UNIVERSITAT ROVIRA i VIRGILI

DESIGN OF A MUTATIONAL LIBRARY TO MEASURE THE STABILITY OF ALL POSSIBLE MUTATIONS IN SOD1

Tomás Quiroga

FINAL DEGREE PROJECT BIOTECHNOLOGY

Academic tutor: Dr. Gerard Pujadas Anguiano, Biotechnology and biochemistry department, gerard.pujadas@gmail.com)

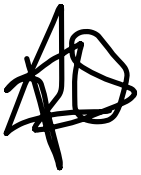
In cooperation with: Institute for Bioengineering of Catalonia (IBEC)

Supervisor: Dr. Benedetta Bolognesi, Protein Phase Transitions in Health and Disease (IBEC) Group leader, bbolognesi@ibecbarcelona.eu

Date of convocation: June, 2023

Jo, Tomás Quiroga , amb DNI AAA317524, sóc coneixedor de la guia de prevenció del plagi a la URV Prevenció, detecció i tractament del plagi en la docència: guia per a estudiants (aprovada el juliol 2017) (http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/) i afirmo que aquet TFG no constitueixen cap de les conductes considerades com a plagi per la URV.

Tarragona, 7 de Juny de 2023



Contents

CENTER DATA

The IBEC mission

The Insitute for Bioengineering of Catalonia is a research centre set up to conduct interdisciplinary research at the cutting edge of knowledge in the bioengineering field which, through the talent it attracts, the creativity associated with scientific progress and the translation thereof, helps to improve health and quality for people and generate prosperity.

The IBEC vision

To earn recognition as one of the world's leading research centres in bioengineering to improve health.

It is a research centre with 24 research groups focused on different fields of bioengineering, such as tissue engineering, protein engineering, bioinformatics and biophysics.

IBEC is in the Parc Cientific de Barcelona, together with other institutions such as the IRB (Institute for Research in Biomedicine), with which it collaborates in its projects.

The group I belonged to, Protein Phase Transitions in Health and Disease, focuses on the study of how mutations affect the stability and activity of proteins involved in neurodegenerative diseases. Specifically, parallel projects focus on Alzheimer's disease, Amyotrophic Lateral Sclerosis and Huntington's disease.

In recent years, they have demonstrated the effectiveness of their studies through the publication of 6 articles in journals such as *Nature Communications* and *eLIFE*.

The group maintains collaborations at national level (CRG, Center for Genomic Regulation, and IRB) and at international level (University of Wollongong, Australia)

ABSTRACT

Superoxide Dismutase 1 (SOD1) is a Reactive Oxygen Species (ROS) neutralising protein that can lead to familial and sporadic ALS when mutated. It is known that mutations in SOD1 can induce misfolding of the protein and affect its functionality, leading to cell death. In this work we aim to determine which mutations are responsible for protein misfolding using a Deep Mutational Scanning (DMS) approach, which involves creating a mutational library of SOD1, selecting it using a high-throughput protein abundance assay (PCA) and quantifying the mutational impact before and after selection, by deep sequencing. Fusing SOD1 to the dihydrofolate reductase (DHFR) enzyme and using it as a reporter of the folded and functional SOD1, we carried out an abundance assay using the SOD1 WT protein and the known destabilizing SOD1 mutation A4V to determine the discriminative power of the approach in quantifying the effect of mutations. We have designed and cloned SOD1 mutational libraries encompassing 6000 distinct SOD1 mutations (substitutions, insertions and deletions) to test their effect on protein folding. We intend to scale up our selections and use the results obtained as evidence to discriminate between pathogenic and non-pathogenic mutations.

Keywords: Superoxide Dismutase 1 (SOD1); mutational library of SOD1; Deep Mutational Scanning (DMS); abundance assay (PCA), SOD1 A4V.

HYPOTHESIS AND GOALS

Hypothesis

The Abundance PCA assay by using a SOD1 mutational library will allow us to know how mutations affect folding of protein, and discriminate between those mutations that can lead to disease and those that cannot.

