UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Mechanisms of protein dysfunction in aminoacyl-tRNA synthetase related to neurological diseases

Diogo Miguel Fernandes Meireles Ferreira

Mestrado em Bioquímica
Especialização em Bioquímica

Dissertação orientada por:
Professora Doutora Bárbara J. Henriques
Professor Doutor Cláudio M. Gomes

2021

Acknowledgements

First of all, I want to thank my supervisors Professor Bárbara Henriques and Professor Cláudio Gomes, for the guidance, patience, support, teachings and wise advices throughout this journey, it was amazing to be part of such a brilliant, talented, motivated and focused team. Thank you for the opportunity to experience, and in some way participate, in such a high-quality investigation environment and be part of an elite squad incredibly led.

I would like to thank Professor Rita Horvath and Dr. Denisa Hathazi (University of Cambridge, UK) for providing the precious hEARS2 and hRARS2 plasmids, an important kick off for this project.

To my kind colleagues at Protein Misfolding and Amyloids in Biomedicine lab, Joana Ribeiro, Guilherme Moreira, Romina Coelho, Filipa Carvalho, António Figueira and Margarida Simões. I want to thank you all, for the always productive and animated coffee-break talks, the friendship and the kindness, the help provided whenever I needed, neither one of you has refused once to help me and for welcoming me since day one, it was a pleasure to meet you and work side by side with each one of you. I will always remember the team work environment I felt in the "cleaning lab day", tough day, but you made it look a lot easier, this is what team work is all about. I would like to thank also André Gonzaga, for the company till late hours, the "Western blot adventure tours" and the friendship.

To all my friends who helped me pass through intense and hard times, especially Inês and Jéssica, you were the catalysts of my happiness in tough moments, over and over again, thank you for that and much more, 5 years of solid friendship and counting.

Last but not the least, a special thanks to my family members that have supported me, especially my mother, who never failed me throughout the entire time. You were absolutely tireless and I will always thank you for it, without you this would never happen. I will miss the days in which I come home and the first question I am asked is "Diogo, do you have protein?".

Abstract

Mitochondrial diseases (MD) are heterogeneous human diseases and one of the most common group of inherited metabolic diseases, with an estimated incidence of 1:5000 individuals. The number of reported mutations associated with this type of diseases has been increasing for the past years, however, due to their complexity, they are still very challenging at clinic and therapeutic level. Therefore, it is important to understand the molecular mechanism underlying these diseases, to develop successful diagnosis and therapies.

Recently, several reports identified mutations on genes coding for mitochondrial aminoacyl-tRNA synthetases (mt-aaRSs), as a cause of different MDs. These enzymes are nuclear encoded and perform their function inside mitochondria. They catalyze the specific attachment of the corresponding amino acid to the respective transfer RNA (tRNA), thus ensuring the fidelity of protein translation, especially of the complexes from the respiratory chain. These complexes are encoded by mitochondrial DNA (mt DNA) and their synthesis occurs inside the mitochondria, therefore, the action of mt-aaRSs is extremely important for their formation. As such, disease-associated variants of mt-aaRSs may impair the correct synthesis of the respiratory chain complexes, which ultimately lead to mitochondrial dysfunction.

The main goal of the project is to clarify the molecular mechanism of MD diseases associated to mt-aaRSs disease variants. Specifically, here we focus on two mt-aaRSs: human mitochondrial glutamyl-tRNA synthetase (hEARS2) and human mitochondrial arginyl-tRNA synthetase (hRARS2), whose variants have been reported to be associated to leukodystrophies. Our aim is to establish for the first-time protocols for heterologous expression and purification of these two proteins, and structurally characterize them.

We successfully established a protocol for expression and purification of hEARS2. For this protein we were able to obtain a significant amount of pure enzyme, which allow us to perform protein structural characterization. Our results showed that heterologous expressed hEARS2 presents an α/β fold, and in the native state tryptophan residues are exposed to the solvent. Stability studies, following loss of secondary structure, showed that hEARS2 thermal unfolding has a cooperative transition with an apparent melting temperature (T_m ^{app}) of ~ 68 ºC. Interestingly, monitoring changes at the tertiary structure level, hEARS2 presented a lower melting temperature, with a T_m ^{app} of 51 ºC. Moreover, in the presence of increasing concentrations of its substrate, glutamate, protein alters its native conformation, however we did not observe major changes in melting temperature.

Regarding hRARS2 we established a protocol for protein expression and purification, however the protocol for protein purification must be further optimized. Nevertheless, we were able to perform preliminary analysis of protein structure and stability. As purified hRARS2 presented an α/β fold with a T_m^{app} of ~ 61 ºC, that value decreased when tryptophan emission was being followed.

Overall, these data represent an important breakthrough on the heterologous production of two human mt-aaRS proteins. Moreover, these data provide an important basis for future studies on disease variants, that will permit to elucidate how mutations lead to disease phenotypes.

Keywords: Bacterial heterologous expression, protein biochemistry, human EARS2, human RARS2, Mitochondrial Diseases (MD)

Resumo

As doenças mitocondriais (MD) representam um grupo de doenças humanas inatas heterogéneas, com elevada heterogeneidade clínica. São consideradas das mais comuns entre as doenças metabólicas com origem genética, com uma prevalência de cerca de 1:5000 indivíduos. Nos últimos anos, o número de estudos nos quais foram identificadas mutações associadas com este tipo de doenças tem vindo a aumentar, no entanto, devido à complexidade das MD, é difícil estabelecer uma correlação entre o genótipo e o fenótipo, e assim estabelecer terapêuticas. Desta forma, torna-se importante perceber o mecanismo molecular associado às MD, e assim, desenvolver estratégias de diagnóstico e de terapias eficientes para fazer face a este problema.

Recentemente, vários estudos reconheceram diferentes MD como resultado de mutações em genes que codificam para as aminoacil-tRNA sintetases mitocondriais (mt-aaRSs), levando à identificação de várias variantes com potencial patológico. A função destas proteínas é catalisar a ligação entre um aminoácido e o seu respetivo RNA de transferência (tRNA) e assim assegurar a correta tradução das proteínas. No caso das mt-aaRSs, elas são codificadas a partir do genoma nuclear, e importadas para a mitocôndria onde desempenham a sua função na matriz mitocondrial. As mt-aaRSs são essenciais para a síntese das proteínas que constituem os complexos da cadeia respiratória, dado que estes complexos são codificados pelo genoma mitocondrial. Como tal, variantes patológicas destas proteínas associadas a MD, podem levar à síntese deficiente dos complexos, que em última instância pode causar disfunção mitocondrial. A grande maioria dos estudos relacionados com estas mutações tem-se focado essencialmente na parte clínica, sendo escassos os estudos ao nível dos aspetos bioquímicos e estruturais das proteínas e das respetivas variantes.

O objetivo principal deste projeto é clarificar o mecanismo molecular das MD associadas a mutações em mt-aaRSs, estabelecendo uma correlação entre genótipo e fenótipo. Especificamente, este projeto de tese foca-se no estudo de duas mt-aaRSs humanas, a glutamil-tRNA sintetase mitocondrial (hEARS2) e a arginil-tRNA sintetase mitocondrial (hRARS2), proteínas cujas variantes proteicas foram reportadas como estando associadas a leucodistrofias. Especificamente, os objetivos deste trabalho foram a implementação de protocolos de expressão heteróloga e purificação de ambas as proteínas, e posteriormente a sua caracterização estrutural *in vitro*.

Para as proteínas em estudo não há disponível, até à data, estrutura cristalográfica, neste sentido e de modo a ajudar nos estudos subsequentes, foram efetuados estudos *in silico* recorrendo a modelação das estruturas proteicas por homologia. Os modelos obtidos foram posteriormente comparados com os modelos obtidos pelo software AlphaFold (2021) e os resultados foram muito semelhantes. Para hEARS2, o modelo foi obtido a partir de EARS de *T. thermophilus*, e apresenta uma estrutura global aparentemente achatada e pouco compacta. Pode verificar-se ainda a existência predominante de hélices α, e também folhas β, mas em menor percentagem. Observa-se também a presença de vários loops. Relativamente a hRARS2, o modelo foi obtido a partir de RARS de *S. cerevisiae* e pela análise do modelo, visualiza-se uma estrutura global mais compacta com os elementos de estrutura secundária hélices α e folhas β presentes.

Seguidamente investigou-se quais as melhores condições de expressão proteica em *E. coli*, para tal usou-se a estirpe Rosetta, adequada à expressão de codões nativos, de acordo com o plasmídeo disponível. Usando sempre o mesmo meio de crescimento (meio Luria Broth (LB)), testaram-se diferentes tempos de indução e quantidade de indutor, isopropil β-D-1-tiogalactopiranósido (IPTG), assim como diferentes temperaturas. Os perfis de expressão para hEARS2 e hRARS2 nas diferentes condições foram analisados

por SDS-PAGE e Western blot. Desta análise observamos que ambas as proteínas são maioritariamente expressas na fração insolúvel, no entanto optou-se por dar preferência às condições que otimizavam a expressão na fração solúvel. Em relação à hEARS2 a condição ótima para expressão foi: crescimento durante 4 horas após indução com 0.5 mM IPTG a uma temperatura de 30 ºC e para hRARS2 foi crescimento durante 6 horas após indução com 1 mM IPTG a uma temperatura de 30 ºC. Posteriormente, nas condições otimizadas realizaram-se crescimentos bacterianos em grandes volumes (4 litros) a fim de obter quantidade suficiente de células para posterior purificação das respetivas proteínas.

A partir das células obtidas preparou-se um extrato proteico solúvel que foi submetido seguidamente a uma cromatografia de afinidade com vista à obtenção de frações puras de cada uma das proteínas. Relativamente a hEARS2, o tampão utilizado para esta cromatografia continha 30 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM-500mM imidazole (gradiente) e 20% glicerol. A fração correspondente a hEARS2 foi submetida seguidamente a uma cromatografia de exclusão molecular por desalting, de forma a trocar o tampão rapidamente e eliminar potenciais efeitos negativos por parte do imidazole e do sal. O tampão utilizado continha 30 mM Tris-HCl pH 8.0, 50 mM NaCl e 20% glicerol e a proteína foi armazenada a -80 ºC, com grau de pureza de cerca de 95%.

No que diz respeito à proteína hRARS2, não nos foi possível estabelecer um protocolo otimizado para a purificação da mesma, sendo que das várias tentativas realizadas não se obteve uma fração com a pureza necessária.

De seguida foram efetuados estudos relativamente à estrutura, ao folding e à estabilidade das proteínas recorrendo a técnicas biofísicas, como dicroísmo circular (CD) e espetroscopia de fluorescência, recorrendo à emissão dos triptofanos da proteína. Relativamente aos estudos estruturais da hEARS2, os resultados obtidos por CD mostram que a proteína possui uma estrutura secundária com α/β fold e uma temperatura de desnaturação aparente (T_m^{app}) de 68 ºC, com uma transição cooperativa. Analisando a estrutura terciária da hEARS2 através de emissão de fluorescência dos triptofanos, os dados indicam que os 5 triptofanos presentes nesta proteína estão expostos ao solvente na sua forma nativa. A temperatura de desnaturação aparente obtida, monitorizando a alteração na estrutura terciaria, foi de 51 ºC, o que sugere que a estrutura terciária é perdida antes da estrutura secundária. Para a hEARS2 testou-se ainda qual o efeito do seu substrato, glutamato, na sua estrutura e estabilidade, para isso analisou-se o sinal de dicroísmo circular na região o far-UV. Os resultados obtidos sugerem que com o aumento das concentrações de substrato, a proteína perde a sua conformação inicial. No entanto, observa-se que a estabilidade proteica é mantida, uma vez que não há alteração do valor da temperatura desnaturação aparente.

Relativamente à hRARS2, os estudos estruturais realizados têm um carater muito preliminar dado o grau de pureza da fração obtida. Os espectros de CD obtidos indicam que esta proteína terá possivelmente uma estrutura secundária α/β fold e uma temperatura de desnaturação aparente de cerca de 61 ºC. Observou-se que a desnaturação térmica seguida por CD apresenta um perfil não cooperativo, provavelmente resultado da mistura de proteínas que compõe a amostra analisada. De acordo com os resultados obtidos da estrutura terciária, os triptofanos encontram-se igualmente expostos ao solvente, e após desnaturação térmica, foi obtida uma transição cooperativa com uma temperatura de desnaturação aparente de 53 ºC.

No geral, efetuamos com sucesso a expressão e purificação de hEARS2, bem como conseguimos ter alguns resultados relativos à sua caracterização estrutural. Para hRARS2 conseguimos expressar a proteína, no entanto, a sua purificação necessita ser otimizada.

Serão ainda necessários estudos mais aprofundados para complementar aspetos estruturais relevantes destas proteínas na sua forma wild-type, mas também estudos relativamente à estrutura e à

conformação das respetivas variantes associadas a MD, em particular a leucodistrofias. O estudo detalhado ao nível da estrutura e função de variantes patológicas irá permitir perceber de que forma essas alterações genéticas se relacionam com estados patológicos. Desta forma, será possível determinar como é que estas mutações levam ao desenvolvimento de um fenótipo de doença, possibilitando também o desenvolvimento de terapias eficientes, capazes de mitigar os efeitos patológicos adjacentes às variantes de mt-aaRSs.

Palavras-chave: Expressão heteróloga por bactérias, bioquímica de proteínas, EARS2 humana, RARS2 humana, Doenças Mitocondriais (MD)

Thesis communications

Part of this work has been presented in the form of poster communication in:

- Diogo M. Ferreira, Denisa Hathazi, Rita Horvath, Cláudio M. Gomes, Bárbara J. Henriques, “On the track of mitochondrial aminoacyl-tRNA synthetase-related neurological diseases - Expression and purification of EARS2 protein”. XVII International SPDM Symposium, 8-10 September 2021, Fátima, Portugal

Table of Contents

List of Figures

List of Tables

Abbreviations

A site	Aminoacyl-tRNA site
aaRSs	Aminoacyl-tRNA synthetases
AARS2	Mitochondrial alanyl-tRNA synthetase
AIMPs	Aminoacyl-tRNA synthetase interacting multifunctional proteins
CD	Circular dichroism
CNS	Central nervous system
Cyt-aaRSs	Cytosolic aminoacyl-tRNA synthetase
DARS2	Mitochondrial aspartyl-tRNA synthetase
EARS	Cytosolic glutamyl-tRNA synthetase
EARS2	Mitochondrial glutamyl-tRNA synthetase
hEARS2	Human mitochondrial glutamyl-tRNA synthetase
EF	Elongation factors
E site	Exit site
GARS	Glycil-tRNA synthetase
HARS2	Mitochondrial histidyl-tRNA synthetase
HSP	Heat shock protein
IF	Initiation factors
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KARS	Lysyl-tRNA synthetase
LARS2	Mitochondrial leucyl-tRNA synthetase
LB	Luria Broth
LTBL	Leukoencephalopathy with thalamus and brainstem involvement and high lactate
MARS2	Mitochondrial methionyl-tRNA synthetase
MD	Mitochondrial Diseases
mdeg	Millidegrees
MiSynPat	Mitochondrial aminoacyl-tRNA Synthetases and Pathology
mRNA	Messenger RNA
MSCs	Multi-synthetase complexes
mt-aaRSS	Mitochondrial aminoacyl-tRNA synthetases
mt-DNA	Mitochondrial DNA
NaCl	Sodium chloride
NARS2	Mitochondrial asparaginyl-tRNA synthetase
NGS	Next generation sequencing
N-helix	Amino-terminal helix
OXPHOS	Oxidative phosphorylation
PCH6	Pontocerebellar Hypoplasia type 6
PMSF	Phenylmethylsulfonyl fluoride
P site	Peptidyl-site
PylRS	Pyrroslysil-tRNA synthetase
QARS2	Mitochondrial glutaminyl-tRNA synthetase

RARS	Cytosolic arginyl-tRNA synthetase
RARS2	Mitochondrial arginyl-tRNA synthetase
hRARS2	Human mitochondrial arginyl-tRNA synthetase
RF	Release factors
rRNA	Ribosomal RNA
SepRS	Phosphoseryl-tRNA synthetase
T_m^{app}	Apparent melting temperature
tRNA	Transfer RNA
WARS2	Mitochondrial tryptophanyl-tRNA synthetase
λ_{max}	Maximum wavelength

1. Introduction

1.1. Protein production in the cell

1.1.1. From protein synthesis to protein folding

A typical eukaryotic cell is made up of several compartments and organelles, each one with specific roles and functions which require a distinct set of macromolecules to function appropriately[1]. From them, proteins stand out due to their diverse biological functions ranging from supporting DNA replication, molecule transport (e.g. oxygen), cytoskeletal structures formation, protein repression, assist other proteins during the folding process, protein synthesis, etc[2]. Also, proteins mediate countless processes taking place in cell, like products synthesis, secretion of unwanted molecules and generation of chemical energy indispensable for the maintenance of all these processes[2]. These macromolecules must be synthesized in response to the cell´s current needs, transported to their cellular location and degraded when no longer needed[1]. For nuclear-encoded proteins, protein synthesis begins in the nucleus and is concluded in the ribosome. Some proteins are encoded by mitochondrial DNA (mt-DNA), and so, their synthesis occur entirely in the mitochondria (in humans, mt-DNA contains genes coding for only 13 proteins, that are subunits of the inner mitochondrial membrane respiratory chain complexes)[3, 4]. Protein synthesis can be divided in three major steps: transcription, translation and posttranslational modifications[3]. Regarding nuclear-encoded proteins, firstly, DNA is activated (**Figure 1.1** – (1) Activation) and RNA polymerase catalyzes the formation of a pre-messenger RNA (pre-mRNA), a RNA strand with a base sequence complementary to one of the DNA strands, except for uracil that is replaced by a thymine (**Figure 1.1** – (2) Transcription). This molecule is then processed (**Figure 1.1** – (3) Processing) and the resulting molecule (mature mRNA) is exported to the cytosolic environment, more specifically to the ribosome, where it is utilized for protein synthesis (**Figure 1.1** – (4) Translation)[1, 3]. The mRNA strand is composed of triplets each coding for a specific amino acid.

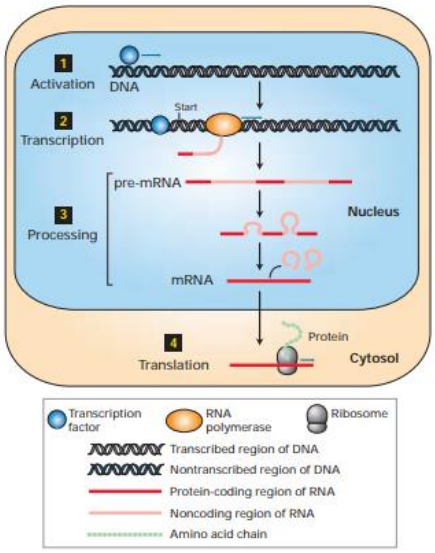


Figure 1.1 – Scheme of conversion of DNA into a protein through a series of processes.
(1): Transcription factors bind to DNA activating a specific gene. (2): RNA polymerase begins transcription of the activated gene, linking nucleotide into a pre-mRNA molecule, using only one strand of DNA as template. (3): The pre-mRNA is processed to remove noncoding sequences, forming a mature mRNA molecule. (4): Mature mRNA then bounds to ribosome in cytoplasmic environment, from which it is formed a linear chain of amino acids linked to each other, resulting in the protein synthesis. From [5].

At the ribosome, translation of mRNA occurs and it is divided in three main stages, initiation, elongation and termination. Before initiating translation, a link must be established between each amino acid and the information in the mRNA, and this is obtained by attaching a transfer RNA (tRNA) to the respective amino acid. This process of attachment of the right amino acid to the right tRNA is critical to obtain a correct protein synthesis, and is catalyzed by a Mg^{2+}-dependent activating enzymes known as aminoacyl-tRNA synthetases (aaRSs) (to be discussed in the next section). After the binding, the tRNA are named as tRNA-charged[1, 5]. At the initiation phase, firstly occurs a dissociation between ribosomal subunits 40S and 60S and the binding of mRNA to 40S subunit. Proteins known as initiation factors (IF) play an important role, as they assist a series of complex reactions of hydrolysis and conformational changes. This result in the formation of an initiation complex, charged with the pair Met-tRNA at the start codon position of the mRNA (P site), and ready to accept the next pair aminoacyl-tRNA, at the next codon position (A site), starting from there on the elongation phase (**Figure 1.2**)[2, 5].

