

6. Analyse des contenus protéiques par Western Blot

L'extraction de protéines totales se fait à partir d'un volume correspondant à $DO_{600nm} = 3$ de culture de levures en phase exponentielle, centrifugées pendant 5 min à 12000 rpm. Le culot est resuspendu dans 500 μ L de H₂O suivi par un ajout de 10 μ L de NaOH 12 M et une incubation 10 min dans la glace, à cela on ajoute 25 μ L de TCA 100% suivi de 10 min dans la glace. Les échantillons sont centrifugés 5 min, 12000 rpm à 4°C et le culot est resuspendu dans 60 μ L de Laemmli 2X (4% SDS, 20% glycérol, 10% β -mercaptoethanol, 0.004% bleu de bromophénol, 120 mM Tris HCl pH 6.8), 15 μ L de tris-base 1 M et 2% de β -mercaptoéthanol. Pour finir, les échantillons sont dénaturés pendant 5 min à 95°C. Le western blot qui s'en suit est réalisé selon un protocole standard. La révélation se fait grâce au substrat chemiluminescent ECL (Interchim) de la peroxydase de raifort (HRP) couplée aux AC II et l'acquisition des images grâce au système G:BOX couplé au logiciel GeneSys.

Voici la liste des anticorps utilisés pour la réalisation de mes Western Blot :

<u>Anticorps</u>	<u>Dilution</u>	<u>Espèces</u>	<u>Fournisseur</u>	<u>Couplage</u>
Anti-lapin	1/10 000	-	Promega	HRP
Anti-HSP60	1/1000	Lapin	-	-
Anti-ATP1	1/5000	Lapin	-	-
Anti-ATP2	1/5000	Lapin	-	-
Anti-PGK1	1/10000	Souris	Abcam	-
Anti-souris	1/10000	-	Promega	HRP

Tableau 4 : Caractéristiques des anticorps utilisés pour l'analyse de Western Blot

7. Test de complémentation fonctionnelle pour l'analyse de la perte potentielle du génome mitochondrial

Les souches de cellules sont croisées avec une souche Rho^o avérée (souche *DI60 Rho^o*) et une souche contrôle délétée de *ATP12*. Quelques colonies sont diluées dans 100 μ l d'H₂O et le croisement est réalisé en superposant deux gouttes de chaque souche sur du milieu riche. Après

une nuit d'incubation les colonies sont repiquées sur milieu YPGly. Seules les colonies ayant réussi la complémentation fonctionnelle auront poussé après une nuit d'incubation.

Références

- Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T., Kohda, D., 2000. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* 100, 551–560. [https://doi.org/10.1016/s0092-8674\(00\)80691-1](https://doi.org/10.1016/s0092-8674(00)80691-1)
- Aksnes, H., Osberg, C., Arnesen, T., 2013. N-Terminal Acetylation by NatC Is Not a General Determinant for Substrate Subcellular Localization in *Saccharomyces cerevisiae*. *PLoS ONE* 8, e61012. <https://doi.org/10.1371/journal.pone.0061012>
- Aksnes, H., Ree, R., Arnesen, T., 2019. Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Mol. Cell* 73, 1097–1114. <https://doi.org/10.1016/j.molcel.2019.02.007>
- Allison, D.S., Schatz, G., 1986. Artificial mitochondrial presequences. *Proc. Natl. Acad. Sci. U. S. A.* 83, 9011–9015.
- Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2, e201900429. <https://doi.org/10.26508/lisa.201900429>
- Amar-Costesec, A., n.d. Découvertes en Cytologie et en Biologie cellulaire.
- Arnesen, T., Van Damme, P., Plevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J.E., Vandekerckhove, J., Lillehaug, J.R., Sherman, F., Gevaert, K., 2009. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci.* 106, 8157–8162. <https://doi.org/10.1073/pnas.0901931106>
- Arnold, R.J., Plevoda, B., Reilly, J.P., Sherman, F., 1999. The Action of N-terminal Acetyltransferases on Yeast Ribosomal Proteins. *J. Biol. Chem.* 274, 37035–37040. <https://doi.org/10.1074/jbc.274.52.37035>
- Avendaño-Monsalve, M.C., Mendoza-Martínez, A.E., Ponce-Rojas, J.C., Poot-Hernández, A.C., Rincón-Heredia, R., Funes, S., 2022. Positively charged amino acids at the N terminus of select mitochondrial proteins mediate early recognition by import proteins $\alpha\beta'$ -NAC and Sam37. *J. Biol. Chem.* 298. <https://doi.org/10.1016/j.jbc.2022.101984>

- Avendaño-Monsalve, M.C., Ponce-Rojas, J.C., Funes, S., 2020. From cytosol to mitochondria: the beginning of a protein journey. *Biol. Chem.* 401, 645–661. <https://doi.org/10.1515/hsz-2020-0110>
- Bachmair, A., Finley, D., Varshavsky, A., 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234, 179–186. <https://doi.org/10.1126/science.3018930>
- Backes, S., Bykov, Y.S., Flohr, T., Räschele, M., Zhou, J., Lenhard, S., Krämer, L., Mühlhaus, T., Bibi, C., Jann, C., Smith, J.D., Steinmetz, L.M., Rapaport, D., Storchová, Z., Schuldiner, M., Boos, F., Herrmann, J.M., 2021. The chaperone-binding activity of the mitochondrial surface receptor Tom70 protects the cytosol against mitoprotein-induced stress. *Cell Rep.* 35, 108936. <https://doi.org/10.1016/j.celrep.2021.108936>
- Backes, S., Hess, S., Boos, F., Woellhaf, M.W., Gödel, S., Jung, M., Mühlhaus, T., Herrmann, J.M., 2018. Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. *J. Cell Biol.* 217, 1369–1382. <https://doi.org/10.1083/jcb.201708044>
- Bartels, T., Choi, J.G., Selkoe, D.J., 2011. α -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* 477, 107–110. <https://doi.org/10.1038/nature10324>
- Bauer, M.F., Sirrenberg, C., Neupert, W., Brunner, M., 1996. Role of Tim23 as voltage sensor and presequence receptor in protein import into mitochondria. *Cell* 87, 33–41. [https://doi.org/10.1016/s0092-8674\(00\)81320-3](https://doi.org/10.1016/s0092-8674(00)81320-3)
- Baysal, C., Pérez-González, A., Eserverri, Á., Jiang, X., Medina, V., Caro, E., Rubio, L., Christou, P., Zhu, C., 2020. Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Res.* 29, 37–52. <https://doi.org/10.1007/s11248-019-00176-9>
- Behnia, R., Panic, B., Whyte, J.R.C., Munro, S., 2004. Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat. Cell Biol.* 6, 405–413. <https://doi.org/10.1038/ncb1120>
- Ben-Saadon, R., Fajerman, I., Ziv, T., Hellman, U., Schwartz, A.L., Ciechanover, A., 2004. The tumor suppressor protein p16(INK4a) and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system. Direct evidence for ubiquitination at the N-terminal residue. *J. Biol. Chem.* 279, 41414–41421. <https://doi.org/10.1074/jbc.M407201200>

- Bose, A., Beal, M.F., 2016. Mitochondrial dysfunction in Parkinson's disease. *J. Neurochem.* 139 Suppl 1, 216–231. <https://doi.org/10.1111/jnc.13731>
- Brix, J., Dietmeier, K., Pfanner, N., 1997. Differential recognition of preproteins by the purified cytosolic domains of the mitochondrial import receptors Tom20, Tom22, and Tom70. *J. Biol. Chem.* 272, 20730–20735. <https://doi.org/10.1074/jbc.272.33.20730>
- Brix, J., Rüdiger, S., Bukau, B., Schneider-Mergener, J., Pfanner, N., 1999. Distribution of Binding Sequences for the Mitochondrial Import Receptors Tom20, Tom22, and Tom70 in a Presequence-carrying Preprotein and a Non-cleavable Preprotein *. *J. Biol. Chem.* 274, 16522–16530. <https://doi.org/10.1074/jbc.274.23.16522>
- Byrd, C., Turner, G.C., Varshavsky, A., 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* 17, 269–277. <https://doi.org/10.1093/emboj/17.1.269>
- Caesar, R., Warringer, J., Blomberg, A., 2006. Physiological importance and identification of novel targets for the N-terminal acetyltransferase NatB. *Eukaryot. Cell* 5, 368–378. <https://doi.org/10.1128/EC.5.2.368-378.2006>
- Chacinska, A., Koehler, C.M., Milenkovic, D., Lithgow, T., Pfanner, N., 2009. Importing mitochondrial proteins: machineries and mechanisms. *Cell* 138, 628–644. <https://doi.org/10.1016/j.cell.2009.08.005>
- Chacinska, A., Lind, M., Frazier, A.E., Dudek, J., Meisinger, C., Geissler, A., Sickmann, A., Meyer, H.E., Truscott, K.N., Guiard, B., Pfanner, N., Rehling, P., 2005. Mitochondrial Presequence Translocase: Switching between TOM Tethering and Motor Recruitment Involves Tim21 and Tim17. *Cell* 120, 817–829. <https://doi.org/10.1016/j.cell.2005.01.011>
- Chang, Y.H., Teichert, U., Smith, J.A., 1992. Molecular cloning, sequencing, deletion, and overexpression of a methionine aminopeptidase gene from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 267, 8007–8011.
- Chen, S., Vetro, J.A., Chang, Y.-H., 2002. The specificity in vivo of two distinct methionine aminopeptidases in *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* 398, 87–93. <https://doi.org/10.1006/abbi.2001.2675>
- Chen, S.-J., Wu, X., Wadas, B., Oh, J.-H., Varshavsky, A., 2017. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* 355, eaal3655. <https://doi.org/10.1126/science.aal3655>
- Ciechanover, A., Ben-Saadon, R., 2004. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol.* 14, 103–106. <https://doi.org/10.1016/j.tcb.2004.01.004>