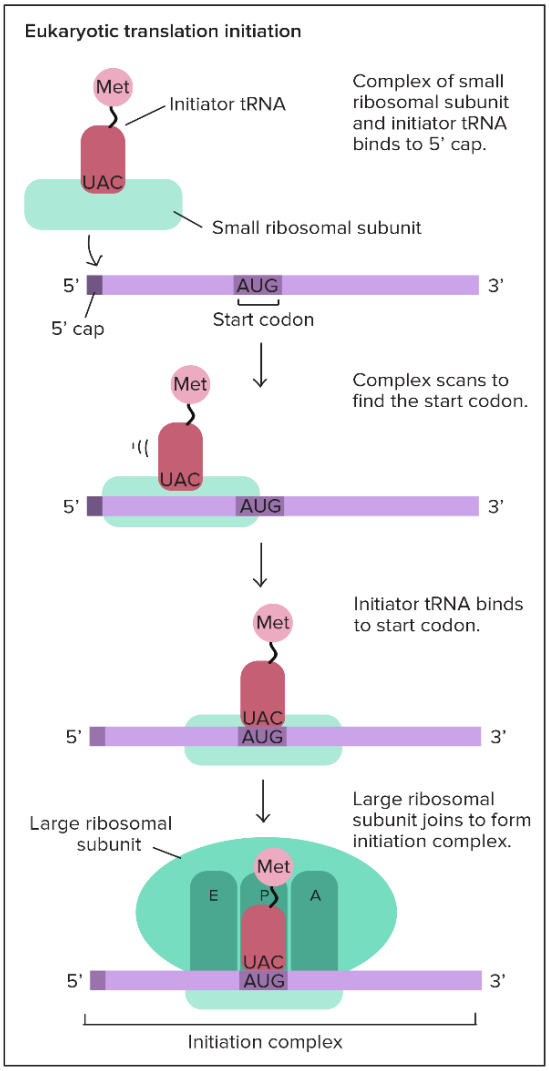


Figure 1.2 – Translation initiation phase
The mRNA containing the information for polypeptide synthesis binds ribosomal subunit 40S. Met-tRNA pair binds mRNA in the start codon region and the larger ribosomal subunit 60S joins the complex to form the initiation complex, signaling the beginning of protein synthesis. Modified from [6].

At the elongation phase, the polypeptide chain grows by successive addition of amino acid residues, to the carboxyl group of the existing polypeptide, in a process assisted by elongation factors. When the codon is recognized, the pair aminoacyl-tRNA binds to it, leading to a peptide bond formation, by a peptidyl-transferase reaction. Ribosome translocate along with the mRNA, allowing the removal of uncharged tRNA and the availability of the A site to accept the next pair aminoacyl-tRNA (**Figure 1.3)**[2].

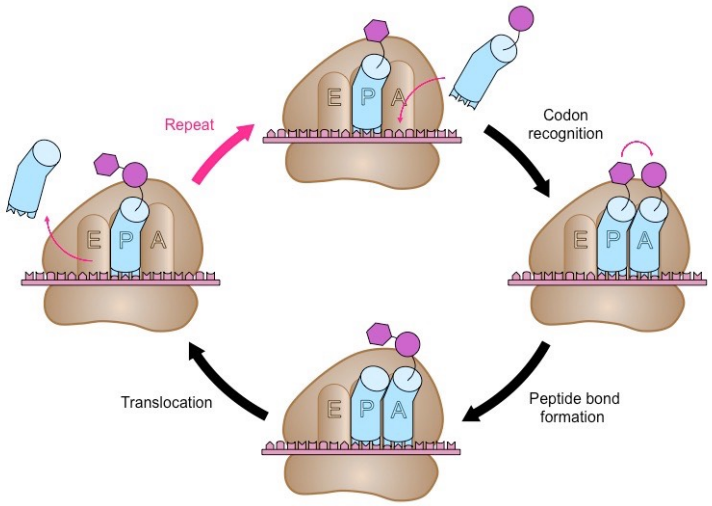


Figure 1.3 – Translation elongation phase
After initiation phase, second codon position in the mRNA prepares to receive another pair aminoacyl-tRNA to bind to this region. A peptidyl-transferase reaction takes place resulting in a peptide bond formation between the amino acids residues involved. This process is followed by ribosome translocation, leaving the next codon position available to receive the following aminoacyl-tRNA pairs. Adapted from [7]

The polypeptide chain keeps elongating until one of the three termination codons, UAA, UAG or UGA appear in the mRNA sequence, which determine the end of protein synthesis. With the aid of release factors (RF) the polypeptide formed is released from the ribosome into the cytosol along with an uncharged tRNA (**Figure 1.4**)[1, 2].

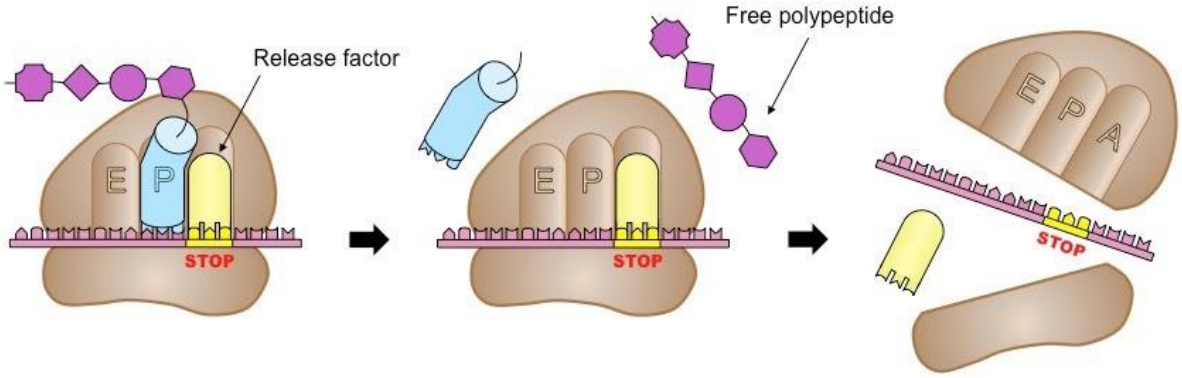


Figure 1.4 – Translation termination phase
The forming polypeptide chain grows until one of the three stop codons is found, and the translation is stopped. The RF stimulate the release of the polypeptide chain and the dissociation between mRNA and the ribosome. In this way, the small subunit, dissociates from the large subunit and gets ready to start all over again the process of initiation. Adapted from [7].

The new polypeptide folds into a proper three-dimensional conformation becoming biologically active. Protein folding is a complex process, through which the linear chain of amino acids acquires a three-dimensional structure, which is biologically functional. The folding process is governed by non-covalent interactions involving the entire molecule, occurring in a time-range of microseconds to seconds[8, 9]. The classic principle of protein folding states that, all the information required for a protein to adopt the correct three-dimensional conformation is provided by its amino acid sequence. This was established by Christian Anfinsen's experiments, demonstrating that denatured RNase can spontaneously refold *in vitro* to its active conformation[10]. This process can be co-translational if it occurs simultaneously with the protein synthesis on the ribosome, or it can be post-translational, if it occurs after the protein synthesis is complete. Post-translational folding predominates over co-translational one within the cell[11]. Besides this, folded proteins can undergo modifications such as: acetylation, phosphorylation, carboxylation, and others, which play a vital role in the function and structure of the protein[1, 2]. Proteins fold into a diversity of three-dimensional structures with diverse topologies and well-defined structural hierarchies, however there is a limited number of unique folds and topologies within the protein universe, as recent studies suggest that the number of proteins fold should converge towards two thousand[12, 13]. Protein folding dynamics and structure, *in vivo*, is affected by the macromolecular crowding effect, decreasing the yield of correctly folded protein by increasing protein aggregation through aberrant non-native interactions, which ultimately will lead to cell toxicity, and so, its formation must be avoided[9]. Cells have managed to develop sophisticated mechanisms for maintaining proteostasis, thus controlling protein synthesis and degradation rates and regulating protein folding and unfolding processes. These processes are mediated by molecular chaperones, proteins that are responsible for assisting others protein folding within the cells ensuring protein correct folding at the right place and time, protease systems, along with cellular clearance mechanisms such as autophagy and lysosomal degradation [14, 15].

1.1.2. Aminoacyl-tRNA synthetase (aaRS)

Aminoacyl-tRNA synthetases (aaRSs) are key enzymes in the translation of the genetic information, as they catalyze the specific attachment of each of the amino acids to its cognate tRNA. This process occurs through a two-step reaction, where they first activate the amino acid with ATP, forming an intermediate aminoacyl-adenylate and then transfer the aminoacyl group to the 3'-end of their own tRNA (**Figure 1.5**)[16, 17]. These enzymes belong to one of the most ancient family of proteins and they were originally characterized and isolated in the late 50s and early 60s and since then they have been extensively studied[18]. Generally, they comprise a core catalytic domain for amino acid activation and transfer and a tRNA anticodon-binding domain responsible for tRNA recognition and binding, however other functional domains were occasionally added throughout evolution, generally at the N or C terminus, displaying non-canonical functions beyond their original translational role (**Figure 1.6**)[19]. As a family of near than 20 enzymes, aaRSs are constrained by evolutionary pressure to preserve its canonical role in the protein synthesis process, but the addition of the supra-referred novel less-conserved domains in eukaryotic aaRSs, opens a new perspective of novel functions associated with them[17].

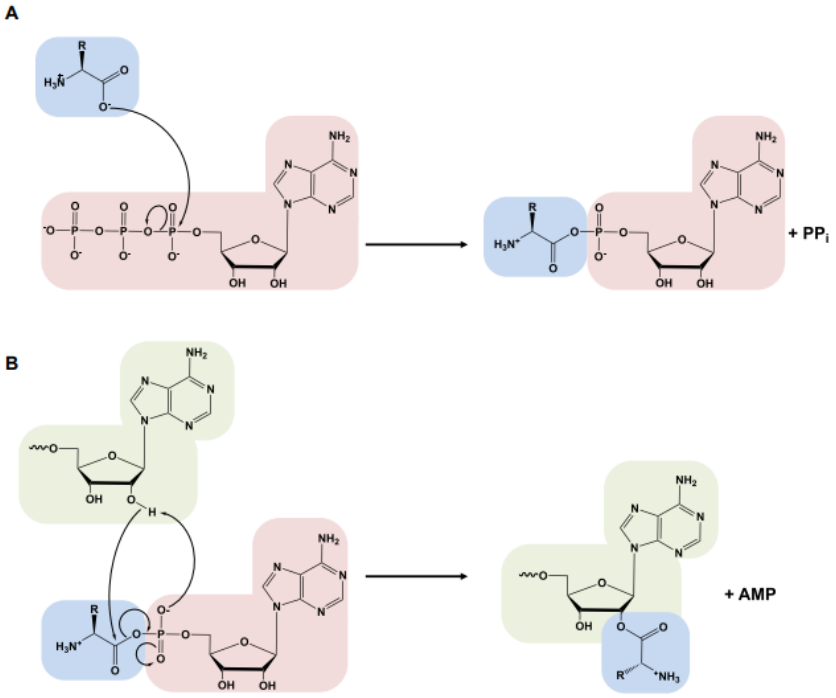


Figure 1.5 - The aminoacylation reaction.
In the first step **(A)**, the amino acid (highlighted in blue) is activated with ATP (highlighted in red) in the aaRs active site (not represented for simplicity), forming the aminoacyl-AMP and releasing inorganic pyrophosphate (PPi). In the second step **(B),** the amino acid is transferred to the tRNA (highlighted in green) and AMP is released. Here is a representation of the transfer to the 3′ -OH characteristic of class I aaRS, while in class II transfer happens at the 2′ -OH with a 3′ -OH attack in this second step. From [20].

Some active sites of aaRSs have difficulties targeting the correct amino acid to its cognate tRNA. These problems are correlated with high similarity, mainly size and/or chemical properties, between correct and incorrect amino acids. Therefore, to prevent mistranslation, some aaRSs have managed to develop an editing domain which performs an editing function, enabling removal of the wrong amino acid from tRNA before it reaches the ribosome, thus ensuring fidelity of the translation[17, 21]. There were found some domains and sequence motifs, which have been progressively added to aaRS, that are not connected neither with aminoacylation nor editing functions. Some examples of shared domains (those in more than one aaRS) include the WHEP domain, which is represented by helix-turn-helix motif, the oligonucleotide binding fold-containing EMAPII domain, present in multi-aaRS complexes, the Leu-zipper motif, the glutathione S-transferase (GST) domain, also used for forming complexes with other proteins and a specialized amino-terminal helix (N-helix)[17]. A total of 23 aaRSs have been described so far, one for each of the 20 proteinogenic amino acids (except for lysine, for which there are two), plus pyrrolysil-tRNA synthetase (PylRS) and phosphoseryl-tRNA synthetase (SepRS) which are only found in some bacterial and archaeal genomes[20]. These enzymes differ from each other in size and oligomeric state, and they are divided in two main classes based on characteristic conserved structural regions in their active site: class I aaRSs contain ‘HIGH’ and ‘KMSKS’ motifs in a Rossmann fold, that displays five parallel β-strands connected via α-helices, whereas class II aaRSs have three homolog motifs in the catalytic domain, with an antiparallel β-sheet flanked by α-helices fold. Most class I aaRSs are monomeric and class II aaRSs are dimeric[4, 22, 23].

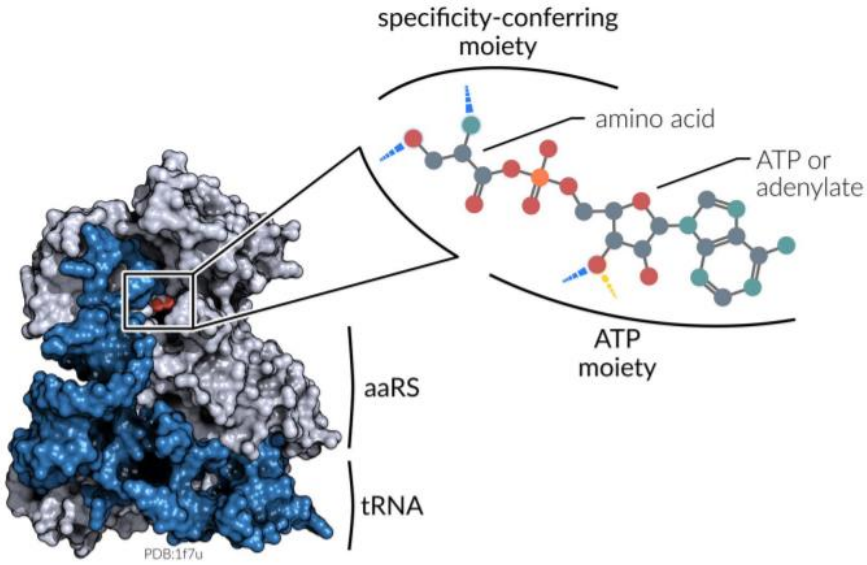


Figure 1.6 – Structural features of complex aaRS·tRNA active site (PDB:1f7u).
The enzyme aaRS catalyzes the binding between an amino acid and a tRNA molecule. The binding site has two moieties, the ATP moiety which fixates ATP and is invariable between each aaRSs and the specificity-conferring moiety which differs between each aaRS and is responsible for the formation of highly specific non-covalent interactions with the respective amino acid ligand. From [24].

In mammals, there are two sets of distinct nuclear encoded genes, coding for both cytosolic (cyt-aaRSs) and mitochondrial aaRSs (mt-aaRSs), with exception of glycil-tRNA synthetase (GARS) and lysil-tRNA synthetase (KARS), for which cytosolic and mitochondrial forms are encoded by single genes, but are generated either from two translational sites or by alternative mRNA splicing[22]. Since the mitochondrial genome does not code for mt-aaRSs, these are first synthetized in the cytosol and then imported into the mitochondria where they perform their function in the mitochondrial translation machinery[22]. Despite both sets being translated in cytosolic environment, the mitochondrial ones possess a N-terminal mitochondrial-specific targeting sequence, that may vary in length from 18 to 54 amino acids, which is responsible for import to the respective organelle[25, 26]. However, there has not been reported yet, a single eukaryotic species capable of encoding 2 complete and unique compartment-specific sets for all the 20 aaRSs, which can theoretically result in an apparent lack of mt-aaRSs[26]. As such, eukaryotes use two different strategies to compensate it: they can generate the aa-tRNA species for which the cognate aaRS is missing or they import the corresponding cytosolic aaRS into mitochondria, enabling a dual-localization, cytosolic and mitochondrial, for the respective enzyme[26]. Eukaryotic cytosolic aaRSs (cyt-aaRSs) subcellular localization can also be influenced by their propensity to form multi-synthetase complexes (MSCs). MSCs were first reported in mammals in the early 70s and initially, it was believed that their formation occurred only in higher eukaryotic level, as all the aaRS isolated either from *Saccharomyces cerevisiae* or *E. coli* were found in a free form, however studies targeting the presence of these complexes showed that they can be found in species ranging from fungi to mammals[18, 26]. These complexes are all composed of cyt-aaRSs and 1 to 3 auxiliary assembly factors, which are called, Aminoacyl-tRNA synthetase Interacting Multifunctional Proteins (AIMPs). The size and composition of the MSCs vary depending on the organism, as MSCs from more complex organisms require more cyt-aaRSs than simpler ones (**Figure 1.7**)[20, 26]. This difference may suggest an evolutionary link between the MSC expansion and the increased complexity of the interaction network of their components[20]. Although the exact function of the MSCs remains unclear,

it has been proposed that the assembly of synthetases may aid the channeling of the recognition of amino acid substrate to the tRNA and further delivery to the ribosomal machinery, enhancing efficiency of tRNA aminoacylation. They are also involved in fundamental processes such as transcription, cell-signaling and tumorigenesis[18, 20, 27]. In the last decade has emerged a consensual concept, which is that MSCs are cytosolic retention platforms for releasable and multifunctional aaRSs, under specific physiological or stress conditions to be relocated into subcellular compartments where they can perform non-conventional aaRSs roles, such as regulation of gene expression, participation in signaling pathways and some others already mentioned[18, 26].

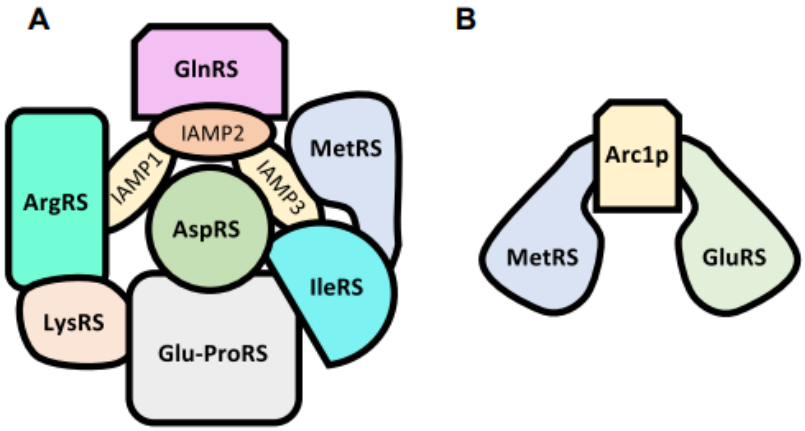


Figure 1.7 – Complexity of different multisynthetase complexes.
Schematic representation of the multisynthetase complex showing the major components for a mammalian (A) and yeast (B) multisynthetase complex (interactions are not shown; neither is the possible homodimerization of some components). Mammalina MSC is a massive complex composed of nine synthetases and three accessory proteins, while otherwise the yeast counterpart is composed of two synthetases and a connecting protein. The scheme is not drawn to scale and spatial arrangements and sizes of the different subunits may not reflect their real positions in the complexes. From [20].

Concerning mt-aaRSs, the first genes coding for mammalian mt-aaRSs were reported in 1994, almost a decade after assignment of the first yeast mt-aaRS[25]. There were only identified 19 mt-aaRSs, unlike the 20 cyt-aaRSs. According with eukaryotic genomic analyses, mitochondrial glutaminyl-tRNA synthetase (QARS2) is not encoded in nuclear genome, consistent with the finding that no activity was detected upon experiments with purified mitochondria from some eukaryotic species. This may indicate that formation of mitochondrial glutaminyl-tRNA^{Gln} occurs via an indirect pathway, first by misaminocylation performed by glutamyl-tRNA synthetase (EARS) forming a glutamyl-tRNA^{Gln} (non-cognate) that is followed by amidation by a glutamyl-amidotransferase to yield the correct glutaminyl-tRNA^{Gln}[16, 25, 28]. Beyond the additional mitochondrial target sequencing, mt-aaRSs might have additional features that render them more suitable to mitochondrial tRNAs, which may suggest that these enzymes have evolved in such way that they use a different tRNA recognition and charging elements compared with the cytosolic ones[16, 29]. Mitochondria are crucial for maintenance of cellular physiology, by controlling the ATP levels via oxidative phosphorylation (OXPHOS) process through the respiratory chain complexes, among other important functions. To execute those functions, approximately a thousand proteins both nuclear (majority) and mitochondrial-encoded (proteins from the respiratory chain) are included in this organelle. The mitochondrial translation machinery is extremely important to produce those mitochondrial-encoded proteins, therefore, mt-aaRSs are the key actors, as they carry on their canonical role, ensuring the correct

translation of the mitochondrial-encoded proteins, involved in the respiratory chain, and so, contribute to cellular energy production[16, 22, 29]. Mutations affecting mt-aaRSs and thus, the mitochondrial translation machinery have been identified as a major cause of mitochondrial diseases and are generating particular interest among this type of diseases (discussed in section 1.2.)[22].