- Claros, M.G., 1995. MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput. Appl. Biosci.* CABIOS 11, 441–447. <https://doi.org/10.1093/bioinformatics/11.4.441>
- Corral-Debrinski, M., Blugeon, C., Jacq, C., 2000. In Yeast, the 3' Untranslated Region or the Presequence of ATM1 Is Required for the Exclusive Localization of Its mRNA to the Vicinity of Mitochondria. *Mol. Cell. Biol.* 20, 7881–7892.
- Daughton, C.G., Ruhoy, I.S., 2013. Lower-dose prescribing: Minimizing “side effects” of pharmaceuticals on society and the environment. *Sci. Total Environ.* 443, 324–337. <https://doi.org/10.1016/j.scitotenv.2012.10.092>
- Dekker, P.J., Ryan, M.T., Brix, J., Müller, H., Hönlinger, A., Pfanner, N., 1998. Preprotein translocase of the outer mitochondrial membrane: molecular dissection and assembly of the general import pore complex. *Mol. Cell. Biol.* 18, 6515–6524. <https://doi.org/10.1128/MCB.18.11.6515>
- del Alamo, M., Hogan, D.J., Pechmann, S., Albanese, V., Brown, P.O., Frydman, J., 2011. Defining the Specificity of Cotranslationally Acting Chaperones by Systematic Analysis of mRNAs Associated with Ribosome-Nascent Chain Complexes. *PLoS Biol.* 9, e1001100. <https://doi.org/10.1371/journal.pbio.1001100>
- Deng, J.M., Behringer, R.R., 1995. An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Res.* 4, 264–269. <https://doi.org/10.1007/BF01969120>
- Deng, S., McTiernan, N., Wei, X., Arnesen, T., Marmorstein, R., 2020. Molecular basis for N-terminal acetylation by human NatE and its modulation by HYPK. *Nat. Commun.* 11, 818. <https://doi.org/10.1038/s41467-020-14584-7>
- Dihanich, M., van Tuinen, E., Lambris, J.D., Marshallsay, B., 1989. Accumulation of viruslike particles in a yeast mutant lacking a mitochondrial pore protein. *Mol. Cell. Biol.* 9, 1100–1108. <https://doi.org/10.1128/mcb.9.3.1100-1108.1989>
- Dikiy, I., Eliezer, D., 2014. N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound α -synuclein and increases its affinity for physiological membranes. *J. Biol. Chem.* 289, 3652–3665. <https://doi.org/10.1074/jbc.M113.512459>
- Dinh, T.V., Bienvenut, W.V., Linster, E., Feldman-Salit, A., Jung, V.A., Meinel, T., Hell, R., Giglione, C., Wirtz, M., 2015. Molecular identification and functional characterization of the first N α -acetyltransferase in plastids by global acetylome profiling. *Proteomics* 15, 2426–2435. <https://doi.org/10.1002/pmic.201500025>

- Doan, K.N., Grevel, A., Mårtensson, C.U., Ellenrieder, L., Thornton, N., Wenz, L.-S., Opaliński, Ł., Guiard, B., Pfanner, N., Becker, T., 2020. The Mitochondrial Import Complex MIM Functions as Main Translocase for α -Helical Outer Membrane Proteins. *Cell Rep.* 31, 107567. <https://doi.org/10.1016/j.celrep.2020.107567>
- Dong, C., Shi, Z., Huang, L., Zhao, H., Xu, Z., Lian, J., 2021. Cloning and characterization of a panel of mitochondrial targeting sequences for compartmentalization engineering in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 118, 4269–4277. <https://doi.org/10.1002/bit.27896>
- Döring, K., Ahmed, N., Riemer, T., Suresh, H.G., Vainshtein, Y., Habich, M., Riemer, J., Mayer, M.P., O'Brien, E.P., Kramer, G., Bukau, B., 2017. Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. *Cell* 170, 298-311.e20. <https://doi.org/10.1016/j.cell.2017.06.038>
- Drazic, A., Myklebust, L.M., Ree, R., Arnesen, T., 2016. The world of protein acetylation. *Biochim. Biophys. Acta* 1864, 1372–1401. <https://doi.org/10.1016/j.bbapap.2016.06.007>
- Drazic, A., Varland, S., 2021. Human NAA30 can rescue yeast *mak3 Δ* mutant growth phenotypes. *Biosci. Rep.* 41, BSR20202828. <https://doi.org/10.1042/BSR20202828>
- Dummitt, B., Micka, W.S., Chang, Y.-H., 2005. Yeast glutamine-fructose-6-phosphate aminotransferase (Gfa1) requires methionine aminopeptidase activity for proper function. *J. Biol. Chem.* 280, 14356–14360. <https://doi.org/10.1074/jbc.M501059200>
- Duttler, S., Pechmann, S., Frydman, J., 2013. Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol. Cell* 50, 379–393. <https://doi.org/10.1016/j.molcel.2013.03.010>
- Edwards, R., Eaglesfield, R., Tokatlidis, K., 2021. The mitochondrial intermembrane space: the most constricted mitochondrial sub-compartment with the largest variety of protein import pathways. *Open Biol.* 11, 210002. <https://doi.org/10.1098/rsob.210002>
- Eliyahu, E., Pnueli, L., Melamed, D., Scherrer, T., Gerber, A.P., Pines, O., Rapaport, D., Arava, Y., 2010. Tom20 mediates localization of mRNAs to mitochondria in a translation-dependent manner. *Mol. Cell. Biol.* 30, 284–294. <https://doi.org/10.1128/MCB.00651-09>
- Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903>

- Emr, S.D., Vassarotti, A., Garrett, J., Geller, B.L., Takeda, M., Douglas, M.G., 1986. The amino terminus of the yeast F1-ATPase beta-subunit precursor functions as a mitochondrial import signal. *J. Cell Biol.* 102, 523–533. <https://doi.org/10.1083/jcb.102.2.523>
- Endo, T., Mitsui, S., Roise, D., 1995. Mitochondrial presequences can induce aggregation of unfolded proteins. *FEBS Lett.* 359, 93–96. [https://doi.org/10.1016/0014-5793\(95\)00015-2](https://doi.org/10.1016/0014-5793(95)00015-2)
- Falkevall, A., Alikhani, N., Bhushan, S., Pavlov, P.F., Busch, K., Johnson, K.A., Eneqvist, T., Tjernberg, L., Ankarcrona, M., Glaser, E., 2006. Degradation of the Amyloid β -Protein by the Novel Mitochondrial Peptidasome, PreP * \blacklozenge . *J. Biol. Chem.* 281, 29096–29104. <https://doi.org/10.1074/jbc.M602532200>
- Fang, Y.-C., Cheng, M., 2002. The effect of C-terminal mutations of HSP60 on protein folding. *J. Biomed. Sci.* 9, 223–233. <https://doi.org/10.1007/BF02256069>
- Forte, G.M.A., Pool, M.R., Stirling, C.J., 2011. N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. *PLoS Biol.* 9, e1001073. <https://doi.org/10.1371/journal.pbio.1001073>
- Frazier, A.E., Dudek, J., Guiard, B., Voos, W., Li, Y., Lind, M., Meisinger, C., Geissler, A., Sickmann, A., Meyer, H.E., Bilanchone, V., Cumsy, M.G., Truscott, K.N., Pfanner, N., Rehling, P., 2004. Pam16 has an essential role in the mitochondrial protein import motor. *Nat. Struct. Mol. Biol.* 11, 226–233. <https://doi.org/10.1038/nsmb735>
- Frey, T.G., Mannella, C.A., 2000. The internal structure of mitochondria. *Trends Biochem. Sci.* 25, 319–324. [https://doi.org/10.1016/S0968-0004\(00\)01609-1](https://doi.org/10.1016/S0968-0004(00)01609-1)
- Friedrich, U.A., Zedan, M., Hessling, B., Fenzl, K., Gillet, L., Barry, J., Knop, M., Kramer, G., Bukau, B., 2021. N α -terminal acetylation of proteins by NatA and NatB serves distinct physiological roles in *Saccharomyces cerevisiae*. *Cell Rep.* 34, 108711. <https://doi.org/10.1016/j.celrep.2021.108711>
- Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P., Imai, K., 2015. MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol. Cell. Proteomics MCP* 14, 1113–1126. <https://doi.org/10.1074/mcp.M114.043083>
- Fünfschilling, U., Rospert, S., 1999a. Nascent polypeptide-associated complex stimulates protein import into yeast mitochondria. *Mol. Biol. Cell* 10, 3289–3299. <https://doi.org/10.1091/mbc.10.10.3289>
- Fünfschilling, U., Rospert, S., 1999b. Nascent polypeptide-associated complex stimulates protein import into yeast mitochondria. *Mol. Biol. Cell* 10, 3289–3299. <https://doi.org/10.1091/mbc.10.10.3289>

- Gadir, N., Haim-Vilmovsky, L., Kraut-Cohen, J., Gerst, J.E., 2011. Localization of mRNAs coding for mitochondrial proteins in the yeast *Saccharomyces cerevisiae*. *RNA* 17, 1551–1565. <https://doi.org/10.1261/rna.2621111>
- Gamerding, M., Kobayashi, K., Wallisch, A., Kreft, S.G., Sailer, C., Schlömer, R., Sachs, N., Jomaa, A., Stengel, F., Ban, N., Deuerling, E., 2019. Early Scanning of Nascent Polypeptides inside the Ribosomal Tunnel by NAC. *Mol. Cell* 75, 996-1006.e8. <https://doi.org/10.1016/j.molcel.2019.06.030>
- Garcia, M., Darzacq, X., Delaveau, T., Jourdain, L., Singer, R.H., Jacq, C., 2007. Mitochondria-associated Yeast mRNAs and the Biogenesis of Molecular Complexes. *Mol. Biol. Cell* 18, 362–368. <https://doi.org/10.1091/mbc.E06-09-0827>
- Garcia, M., Delaveau, T., Goussard, S., Jacq, C., 2010. Mitochondrial presequence and open reading frame mediate asymmetric localization of messenger RNA. *EMBO Rep.* 11, 285–291. <https://doi.org/10.1038/embor.2010.17>
- Gautschi, M., Just, S., Mun, A., Ross, S., Rücknagel, P., Dubaquié, Y., Ehrenhofer-Murray, A., Rospert, S., 2003. The Yeast α -Acetyltransferase NatA Is Quantitatively Anchored to the Ribosome and Interacts with Nascent Polypeptides. *Mol. Cell. Biol.* 23, 7403–7414. <https://doi.org/10.1128/MCB.23.20.7403-7414.2003>
- Gebert, M., Schrempp, S.G., Mehnert, C.S., Heißwolf, A.K., Oeljeklaus, S., Ieva, R., Bohnert, M., von der Malsburg, K., Wiese, S., Kleinschroth, T., Hunte, C., Meyer, H.E., Haferkamp, I., Guiard, B., Warscheid, B., Pfanner, N., van der Laan, M., 2012. Mgr2 promotes coupling of the mitochondrial presequence translocase to partner complexes. *J. Cell Biol.* 197, 595–604. <https://doi.org/10.1083/jcb.201110047>
- George, R., Beddoe, T., Landl, K., Lithgow, T., 1998. The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2296–2301. <https://doi.org/10.1073/pnas.95.5.2296>
- George, R., Walsh, P., Beddoe, T., Lithgow, T., 2002. The nascent polypeptide-associated complex (NAC) promotes interaction of ribosomes with the mitochondrial surface in vivo. *FEBS Lett.* 516, 213–216. [https://doi.org/10.1016/s0014-5793\(02\)02528-0](https://doi.org/10.1016/s0014-5793(02)02528-0)
- Gold, V.A., Chroscicki, P., Bragoszewski, P., Chacinska, A., 2017. Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep.* 18, 1786–1800. <https://doi.org/10.15252/embr.201744261>
- Grauslund, M., Rønnow, B., 2000. Carbon source-dependent transcriptional regulation of the mitochondrial glycerol-3-phosphate dehydrogenase gene, GUT2, from *Saccharomyces cerevisiae*. *Can. J. Microbiol.* 46, 1096–1100. <https://doi.org/10.1139/w00-105>

- Gray, M.W., 2012. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* 4, a011403. <https://doi.org/10.1101/cshperspect.a011403>
- Greenfield, N.J., Stafford, W.F., Hitchcock-DeGregori, S.E., 1994. The effect of N-terminal acetylation on the structure of an N-terminal tropomyosin peptide and alpha alpha-tropomyosin. *Protein Sci. Publ. Protein Soc.* 3, 402–410. <https://doi.org/10.1002/pro.5560030304>
- Grunwald, S., Hopf, L.V.M., Bock-Bierbaum, T., Lally, C.C.M., Spahn, C.M.T., Daumke, O., 2020. Divergent architecture of the heterotrimeric NatC complex explains N-terminal acetylation of cognate substrates. *Nat. Commun.* 11, 5506. <https://doi.org/10.1038/s41467-020-19321-8>
- Günsel, U., Mokranjac, D., 2019. A journey along the TIM23 complex, the major protein translocase of the mitochondrial inner membrane. <https://doi.org/10.5281/ZENODO.3532063>
- Hamilton, R., Watanabe, C.K., de Boer, H.A., 1987. Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs. *Nucleic Acids Res.* 15, 3581–3593.
- Hammen, P.K., Heard, T.S., Waltner, M., Weiner, H., 1999. The loss in hydrophobic surface area resulting from a Leu to Val mutation at the N-terminus of the aldehyde dehydrogenase presequence prevents import of the protein into mitochondria. *Protein Sci. Publ. Protein Soc.* 8, 890–896.
- Hershko, A., Heller, H., Eytan, E., Kaklij, G., Rose, I.A., 1984. Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad. Sci. U. S. A.* 81, 7021–7025. <https://doi.org/10.1073/pnas.81.22.7021>
- Holmes, W.M., Mannakee, B.K., Gutenkunst, R.N., Serio, T.R., 2014. Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nat. Commun.* 5, 4383. <https://doi.org/10.1038/ncomms5383>
- Hong, H., Cai, Y., Zhang, S., Ding, H., Wang, H., Han, A., 2017. Molecular Basis of Substrate Specific Acetylation by N-Terminal Acetyltransferase NatB. *Struct. Lond. Engl.* 1993 25, 641-649.e3. <https://doi.org/10.1016/j.str.2017.03.003>
- Horwich, A.L., Kalousek, F., Fenton, W.A., Pollock, R.A., Rosenberg, L.E., 1986. Targeting of pre-ornithine transcarbamylase to mitochondria: Definition of critical regions and residues in the leader peptide. *Cell* 44, 451–459. [https://doi.org/10.1016/0092-8674\(86\)90466-6](https://doi.org/10.1016/0092-8674(86)90466-6)

- Hoseini, H., Pandey, S., Jores, T., Schmitt, A., Franz-Wachtel, M., Macek, B., Buchner, J., Dimmer, K.S., Rapaport, D., 2016. The cytosolic cochaperone Sti1 is relevant for mitochondrial biogenesis and morphology. *FEBS J.* 283, 3338–3352. <https://doi.org/10.1111/febs.13813>
- Hou, F., Chu, C.-W., Kong, X., Yokomori, K., Zou, H., 2007. The acetyltransferase activity of San stabilizes the mitotic cohesin at the centromeres in a shugoshin-independent manner. *J. Cell Biol.* 177, 587–597. <https://doi.org/10.1083/jcb.200701043>
- Huang, S., Taylor, N.L., Whelan, J., Millar, A.H., 2009. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol.* 150, 1272–1285. <https://doi.org/10.1104/pp.109.137885>
- Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O’Shea, E.K., 2003. Global analysis of protein localization in budding yeast. *Nature* 425, 686–691. <https://doi.org/10.1038/nature02026>
- Hurt, E.C., Pesold-Hurt, B., Schatz, G., 1984. The amino-terminal region of an imported mitochondrial precursor polypeptide can direct cytoplasmic dihydrofolate reductase into the mitochondrial matrix. *EMBO J.* 3, 3149–3156. <https://doi.org/10.1002/j.1460-2075.1984.tb02272.x>
- Hwang, C.-S., Shemorry, A., Varshavsky, A., 2010. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* 327, 973–977. <https://doi.org/10.1126/science.1183147>
- Imai, K., Fujita, N., Gromiha, M.M., Horton, P., 2011. Eukaryote-wide sequence analysis of mitochondrial β -barrel outer membrane proteins. *BMC Genomics* 12, 79. <https://doi.org/10.1186/1471-2164-12-79>
- Imai, K., Nakai, K., 2020. Tools for the Recognition of Sorting Signals and the Prediction of Subcellular Localization of Proteins From Their Amino Acid Sequences. *Front. Genet.* 11.
- Isaya, G., Kalousek, F., Rosenberg, L.E., 1992. Sequence analysis of rat mitochondrial intermediate peptidase: similarity to zinc metallopeptidases and to a putative yeast homologue. *Proc. Natl. Acad. Sci. U. S. A.* 89, 8317–8321. <https://doi.org/10.1073/pnas.89.17.8317>
- Itami, N., Shiratsuki, S., Shirasuna, K., Kuwayama, T., Iwata, H., 2015. Mitochondrial biogenesis and degradation are induced by CCCP treatment of porcine oocytes. *Reproduction* 150, 97–104. <https://doi.org/10.1530/REP-15-0037>