1.1.3. EARS2 and RARS2

EARS2 (mitochondrial glutamyl-tRNA) is a nuclear-encoded mitochondrial protein, which belongs to class I family of mt-aaRSs. As such, it has a N-terminal mitochondrial targeting sequence, which is represented for the first 41 amino acids of the sequence[28, 30]. It presents a 'HIGH' motif region at amino acids 45-53 and a 'KMSKS' motif at amino acids 284-288, classic from this subfamily. The catalytic domain is localized between amino acids 36 and 346 and finally an anticodon binding domain is also found in amino acids from 347-523[31, 32]. This protein canonical function is to catalyze the reaction of aminoacylation of mitochondrial $tRNA^{Glu}$ with glutamate. However, since EARS2 is a non-discriminating protein, it can misaminoacylate mitochondrial $tRNA^{Gln}$, leading to the formation of QARS2, via an indirect pathway, already described in previous section. Only EARS2 from yeast is reported as not being able to perform this function[28, 30]. As most mitochondrial proteins are homologous to their bacterial counterparts, it is possible to compare by sequence alignment putative human EARS2 with other bacterial EARS (**Figure 1.8**)[28].

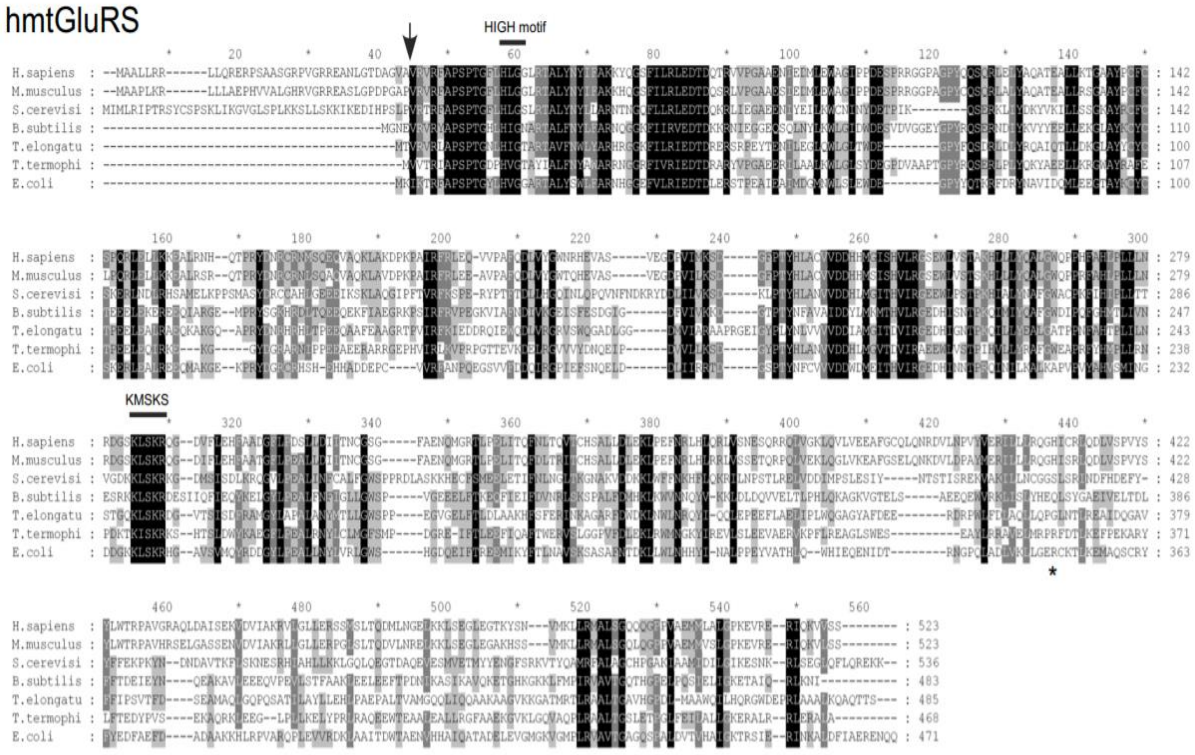


Figure 1.8 – Multiple sequence alignment of human EARS2 with mouse EARS2 and EARS from other organisms.
Sequence alignment of EARS2 proteins from human, mouse with EARS from *S.* cerevisiae and bacterial EARS proteins from, *B. subtilis*, *T. elongatus*, *T. thermophilus*, and *E. coli*. Arrow points towards the predicted cleavage site of the N-terminal mitochondrial targeting sequence. Signature motifs of class I aaRS (HIGH and KMSKS) are indicated for each sequence. The black and gray boxes indicate regions of sequence identity (black) and sequence similarity (gray). The position corresponding to Arg-358 in *T. thermophilus* EARS is indicated by an asterisk. From [28].

RARS2 (mitochondrial arginyl-tRNA synthetase) is also a nuclear-encoded mitochondrial protein, and likewise EARS2, it also belongs to the class I family of mt-aaRSs[33]. In its sequence this protein has a N-terminal signal peptide, which correspond to the first 16 amino acids of the sequence and a 'HIGH' motif region represented by amino acids 133 to 144, which also comprised within a L-arginine binding domains from amino acid 133 to 135[34]. RARS2 is responsible for catalyzing the aminoacylation of mitochondrial $tRNA^{Arg}$ with arginine[33]. Uniquely, among others class I family aaRSs, most of RARS2 species does not have a 'KMSKS' motif, which is an essential component of the catalytic site. This way, the mechanism underlying the catalytic reaction of this enzyme is different from the others[35]. A multiple sequence alignment analysis was executed between RARS2 from human, other mammals and bacterial RARS and is present in **Figure 1.9**.

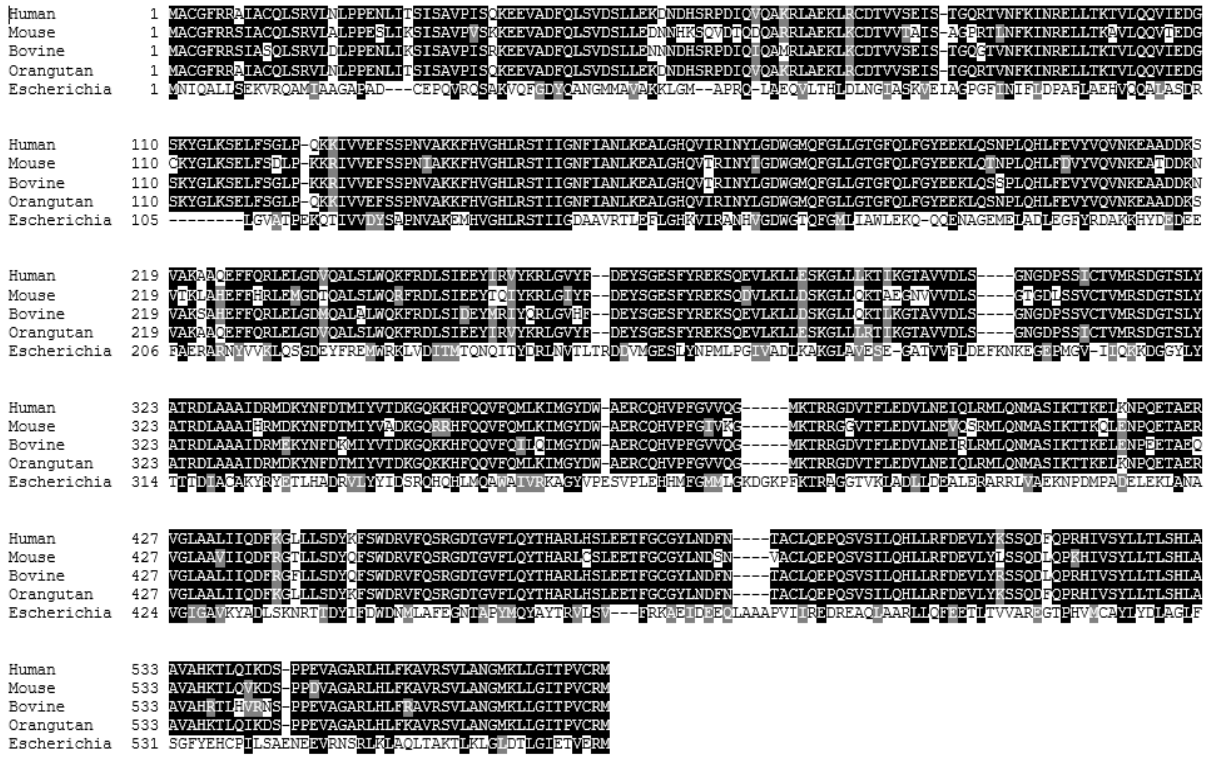


Figure 1.9 - Sequence alignment of human RARS2 with bovine, mouse, orangutan RARS2 and bacterial RARS.
Amino acid sequence alignment of RARS2 proteins from human (*H. sapiens*), mouse (*M. musculus*), bovine (*B. aurus*), orangutan (*P. abelli*), with *E.coli* RARS protein. The black and gray boxes indicate regions of sequence identity (black) and sequence similarity (gray). Sequence alignment obtained by boxshade server.

For both human EARS2 and RARS2 there is not a 3D structure experimentally determined, however, recently, a computational method which predicts three dimensional structures of proteins, named AlphaFold, has been solving countless unknown human protein´s structures at an unbelievable pace and with a great accuracy[36]. Among them hEARS2 and hRARS2 structures are examples of proteins which have already been predicted through this accurate computational approach (**Figure 1.10**). This represents a huge breakthrough in the analysis of the protein structure and allows us to speculate on the overall structure of both of them, analyze and eventually compare with *in vitro* assays performed in the future.

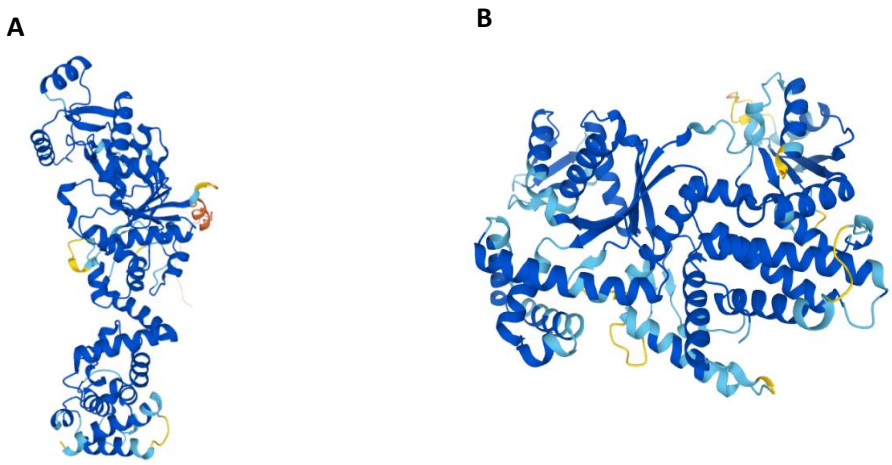


Figure 1.10 – AlphaFold predicted structures of human EARS2 (A) and human RARS2 (B). Both structures were obtained from AlphaFold Protein Structure Database.

1.2. Mitochondrial disorders

1.2.1. General concepts

In the year of 1962 it was described the first mitochondrial disease, Luft syndrome, and 26 years later, it was discovered the first pathogenic mitochondrial DNA mutations and from there on, the field of mitochondrial disorders have been expanding at an unbelievable pace[37] .The term mitochondrial disorder is referred to a group of clinically and genetically heterogeneous inborn human diseases, that are caused by OXPHOS and mitochondrial respiratory chain dysfunction, directly affecting the energy metabolism, which takes place in mitochondria in the OXPHOS process. Despite being rare, these diseases are among the most common inherited human diseases, with a prevalence of 1:5000 individuals. They present a wide variety of symptoms, affecting any organ (especially the ones with a highest need of energy), at any age and with different modes of inheritance[16, 38, 39]. Mitochondrial dysfunctions are associated with some human diseases such as neurodegenerative disorders, cardiovascular disorders, neurometabolic disorders, cancer, obesity, among others[40]. As mentioned before, mitochondria plays a critical role in the production of energy in the OXPHOS process. This process relies on the transport of electrons generated from metabolite oxidation, that are sequentially transferred to molecular oxygen via enzymatic complexes present in mitochondrial inner membrane[41]. The formation of the complexes is under control of two genome, the nuclear and the mitochondrial, as 4 out of the 5 respiratory chain complexes contain both nuclear and mitochondrial-encoded proteins[42]. Since mitochondrial disorders have genetic origin, they can arise by mutations either in nuclear or mitochondrial genes, that encode for proteins belonging to the respiratory chain or related to it, and thus, the resulting pathogenic variants may impair the energy production, ultimately leading to disease[38, 43]. As 99% of mitochondrial proteins are nuclear-encoded, it is expected that most of the mitochondrial disorders have origin in mutations in the nuclear genome[38]. Nuclear genome inheritance can be autosomal-dominant, autosomal recessive or X-linked, but *de novo* sporadic mutations are not uncommon[38, 44]. Unlike nuclear genome, mitochondrial one is exclusively inherited from maternal line, and as such, some mitochondrial disorders have a maternal inheritance pattern[38]. Normally, the mitochondrial genome of an individual is composed by a single mtDNA species, condition known as homoplasmy, principally present at birth, however, this DNA can randomly mutate, determining a transitory condition known as heteroplasmy[42]. The variability in clinical presentation and underlying causative mutations make the diagnosis very challenging, involving extensive clinical and specialized laboratory evaluation, however, no reliable diagnostic screening or biomarker is available that is both sensitive and specific for mitochondrial disorders. Problems in biochemical characterization of mitochondrial disorders, such as variability in tissue manifestation, difficulty in establishing realistic reference ranges, inability of enzyme assays to detect some defects and a lack of widely accepted diagnostic criteria, make the diagnosis even harder, and some patients die even without being diagnosed[39]. However, development of next-generation sequencing (NGS) has offered an excellent new tool to identify molecular backgrounds for this disease groups[38]. Presently, there is no cure for most mitochondrial disorders, but some treatments towards the relieving of the symptoms have been developed. Molecular mechanisms underlying mitochondrial disorders and correlation between genotype-phenotype are not fully understood, since same genotypes can cause different phenotypes, or the same phenotype can be caused by a large number of different genotypes (for example in the Leigh syndrome)[40, 41]. Examples of mitochondrial diseases due to mutations on nuclear DNA are: encephalopathic disorder with ataxia and cerebellar atrophy, myopathic disorder with recurrent myoglobinuria, generalized mitochondrial encephalomyopathy, glutaric aciduria types I and II, leukoencephalopathy with thalamus and

brainstem involvement and high lactate (LTBL), pontocerebellar hypoplasia type 6 (PCH6), among others[33, 37, 45, 46].

1.2.2. mt-aaRSs and disease

Mutations in mt-aaRSs are also one of the causes of some types of mitochondrial disorders. The first case described of a mutation in the genes coding for these proteins, causing a disease date back to 2007, when mutations within the gene coding for mitochondrial aspartyl-tRNA synthetase (DARS2), were associated with a leukoencephalopathy. Today, all 19 mt-aaRSs-encoding genes have been reported to be affected, mainly using approaches such as, genetic linkage studies, homozygosity mapping and more recently next-generation whole-exome sequencing[33, 47]. From there on, the discovery of disorders caused by mutations in genes coding for mt-aaRSs generated particular interest, due to their predilection and damage effects in the central nervous system (CNS)[22, 48]. But not all mt-aaRSs disorders affect the CNS, as it were already observed sporadic manifestations in skeletal muscle, kidney, lung, and/or heart, in general, high-energy demand tissues can be affected[49]. Diverse clinical manifestations arise from these mutations, usually with an early onset and transmitted as autosomal recessive traits, leading to autosomal recessive disorders. The exception are the GARS and KARS, whose mutations exhibit dominant inheritance. Such a broad clinical spectrum can be partially explained by the multiple and still to be figured out functions of mt-aaRSs[16, 22]. From the phenotypical point of view, mt-aaRSs-encoding gene mutations can lead to leukodystrophies (genes coding for DARS2, EARS2, AARS2 and MARS2), to Perrault syndrome (for HARS2 and LARS2), to deafness (for MARS2 and NARS2), to intellectual disability (for RARS2 and WARS2) and can lead to the development of cardiomyopathies (for AARS2, GARS and KARS)[22, 48]. The molecular mechanisms behind this selective tissue involvement and disease phenotype for specific mt-aaRSs disorders are currently poorly understood[50].

New mutations in these enzymes are being continuously discovered, therefore it became preponderant to gather all the existing and emerging data about them in order to have all information accessible to researchers, clinicians and patient families. For that reason, it was developed a knowledge database and a Web server, named MiSynPat (Mitochondrial aminoacyl-tRNA Synthetases and Pathologies), which is constantly updating. This platform allows a focused search for disease-causing mutations, protein sequence alignments, 3D structures, literature and statistics[47].

In particular for hEARS2, genetic studies have identified pathogenic variants linked to a mitochondrial disorder named leukoencephalopathy with thalamus and brainstem involvement and high lactate (LTBL)[30]. Brain MRI features found on patients with these mutations showed unique patterns of symmetrical abnormalities of cerebral white matter, thalamus, midbrain, pons, medulla oblongata and cerebellar white matter. This disease, caused by either compound heterozygous and homozygous variants, has infantile onset and rapid progression with a striking partial recovery and regaining of milestones in some patients and no improvement in others[33, 46]. However, the relationship between EARS2 variants and the phenotypic characteristics of LTBL remain poorly understood, and further studies are needed to elucidate the molecular mechanisms underlying this disease[46]. A recent case study has reported that near 46% of mutations in EARS2, which are linked to disease, occur in arginine residues. The authors were able to estimate a 3D structure of hEARS2 by SWISS-MODEL homology modelling. They found that some of these arginine residues, were located in the catalytic domain near the ligands, in the glutamate binding region and also in the ATP binding region (**Figure 1.11**)[32].

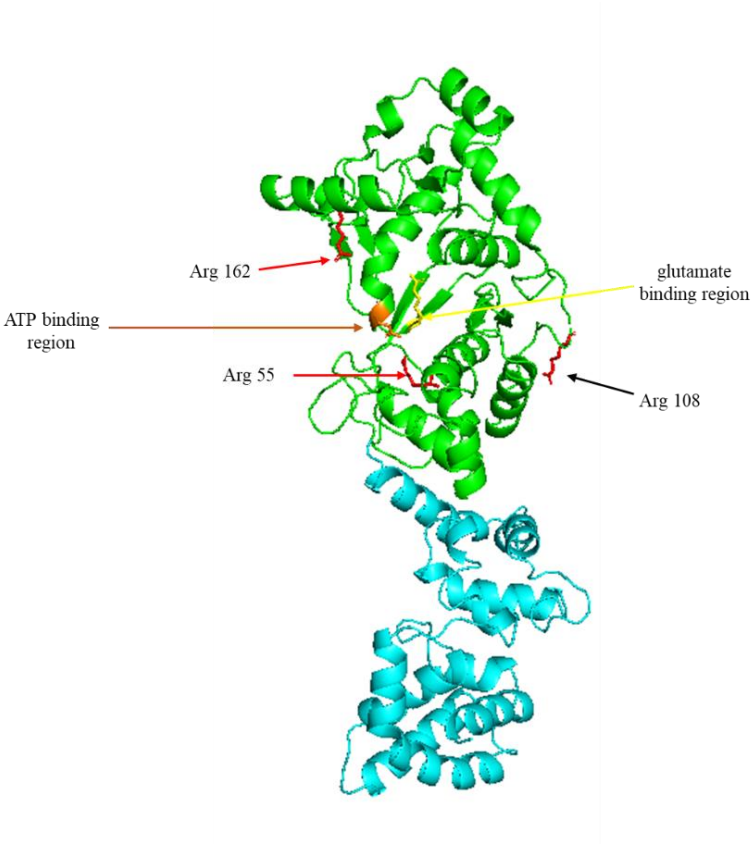


Figure 1.11 – 3D predicted structure of hEARS2.
Estimated 3D structure model of hEARS2 was obtained using SWISS-MODEL Homology Modelling and EARS from *T. Thermophilus* as template. Green color represents the catalytic site and blue cyan represents the anticodon binding region. Black arrow represents arginine residue in position 108, which corresponds to a site of frequently reported variants. Red arrows indicate arginine residues, whose variants have been reported in a case study, p.Arg55Leu and p.Arg162Trp. Yellow arrow indicates the region of binding of glutamate. Orange arrow point towards the ATP binding region. Modified from [32]

Mutations in the gene coding for RARS2 protein were identified in patients suffering from a severe infantile pontocerebellar hypoplasia type 6 (PCH6)[33]. Phenotypically, this disease is presented with cerebral cortex and white matter involvement. This severe clinical picture is characterized by lethargy, seizures with apneic spells, hypotonia, and premature death[16]. The reported RARS2 mutations were all compound heterozygous and it has been shown that they cause reduced cellular abundance of RARS2 mRNA or RARS2 protein, impaired mt-tRNA for arginine stability or protein activity, but no clear correlation to clinical severity has been described to these alterations[16, 33]. However, it is not fully understood why mutations in this gene cause a specific pattern of neurodegeneration leading to PCH6[51].

1.3. Objectives

Most of studies about mitochondrial disorders, which have origin in mutations on mt-aaRSs, have been focused on the genomic analysis of these diseases. It has been reported that the localization of missense mutations in mt-aaRS may influence disease phenotype, but information regarding enzymes structure, biochemical and biophysical properties remain very scarce. Therefore, it is mandatory to fully characterize these enzymes to better understand the molecular mechanism of disease.

We hypothesize that mutations in specific structural domains will have different conformational impacts in mt-aaRS structure and function and possibly, will determine the severity of the disease phenotype. The major aim of the project is to discriminate on possible genotype-phenotype links to increase the knowledge on the molecular mechanism of these diseases. For this we first need to implement protocols for heterologous expression and purification of human mt-aaRSs wild-type proteins.

Specifically in my master´s project the objectives are:

1 – Establish protein heterologous expression protocols for human EARS2 and human RARS2;

2 – Set up purification protocols for both proteins;

3 – Characterize wild-type hEARS2 and hRARS2 conformational fold and thermal stability.

To accomplish these goals, I will use several biochemical techniques that go from, heterologous protein expression in *E. coli* system, to protein purification using different chromatographic techniques, and protein structural characterization using spectroscopic techniques such as CD and fluorescence.

2. Materials and Methods

2.1. Chemicals

All reagents were of the highest purity grade commercially available. Isopropyl β-D-1-thiogalactopyranoside (IPTG), NZY bacterial lysis buffer and glycerol (99.5%) were purchased from Nzytech, monoclonal anti-polyHistidine-peroxidase conjugate antibody and L-glutamate were purchased from Sigma, phenylmethylsulfonylfluoride (PMSF) was purchased from Roth, imidazole was purchased from Applichem.

2.2. *In silico* analysis

Human EARS2 and RARS2 structures were predicted using SWISS-MODEL Homology Modelling web server. It was used bacterial EARS as template for hEARS2 structure prediction and *S. cerevisiae* RARS as template for prediction of hRARS2 structure. Multiple sequence alignments were performed in T-coffee Multiple Sequence Alignment software and further improved by ESPript program and Boxshade software. Proteins biochemical characteristics (molecular weight, theoretical pI and molar extinction coefficient) were computed in ProtParam from Expasy online web server.

2.3. Protein expression

The cDNAs encoding for human EARS2 or RARS2 were inserted into pET-28a(+) vector (Novagen) (obtained from Genscript). These plasmids code for the mature forms of the proteins, without the signal peptide, with a 6xhis-tag attached to the C-terminal region. The plasmids contain a lac promotor and resistance to kanamycin. *E. coli* Rosetta cells were transformed with hEARS2 or hRARS2 plasmid and grown in LB (Luria Broth) medium (10 g Bacto tryptone, 5 g Bacto yeast extract and 10 g NaCl per liter) supplemented with 10 $\mu g.mL^{-1}$ kanamycin at 37 ºC in a shaking incubator until OD_{600nm} 0.5-0.7 was reached. Protein expression was then induced under different conditions 0.2 mM, 0.5 mM and 1 mM IPTG at 22 ºC, 30 ºC, 37 ºC. To evaluate protein expression samples (1 mL) from the growth were withdrawn before induction (t0) and after induction 1H, 3H, 4.5H, 5H, 6H and during overnight, harvested by centrifugation and stored at -20 ºC until further processing. Cell pellets were treated with NZY bacterial cell lysis buffer (Nzytech) accordingly to manufacturer instructions, and protein soluble and insoluble fractions were analyzed by SDS-PAGE and Western blot. Briefly, Tris-HCl gels with 12% acrylamide were run at 200 V for 45 minutes and stained with Coomassie Blue. Similar gels were made to perform Western blot methodology, and for this, proteins in the gel were transferred to a nitrocellulose membrane, using a semi-dry apparatus (BioRad). Membrane was blocked with 5% skimmed milk for 1 hour and incubated with monoclonal anti-polyHistidine-peroxidase conjugate (A7058, Sigma Aldrich) for 2 hours, according with fabricant instructions. Chemiluminescence was detected in a BioRad chemidoc XRS or an Amersham Imager 680 RGB instrument.

2.4. hEARS2 Purification

To produce sufficient material for protein purification, Rosetta cells transformed with hEARS2 plasmid were grown in 4 L of LB containing 10 $\mu g.mL^{-1}$ of kanamycin, at 37 ºC until OD_{600nm} reached values between 0.5-0.7. After this, cells were induced with 0.5 mM of IPTG, and grown at 30ºC for 4 hours. Cells were harvested by centrifugation, and the resulting pellet was then kept at -20ºC for further

purification. For protein purification, cells were defrosted, resuspended in buffer A (30 mM Tris-HCl, pH 8.0, 300 mM sodium chloride (NaCl), 20 mM imidazole, 0.5 mM PMSF and 20% glycerol), in presence of DNAse I (AppliChem) and disrupted by sonication using a UP200S sonicator (20 minutes, 0.5 seconds pulse, 70% amplitude). Soluble extract was obtained through a series of centrifugations at 4 ºC, first at 18 000 rpm JA-20 rotor (Beckman) for 30 minutes, secondly the resulting supernatant was ultracentrifuge at 42 000 rpm (in a Ti-45 rotor (Beckman)) for 60 minutes.

The soluble extract was loaded into a 5 mL His-trap HP (GE Healthcare) for affinity chromatography separation using a Chromatographic System AKTA Purifier UPC 10 (GE Healthcare). This separation technique is based on interaction between an immobilized metal ion usually, Cu(II), Ni(II), Zn(II), Co(II) or Fe(II), and electron donor groups located on the surface of proteins[52]. In this work were used nickel ions and the electron donor groups were the 6-His-tag attached to hEARS2. To efficiently separate the protein of interest from others, His-trap column was equilibrated in approximately 5 volumes of buffer A. The low concentration of imidazole and the presence of salt prevent unspecific interactions between the ligand and the other proteins in the soluble extract. Bound proteins were eluted by performing an imidazole gradient, from 20 mM to 500 mM, with buffer B (30 mM Tris-HCl, pH 8.0, 300 mM NaCl, 500 mM imidazole, 0.5 mM PMSF and 20% glycerol). Chromatography was followed at 280 nm, separated peaks were collected and protein content analyzed by SDS-PAGE, to characterize the protein elution pattern. The fractions corresponding to hEARS2 (with the respective molecular weight of 58.5 kDa) were pooled together and loaded into a 5 mL Hi-trap desalting G-25 column (GE Healthcare) in buffer C (30 mM Tris-HCl, pH 8.0, 50 mM NaCl and 20% glycerol). Desalting chromatography is a type of gel filtration, which has the aim to exchange buffer, based on the size difference between protein and the solvent molecules. This technique allows the removal of molecules that can have negative effects on the protein stability and function, allowing storage of purified protein under optimized conditions[53]. hEARS2 fractions from desalting were added together, concentrated in an amicon (Merck Millipore) with a 10kDa cut-off, aliquoted and stored at -80 ºC for further characterization studies. Protein purity was confirmed by SDS-PAGE Coomassie Blue stained and Western blot. Protein concentration was determined spectrophotometrically using hEARS2 theoretical molar extinction coefficient value at 280 nm (48 860 M^{-1} cm^{-1}) and Bradford assay.

2.5. hRARS2 Purification

Rosetta cells containing hRARS2 plasmid were grown in 4L LB medium, in the presence of 10 µg.mL^{-1} of kanamycin at 37 ºC until cells OD_{600nm} reached 0.5-0.7. Then, cells were induced with 1 mM IPTG and incubated for 6 hours at 37 ºC. Cells were further harvested by centrifugation and stored at -20 ºC until purification. For protein purification, cells were defrosted, resuspended in buffer A (the same as used for hEARS2), in presence of DNAse I (AppliChem) and disrupted by sonication using a UP200S sonicator (20 minutes, 0.5 seconds pulse, 70% amplitude). The soluble extract was then submitted to centrifugation 18 000 rpm JA-20 rotor (Beckman) for 30 minutes and the resulting supernatant was ultracentrifuged 42 000 rpm Ti-45 rotor (Beckman) for 60 minutes.

Similar to hEARS2, hRARS2 was purified from soluble extract, using a 5 mL His-trap column (GE Healthcare). The column was previously equilibrated in 5 volumes of buffer A, and bound proteins eluted in a gradient of 20 mM to 500 mM imidazole with buffer B (same as in section 2.4). The peaks corresponding to hRARS2 fraction were identified by SDS-PAGE. hRARS2 peaks were pooled together, and the resulting pool was loaded into a 5 mL Hi-trap desalting G-25 column (GE Healthcare), equilibrated in buffer C (equal

to the one used in hEARS2 purification). Fractions corresponding to the protein were finally collected, added together, concentrated in an amicon (Merck Millipore) with a cut-off of 10 kDa, aliquoted and stored at -80 ºC, for further characterization studies. Protein purity was then assessed by SDS-PAGE and Western blot. Protein concentration was determined spectrophotometrically using hRARS2 theoretical corresponding molar extinction coefficient value at 280 nm of 49 329 M^{-1} cm^{-1}, and by Bradford assay.

2.6. Circular Dichroism Spectroscopy

Circular Dichroism (CD) spectroscopy is a technique used in structural biology which provides structural information on biomolecules containing chiral chromophores. This information is obtained in terms of degrees of ellipticity, by measuring the difference between absorption of left-handed and right-handed polarized light. In the study of proteins, there are different types of relevant chromophores to study protein structural properties, such as, peptide bond (absorption below 240 nm), aromatic amino acid side chains (absorption in the range of 260-320 nm) and disulphide bonds (absorption around 260 nm)[54]. Far-UV CD spectra (180-250 nm) gives information about protein secondary structure, based on the excitation of electronic transitions in amide groups. So, each type of secondary structure will have a characteristic CD spectrum in this region. Information regarding α-helical structure can be estimated by the values of ellipticity at 208 nm and 222 nm, whereas β-sheet structures present a fingerprint around 217-220 nm[54, 55]. Near-UV CD spectra (260-320 nm) of a protein arise from its aromatic amino acids, and thus, provides valuable information about protein tertiary structure. Taking advantage of these properties it is possible to follow protein structural changes upon interaction with a ligand and to study protein chemical or thermal stability, using a denaturing agent or increasing temperature, respectively, as we can monitor the loss of the characteristic CD signal of the protein, by comparing the initial signal to the obtained after exposure to unfavorable conditions[54].

CD spectra were recorded on a Jasco-1500 Circular Dichroism Spectrophotometer with a cell holder thermostatically controlled with a Peltier. To study the secondary structure of proteins CD spectra were performed at far-UV region (200-260 nm), using a 1 mm path length quartz cuvette (Hellma) and typically 0.1 $mg.mL^{-1}$ of protein.

2.7. Fluorescence Spectroscopy

Fluorescence spectroscopy is a central methodology used in biological sciences, not only for being highly sensitive, but also because it does not require huge amounts of sample to be performed. This technique is based on the emission of energy caused by the return of the electrons into their ground-state, after being at a higher electronic state (excited state). In proteins intrinsic fluorescence typically occurs at the level of the side chain of aromatic residues, or in some cases from cofactors such as flavin mononucleotide (FMN) and flavin dinucleotide (FAD), and we can take advantage of these feature and use this technique to analyze the tertiary structure of a protein. In the case of the aromatic residues contribution to protein´s intrinsic fluorescence is highly dependent on their environment[56], in particular for tryptophan residues, which are the ones that highly contribute to protein´s intrinsic fluorescence, due to their high quantum yield. The indole groups of these residues are very good reporters of the surrounding environment, because of their high sensitiveness to solvent polarity, thus allowing to monitor protein conformational changes, polarity changes and tertiary structure[56, 57]. This information can be obtained by monitoring tryptophan fluorescence emission using excitation wavelength of 280 nm[57].

Fluorescence spectra were recorded on a Jasco FP-8200 Spectrofluorometer and it was utilized a 10 mm path length quartz cell (Hellma) and usually 1 µM protein concentration. The tertiary structure of proteins was studied by following tryptophan emission between 300 nm and 450 nm using an excitation wavelength of 280 nm. Typically, slits were set at 5 nm and 10 nm for excitation and emission, respectively.

2.8. Thermal stability studies

To assess protein stability, we performed thermal denaturation assays using the available spectroscopic techniques, CD and fluorescence spectroscopy, to evaluate the thermal denaturation pattern of proteins.

Studies were conducted from 25 ºC to 90 ºC, with a heating rate of 1 ºC.minute^{-1}. Secondary structure was followed by monitoring CD signal at 208 nm and 222 nm. Fluorescence spectroscopy was used to follow tryptophan´s protein emission at excitation wavelength of 280 nm and emission wavelengths of 330 nm and 350 nm.

Thermal denaturation data were analyzed according to a two-state model and the transition was fitted to a sigmoid curve with OriginPro2019b software, which allowed the determination of the apparent melting temperature (T_m^{app}) of the proteins, which corresponds to the midpoint transition temperature, i.e., the temperature at which there 50 % of folded and unfolded states of the protein.

3. Results and Discussion

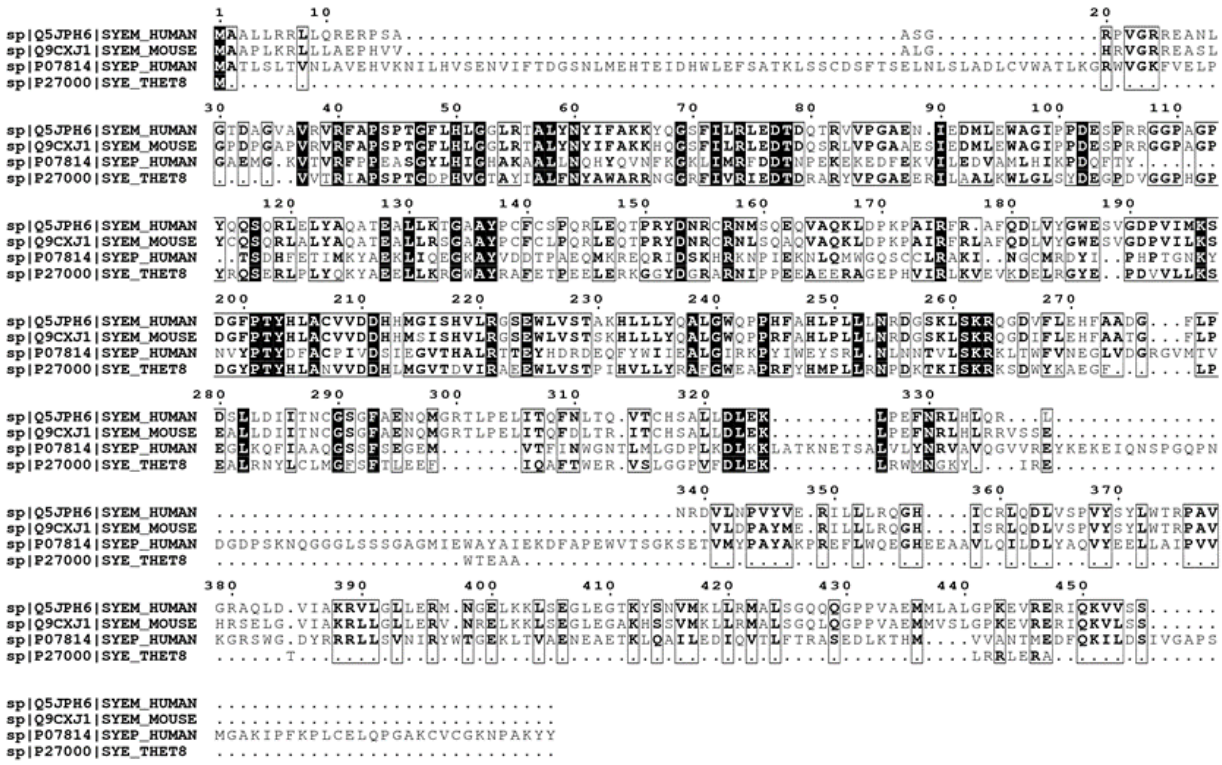
With the increase in genomic studies on patients with MD, the identification of mutations in mt-aaRSs has been increasing, however studies regarding the effects at the protein level are almost inexistent, representing a major gap in the knowledge of these diseases. Disease phenotype can be influenced by the localization of mutations in mt-aaRS structure, so it is of extreme importance to characterize these proteins at structural, biochemical and biophysical level. This characterization can further allow us to understand the molecular mechanism of disease, by establishing a genotype-phenotype links on the disease.

To tackle this challenge, we will establish protocols for heterologous expression and purification of two human mt-aaRSs, hEARS2 and hRARS2 (whose mutations have been reported to cause MD), and perform their structural characterization as describe in the following section.

3.1. Bioinformatic tools to study protein features

Since there is no information about human EARS2 and RARS2 structures, we resorted to *in silico* analysis, to try to predict their structures and attempt to have a first insight into a probable structure of these proteins and hopefully, guide our approach according to a protein which has a similar sequence. Firstly, we compared the sequences of both human proteins with other organisms´ and also with the cytosolic form of each by performing a multiple sequence alignment. The results are shown in **Figure 3.1**.

A



B

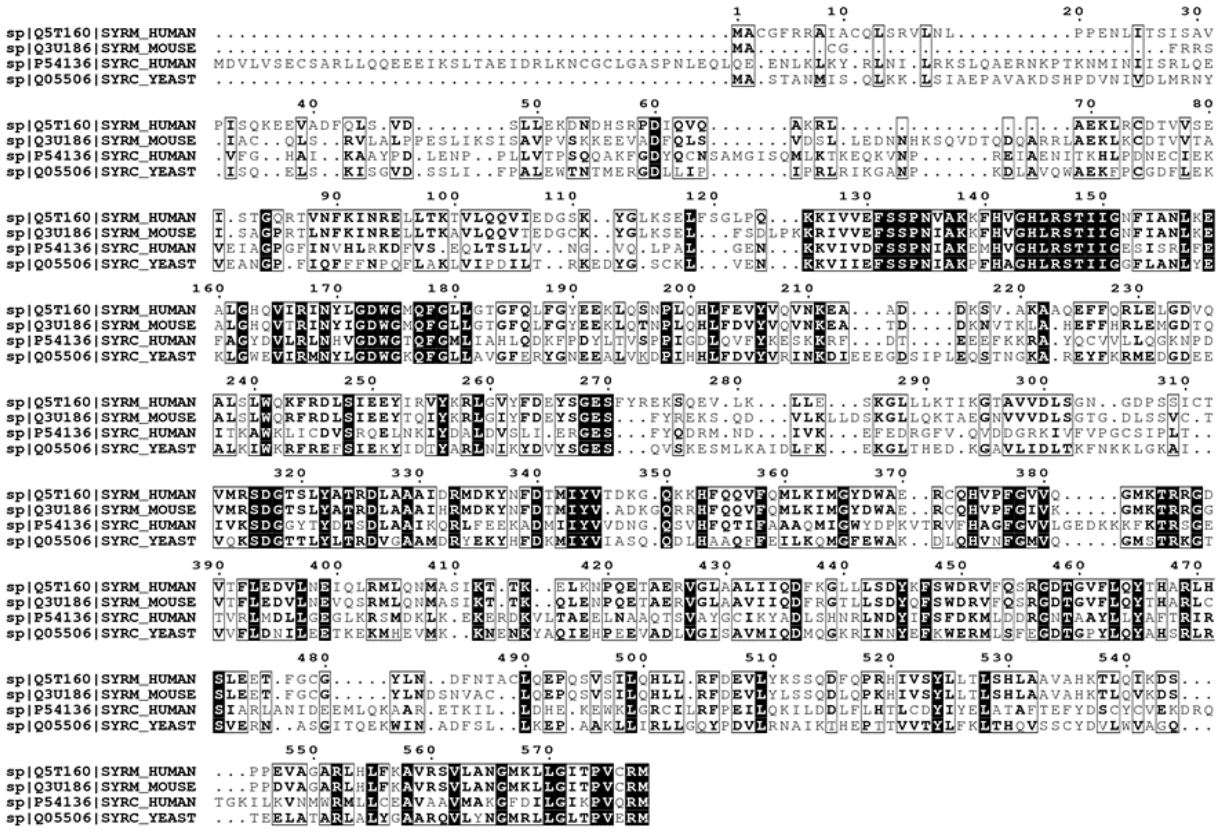


Figure 3.1 – Multiple sequence alignments of (A) human EARS2 and (B) human RARS2 sequences with other sequences. **(A)** – Sequence alignment of human EARS2 (first line), mouse EARS2 (second line), human EARS (third line) and *T. thermophilus* EARS (fourth line). **(B)** – Sequence alignment of human RARS2 (first line), mouse RARS2 (second line), human RARS (third line) and *S. cerevisiae* EARS (fourth line). Black color represents sequence identity between all proteins, grey color represents sequence similarity between proteins and white color represents no identity at all. Alignments were performed in T-coffee software and treated by ESPript software. Information in [58].

As we can observe in **Figure 3.1A** hEARS2 alignment, overall, there are few identical regions, as can be seen by the low number of black squares. However, EARS2 from mouse was the protein with the most identical positions accounting a total of 445 amino acids which corresponds to a total of 85.1% identity to hEARS2. Sequence similarity was calculated to be around 90 %. In the other hand, the human cytosolic form of EARS, which has 3 times more amino acids than its mitochondrial pair, accounts only 112 identical positions corresponding to a 7.23% identity. For this reason, using this protein for future studies comparison may not be a good approach. For EARS from *T. thermophilus*, this protein presents 31.4% identity (168 identical positions) and a sequence similarity of around 37% comparing with hEARS2. Despite the lower identity percentage, comparing with EARS2 from mouse, we decided to use EARS from *T. thermophilus* has a template to obtain a model from homology for hEARS2, as this protein has already a 3D structure available. However, given the lack of identity between sequences, probable future comparison between these proteins, regarding biophysical and biochemical studies, may not be the most accurate procedure.