- Jarvis, J.A., Ryan, M.T., Hoogenraad, N.J., Craik, D.J., Høj, P.B., 1995. Solution structure of the acetylated and noncleavable mitochondrial targeting signal of rat chaperonin 10. *J. Biol. Chem.* 270, 1323–1331. <https://doi.org/10.1074/jbc.270.3.1323>
- Johnson, S.C., Yanos, M.E., Kayser, E.-B., Quintana, A., Sangesland, M., Castanza, A., Uhde, L., Hui, J., Wall, V.Z., Gagnidze, A., Oh, K., Wasko, B.M., Ramos, F.J., Palmiter, R.D., Rabinovitch, P.S., Morgan, P.G., Sedensky, M.M., Kaeberlein, M., 2013. mTOR Inhibition Alleviates Mitochondrial Disease in a Mouse Model of Leigh Syndrome. *Science* 342, 1524–1528. <https://doi.org/10.1126/science.1244360>
- Jörnvall, H., 1975. Acetylation of Protein N-terminal amino groups structural observations on alpha-amino acetylated proteins. *J. Theor. Biol.* 55, 1–12. [https://doi.org/10.1016/s0022-5193\(75\)80105-6](https://doi.org/10.1016/s0022-5193(75)80105-6)
- Kats, I., Khmelinskii, A., Kschonsak, M., Huber, F., Knieß, R.A., Bartosik, A., Knop, M., 2018. Mapping Degradation Signals and Pathways in a Eukaryotic N-terminome. *Mol. Cell* 70, 488-501.e5. <https://doi.org/10.1016/j.molcel.2018.03.033>
- Kats, I., Reinbold, C., Kschonsak, M., Khmelinskii, A., Armbruster, L., Ruppert, T., Knop, M., 2021. Up-regulation of ubiquitin–proteasome activity upon loss of NatA-dependent N-terminal acetylation. *Life Sci. Alliance* 5, e202000730. <https://doi.org/10.26508/lisa.202000730>
- Kellems, R.E., Allison, V.F., Butow, R.A., 1974. Cytoplasmic Type 80 S Ribosomes Associated with Yeast Mitochondria: II. EVIDENCE FOR THE ASSOCIATION OF CYTOPLASMIC RIBOSOMES WITH THE OUTER MITOCHONDRIAL MEMBRANE IN SITU. *J. Biol. Chem.* 249, 3297–3303. [https://doi.org/10.1016/S0021-9258\(19\)42672-0](https://doi.org/10.1016/S0021-9258(19)42672-0)
- Kim, H.-K., Kim, R.-R., Oh, J.-H., Cho, H., Varshavsky, A., Hwang, C.-S., 2014. The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell* 156, 158–169. <https://doi.org/10.1016/j.cell.2013.11.031>
- Kirstein-Miles, J., Scior, A., Deuerling, E., Morimoto, R.I., 2013. The nascent polypeptide-associated complex is a key regulator of proteostasis. *EMBO J.* 32, 1451–1468. <https://doi.org/10.1038/emboj.2013.87>
- Klimpel, A., Neundorff, I., 2018. Bifunctional peptide hybrids targeting the matrix of mitochondria. *J. Controlled Release* 291, 147–156. <https://doi.org/10.1016/j.jconrel.2018.10.029>
- Knorr, A.G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., Beckmann, R., 2019. Ribosome-NatA architecture reveals that rRNA expansion segments

- coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.* 26, 35–39. <https://doi.org/10.1038/s41594-018-0165-y>
- Kowalski, L., Bragoszewski, P., Khmelinskii, A., Glow, E., Knop, M., Chacinska, A., 2018. Determinants of the cytosolic turnover of mitochondrial intermembrane space proteins. *BMC Biol.* 16, 66. <https://doi.org/10.1186/s12915-018-0536-1>
- Krimmer, T., Rassow, J., Kunau, W.H., Voos, W., Pfanner, N., 2000. Mitochondrial protein import motor: the ATPase domain of matrix Hsp70 is crucial for binding to Tim44, while the peptide binding domain and the carboxy-terminal segment play a stimulatory role. *Mol. Cell. Biol.* 20, 5879–5887. <https://doi.org/10.1128/MCB.20.16.5879-5887.2000>
- Kumamoto, T., MOROHASHI, K., ITO, A., OMURA, T., 1987. Site-Directed Mutagenesis of Basic Amino Acid Residues in the Extension Peptide of P-450(SCC) Precursor: Effects on the Import of the Precursor into Mitochondria. *J. Biochem. (Tokyo)* 102, 833–838. <https://doi.org/10.1093/oxfordjournals.jbchem.a122122>
- Kuo, M.-L., den Besten, W., Bertwistle, D., Roussel, M.F., Sherr, C.J., 2004. N-terminal polyubiquitination and degradation of the Arf tumor suppressor. *Genes Dev.* 18, 1862–1874. <https://doi.org/10.1101/gad.1213904>
- Kutik, S., Stojanovski, D., Becker, L., Becker, T., Meinecke, M., Krüger, V., Prinz, C., Meisinger, C., Guiard, B., Wagner, R., Pfanner, N., Wiedemann, N., 2008. Dissecting Membrane Insertion of Mitochondrial β -Barrel Proteins. *Cell* 132, 1011–1024. <https://doi.org/10.1016/j.cell.2008.01.028>
- Lee, D.W., Lee, S., Min, C.-K., Park, C., Kim, J.-M., Hwang, C.-S., Park, S.K., Cho, N.-H., Hwang, I., 2020. Cross-Species Functional Conservation and Possible Origin of the N-Terminal Specificity Domain of Mitochondrial Presequences. *Front. Plant Sci.* 11.
- Lee, K.-E., Heo, J.-E., Kim, J.-M., Hwang, C.-S., 2016. N-Terminal Acetylation-Targeted N-End Rule Proteolytic System: The Ac/N-End Rule Pathway. *Mol. Cells* 39, 169–178. <https://doi.org/10.14348/molcells.2016.2329>
- Lee, Y.J., Wickner, R.B., 1992. Mak10, a Glucose-Repressible Gene Necessary for Replication of a Dsrna Virus of *Saccharomyces Cerevisiae*, Has T Cell Receptor α -Subunit Motifs. *Genetics* 132, 87–96.
- Lesnik, C., Cohen, Y., Atir-Lande, A., Schuldiner, M., Arava, Y., 2014. OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nat. Commun.* 5, 5711. <https://doi.org/10.1038/ncomms6711>

- Li, X., Chang, Y.H., 1995. Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. U. S. A.* 92, 12357–12361. <https://doi.org/10.1073/pnas.92.26.12357>
- Liszczyk, G., Goldberg, J.M., Foyn, H., Petersson, E.J., Arnesen, T., Marmorstein, R., 2013. Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nat. Struct. Mol. Biol.* 20, 1098–1105. <https://doi.org/10.1038/nsmb.2636>
- Liu, Y., Wang, X., Coyne, L.P., Yang, Y., Qi, Y., Middleton, F.A., Chen, X.J., 2019. Mitochondrial carrier protein overloading and misfolding induce aggregates and proteostatic adaptations in the cytosol. *Mol. Biol. Cell* 30, 1272–1284. <https://doi.org/10.1091/mbc.E19-01-0046>
- Maccacchini, M.L., Rudin, Y., Blobel, G., Schatz, G., 1979. Import of proteins into mitochondria: precursor forms of the extramitochondrially made F1-ATPase subunits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 76, 343. <https://doi.org/10.1073/pnas.76.1.343>
- Malina, C., Larsson, C., Nielsen, J., 2018. Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Res.* 18, foy040. <https://doi.org/10.1093/femsyr/foy040>
- Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., Jacq, C., 2002. Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep.* 3, 159–164. <https://doi.org/10.1093/embo-reports/kvf025>
- Margeot, A., Garcia, M., Wang, W., Tetaud, E., di Rago, J.P., Jacq, C., 2005. Why are many mRNAs translated to the vicinity of mitochondria: a role in protein complex assembly? *Gene* 354, 64–71. <https://doi.org/10.1016/j.gene.2005.04.022>
- Markesich, D.C., Gajewski, K.M., Nazimiec, M.E., Beckingham, K., 2000. bicaudal encodes the *Drosophila* beta NAC homolog, a component of the ribosomal translational machinery*. *Development* 127, 559–572. <https://doi.org/10.1242/dev.127.3.559>
- Mattiazzi, M., D'Aurelio, M., Gajewski, C.D., Martushova, K., Kiaei, M., Beal, M.F., Manfredi, G., 2002. Mutated human SOD1 causes dysfunction of oxidative phosphorylation in mitochondria of transgenic mice. *J. Biol. Chem.* 277, 29626–29633. <https://doi.org/10.1074/jbc.M203065200>
- McMichael, G., Bainbridge, M.N., Haan, E., Corbett, M., Gardner, A., Thompson, S., van Bon, B.W.M., van Eyk, C.L., Broadbent, J., Reynolds, C., O'Callaghan, M.E., Nguyen, L.S., Adelson, D.L., Russo, R., Jhangiani, S., Doddapaneni, H., Muzny, D.M., Gibbs, R.A., Gecz, J., MacLennan, A.H., 2015. Whole-exome sequencing points to considerable