As what concerns hRARS2 alignment, **Figure 3.1B**, we see a significant number of black squares meaning that, overall, there are regions of the sequences with full identity. For example, for RARS2 from mouse we have an identity of 86.5% and a sequence similarity of 93%, identical as observed for hEARS2, and as can be expected from a mammalian organism, unfortunately there is no 3D structure for this protein. Looking into the human cytosolic form of RARS, we have around 35.14% identity (around 170 identical positions) with a sequence similarity of 37%. Given the difference between the sequences, it is not likely to further compare studies between them, despite having a similar function. For cytosolic RARS from yeast *S. cerevisiae* we obtained an identity percentage of 40.6% with a sequence similarity of 58% comparing with hRARS2, better than the obtained for human cytosolic RARS. Given these results and knowing that 3D structures for human and yeast RARS are available, we decided to use *S. cerevisiae* as a template for prediction of a model for hRARS2 structure.

As such, as a second approach, both protein´s structures, human EARS2 and human RARS2, were modelled by homology, in accordance with data from multiple sequence alignments, so that, at this point we could visualize a probable structure for each protein. The resulting models can be found in **Figure 3.2**.

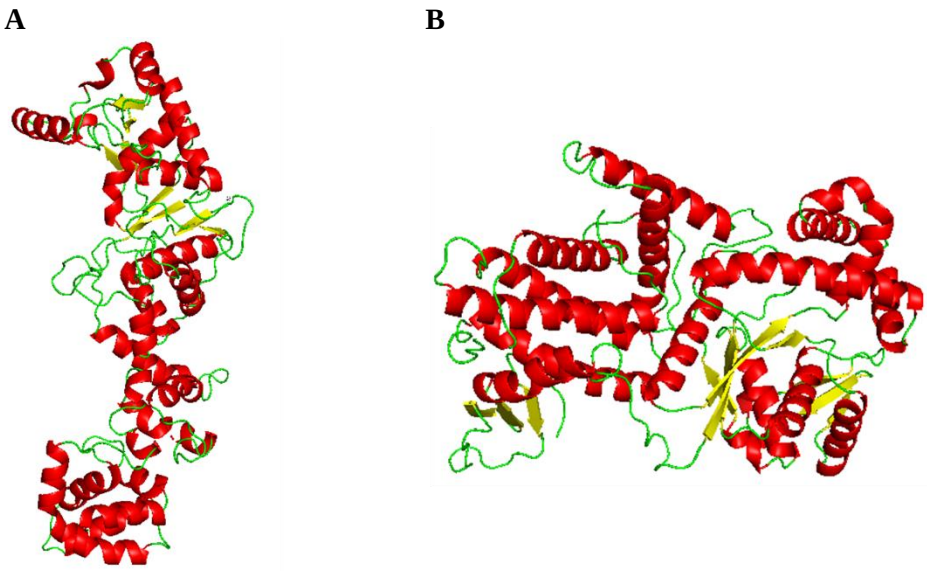


Figure 3.2 – Structural models of (A) hEARS2 and (B) hRARS2 proteins
(A) – hEARS2 model obtained by homology modelling using *T. thermophilus* EARS as template. **(B)** – RARS2 model obtained by homology modelling using *S. cerevisiae* RARS as template. Both structures were obtained by SWISS MODEL webserver and analyzed by Pymol program. Red color represents α-helical content, yellow color represents β-sheet content and green color represents loops.

Observing **Figure 3.2A**, we can see that hEARS2 model obtained has an overall flat-like structure, which can possibly be indicative of instability. We can see as well that protein is composed mainly by α-helical content, despite having some β-sheet content and loops. Interestingly, the model obtained is similar to the one predicted by AlphaFold algorithm (software available in October 2021, 1 year after we started this project) represented in introduction **Figure 1.10** which can indicate that our prediction could be a good approximation of the overall structure of the protein. As for hRARS2 model (**Figure 3.2B**), it has a mix of α-helical and β-sheet (mainly α-helical) with an overall globular-like shape, which can suggest a more compact and stable form. Just like for hEARS2, hRARS2 model is similar to the predicted by AlphaFold

algorithm, validating our model. These models can give us a new insight on protein´s global structure, and help us to look with criticism to the results obtained by biophysical approaches.

3.2. hEARS2

3.2.1. Expression tests

For optimization of recombinant heterologous expression of human EARS2, we transformed *E. coli* Rosetta cells with the mature form of human EARS2 plasmid. We chose these cells, because they are ideal for rare codon gene expression, and hEARS2 was designed with the native codon sequence. Cells were grown in LB medium and protein expression induction was performed with 0.5 mM IPTG at different temperatures of cell growth (22 ºC, 30 ºC and 37 ºC). hEARS2 expression at the different conditions was assessed in both soluble and insoluble fractions. **Figure 3.3** shows the profile of hEARS2 expression followed by SDS-PAGE blue stained gels and Western blot.

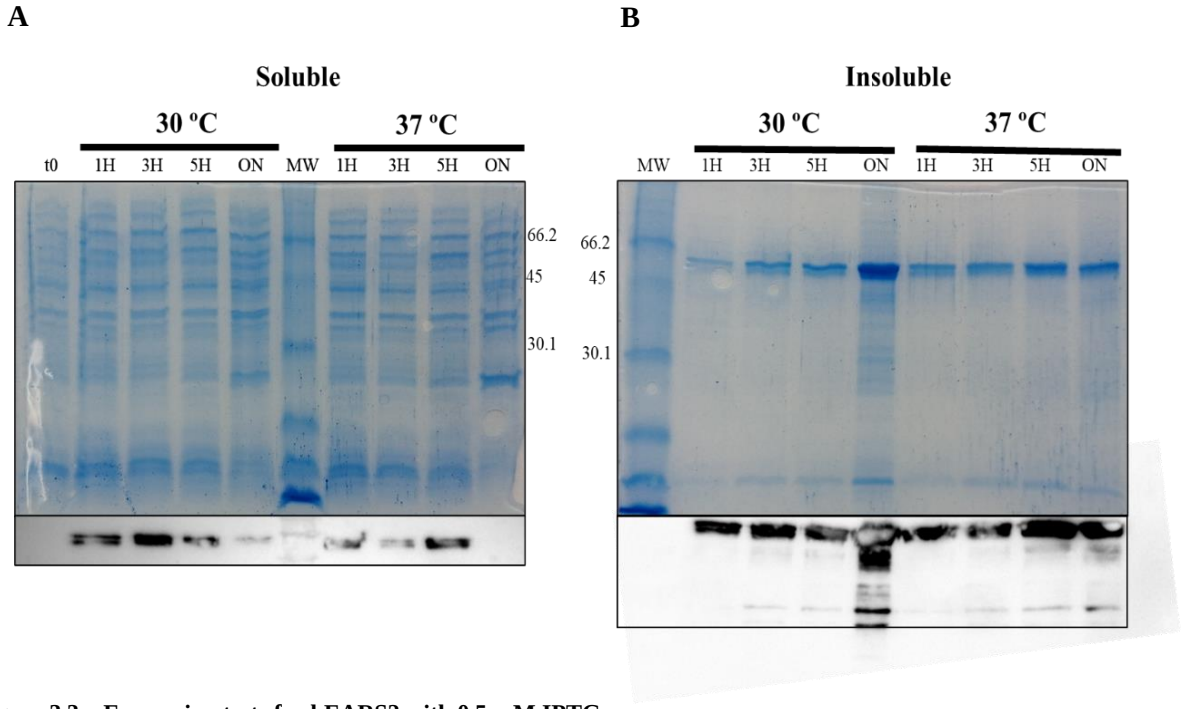


Figure 3.3 – Expression tests for hEARS2 with 0.5 mM IPTG
(A) Soluble fraction - Upper panel: Blue stained SDS-PAGE gel for timepoints after induction with 0.5 mM IPTG at 30 ºC and 37 ºC; Lower panel: Western blot membrane, portion around 60 kDa, for the respective timepoints with anti-his tag antibody; **(B) Insoluble fraction –** Upper panel: Blue stained SDS-PAGE gel for timepoints after induction with 0.5 mM IPTG at 30 ºC and 37 ºC; Lower panel: Western blot membrane, portion around 60 kDa, for the respective timepoints with anti-his tag antibody. MW stands for molecular weight size marker, t0 represents timepoint before IPTG induction, 1H represents timepoint after 1 hour of induction, 3H represents timepoint after 3 hours of induction, 5H represents timepoint after 5 hours of induction and ON represents timepoint after overnight induction.

According to the **Figure 3.3** we can see that hEARS2 is expressed in soluble fraction but more abundantly in insoluble fraction, where we can observe a dominant band around the expected molecular weight of the protein in the blue stained gel. In the soluble fraction, analyzing the western blot, we can see for 30 ºC and 37 ºC conditions, that incubation overnight results in loss of protein, possibly by degradation or by moving to inclusion bodies. Protein is expressed at a good level for condition 3 hours 30 ºC or 5 hours 37 ºC. In an attempt to further optimize protein expression, we evaluate the effect of low temperature on protein expression in soluble fraction. For this, we assessed protein expression at 22 ºC overnight.

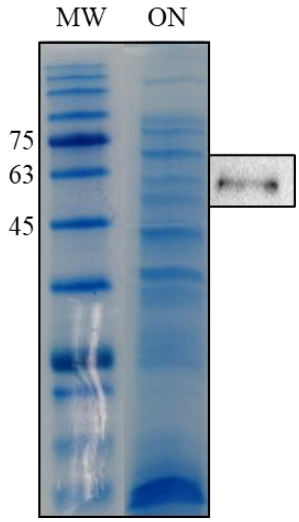


Figure 3.4 – Expression test for hEARS2 at 22 ºC ON at 0.5 mM IPTG.
Blue stained SDS-PAGE gel for ON timepoint after 0.5 mM IPTG induction at 22 ºC with Western blot band corresponding to molecular weight of hEARS2 marked with antibody anti-his tag.

Figure 3.4 is representative of the profile obtained for hEARS2 expression with 0.5 mM IPTG during overnight growth at 22 ºC, which indicates that although there is no high expression level as evaluated by the blue stain gel, at lower temperatures of incubation, protein is not degraded overnight (confirmed by Western blot, side panel), at least at the same extent, as at 30 ºC and 37 ºC temperature growth.

From the expression tests at different conditions, we always observed that hEARS2 is highly expressed in inclusion bodies (insoluble fraction). Protocols for purification from inclusion bodies, use denaturing agents, which for some protein can be harsh, and the renaturation process can result in a protein conformation different from the native form. Therefore, we proceed for protein purification from the soluble extract, keeping in mind the low yield of expression. Also, we decided to use as reference a protocol reported for human cytosolic RARS[23], whose purification was performed from the soluble fraction.

Therefore, the ideal conditions for protein expression for further scale-up and purification from soluble fraction is induction for 3 hours at 30 ºC, but overnight induction at 22 ºC can also be a good alternative, if we want to have higher mass of cells.

3.2.2. Purification

Following optimization of the protein expression we move forward to purify hEARS2. As already mentioned, as a first approach we adapted our protocol to the one reported for class I human cytosolic RARS[23], once, as far as we know, there is no report on the purification of heterologous expressed mt-aaRSs. As such, we took advantage of the C-terminal His-tag and performed an affinity chromatography using a His-trap column. Soluble extract was prepared in 30 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole and 0.5 mM PMSF (Buffer A) and applied in the His-trap column. Bound proteins eluted in gradient from 20 mM to 500 mM imidazole in the same buffer. Chromatogram obtained can be seen in **Figure 3.5A**, with the arrow representing the fraction collected.

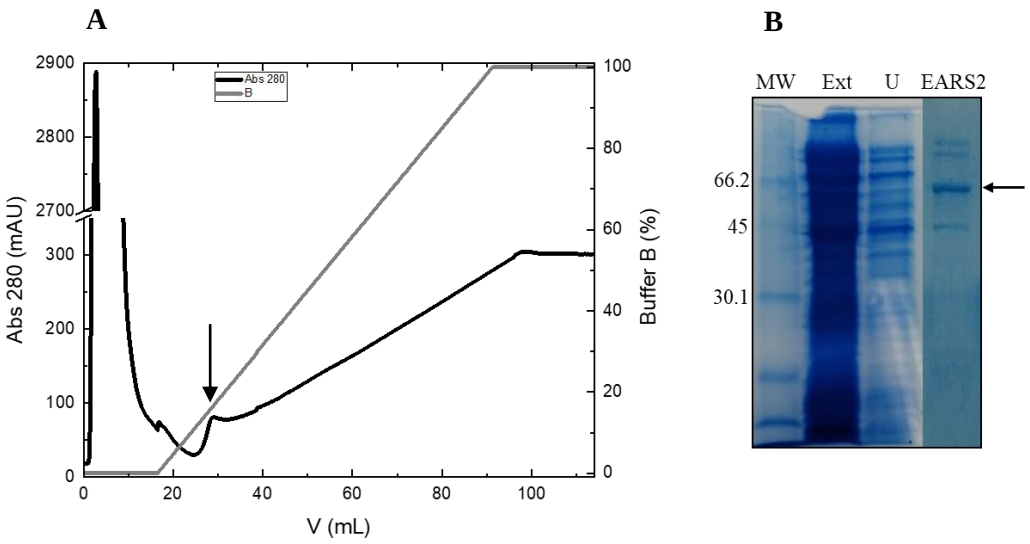


Figure 3.5 – Purification of hEARS2.
(A) Affinity chromatogram (5 mL His-trap column equilibrated in 5 volumes of buffer A and with imidazole gradient in buffer B) with arrow representing peak where protein eluted. **(B)** SDS-PAGE gel for fraction analysis: MW – molecular weight size marker; Ext – soluble extract; U – Unbound fraction; EARS2 – Peak corresponding to EARS2 elution (arrow represents band corresponding to EARS2)

The bound protein eluted around 20 % buffer B (100 mM imidazole). We analyzed the eluted fraction by SDS-PAGE (**Figure 3.5B**), to confirm the presence of the protein of interest. As can be seen in fraction 1, there is a band with molecular weight of hEARS2 as expected, although we can identify also other bands but with lower yields.

Protein fraction was cleared from imidazole that could have damaging effect on protein folding during long storage. This procedure was made by a series of dilution and concentration using amicon membranes. Final protein was stored in 30 mM Tris-HCl, pH 8.0, and protein concentration was determined. Using this procedure, we obtained approximately 0.26 mg protein per gram of cells around 70% pure.

To improve protein purification protocol, we changed the amount of NaCl in the binding buffer. High amounts will diminish unspecific interactions of soluble extract proteins to the column and also between hEARS2 and these proteins. Therefore, binding buffer was equal to the above buffer A but NaCl concentration was increased to 300 mM. Moreover, to start with higher amount of material we culture cells at 22 ºC overnight. Soluble extract preparation was similar to the previous one, and it was applied in the His-trap column. The elution profile can be observed in **Figure 3.6A**.

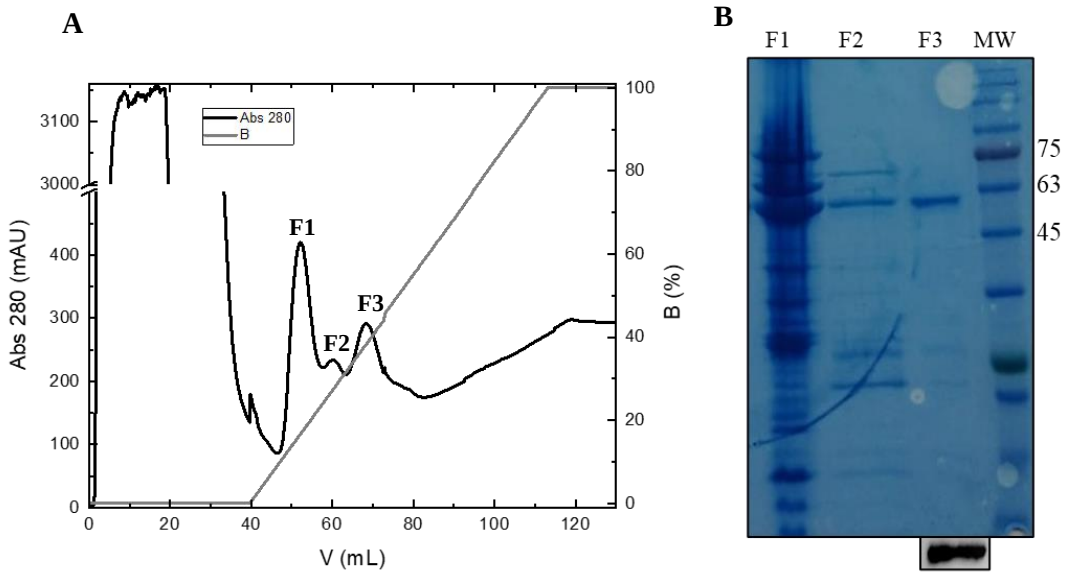


Figure 3.6 - **Purification of hEARS2.**
(A) Affinity chromatogram of hEARS2, using buffer A and buffer B to perform an imidazole gradient (300 mM NaCl): F1 – first peak eluted; F2 – second peak eluted; F3 – third peak eluted. **(B)** SDS-PAGE of the corresponding fractions to assess EARS2 purity and further presence by Western blot.

This time instead of one peak we obtained 3 peaks, fraction 1 (F1), fraction 2 (F2) and Fraction 3 (F3). SDS-PAGE (**Figure 3.6B**), showed that hEARS2 eluted in fraction at around 30% buffer B (150 mM imidazole), with a purity grade around 95%. After verifying that the fraction corresponds to the protein of interest by western blot, we proceeded with the removal of imidazole. Again, we used the methodology of dilute and concentrate in buffer 30 mM Tris-HCl, pH 8.0 without imidazole or salt, using the amicons. However, during the centrifugation cycles, most of the protein precipitated. Although with this protocol we improved EARS2 purity, the yield of the purification was reduced (**Table 1** annex). Also, we realized that the protein that was stored at -20 ºC, precipitated when defrosted. It was mandatory to resolve this protein instability problem. To do so, we decided to add to the buffer a well-known protein stabilizer, glycerol. Also, we sought for an alternative method to the dilution/concentration cycles for removal of imidazole and NaCl. We tested a desalting column, that permitted quickly change the protein buffer.

We repeated the purification procedure from a cell growth at 30 ºC for 4 hours, using 20% glycerol in all buffers, and maintaining the 300 mM NaCl, that provided a better separation profile. After the affinity chromatography, that presented an elution profile similar to **Figure 3.6A**, hEARS2 fraction was injected into a 5 mL Hi-trap Desalting G25, equilibrated in 30 mM Tris-HCl, pH 8.0, 50 mM NaCl and 20% glycerol and analyzed by SDS-PAGE (**Figure 3.7**). This procedure prevented precipitation during purification steps. Moreover, pure fractions were stored at -80 ºC. The storage condition also, result in negligible protein precipitation. Purification yield obtained was 0.49 mg of protein per gram of cells, the better obtained so far.

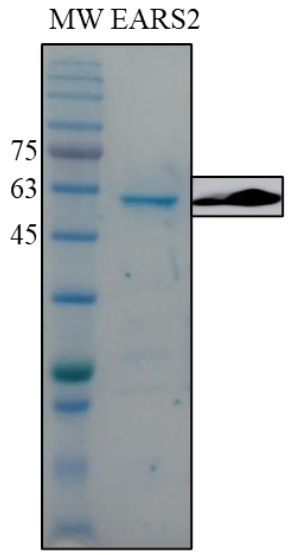


Figure 3.7 – hEARS2 Purification.
SDS-PAGE blue stain gel of hEARS2 final fraction purified and Western blot of hEARS2 marked with anti-his tag antibody at molecular weight corresponding to the band observed in the blue stain gel.

3.2.3. Structural characterization

Having a considerable amount of pure hEARS2 we proceed to structurally characterize the protein, using different spectroscopic techniques. We have analyzed protein secondary structure by circular dichroism (CD). hEARS2 spectrum was performed in the Far-UV region.

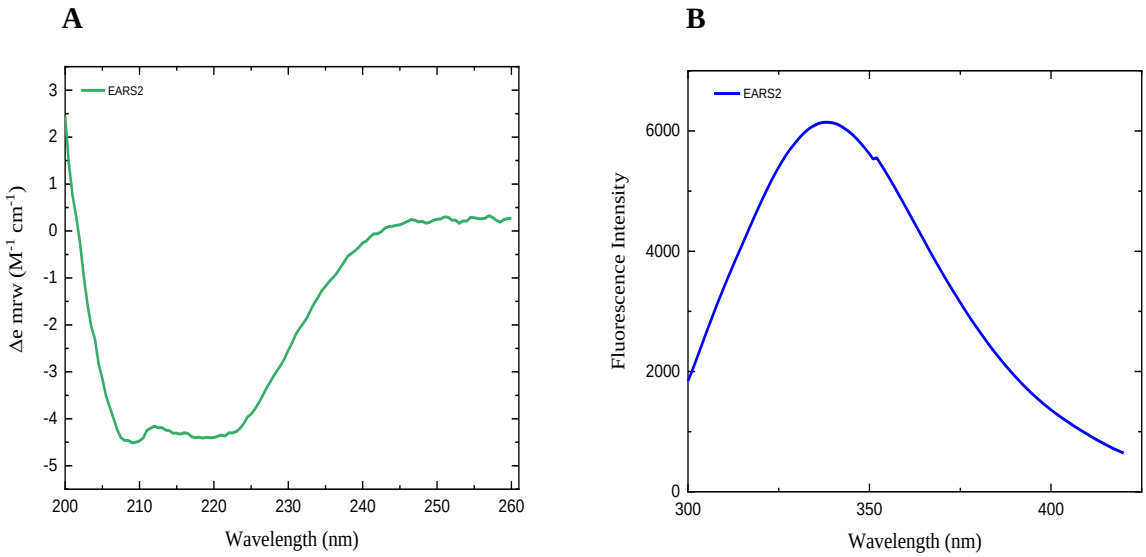


Figure 3.8 – hEARS2 structural characterization.
(A) Far-UV CD spectrum of hEARS2 (protein concentration was 0.1 mg.mL^{-1}); **(B)** Tryptophan fluorescence emission spectrum of hEARS2 (protein concentration was 1µM). The buffer contained 30 mM Tris-HCl, 50 mM NaCl and 20 % glycerol.

Figure 3.8A shows that the protein has an α/β characteristic fold. The two minimums at 208 nm and 222 nm indicate that it is mostly composed of a α-helical structure. This spectrum is characteristic of a protein in a folded state, confirming that we were able to purify and storage hEARS2 in the native form. Further, we used tryptophan fluorescence emission to evaluate the tertiary structure of the protein. **Figure 3.8B** is a representation of the obtained spectrum, with a maximum at 340 nm indicating that tryptophans are slightly exposed to solvent although in a folded state.

3.2.4. Thermal stability

Taking advantage of the spectroscopic techniques available we evaluated hEARS2 thermal stability by performing a thermal denaturation, which would give us information about the melting temperature (T_m) of the protein, i.e., the temperature at which folded and unfolded states are equally populated. In **Figure 3.9** are the curves obtained by CD and tryptophan emission when we increase temperature from 25 ºC to 90 ºC at 1 ºC.min^{-1}.

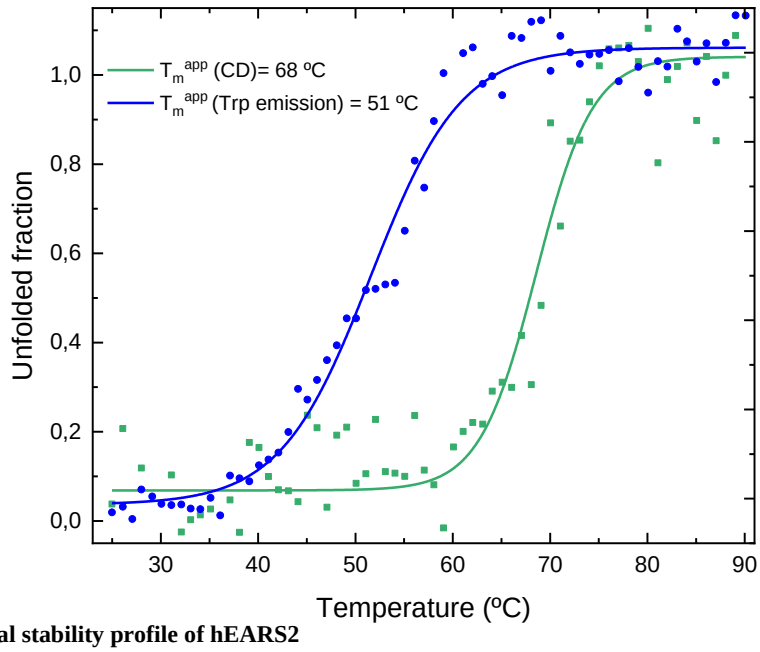


Figure 3.9 – Thermal stability profile of hEARS2
Thermal denaturation curves followed by CD (green squares) and tryptophan emission fluorescence (blue circles). The solid curves (green and blue, respectively) represent two state sigmoidal fit from which the apparent midpoint denaturation temperatures were determined, T_m^{app} (CD) was 68 ºC and T_m^{app} (tryptophan emission) was 51 ºC.

We also performed the reverse curve (renaturation profile), but it resulted in no major changes in CD signal, and inclusively at 90 ºC, we were able to observe protein precipitation. In this case, as the process of unfolding is not reversible, we were not able to driven thermodynamic parameters. Moreover, the midpoint transition must be considered as an apparent melting temperature (T_m^{app}), that can still be used as a measure of protein stability. Following unfolding by secondary structure, we obtained a T_m^{app} of 68 ºC. The unfolding curve yield a cooperative transition between folded and unfolded state.

By measuring fluorescence emission, we follow changes in the tryptophan exposure to the solvent, through a combine change of intensity variation with a shift of the maximum wavelength (λ_{max}) to red (355-360 nm).

Although no major red shift was observed, there was a change in fluorescence intensity, yielding a cooperative unfolding curve with a T_m^{app} of ~51 ºC. These results suggest that tryptophan´s in the purified fraction of hEARS2 are slight exposed to the solvent.

Interestingly is the fact that we obtained different T_m^{app} values by the two techniques, which could indicate that first there is a loss in tertiary structure (lower T_m^{app} by tryptophan emission) and only then, secondary structure rearrangements are lost.

3.2.5. Titration with glutamate

Further, we tested the effect of glutamate, one of the substrates of hEARS2, in the structure of the protein. To do so, we followed a titration experiment by CD, increasing glutamate concentration from 0.25x up to 1.5x concentration of the protein. In **Figure 3.10A** we can observe the evolution of hEARS2 structure at each condition of the titration.

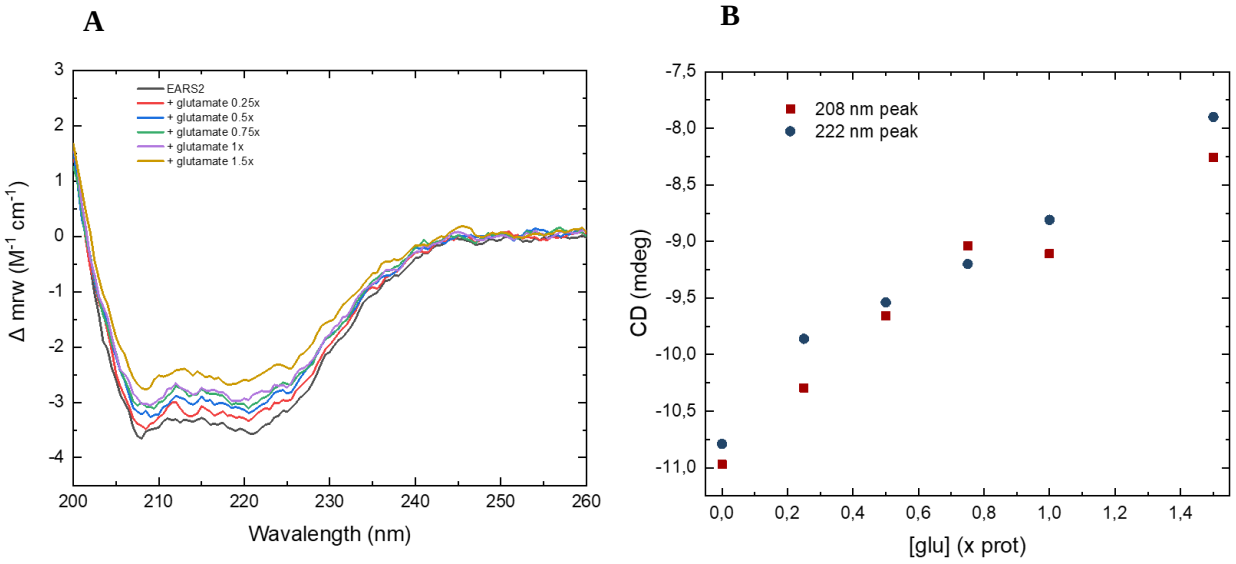


Figure 3.10 - Titration of hEARS2 with glutamate.
(A) - Far-UV CD spectra of hEARS2 in the absence of glutamate (solid black line) and hEARS2 in the presence of increasing glutamate concentrations: 0.25X EARS2 (solid red line), 0.5X EARS2 (solid blue line), 0.75X EARS2 (solid green line), 1X EARS2 (solid purple line) and 1.5X EARS2 (solid roasted yellow line). **(B)** – Variation of the intensity of the characteristic minimums at 208 nm and 222 nm with the increase of glutamate concentration.

To analyze these results we plotted the values of the minimums 208 nm and 222 nm in mdeg in function of glutamate concentration (**Figure 3.10B**). From the observation of the figures, we can conclude that as glutamate concentration increases, protein losses its global structure, as we can see by the increase of both minimums´ intensity. Interestingly, the ratio between the values of the characteristic minimums does not alter (**Figure 6.3** annex), indicating that the global fold does not change for each glutamate concentration. These results suggest that glutamate has a concentration dependent negative effect in hEARS2 structure.

We also evaluated the thermal stability in the presence of saturated concentration of substrate (1.5x concentration of hEARS2). In **Figure 3.11**, we can see the comparison of the thermal denaturation curves of hEARS2 in presence of 1.5x concentration of glutamate and absence. The shape of the unfolding curves changes, although a similar T_m^{app} was obtained. The unfolding curve in presence of substrate presented a less cooperative transition, which correlated with previous observations, and leads us to conclude that at this glutamate concentration the protein has already loss part of its secondary structure.

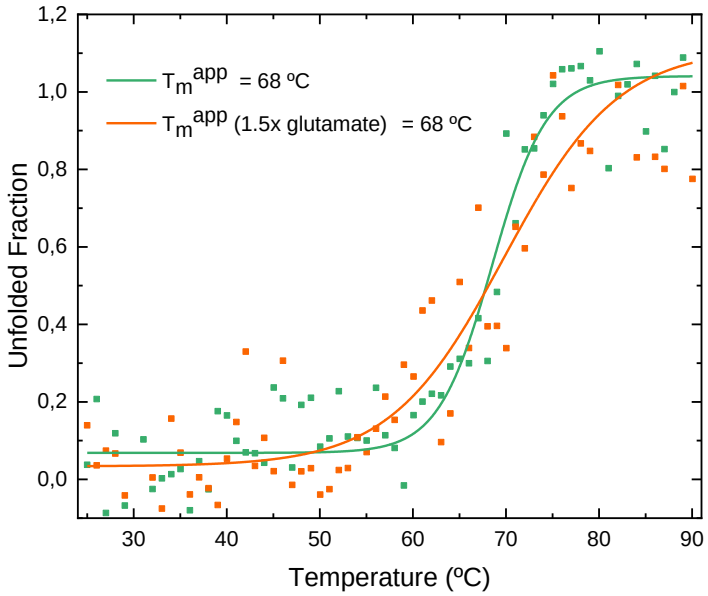


Figure 3.11 – Comparison of thermal stability profile between hEARS2 in the absence and presence of glutamate 1.5X [hEARS2]
Thermal denaturation curves in the absence of glutamate (green squares) and in the presence of glutamate 1.5x [hEARS2] (orange squares). Solid curves (green and orange, respectively) represent a two state sigmoidal fit to hEARS2 in the absence and in the presence of glutamate 1.5x more concentrated. From both fits it was determined the apparent melting temperature, T_m^{app}, which was 68 ºC for the two of them.

3.3. hRARS2

3.3.1. Expression tests

To optimize the recombinant heterologous expression of human RARS2, initially, *E. coli* Rosetta cells were transformed with plasmid encoding for mature form of human RARS2. To determine the optimal protein expression condition, we tested different IPTG concentrations (0.5 mM and 1 mM) and different temperatures for cell growth. Protein expression was evaluated in soluble and insoluble fractions from the extracts obtained at the different conditions. **Figure 3.12** shows the blue stained gels of the soluble and insoluble fractions.

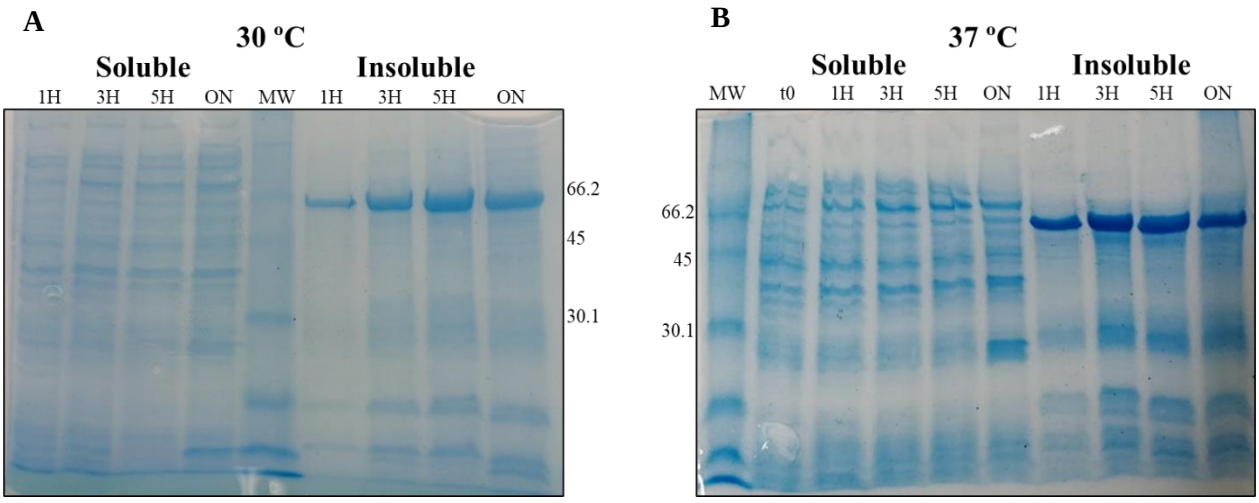


Figure 3.12 – Expression tests for hRARS2 with 1 mM IPTG
(A) – Blue sained SDS-PAGE gel for timepoints after induction with 1 mM IPTG at 30 ºC. **(B)** – Blue sained SDS-PAGE gel for timepoints after induction with 1 mM IPTG at 37 ºC. t0 represents timepoint before IPTG induction, 1H, 3H, 5H and ON represent respectively timepoints 1 hour, 3 hours, 5 hours and overnight after IPTG induction.

These results point towards a high expression of RARS2 in insoluble fraction. To better analyze the expression in soluble fraction and to confirm the presence of the protein of interest in soluble and insoluble parts we performed a Western blot using an antibody that recognizes his-tag sequences. **Figure 3.13** shows the Western blot membranes for soluble fraction with 0.5 mM and 1 mM IPTG at 30 ºC and 37 ºC conditions.

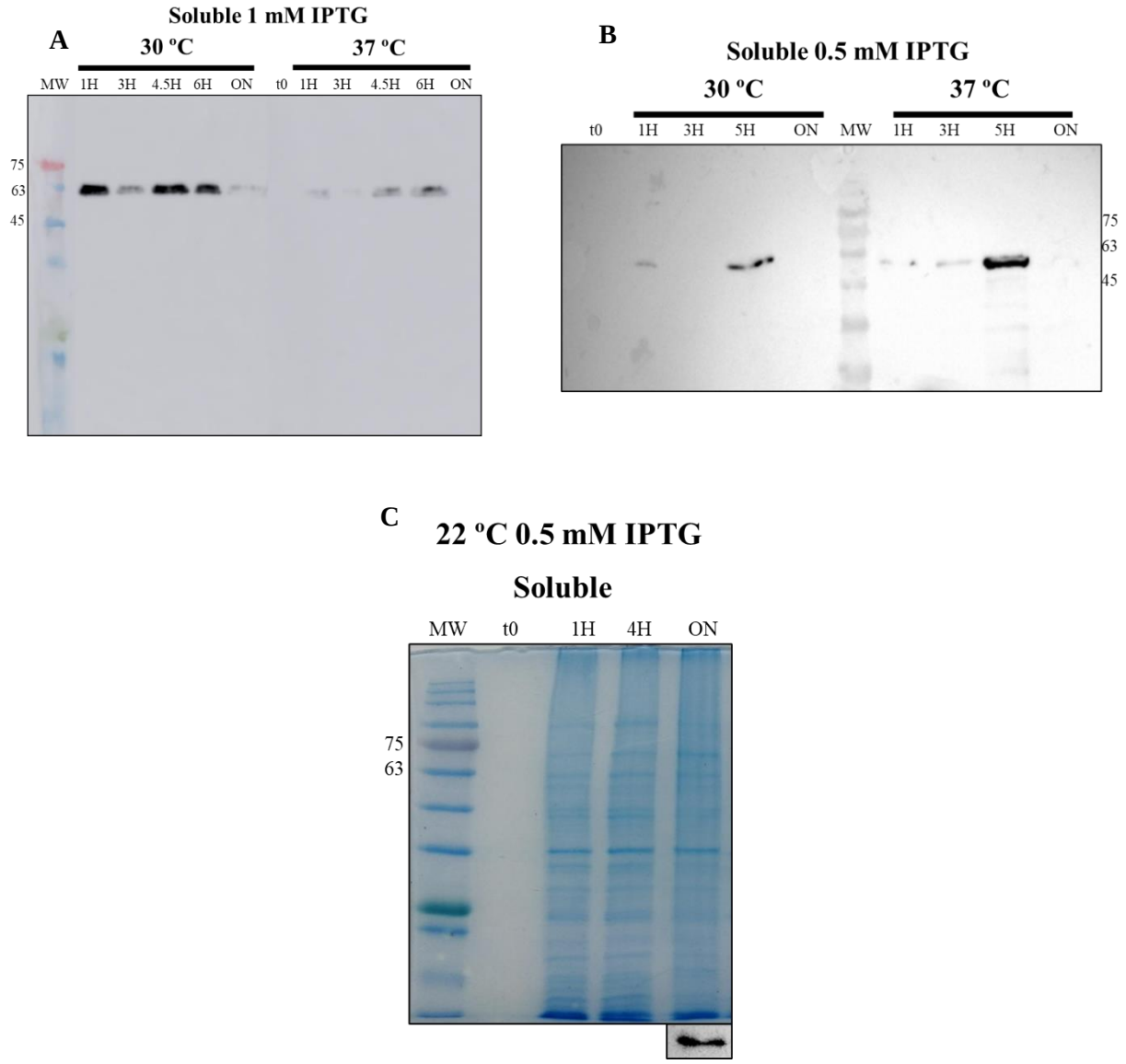


Figure 3.13 – Expression tests for hRARS2 0.5 mM and 1 mM IPTG.
(A) – Western blot membrane for hRARS2 expression assessment at different timepoints in soluble fraction at 1 mM IPTG. **(B)** – Western blot membrane for hRARS2 for timepoints in soluble fraction at 0.5 mM IPTG. **(C) –** Blue stained gel for hRARS2 expression after induction with 0.5 mM IPTG at 22 ºC for soluble fraction and Western blot portion for ON timepoint. Protein was detected using an anti-his tag antibody.

We observed that RARS2 is also expressed in soluble fraction, although at a lower yield. Higher concentrations of IPTG seem to better induce protein expression in soluble fraction. Moreover, after overnight culture RARS2 does not appear in soluble fraction, suggesting that it was degraded or/and rescued for inclusion bodies (insoluble fraction), except for overnight growth at 22 ºC.

3.3.2. Purification

After optimization of hRARS2 expression conditions, we pass to establish a purification protocol for this protein. As expression at lower timepoints was reduced we used a condition with 0.5 mM IPTG during overnight growth at 22 ºC, so that we can compensate lower expression with higher amounts of cell material. Likewise, we adapted purification protocol from[23] and from results obtained with hEARS2. As hRARS2 has a C-terminal His-tag an affinity chromatography using a 5 mL His-trap column was performed. Soluble extract injected was prepared similar as described for hEARS2, in buffer A with 300 mM NaCl and 20% glycerol. Proteins that bound to the column eluted in a gradient of imidazole from 20 mM up to 500 mM. **Figure 3.14A** shows the chromatogram we obtained.