- genetic heterogeneity of cerebral palsy. *Mol. Psychiatry* 20, 176–182. <https://doi.org/10.1038/mp.2014.189>
- Mehawej, C., Delahodde, A., Legeai-Mallet, L., Delague, V., Kaci, N., Desvignes, J.-P., Kibar, Z., Capo-Chichi, J.-M., Chouery, E., Munnich, A., Cormier-Daire, V., Mégarbané, A., 2014. The Impairment of MAGMAS Function in Human Is Responsible for a Severe Skeletal Dysplasia. *PLOS Genet.* 10, e1004311. <https://doi.org/10.1371/journal.pgen.1004311>
- Miyasaka, H., 1999. The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*. *Yeast* 15, 633–637. [https://doi.org/10.1002/\(SICI\)1097-0061\(19990615\)15:8<633::AID-YEA407>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0061(19990615)15:8<633::AID-YEA407>3.0.CO;2-O)
- Miyata, N., Steffen, J., Johnson, M.E., Fargue, S., Danpure, C.J., Koehler, C.M., 2014. Pharmacologic rescue of an enzyme-trafficking defect in primary hyperoxaluria 1. *Proc. Natl. Acad. Sci.* 111, 14406–14411. <https://doi.org/10.1073/pnas.1408401111>
- Moczko, M., Bömer, U., Kübrich, M., Zufall, N., Hönlinger, A., Pfanner, N., 1997. The intermembrane space domain of mitochondrial Tom22 functions as a trans binding site for preproteins with N-terminal targeting sequences. *Mol. Cell. Biol.* 17, 6574–6584.
- Moerschell, R.P., Hosokawa, Y., Tsunasawa, S., Sherman, F., 1990. The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes c created by oligonucleotide transformation. *J. Biol. Chem.* 265, 19638–19643.
- Moiso, N., Klupsch, K., Fedele, V., East, P., Sharma, S., Renton, A., Plun-Favreau, H., Edwards, R.E., Teismann, P., Esposti, M.D., Morrison, A.D., Wood, N.W., Downward, J., Martins, L.M., 2009. Mitochondrial dysfunction triggered by loss of HtrA2 results in the activation of a brain-specific transcriptional stress response. *Cell Death Differ.* 16, 449–464. <https://doi.org/10.1038/cdd.2008.166>
- Monda, J.K., Scott, D.C., Miller, D.J., Lydeard, J., King, D., Harper, J.W., Bennett, E.J., Schulman, B.A., 2013. Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes. *Struct. Lond. Engl.* 21, 42–53. <https://doi.org/10.1016/j.str.2012.10.013>
- Mossmann, D., Vögtle, F.-N., Taskin, A.A., Teixeira, P.F., Ring, J., Burkhart, J.M., Burger, N., Pinho, C.M., Tadic, J., Loreth, D., Graff, C., Metzger, F., Sickmann, A., Kretz, O., Wiedemann, N., Zahedi, R.P., Madeo, F., Glaser, E., Meisinger, C., 2014. Amyloid- β

- Peptide Induces Mitochondrial Dysfunction by Inhibition of Preprotein Maturation. *Cell Metab.* 20, 662–669. <https://doi.org/10.1016/j.cmet.2014.07.024>
- Moulin, C., Caumont-Sarcos, A., Ieva, R., 2019. Mitochondrial presequence import: Multiple regulatory knobs fine-tune mitochondrial biogenesis and homeostasis. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1866, 930–944. <https://doi.org/10.1016/j.bbamcr.2019.02.012>
- Mughal, A.A., Grieg, Z., Skjellegrind, H., Fayzullin, A., Lamkhannat, M., Joel, M., Ahmed, M.S., Murrell, W., Vik-Mo, E.O., Langmoen, I.A., Stangeland, B., 2015. Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells. *Mol. Cancer* 14, 160. <https://doi.org/10.1186/s12943-015-0432-z>
- Mukhopadhyay, A., Yang, C.-S., Weiner, H., 2006. Binding of mitochondrial leader sequences to Tom20 assessed using a bacterial two-hybrid system shows that hydrophobic interactions are essential and that some mutated leaders that do not bind Tom20 can still be imported. *Protein Sci. Publ. Protein Soc.* 15, 2739–2748. <https://doi.org/10.1110/ps.062462006>
- Mullen, J.R., Kayne, P.S., Moerschell, R.P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F., Sternglanz, R., 1989. Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast. *EMBO J.* 8, 2067–2075.
- Neupert, W., Herrmann, J.M., 2007. Translocation of Proteins into Mitochondria. *Annu. Rev. Biochem.* 76, 723–749. <https://doi.org/10.1146/annurev.biochem.76.052705.163409>
- Nguyen, K.T., Mun, S.-H., Lee, C.-S., Hwang, C.-S., 2018. Control of protein degradation by N-terminal acetylation and the N-end rule pathway. *Exp. Mol. Med.* 50, 1–8. <https://doi.org/10.1038/s12276-018-0097-y>
- Oh, J.-H., Chen, S.-J., Varshavsky, A., 2017. A reference-based protein degradation assay without global translation inhibitors. *J. Biol. Chem.* 292, 21457–21465. <https://doi.org/10.1074/jbc.M117.814236>
- Opaliński, Ł., Song, J., Priesnitz, C., Wenz, L.-S., Oeljeklaus, S., Warscheid, B., Pfanner, N., Becker, T., 2018. Recruitment of Cytosolic J-Proteins by TOM Receptors Promotes Mitochondrial Protein Biogenesis. *Cell Rep.* 25, 2036-2043.e5. <https://doi.org/10.1016/j.celrep.2018.10.083>
- Osellame, L.D., Blacker, T.S., Duchon, M.R., 2012. Cellular and molecular mechanisms of mitochondrial function. *Best Pract. Res. Clin. Endocrinol. Metab., Mitochondria in Endocrinology* 26, 711–723. <https://doi.org/10.1016/j.beem.2012.05.003>

- Palmer, C.S., Anderson, A.J., Stojanovski, D., 2021. Mitochondrial protein import dysfunction: mitochondrial disease, neurodegenerative disease and cancer. *FEBS Lett.* 595, 1107–1131. <https://doi.org/10.1002/1873-3468.14022>
- Persson, B., Flinta, C., von Heijne, G., Jörnvall, H., 1985. Structures of N-terminally acetylated proteins. *Eur. J. Biochem.* 152, 523–527. <https://doi.org/10.1111/j.1432-1033.1985.tb09227.x>
- Polevoda, B., 1999. Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J.* 18, 6155–6168. <https://doi.org/10.1093/emboj/18.21.6155>
- Polevoda, B., Brown, S., Cardillo, T.S., Rigby, S., Sherman, F., 2008. Yeast N α -terminal acetyltransferases are associated with ribosomes. *J. Cell. Biochem.* 103, 492–508. <https://doi.org/10.1002/jcb.21418>
- Polevoda, B., J, H., F, S., 2009. Properties of Nat4, an N(alpha)-acetyltransferase of *Saccharomyces cerevisiae* that modifies N termini of histones H2A and H4. *Mol. Cell. Biol.* 29. <https://doi.org/10.1128/MCB.00147-08>
- Polevoda, B., Sherman, F., 2003. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J. Mol. Biol.* 325, 595–622. [https://doi.org/10.1016/s0022-2836\(02\)01269-x](https://doi.org/10.1016/s0022-2836(02)01269-x)
- Polevoda, B., Sherman, F., 2001. NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. *J. Biol. Chem.* 276, 20154–20159. <https://doi.org/10.1074/jbc.M011440200>
- Ponce-Rojas, J.C., Avendaño-Monsalve, M.C., Yañez-Falcón, A.R., Jaimes-Miranda, F., Garay, E., Torres-Quiroz, F., DeLuna, A., Funes, S., 2017. $\alpha\beta'$ -NAC cooperates with Sam37 to mediate early stages of mitochondrial protein import. *FEBS J.* 284, 814–830. <https://doi.org/10.1111/febs.14024>
- Popov-Čeleketić, D., Mapa, K., Neupert, W., Mokranjac, D., 2008. Active remodelling of the TIM23 complex during translocation of preproteins into mitochondria. *EMBO J.* 27, 1469–1480. <https://doi.org/10.1038/emboj.2008.79>
- Prchal, J.T., Cashman, D.P., Kan, Y.W., 1986. Hemoglobin Long Island is caused by a single mutation (adenine to cytosine) resulting in a failure to cleave amino-terminal methionine. *Proc. Natl. Acad. Sci. U. S. A.* 83, 24–27. <https://doi.org/10.1073/pnas.83.1.24>
- Ree, R., Varland, S., Arnesen, T., 2018. Spotlight on protein N-terminal acetylation. *Exp. Mol. Med.* 50, 90. <https://doi.org/10.1038/s12276-018-0116-z>

- Reid, G.A., Schatz, G., 1982. Import of proteins into mitochondria. Yeast cells grown in the presence of carbonyl cyanide m-chlorophenylhydrazone accumulate massive amounts of some mitochondrial precursor polypeptides. *J. Biol. Chem.* 257, 13056–13061.
- Reinders, J., Zahedi, R.P., Pfanner, N., Meisinger, C., Sickmann, A., 2006. Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J. Proteome Res.* 5, 1543–1554. <https://doi.org/10.1021/pr050477f>
- Roger, A.J., Muñoz-Gómez, S.A., Kamikawa, R., 2017. The Origin and Diversification of Mitochondria. *Curr. Biol.* 27, R1177–R1192. <https://doi.org/10.1016/j.cub.2017.09.015>
- Roise, D., Horvath, S.J., Tomich, J.M., Richards, J.H., Schatz, G., 1986. A chemically synthesized pre-sequence of an imported mitochondrial protein can form an amphiphilic helix and perturb natural and artificial phospholipid bilayers. *EMBO J.* 5, 1327–1334.
- Saint-Georges, Y., Garcia, M., Delaveau, T., Jourden, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., Jacq, C., 2008. Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PloS One* 3, e2293. <https://doi.org/10.1371/journal.pone.0002293>
- Saitoh, T., Igura, M., Obita, T., Ose, T., Kojima, R., Maenaka, K., Endo, T., Kohda, D., 2007. Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J.* 26, 4777–4787. <https://doi.org/10.1038/sj.emboj.7601888>
- Sakhrani, N.M., Padh, H., 2013. Organelle targeting: third level of drug targeting. *Drug Des. Devel. Ther.* 7, 585–599. <https://doi.org/10.2147/DDDT.S45614>
- Savojardo, C., Martelli, P.L., Fariselli, P., Casadio, R., 2015. TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. *Bioinforma. Oxf. Engl.* 31, 3269–3275. <https://doi.org/10.1093/bioinformatics/btv367>
- Schneider, A., 2018. Mitochondrial protein import in trypanosomatids: Variations on a theme or fundamentally different? *PLOS Pathog.* 14, e1007351. <https://doi.org/10.1371/journal.ppat.1007351>
- Shemesh, R., Novik, A., Cohen, Y., 2010. Follow the leader: preference for specific amino acids directly following the initial methionine in proteins of different organisms. *Genomics Proteomics Bioinformatics* 8, 180–189. [https://doi.org/10.1016/S1672-0229\(10\)60020-4](https://doi.org/10.1016/S1672-0229(10)60020-4)

- Shemorry, A., Hwang, C.-S., Varshavsky, A., 2013. Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* 50, 540–551. <https://doi.org/10.1016/j.molcel.2013.03.018>
- Shen, K., Gamberdinger, M., Chan, R., Gense, K., Martin, E.M., Sachs, N., Knight, P.D., Schlömer, R., Calabrese, A.N., Stewart, K.L., Leiendecker, L., Baghel, A., Radford, S.E., Frydman, J., Deuerling, E., 2019. Dual Role of Ribosome-Binding Domain of NAC as a Potent Suppressor of Protein Aggregation and Aging-Related Proteinopathies. *Mol. Cell* 74, 729-741.e7. <https://doi.org/10.1016/j.molcel.2019.03.012>
- Sherman, F., Stewart, J.W., Tsunasawa, S., 1985. Methionine or not methionine at the beginning of a protein. *BioEssays* 3, 27–31. <https://doi.org/10.1002/bies.950030108>
- Shiota, T., Imai, K., Qiu, J., Hewitt, V.L., Tan, K., Shen, H.-H., Sakiyama, N., Fukasawa, Y., Hayat, S., Kamiya, M., Elofsson, A., Tomii, K., Horton, P., Wiedemann, N., Pfanner, N., Lithgow, T., Endo, T., 2015. Molecular architecture of the active mitochondrial protein gate. *Science* 349, 1544–1548. <https://doi.org/10.1126/science.aac6428>
- Shiota, T., Mabuchi, H., Tanaka-Yamano, S., Yamano, K., Endo, T., 2011. In vivo protein-interaction mapping of a mitochondrial translocator protein Tom22 at work. *Proc. Natl. Acad. Sci.* 108, 15179–15183. <https://doi.org/10.1073/pnas.1105921108>
- Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4, 1581–1590. <https://doi.org/10.1002/pmic.200300776>
- Solbiati, J., Chapman-Smith, A., Miller, J.L., Miller, C.G., Cronan, J.E., 1999. Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. *J. Mol. Biol.* 290, 607–614. <https://doi.org/10.1006/jmbi.1999.2913>
- Song, O., Wang, X., Waterborg, J.H., Sternglanz, R., 2003. An N α -Acetyltransferase Responsible for Acetylation of the N-terminal Residues of Histones H4 and H2A *. *J. Biol. Chem.* 278, 38109–38112. <https://doi.org/10.1074/jbc.C300355200>
- Starheim, K.K., Gromyko, D., Evjenth, R., Rynningen, A., Varhaug, J.E., Lillehaug, J.R., Arnesen, T., 2009. Knockdown of human N alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. *Mol. Cell. Biol.* 29, 3569–3581. <https://doi.org/10.1128/MCB.01909-08>
- Stein, K.C., Kriel, A., Frydman, J., 2019. Nascent Polypeptide Domain Topology and Elongation Rate Direct the Cotranslational Hierarchy of Hsp70 and TRiC/CCT. *Mol. Cell* 75, 1117-1130.e5. <https://doi.org/10.1016/j.molcel.2019.06.036>

- Stenström, C.M., Jin, H., Major, L.L., Tate, W.P., Isaksson, L.A., 2001. Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene* 263, 273–284. [https://doi.org/10.1016/S0378-1119\(00\)00550-3](https://doi.org/10.1016/S0378-1119(00)00550-3)
- Suissa, M., Schatz, G., 1982. Import of proteins into mitochondria. Translatable mRNAs for imported mitochondrial proteins are present in free as well as mitochondria-bound cytoplasmic polysomes. *J. Biol. Chem.* 257, 13048–13055.
- Tabrizi, S.J., Workman, J., Hart, P.E., Mangiarini, L., Mahal, A., Bates, G., Cooper, J.M., Schapira, A.H., 2000. Mitochondrial dysfunction and free radical damage in the Huntington R6/2 transgenic mouse. *Ann. Neurol.* 47, 80–86. [https://doi.org/10.1002/1531-8249\(200001\)47:1<80::aid-ana13>3.3.co;2-b](https://doi.org/10.1002/1531-8249(200001)47:1<80::aid-ana13>3.3.co;2-b)
- Tahirov, T.H., Oki, H., Tsukihara, T., Ogasahara, K., Yutani, K., Libeu, C.P., Izu, Y., Tsunasawa, S., Kato, I., 1998. High-resolution crystals of methionine aminopeptidase from *Pyrococcus furiosus* obtained by water-mediated transformation. *J. Struct. Biol.* 121, 68–72. <https://doi.org/10.1006/jsbi.1997.3940>
- Takeda, T., Shirai, K., Kim, Y.-W., Higuchi-Takeuchi, M., Shimizu, M., Kondo, T., Ushijima, T., Matsushita, T., Shinozaki, K., Hanada, K., 2022. A de novo gene originating from the mitochondria controls floral transition in *Arabidopsis thaliana*. *Plant Mol. Biol.* <https://doi.org/10.1007/s11103-022-01320-6>
- Tasaki, T., Zakrzewska, A., Dudgeon, D.D., Jiang, Y., Lazo, J.S., Kwon, Y.T., 2009. The Substrate Recognition Domains of the N-end Rule Pathway. *J. Biol. Chem.* 284, 1884–1895. <https://doi.org/10.1074/jbc.M803641200>
- Tats, A., Remm, M., Tenson, T., 2006. Highly expressed proteins have an increased frequency of alanine in the second amino acid position. *BMC Genomics* 7, 28. <https://doi.org/10.1186/1471-2164-7-28>
- Tercero, J.C., Dinman, J.D., Wickner, R.B., 1993. Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus. *J. Bacteriol.* 175, 3192–3194.
- Tercero, J.C., Wickner, R.B., 1992. MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly. *J. Biol. Chem.* 267, 20277–20281.
- Trexler, A.J., Rhoades, E., 2012. N-Terminal acetylation is critical for forming α -helical oligomer of α -synuclein. *Protein Sci. Publ. Protein Soc.* 21, 601–605. <https://doi.org/10.1002/pro.2056>