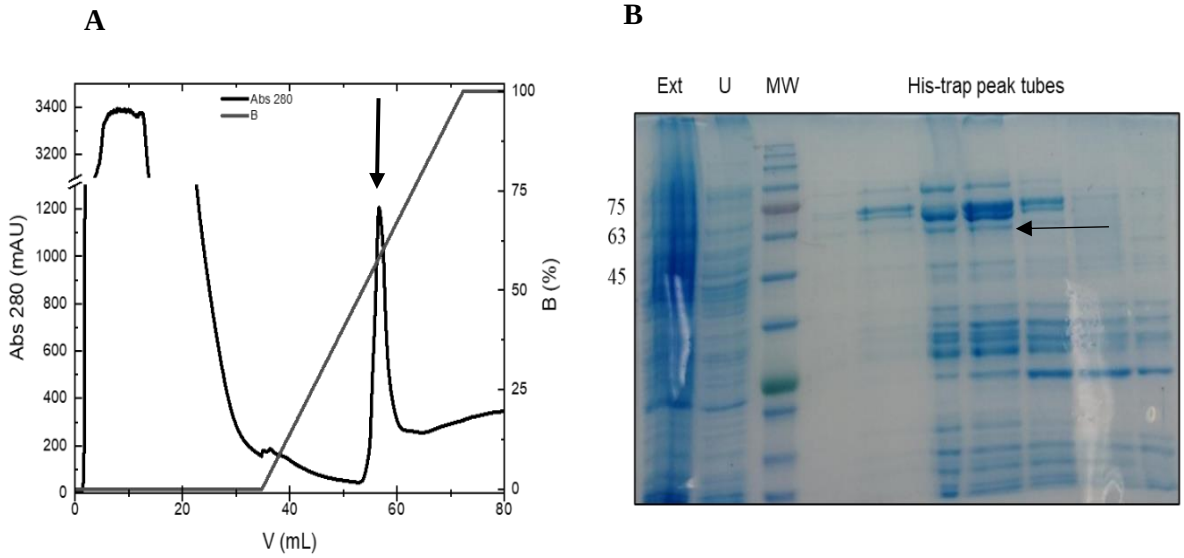


Figure 3.14 - Purification of hRARS2.
(A) Affinity chromatogram of hRARS2 using buffer A (300 mM NaCl and 20% glycerol) and buffer B for imidazole gradient. **(B)** SDS-PAGE gel for fraction analysis: MW – molecular weight size marker; Ext – extract; U – unbound; His-trap peak tubes – tubes corresponding to fractions of the chromatogram peak (represented with an arrow in (A)).

Bound proteins eluted ~50% of buffer B (250 mM imidazole) and the fraction corresponding to the peak (represented with an arrow in **Figure 3.14A)** were collected and analyzed by SDS-PAGE (**Figure 3.14B**), to confirmed the presence of the interest protein. Looking at **Figure 3.14B** we can see that the fraction collected had many contaminants. Still, we observed a band with the molecular weight around the expected for hRARS2 (~63 kDa), represented by an arrow. However, we note that other proteins with a higher molecular weight appear with a higher signal than hRARS2 band. Imidazole and NaCl were removed of the fraction by dilution and concentration series with amicons and final fraction was stored in buffer 30 mM Tris-HCl pH 8.0, 20% glycerol.

Taking these results into account, it was important to improve the protocol employed for hRARS2 purification. Therefore, we selected another condition of expression of hRARS2, induction with 1 mM IPTG for 6H at 37 ºC, as it was previously observed, that this condition yielded a good expression of the protein. We maintained the same buffer changing only NaCl concentration from 300 mM to 500 mM, so that we could decrease even more interactions between the proteins with column and the proteins with hRARS2, and hopefully, improve separation of protein eluted. Soluble extract was prepared similar way as described

above. Using these conditions, we now performed an affinity chromatography and obtained a similar chromatographic profile.

Analysis of the fraction collected was performed by SDS-PAGE (**Figure 3.15A**) and Western blot. We were able to detect the presence of the protein of interest, however with the same contaminants observed before. To avoid protein precipitation, we used a 5 mL Hi-trap Desalting G25 column in buffer C (details in section 3.2.2), to interchange buffer more rapidly. Final fraction eluted from this column was analyzed by SDS-PAGE and Western blot, **Figure 3.15B**.

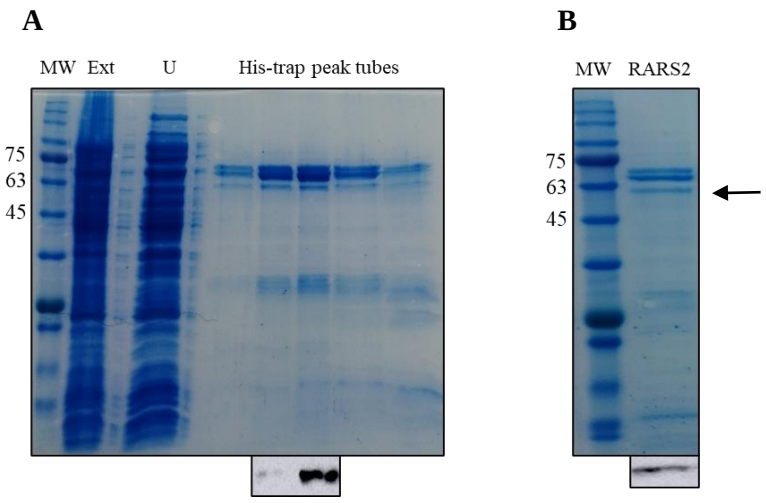


Figure 3.15 – Purification of hRARS2.
(A) – SDS-PAGE gel for fraction of chromatogram analysis and Western blot membrane portion around 60 kDa: MW – molecular weight size marker; Ext – extract; U – unbound; His-trap peak tubes – tubes corresponding to fractions of the chromatogram peak (represented with an arrow in (A)). **(B)** SDS-PAGE and Western blot corresponding to 60 kDa of the final hRARS2 fraction.

Protein presence was confirmed, although hRARS2 was not the major component of the mixture obtained, being less than 50% of all fraction. Despite the presence of contaminants, we stored the protein in buffer C at -80 ºC for further preliminary characterization studies. Fraction yield was determined to be 0.14 mg of protein per gram of cells, an overestimated value as hRARS2 is not even representative of half of the mixture.

3.3.3. Preliminary structural characterization

hRARS2 fractions structural characterization was conducted, taking advantage of the available spectroscopic techniques. Secondary structure was assessed by CD in the Far-UV region and the spectrum obtained can be observed in **Figure 3.16A**.

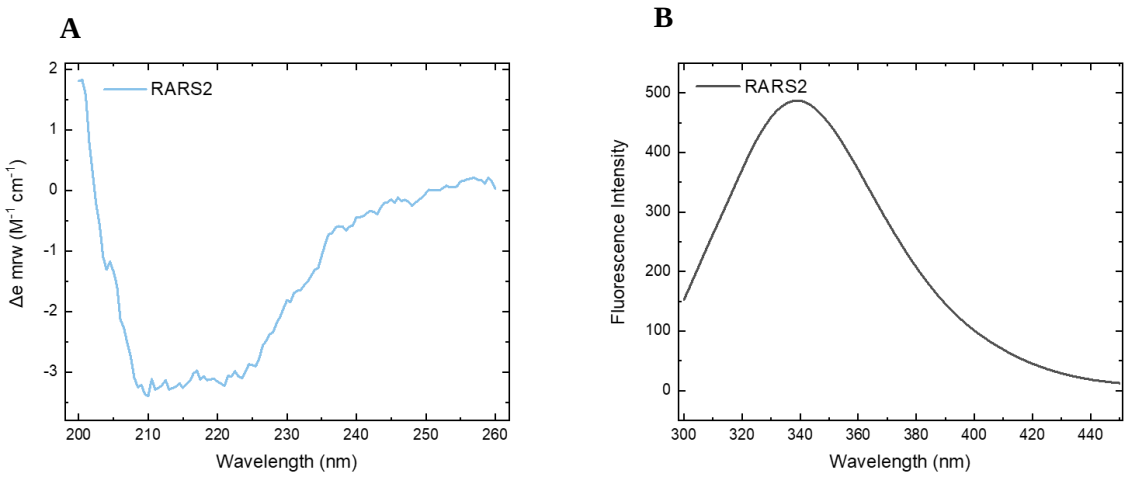


Figure 3.16 - hRARS2 structural characterization.
(A) Far-UV CD spectrum of hRARS2 (protein concentration of 0.1 mg.mL^{-1}); **(B)** Tryptophan fluorescence emission spectrum of RARS2 (protein concentration of 1 µM). Protein used to obtain these spectra was in buffer 30 mM Tris-HCl pH 8.0, 50 mM NaCl and 20% glycerol

Analyzing the spectrum, we see a α/β fold profile, with minimums at 208 nm 222 nm, however signal to noise ratio is high, and the ellipticity of the solution at the typical concentration of 0.1 mg.mL^{-1} is low. This likely occurs due to presence of a mixture of conformers, and not only hRARS2, as could be expected from the purity of the fraction, with the contaminants interfering in the spectrum. Taking advantage of the 4 tryptophans of hRARS2, we used fluorescence spectroscopy to assess tertiary structure of the protein. The result is depicted in **Figure 3.16B**, and we can see a maximum in emission at 336 nm indicating that tryptophans are slightly exposed to solvent.

Thermal stability of hRARS2 fractions was also evaluated by thermal denaturation assays using the same techniques as used for structural characterization. The melting curves can be found in **Figure 3.17**.

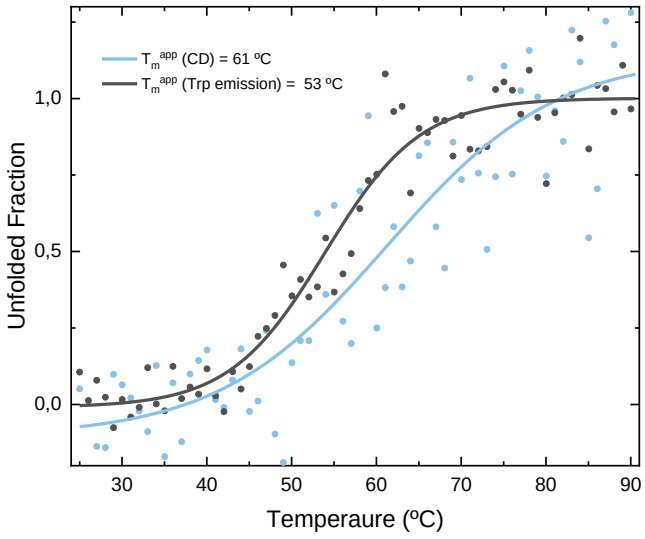


Figure 3.17 – Thermal stability profile of hRARS2
Thermal denaturation curves followed by CD (blue squares) and by tryptophan emission fluorescence (grey circles). Solid blue and grey curves are the sigmoidal fit to the data. T_m^{app} (CD) was 61 ºC and T_m^{app} (Tryptophan emission) was 53 ºC.

The CD thermal denaturation followed loss in secondary structure content, and in agreement with the obtained spectrum, the unfolding transition was not cooperative, a profile that is many times observed for mixture of proteins. Even though we fitted the data to a sigmoidal curve and obtained a T_m^{app} of 61 ºC. Moreover, at the end of the unfolding at 90 ºC we observed high amount of precipitated protein in the cuvette. Changes in the tryptophan exposure due to increase in temperature were monitored, by fluorescence spectroscopy. **Figure 3.17** in grey is representative of thermal unfolding obtained by fluorescence. Following changes in tertiary structure yielded a cooperative curve with a T_m^{app} of around 53 ºC. Nevertheless, the obtained T_m^{app} values, although reproducible between the two purifications, must be considered as very preliminary data due to the protein solution analyzed.

4. Conclusions

Recently, several studies have been reporting mutations in genes coding for mt-aaRSs, proteins which are involved in the synthesis of mitochondrial respiratory chain complexes. Some of these mutations give rise to the appearance of pathogenic variants, which impair energy production and mitochondrial respiratory process, leading to the development of some MD, like leukodystrophies [14, 36]. The underlying mechanisms through which variants of these proteins cause MD are not yet fully understood, and so, studies focusing the molecular level need to be conducted, to give new insights in terms of proper diagnosis and treatment[40]. In this project we aimed to establish protocols for heterologous expression and purification of two wild-type human mt-aaRSs, hEARS2 and hRARS2. Furthermore, we also wanted to study both proteins´ structure, folding and stability properties. This initial approach is very important, because a successful purification of these proteins allows their structural characterization and eventually, the production of mt-aaRSs disease associated variants whose respective structural characterization may give clues about structural and functional differences between wild-type and variant protein, just like recently reported for human TARS2 case[59]. So far, reports on heterologously expressed, purified and characterized human mt-aaRS are exclusive for AARS2, HARS2, LARS2 and TARS2 cases[60].

First of all, we successfully established hEARS2 and hRARS2 heterologous expression in *E. coli* protocols , unlike other cases of human mt-aaRSs reported in which a different expression system was used for heterologous expression, for example a baculovirus expression system[61] or even in modified vaccinia Ankara-infected BHK 21 cells[62]. This is an important breakthrough as expression in *E. coli* is simpler and less expensive, thus providing an excelent tool for protein production. Then we moved on to establish purification protocols for each protein by affinity chromatography and then by desalting chromatography obtaining 95 % purity grade for hEARS2 and a fraction with a purity not higher than 50% for hRARS2. For hRARS2 this process needs to be optimized, a possible hypothesis to optimize it, is to purify the protein from inclusion bodies, as the yield for protein expression is higher than in soluble fraction. There has been reported for heterologous expression in *E. coli* and purification from inclusion bodies of human FARS2[63], however due to the complexity of this enzyme (chimera), and the fact that it is a class II aaRSs we decided to base our approach for hRARS2 in protocols described in for human RARS[23]. A scheme of the optimized protocols is described in **Figure 4.1**.

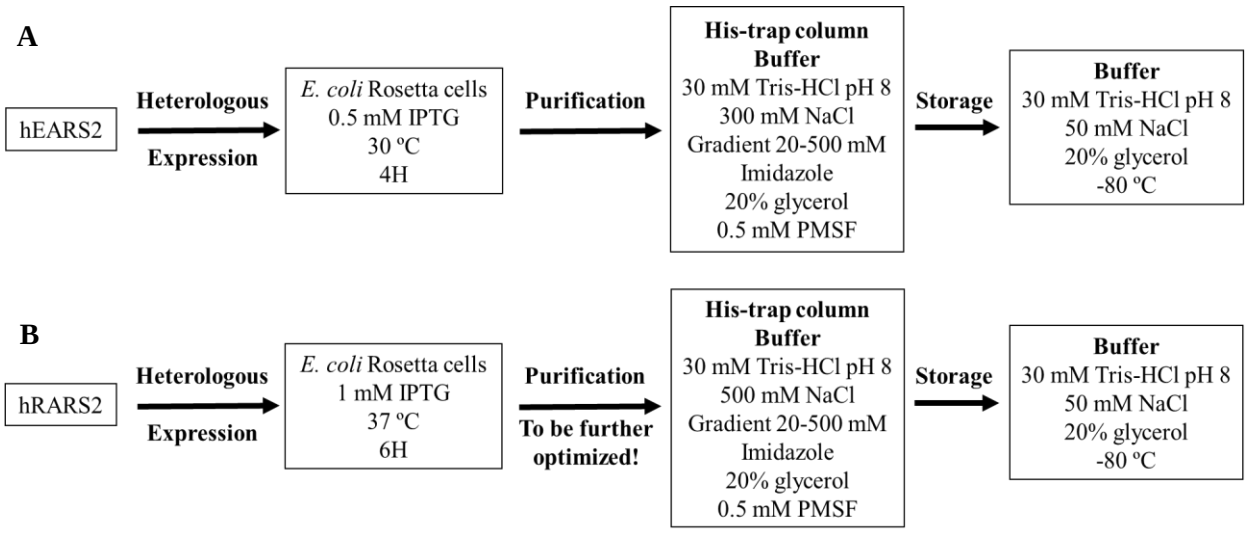


Figure 4.1 – Schematic representation of hEARS2 (A) and hRARS2 (B) expression and purification protocols and storage conditions

Taking advantage of the biophysical techniques available, CD and fluorescence spectroscopy and comparing it with the *in silico* studies, we evaluated secondary and tertiary structure of both mt-aaRSs. Our results for EARS2 suggest that this protein has a α/β fold, observed by the characteristic CD spectrum minimums at 208 nm and 222 nm. This result is corroborated by the *in silico* analysis performed, by which we see the presence of both secondary structure signatures, with a prevalence of α-helical content. Homology modelling results also suggest that EARS2 has a flat-like structure, which can explain the protein instability observed during protein purification optimization. By tertiary structure analysis, obtained from tryptophan emission spectrum, we observed by the maximum wavelength emission that tryptophans are likely to be exposed to the solvent. This result is corroborated by the analysis of the spectrum after fully desnaturation by temperature **Figure 6.2B** (annex), in which there is no red-shift in maximum wavelength. Furthermore we studied thermal stability of hEARS2 and our results were indicative of a T_m^{app} = 68 ºC obtained by CD and T_m^{app} = 51 ºC obtained by tryptophan fluorescence emission. This difference suggest that tertiary structure is firstly disrupted and only then, the protein loses its secondary structure arrangement. For hRARS2, our preliminary results showed as well a α/β fold with a globular-like structure, with α-helices being the major contributor for protein´s secondary structure. Tryptopahn fluorescence emission results also showed that protein´s tryptophans are exposed to the solvent. In terms of hRARS2 thermal stability our results showed a T_m^{app} = 61 ºC b y CD and T_m^{app} = 53 ºC by tryptophan fluorescence emission.

Some studies *in vitro* conducted in mt-aaRSs, for example for LARS2[64] or for TARS2[65] were based on the editing function of mt-aaRSs, as in some cases this function can be impaired (LARS2 case) or it can be robust (TARS2 case). Future editing function assays on hEARS2 and hRARS2 can be a good approach to complement structural characterization assays. Moreover, future comparison of tRNA aminoacylation rates between wild-type and mutant forms of hEARS2 and hRARS2, may also be an important functional study, as in some mt-aaRSs pathogenic variants it can range from no effect at all, to a decrease of around 80 fold in tRNA aminoacylation rate[60].

The work in here depicted, represents a major step towards the increase of knowledge regarding molecular properties and features of wild-type mt-aaRSs, whose variants are associated with MD. Moreover, this work in an important first step into the deepest of mt-aaRSs structure and from it, more targeted functional and structural studies can be conducted in order to becom closer to understand the molecular mechanism of MD. Knowing the secrets behind the structures of human wild-type mt-aaRS, may allow us to step up into the mt-aaRS disease associated variants core, where we can, hopefully, futurely characterize structural and functional changes. These studies can ultimately lead us to establish the correlation between genotype properties and phenotype manifestations. For future work, structural characterization studies on wild-type and disease-associated variants hEARS2 and hRARS2 will be continued.

5. Bibliography

1. David L. Nelson, M.M.C., *Lehninger Principles Of Biochemistry*. fifth ed. 2008.
2. Whitford, D., *Proteins: Structure and Function*. 2005.
3. Litwack, G., *Human biochemistry*. 2018.
4. Bonnefond, L., et al., *Toward the full set of human mitochondrial aminoacyl-tRNA synthetases: characterization of AspRS and TyrRS.* Biochemistry, 2005. **44**(12): p. 4805-16.
5. Harvey Lodish, A.B., Paul Matsudaira, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Lawrence Zipursky, James Darnell, *Molecular Cell Biology*. fifth ed. 2004.
6. Berg, J.M.T.J.L.S.L.S.L.N.C.f.B.I., *Biochemistry.* 2002.
7. *Translation* [cited 2021 12/2021]; Available from: https://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/73-translation/translation-hl.html.
8. Rao, F. and A. Caflisch, *The protein folding network.* J Mol Biol, 2004. **342**(1): p. 299-306.
9. Gomes, C. and P. Faísca, *Protein Folding: An Introduction*. 2019. 1-63.
10. Cooper, G.M., *The Cell: A Molecular Approach*. second ed. 2000.
11. Ellgaard, L., et al., *Co- and Post-Translational Protein Folding in the ER.* Traffic, 2016. **17**(6): p. 615-38.
12. Dagget, R.D.S.a.V., *Protein folds and protein folding.* Protein Engineering, Design and Selection, 2011: p. 11-19.
13. Hou, J., et al., *A global representation of the protein fold space.* Proceedings of the National Academy of Sciences, 2003. **100**(5): p. 2386-2390.
14. Saibil, H., *Chaperone machines for protein folding, unfolding and disaggegation.* Nature Reviews Mol Cell Biol, 2013: p. 630-642.
15. Ivet Bahar, R.L.J., Ken A Dill, *Protein Action - Principles and Modelling*. 2017.
16. Diodato, D., D. Ghezzi, and V. Tiranti, *The Mitochondrial Aminoacyl tRNA Synthetases: Genes and Syndromes.* Int J Cell Biol, 2014. **2014**: p. 787956.
17. Guo, M., X.L. Yang, and P. Schimmel, *New functions of aminoacyl-tRNA synthetases beyond translation.* Nat Rev Mol Cell Biol, 2010. **11**(9): p. 668-74.
18. Laporte, D., et al., *Exploring the evolutionary diversity and assembly modes of multi-aminoacyl-tRNA synthetase complexes: Lessons from unicellular organisms.* FEBS Letters, 2014. **588**(23): p. 4268-4278.
19. Vasu, K., et al., *The zinc-binding domain of mammalian prolyl-tRNA synthetase is indispensable for catalytic activity and organism viability.* iScience, 2021. **24**(3): p. 102215.
20. Gomez, M.A.R. and M. Ibba, *Aminoacyl-tRNA synthetases.* Rna, 2020. **26**(8): p. 910-936.
21. Chen, M., et al., *Cross-editing by a tRNA synthetase allows vertebrates to abundantly express mischargeable tRNA without causing mistranslation.* Nucleic Acids Res, 2020. **48**(12): p. 6445-6457.
22. Sissler, M., L.E. Gonzalez-Serrano, and E. Westhof, *Recent Advances in Mitochondrial Aminoacyl-tRNA Synthetases and Disease.* Trends Mol Med, 2017. **23**(8): p. 693-708.
23. Kim, H.S., et al., *The crystal structure of arginyl-tRNA synthetase from Homo sapiens.* FEBS Lett, 2014. **588**(14): p. 2328-34.
24. Kaiser, F., et al., *The structural basis of the genetic code: amino acid recognition by aminoacyl-tRNA synthetases.* Sci Rep, 2020. **10**(1): p. 12647.
25. Sissler M, P.J., Fasiolo F, et al., *Mitochondrial Aminoacyl-tRNA Synthetases. In: Madame Curie Bioscience Database [Internet].* . 2013: Landes Bioscience.
26. Yakobov, N., et al., *Cytosolic aminoacyl-tRNA synthetases: Unanticipated relocations for unexpected functions.* Biochim Biophys Acta Gene Regul Mech, 2018. **1861**(4): p. 387-400.
27. Mirande, M., *The Aminoacyl-tRNA Synthetase Complex.* Subcell Biochem, 2017. **83**: p. 505-522.
28. Nagao, A., et al., *Biogenesis of glutaminyl-mt tRNAGln in human mitochondria.* Proc Natl Acad Sci U S A, 2009. **106**(38): p. 16209-14.