- Truscott, K.N., Kovermann, P., Geissler, A., Merlin, A., Meijer, M., Driessen, A.J., Rassow, J., Pfanner, N., Wagner, R., 2001. A presequence- and voltage-sensitive channel of the mitochondrial preprotein translocase formed by Tim23. *Nat. Struct. Biol.* 8, 1074–1082. <https://doi.org/10.1038/nsb726>
- Van Damme, P., Hole, K., Gevaert, K., Arnesen, T., 2015. N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. *Proteomics* 15, 2436–2446. <https://doi.org/10.1002/pmic.201400575>
- Van Damme, P., Hole, K., Pimenta-Marques, A., Helsens, K., Vandekerckhove, J., Martinho, R.G., Gevaert, K., Arnesen, T., 2011. NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet.* 7, e1002169. <https://doi.org/10.1371/journal.pgen.1002169>
- Van Damme, P., Kalvik, T.V., Starheim, K.K., Jonckheere, V., Myklebust, L.M., Menschaert, G., Varhaug, J.E., Gevaert, K., Arnesen, T., 2016. A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity. *Mol. Cell. Proteomics MCP* 15, 3361–3372. <https://doi.org/10.1074/mcp.M116.061010>
- Van Damme, P., Lasa, M., Polevoda, B., Gazquez, C., Elosegui-Artola, A., Kim, D.S., De Juan-Pardo, E., Demeyer, K., Hole, K., Larrea, E., Timmerman, E., Prieto, J., Arnesen, T., Sherman, F., Gevaert, K., Aldabe, R., 2012. N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proc. Natl. Acad. Sci. U. S. A.* 109, 12449–12454. <https://doi.org/10.1073/pnas.1210303109>
- van Loon, A.P., Brändli, A.W., Schatz, G., 1986. The presequences of two imported mitochondrial proteins contain information for intracellular and intramitochondrial sorting. *Cell* 44, 801–812. [https://doi.org/10.1016/0092-8674\(86\)90846-9](https://doi.org/10.1016/0092-8674(86)90846-9)
- van Loon, A.P., Young, E.T., 1986. Intracellular sorting of alcohol dehydrogenase isoenzymes in yeast: a cytosolic location reflects absence of an amino-terminal targeting sequence for the mitochondrion. *EMBO J.* 5, 161–165. <https://doi.org/10.1002/j.1460-2075.1986.tb04191.x>
- Varshavsky, A., 1996. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12142–12149. <https://doi.org/10.1073/pnas.93.22.12142>
- Vögtle, F.-N., Burkhart, J.M., Gonczarowska-Jorge, H., Kücükköse, C., Taskin, A.A., Kopczynski, D., Ahrends, R., Mossmann, D., Sickmann, A., Zahedi, R.P., Meisinger, C., 2017. Landscape of submitochondrial protein distribution. *Nat. Commun.* 8, 290. <https://doi.org/10.1038/s41467-017-00359-0>

- Vögtle, F.-N., Prinz, C., Kellermann, J., Lottspeich, F., Pfanner, N., Meisinger, C., 2011. Mitochondrial protein turnover: role of the precursor intermediate peptidase Oct1 in protein stabilization. *Mol. Biol. Cell* 22, 2135–2143. <https://doi.org/10.1091/mbc.E11-02-0169>
- Vögtle, F.-N., Wortelkamp, S., Zahedi, R.P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., Meisinger, C., 2009. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* 139, 428–439. <https://doi.org/10.1016/j.cell.2009.07.045>
- von Heijne, G., 1986. Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J.* 5, 1335–1342.
- Vyas, S., Zaganjor, E., Haigis, M.C., 2016. Mitochondria and Cancer. *Cell* 166, 555–566. <https://doi.org/10.1016/j.cell.2016.07.002>
- Wang, F., Durfee, L.A., Huibregtse, J.M., 2013. A cotranslational ubiquitination pathway for quality control of misfolded proteins. *Mol. Cell* 50, 368–378. <https://doi.org/10.1016/j.molcel.2013.03.009>
- Wang, Y., Xu, E., Musich, P.R., Lin, F., 2019. Mitochondrial dysfunction in neurodegenerative diseases and the potential countermeasure. *CNS Neurosci. Ther.* 25, 816–824. <https://doi.org/10.1111/cns.13116>
- Weill, U., Yofe, I., Sass, E., Stynen, B., Davidi, D., Natarajan, J., Ben-Menachem, R., Avihou, Z., Goldman, O., Harpaz, N., Chuartzman, S., Kniazev, K., Knoblach, B., Laborenz, J., Boos, F., Kowarzyk, J., Ben-Dor, S., Zalckvar, E., Herrmann, J.M., Rachubinski, R.A., Pines, O., Rapaport, D., Michnick, S.W., Levy, E.D., Schuldiner, M., 2018. Genome-wide SWAp-Tag yeast libraries for proteome exploration. *Nat. Methods* 15, 617–622. <https://doi.org/10.1038/s41592-018-0044-9>
- Wiedemann, N., Kozjak, V., Chacinska, A., Schönfisch, B., Rospert, S., Ryan, M.T., Pfanner, N., Meisinger, C., 2003. Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* 424, 565–571. <https://doi.org/10.1038/nature01753>
- Wiedmann, B., Sakai, H., Davis, T.A., Wiedmann, M., 1994. A protein complex required for signal-sequence-specific sorting and translocation. *Nature* 370, 434–440. <https://doi.org/10.1038/370434a0>
- Williams, C.C., Jan, C.H., Weissman, J.S., 2014. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* 346, 748–751. <https://doi.org/10.1126/science.1257522>

- Wrobel, L., Topf, U., Bragoszewski, P., Wiese, S., Sztolsztener, M.E., Oeljeklaus, S., Varabyova, A., Lirski, M., Chroscicki, P., Mroczek, S., Januszewicz, E., Dziembowski, A., Koblowska, M., Warscheid, B., Chacinska, A., 2015. Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* 524, 485–488. <https://doi.org/10.1038/nature14951>
- Xu, Y.-F., Zhang, Y.-J., Lin, W.-L., Cao, X., Stetler, C., Dickson, D.W., Lewis, J., Petrucelli, L., 2011. Expression of mutant TDP-43 induces neuronal dysfunction in transgenic mice. *Mol. Neurodegener.* 6, 73. <https://doi.org/10.1186/1750-1326-6-73>
- Yamano, K., Yatsukawa, Y.-I., Esaki, M., Hobbs, A.E.A., Jensen, R.E., Endo, T., 2008. Tom20 and Tom22 share the common signal recognition pathway in mitochondrial protein import. *J. Biol. Chem.* 283, 3799–3807. <https://doi.org/10.1074/jbc.M708339200>
- Yan, J., Bao, E., Yu, J., 2009. Heat shock protein 60 expression in heart, liver and kidney of broilers exposed to high temperature. *Res. Vet. Sci.* 86, 533–538. <https://doi.org/10.1016/j.rvsc.2008.09.002>
- Young, J.C., Hoogenraad, N.J., Hartl, F.U., 2003. Molecular Chaperones Hsp90 and Hsp70 Deliver Preproteins to the Mitochondrial Import Receptor Tom70. *Cell* 112, 41–50. [https://doi.org/10.1016/S0092-8674\(02\)01250-3](https://doi.org/10.1016/S0092-8674(02)01250-3)
- Zanphorlin, L.M., Lima, T.B., Wong, M.J., Balbuena, T.S., Minetti, C.A.S.A., Remeta, D.P., Young, J.C., Barbosa, L.R.S., Gozzo, F.C., Ramos, C.H.I., 2016. Heat Shock Protein 90 kDa (Hsp90) Has a Second Functional Interaction Site with the Mitochondrial Import Receptor Tom70. *J. Biol. Chem.* 291, 18620–18631. <https://doi.org/10.1074/jbc.M115.710137>
- Zhang, Z., Tsukikawa, M., Peng, M., Polyak, E., Nakamaru-Ogiso, E., Ostrovsky, J., McCormack, S., Place, E., Clarke, C., Reiner, G., McCormick, E., Rappaport, E., Haas, R., Baur, J.A., Falk, M.J., 2013. Primary Respiratory Chain Disease Causes Tissue-Specific Dysregulation of the Global Transcriptome and Nutrient-Sensing Signaling Network. *PLoS ONE* 8, e69282. <https://doi.org/10.1371/journal.pone.0069282>
- Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M., Hall, T.M.T., 2009. A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20192–20197. <https://doi.org/10.1073/pnas.0812079106>
- Zubko, E.I., Zubko, M.K., 2014. Deficiencies in mitochondrial DNA compromise the survival of yeast cells at critically high temperatures. *Microbiol. Res.* 169, 185–195. <https://doi.org/10.1016/j.micres.2013.06.011>

Table des figures :

Figure 1 : Schéma simplifié d'une mitochondrie avec ses 4 sous-compartiments.....	9
Figure 2 : Répartition des 818 protéines identifiées en fonction de leur compartiment	11
Figure 3 : Les différents signaux d'adressage des protéines mitochondriales	14
Figure 4 : Organisation des différentes protéines du complexe TOM.	17
Figure 5 : Représentation schématique du complexe TIM23	18
Figure 6 : : La voie d'adressage des protéines avec MTS vers la membrane interne et la matrice mitochondriale.....	25
Figure 7 : Les différents sites de clivages des MTS.....	27
Figure 8 : Les différentes étapes de l'import des protéines mitochondriales qui ont pour destination la membrane interne	29
Figure 9 : Schéma simplifié de la machine d'importation des protéines mitochondriales dans la membrane externe	30
Figure 10 : Schéma de l'import des protéines mitochondriales avec MTS	32
Figure 11 : Importation et repliement des protéines riches en cystéine dans l'espace intermembranaire de la mitochondrie.	33
Figure 12 : Une boucle de rétroaction pour inhiber le signal mPOS	35
Figure 13 : Schéma de l'import post et cotraductionnelle des protéines mitochondriales.....	41
Figure 14 : La modification chimique de la N acétylation.....	44
Figure 15 : La technique du COFRADIC	46
Figure 16 : Tableau récapitulatif des différentes NATs chez la levure.....	49
Figure 17 : Test de croissance en gouttes de différents mutants sur YPD et YPGly	49
Figure 18 : La règle de la N end Rule	51
Figure 19 : Les protéines sont ciblées par les Nats en fonction de leur résidu N-terminal.....	53
Figure 20 : Ce diagramme représente la complémentarité fonctionnelle entre les voies de la règle Arg/N-end et de la règle Ac/N-end	54
Figure 21 : Résumé schématique des différentes conséquences d'une acétylation N-terminale des protéines.....	56
Figure 22 : Fréquences des 20 acides aminés en position 2 chez différentes espèces	58
Figure 23 : Histogramme obtenu à partir d'analyse in silico qui représente la fréquence des 20 acides aminés en position 2 par rapport à leurs fréquences dans le protéome	59

Figure 24 : Analyse de l'enrichissement des différents acides aminés en position 2 selon le compartiment cellulaire dans le protéome de <i>S. cerevisiae</i>	60
Figure 25 : Diagramme circulaire représentant pour chaque acide aminé la part des termes GO sélectionné par l'analyse de biais au niveau du RNt.....	61
Figure 26 : Le selTrap révéla le rôle potentiel de NatC sur les précurseurs mitochondriaux ..	63
Figure 27 : Test de croissance en gouttes des mutants souches YPH499 et YPH-pam16Δ-hpam16(N76D) sur milieu riche (YPD) et sur milieu avec une source de carbone non fermentable (YPGly). Le test de croissance est réalisé sur 3 jours d'incubation.....	67
Figure 28 : Western Blot des souches mutantes au niveau du deuxième acide aminé pour la protéine Hsp60	118
Figure 29 : Western blot des souches délétées de NatC dans le fond génétique <i>YPH-pam16Δ-hpam16(N76D)</i>	120
Figure 30 : Test de croissance en gouttes de souches mutantes de NatC sur YPD et YPGly	122
Figure 31 : Test de vérification des souches Rho ^o	123
Figure 32 : Test de croisance des souches mutantes de NatC	126
Figure 33 : Western blot des souches délétées de NATC dans le fond génétique YPH499 ..	127
Figure 34 : Western blot des souches délétées de NATC dans le fond génétique <i>YPH-pam16Δ-hpam16(N76D)</i>	128
Figure 35 : Test de croissance en gouttes de souches mutantes de NatC et/ou de Ubr1 sur YPD, YPGly et YPE	131
Figure 36 : Western blot des souches délétées de NATC et/ou d'UBR1 dans le fond génétique <i>YPH499</i>	132
Figure 37 : Western blot des souches délétées dans le fond génétique YPH499 exposées à du CCCP.....	133
Figure 38 : Photographies des boîtes de pétri après transformation avec des plasmides pHSPmyc et pHSPmyc(L>S).....	134
Figure 39 : Western blot des souches délétées de NATC dans le fond génétique <i>YPH-pam16Δ-hpam16(N76D)</i>	136
Figure 40 : Western blot de 5 souches délétées de NatC dans le fond génétique <i>YPH-pam16Δ-hpam16(N76D)</i>	137
Figure 41 : Western blot des souches délétées de NatC dans le fond génétique <i>YPH-pam16Δ-hpam16(N76D)</i> avec un plasmide pMak1 #1 ou 2.....	138

Table des tableaux

Tableau 1 : Tableau récapitulatif de l'ensemble des souches obtenues dans les fonds génétiques <i>YPH499</i> et <i>YPH-pam16Δ-hpam16(N76D)</i>	130
Tableau 2 : Liste des souches utilisées lors de mon stage.....	150
Tableau 3 : Séquences des oligonucléotides utilisés pour les PCR.....	153
Tableau 4 : Caractéristiques des anticorps utilisés pour l'analyse de Western Blot.....	154

Résumé en anglais :

Mitochondria are complex organelles involving a thousand proteins, most of which are encoded in the nuclear genome. Their biogenesis has required the evolutionary development of efficient protein addressing and import systems, and failures of these systems are associated with serious pathologies, neuropathies, cardiovascular disorders, myopathies, neurodegenerative diseases and cancers.

Many mitochondrial proteins have an N-terminal addressing sequence called MTS (Mitochondrial Targeting Sequence) which forms an amphiphilic alpha helix essential for their mitochondrial import. However, the sequence of the various MTSs is highly variable and their critical characteristics are not yet well understood. The starting point of my thesis was the discovery in yeast of an overrepresentation of 4 hydrophobic amino acids (F, L, I, W) at position 2 of the MTSs sequences. During my thesis, I was able to confirm the critical role of the nature of the residue in position 2 of the MTSs through directed mutagenesis experiments. Indeed, thanks to the development of an innovative system for screening import defects based on the functional rescue of the toxicity of a mitochondrial protein, I was able to observe that only residues overrepresented at position 2 of mitochondrial proteins allowed efficient import. My work has thus demonstrated the existence of strong evolutionary constraints at position 2 of MTSs, the understanding of which could ultimately be useful for optimising the mitochondrial addressing of therapeutic proteins in patients suffering from mitochondrial diseases.

Key words : Position 2 in proteins, N-terminal modification enzymes, MTS, NatC, *S. cerevisiae*

Résumé en français :

Les mitochondries sont des organites complexes impliquant un millier de protéines, la plupart codées dans le génome nucléaire. Leur biogenèse a nécessité au cours de l'évolution la mise en place de systèmes efficaces d'adressage et d'import protéique, et des défaillances de ces systèmes sont associées à des pathologies graves, neuropathies, troubles cardiovasculaires, myopathies, maladies neurodégénératives ainsi que cancers.

De nombreuses protéines mitochondriales possèdent en N-terminal une séquence d'adressage appelée MTS (Mitochondrial Targeting sequence) qui forme une hélice alpha amphiphile essentielle pour leur import mitochondrial. La séquence des divers MTSs est néanmoins très variables et leur caractéristiques critiques ne sont pas encore bien comprises. Le point de départ de ma thèse est la découverte, chez les levures, d'une surreprésentation en position 2 des séquences MTSs de 4 acides aminés hydrophobes (F, L, I, W). Au cours de mes années de thèse, j'ai pu confirmer, par des expériences de mutagenèse dirigée, le rôle critique de la nature du résidu en position 2 des MTSs. En effet, grâce au développement d'un système novateur de criblage des défauts d'import basé sur le sauvetage fonctionnel de la toxicité d'une protéine mitochondriale, j'ai pu observer que seuls les résidus surreprésentés en position 2 des protéines mitochondriales permettaient un import efficace. Mon travail a ainsi démontré l'existence de fortes contraintes évolutives s'exerçant au niveau de la position 2 des MTSs dont la compréhension pourrait à terme être utile pour optimiser l'adressage mitochondrial de protéines thérapeutiques chez des patients atteints de maladies mitochondriales.

Mots clefs : Position 2 des protéines, Enzymes de modifications N-terminales, MTS, NatC, *S. cerevisiae*