29. Garin, S., et al., *Localization and RNA Binding of Mitochondrial Aminoacyl tRNA Synthetases.* Genes (Basel), 2020. **11**(10).
30. Ni, M., et al., *Metabolic impact of pathogenic variants in the mitochondrial glutamyl-tRNA synthetase EARS2.* J Inherit Metab Dis, 2021. **44**(4): p. 949-960.
31. *Uniprot*. at 10/2021 2021; Available from: https://www.uniprot.org/uniprot/Q5JPH6.
32. Sawada, D., et al., *Remitting and exacerbating white matter lesions in leukoencephalopathy with thalamus and brainstem involvement and high lactate.* Brain Dev, 2021. **43**(7): p. 798-803.
33. Konovalova, S. and H. Tyynismaa, *Mitochondrial aminoacyl-tRNA synthetases in human disease.* Mol Genet Metab, 2013. **108**(4): p. 206-11.
34. *Human RARS2 Uniprot*. at 10/2021]; Available from: https://www.uniprot.org/uniprot/Q5T160.
35. Igloi, G.L., *Molecular evidence for the evolution of the eukaryotic mitochondrial arginyl-tRNA synthetase from the prokaryotic suborder Cystobacterineae.* FEBS Lett, 2020. **594**(5): p. 951-957.
36. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
37. DiMauro, S. and M. Hirano, *Pathogenesis and Treatment of Mitochondrial Disorders*, in *Inherited Neuromuscular Diseases: Translation from Pathomechanisms to Therapies*, C. Espinós, V. Felipo, and F. Palau, Editors. 2009, Springer Netherlands: Dordrecht. p. 139-170.
38. Carroll, C.J., V. Brilhante, and A. Suomalainen, *Next-generation sequencing for mitochondrial disorders.* British Journal of Pharmacology, 2014. **171**(8): p. 1837-1853.
39. Vasta, V., et al., *Next generation sequence analysis for mitochondrial disorders.* Genome Medicine, 2009. **1**(10).
40. Khan, N.A., et al., *Mitochondrial disorders: challenges in diagnosis & treatment.* Indian J Med Res, 2015. **141**(1): p. 13-26.
41. Kanungo, S., et al., *Mitochondrial disorders.* Ann Transl Med, 2018. **6**(24): p. 475.
42. Zeviani, M. and S. Di Donato, *Mitochondrial disorders.* Brain, 2004. **127**(Pt 10): p. 2153-72.
43. Chinnery, P.F., *Mitochondrial Disorders Overview*. 1993: University of Washington, Seattle, Seattle (WA).
44. Fernandez-Vizarra, E. and M. Zeviani, *Mitochondrial disorders of the OXPHOS system.* FEBS Lett, 2021. **595**(8): p. 1062-1106.
45. Ribeiro, J.V., C.M. Gomes, and B.J. Henriques, *Functional Recovery of a GCDH Variant Associated to Severe Deflavinylation-Molecular Insights into Potential Beneficial Effects of Riboflavin Supplementation in Glutaric Aciduria-Type I Patients.* Int J Mol Sci, 2020. **21**(19).
46. Barbosa-Gouveia, S., et al., *Identification of a Novel Variant in EARS2 Associated with a Severe Clinical Phenotype Expands the Clinical Spectrum of LTBL.* Genes (Basel), 2020. **11**(9).
47. Moulinier, L., et al., *MiSynPat: An integrated knowledge base linking clinical, genetic, and structural data for disease-causing mutations in human mitochondrial aminoacyl-tRNA synthetases.* Human Mutation, 2017. **38**(10): p. 1316-1324.
48. Fine, A.S., et al., *Mitochondrial aminoacyl-tRNA synthetase disorders: an emerging group of developmental disorders of myelination.* J Neurodev Disord, 2019. **11**(1): p. 29.
49. Schwenzer, H., et al., *Pathogenic implications of human mitochondrial aminoacyl-tRNA synthetases.* Top Curr Chem, 2014. **344**: p. 247-92.
50. Webb, B.D., G.A. Diaz, and P. Prasun, *Mitochondrial translation defects and human disease.* J Transl Genet Genom, 2020. **4**: p. 71-80.
51. Lax, N.Z., et al., *Neuropathologic Characterization of Pontocerebellar Hypoplasia Type 6 Associated With Cardiomyopathy and Hydrops Fetalis and Severe Multisystem Respiratory Chain Deficiency due to Novel RARS2 Mutations.* Journal of Neuropathology & Experimental Neurology, 2015. **74**(7): p. 688-703.
52. Cheung, R.C., J.H. Wong, and T.B. Ng, *Immobilized metal ion affinity chromatography: a review on its applications.* Appl Microbiol Biotechnol, 2012. **96**(6): p. 1411-20.
53. Hagel, L. and L. Haneskog, *Size-exclusion Chromatography*, in *Encyclopedia of Life Sciences*. 2010.

54. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2005. **1751**(2): p. 119-139.
55. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy.* Proc Natl Acad Sci U S A, 2015. **112**(24): p. E3095-103.
56. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. third edition ed. 2006.
57. Cristovao, J.S., B.J. Henriques, and C.M. Gomes, *Biophysical and Spectroscopic Methods for Monitoring Protein Misfolding and Amyloid Aggregation.* Methods Mol Biol, 2019. **1873**: p. 3-18.
58. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server*. Nucleic Acids Research, 2014. **42**(W1): p. W320-W324.
59. Zheng, W.-Q., et al., *Elucidating the molecular mechanisms associated with TARS2-related mitochondrial disease.* Human Molecular Genetics, 2021.
60. González-Serrano, L.E., J.W. Chihade, and M. Sissler, *When a common biological role does not imply common disease outcomes: Disparate pathology linked to human mitochondrial aminoacyl-tRNA synthetases.* The Journal of biological chemistry, 2019. **294**(14): p. 5309-5320.
61. Jorgensen, R., et al., *Identification and characterization of human mitochondrial tryptophanyl-tRNA synthetase.* J Biol Chem, 2000. **275**(22): p. 16820-6.
62. Gonzalez-Serrano, L.E., et al., *Three human aminoacyl-tRNA synthetases have distinct sub-mitochondrial localizations that are unaffected by disease-associated mutations.* J Biol Chem, 2018. **293**(35): p. 13604-13615.
63. Levin, I., et al., *Purification, crystallization and preliminary X-ray characterization of a human mitochondrial phenylalanyl-tRNA synthetase.* Acta crystallographica. Section F, Structural biology and crystallization communications, 2007. **63**(Pt 9): p. 761-764.
64. Lue, S.W. and S.O. Kelley, *An aminoacyl-tRNA synthetase with a defunct editing site.* Biochemistry, 2005. **44**(8): p. 3010-6.
65. Wang, Y., et al., *A Human Disease-causing Point Mutation in Mitochondrial Threonyl-tRNA Synthetase Induces Both Structural and Functional Defects.* J Biol Chem, 2016. **291**(12): p. 6507-20.

6. Annex

Table 1 – hEARS2 purification yields

	Expression Conditions	Mass cells (g)	[EARS2] ($mg.mL^{-1}$)	Total EARS2 (mg)	Purity (%)	mg EARS2/g cells
hEARS2 purifications	4H 30ºC	1	0.18	0.27	80	0.26
	ON 22ºC	10	0.1	0.14	90	0.02
	4H 30ºC *	4	0.19	1.99	95	0.49

Note: First and third expression conditions (both 4 hours at 30 ºC incubation) differed in the buffer A used (for the first it was used the buffer with 150 mM NaCl and for the third it was used buffer with 300 mM NaCl and 20% glycerol) and in the fact that for the third expression condition it was performed a size exclusion desalting chromatography for buffer exchange, after affinity chromatography.

*Protein purification performed with optimized protocol.

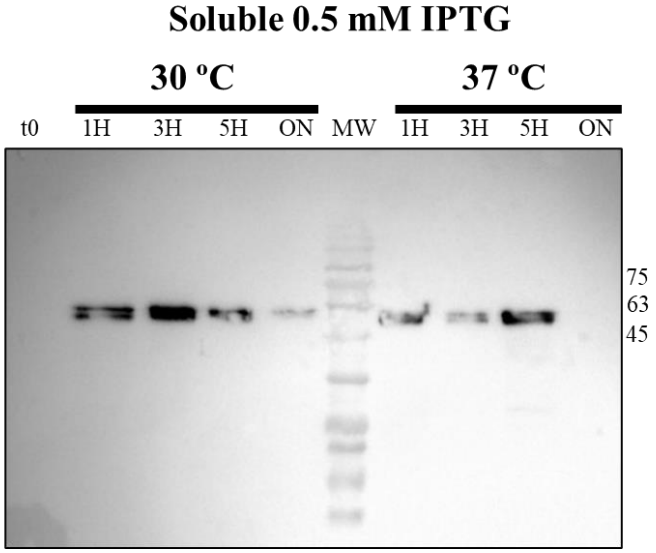


Figure 6.1 – Expression tests for hEARS2 with 0.5 mM IPTG.
Full size Western blot membrane hEARS2 expression at different timepoints for soluble fraction after induction with 0.5 mM IPTG at 30 ºC and 37 ºC. Protein detection was accomplished with the use of an anti-his tag antibody.

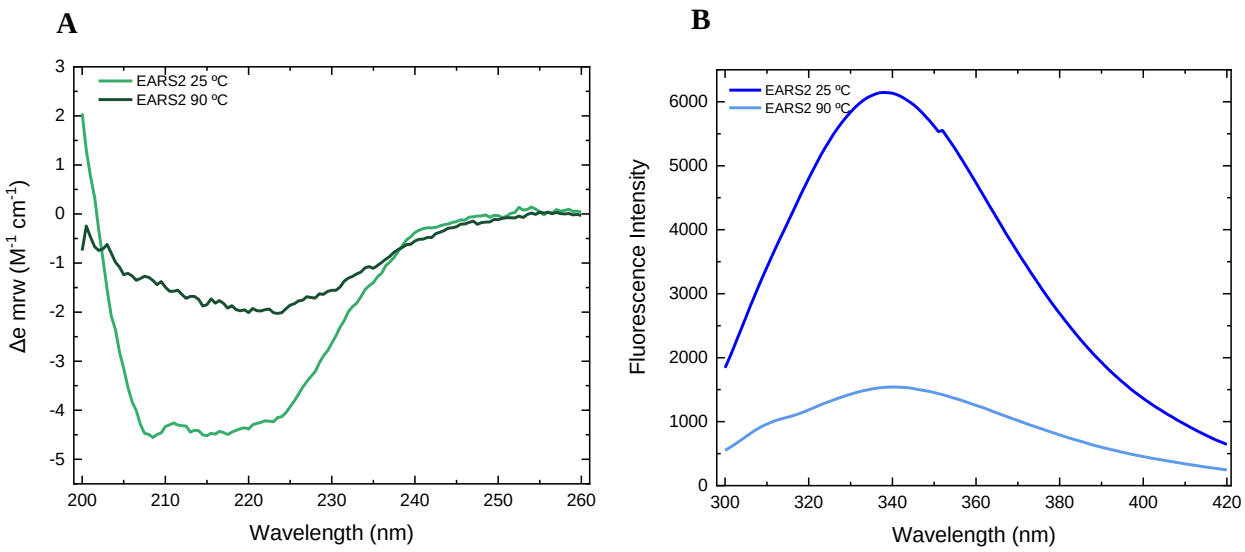


Figure 6.2 - hEARS2 variation in spectra before and after thermal denauration.
(A) – hEARS2 far-UV CD spectra at 25 ºC (solid light green line) and at 90 ºC (solid dark green line); **(B) –** hEARS2 tryptophan emission spectra at 25 ºC (solid blue line) and at 90 ºC (solid light blue line).

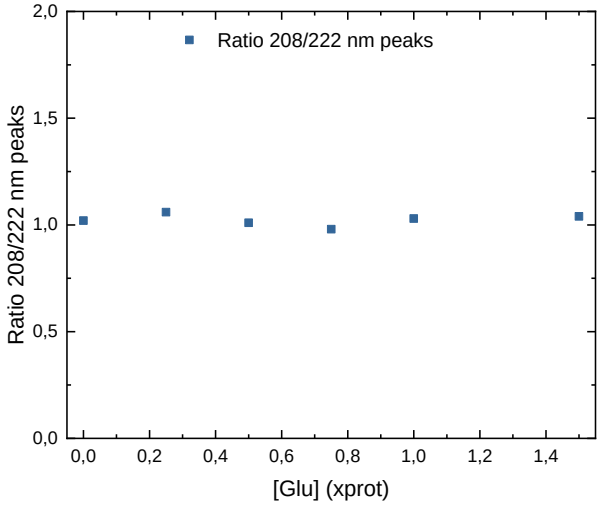


Figure 6.3 – hEARS2 titration with glutamate analysis.
Evolution of CD minimums ratio 208 nm and 222 nm with increase of glutamate concentration.

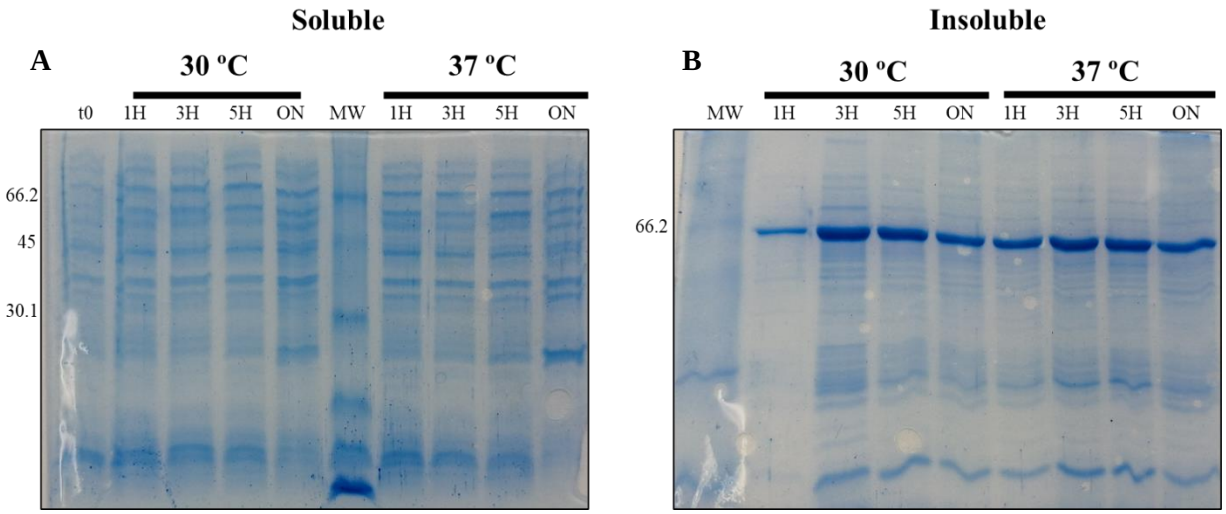


Figure 6.4 – hRARS2 expression tests after induction with 0.5 mM IPTG
(A) – Blue stained SDS-PAGE gel of insoluble fraction of hRARS2 expressions tests for 0.5 mM IPTG at 30 ºC and 37 ºC; **(B) –** Blue stained SDS-PAGE gel of soluble fraction of hRARS2 expressions tests for 0.5 mM IPTG at 30 ºC and 37 ºC.

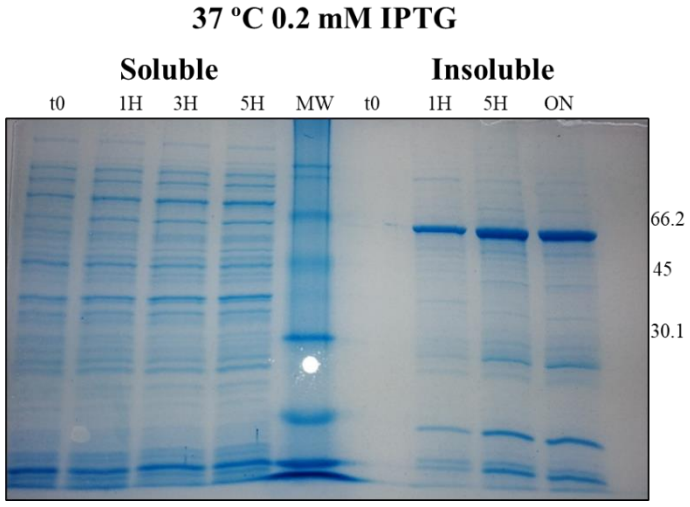


Figure 6.5 – hRARS2 expression tests
Blue stained gel for hRARS2 expression in soluble and insoluble fraction after induction with 0.2 mM IPTG a 37 ºC for different timepoints.

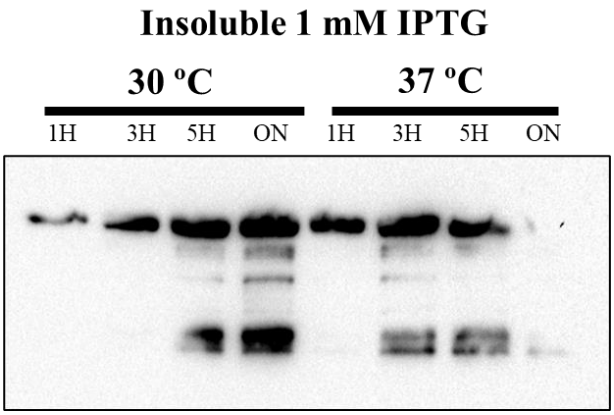


Figure 6.6 - hRARS2 expression tests
Western blot membrane for hRARS2 expression assessment at different timepoints in insoluble fraction at 1 mM IPTG. Protein was detected with anti his tag antibody

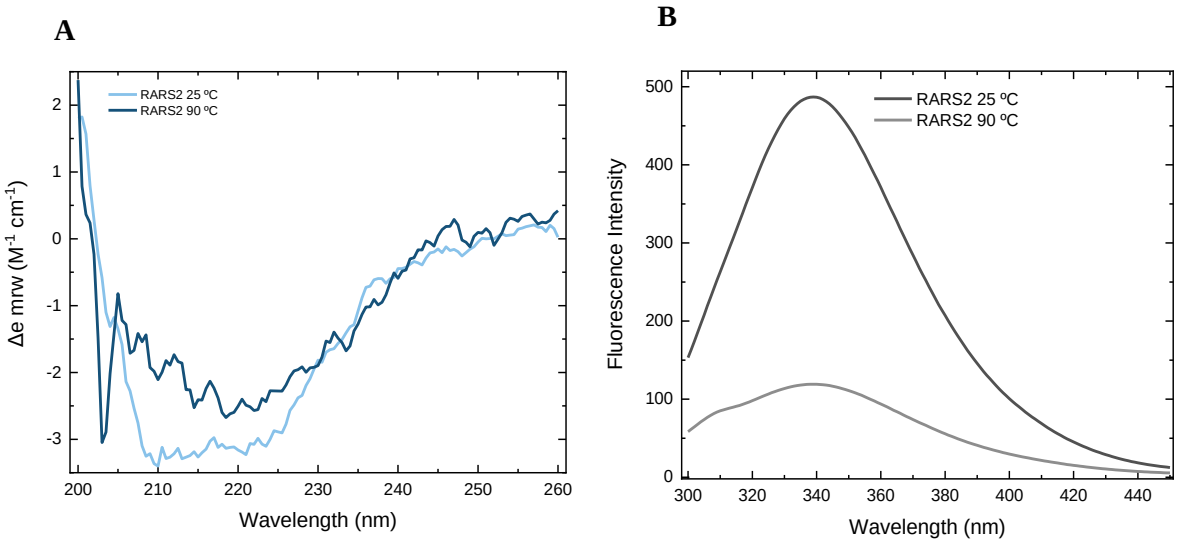


Figure 6.7 - hRARS2 variation in spectra before and after thermal denauration.
(A) – hRARS2 far-UV CD spectra at 25 ºC (solid light blue line) and at 90 ºC (solid dark blue line); **(B) –** hRARS2 tryptophan emission spectra at 25 ºC (solid dark grey line) and at 90 ºC (solid light grey line).