Goals

- Determining at which end of DHFR (C-terminal or N-terminal) the SOD1 insertion allows for greater discrimination of variants.
- Using the abundance assay to assess the growth of WT SOD1 and the A4V mutant
- Design and clone plasmid constructs to carry out the abundance assay with the mutational libraries.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a terminal neurodegenerative disease caused by the loss of function of motor neurons. This dysfunctionality is manifested by a progressive loss of muscle capacity leading to loss of respiratory capacity and, therefore, the death. (Masrori & Van Damme, 2020; Mejzini et al., 2019). ALS can be familial or sporadic. Familial ALS (fALS) accounts for 10% of ALS patients and is directly related to family ancestors who have had the disease, while sporadic ALS (sALS) accounts for 90% of all cases and is not due to the presence of the disease in family ancestors. ALS usually appears in adulthood (45-75 years) and is a lethal disease, with a life expectancy of 3-5 years after diagnosis (Di Gregorio & Duennwald, 2018). However, the age of onset is closely related to the type of ALS the patient has, with the range of 58-63 years being common in sALS patients and 40-60 years in fALS patients. In addition, there is a huge variation also for people with the same mutation. It has an incidence of 1.75-3 per 100 000 people per year and a prevalence of 10-12 per 100 000 in Europe (Masrori & Van Damme, 2020). While there is still no treatment that eradicates the disease, all current therapies are based on alleviating symptoms as the disease progresses.

The causes of ALS are still unknown, but studies over the last 30 years have demonstrated the close association of certain genes with the disease, such as the TAR DNA-binding protein 43 (*TARDBP,* which encodes to the TDP-43), fused in sarcoma (*FUS*) and Cu, Zn Superoxide Dismutase 1 (*SOD1*), among 20 other genes involved (Di Gregorio & Duennwald, 2018; Kaur et al., 2016; Masrori & Van Damme, 2020; Mejzini et al., 2019; Peggion et al., 2022). Given the native function, dysfunctions in each protein affect the cell in a different way. For example, TDP-43 is a DNA-binding protein whose basal cellular location is in the nucleus. However, in ALS patients it has been observed in the cytosol, in the form of aggregates.

Protein aggregation behavior is a very common phenomena in neurodegenerative diseases, as occurs, for example, in Alzheimer's disease with the formation of amyloid-beta 42 plates. It is a process by which a peptide undergoes a conformational change that affects the arrangement of its amino acids, either by mutations or by external

factors. Such changes can expose amino acids that were previously inaccessible, thus enabling the protein to carry out new interactions with proteins of the same or even different types. These changes result in a loss of solubility due to a tendency to increase hydrophobicity. By becoming insoluble, the affected peptides can interact with peptides with which they did not previously interact, and form aggregates. In the case of ALS, widespread aggregation of several proteins has been observed in the motor neurons of patients. Although there is still some debate as to whether aggregation occurs for all proteins involved in ALS, such as SOD1 (Hayashi et al., 2016), and whether it has a direct toxic effect at the cellular level, several studies have shown a close relationship between SOD1 aggregation and the development of ALS (Kaur et al., 2016; Münch & Bertolotti, 2010; Sheng et al., 2013; Workman, 2020).

SOD1, located on human chromosome 21, is one of the most highly expressed proteins in mammalian tissues, particularly in liver and central nervous system cells (Peggion et al., 2022). It is one of the proteins most involved in ALS, being the second most responsible for causing fALS and sALS (Andersen, 2006). It is a 153 amino acid long peptide in homodimer form, comprised of five exons that are separated by four introns, whose subunits are linked by a disulphide bridge (PDB: 1PU0). A metal core is in each of these subunits, allowing them to bind to their activating ligands: zinc and copper ions. Under the governance of this binding, SOD1 carries out its enzymatic activity: the conversion of superoxide anions into molecular oxygen and H_2O_2. However, activating SOD1 is not the only function of these two ions; they have also been found to play an indispensable role in the formation and maintenance of its quaternary structure, as demonstrated in a recent study with a mouse model, which suggests that without copper metalation on SOD1, the protein does not fold properly and results in aggregates (Hilton et al., 2016).

While the causative link between SOD1 aggregation and ALS is not entirely clear, there are 150 reported mutations in SOD1 that causes ALS. Hence human genetics identifies a strong causality between SOD1 alterations and disease onset.

Although the effects of the mutations in SOD1 are still unknown, it has been observed that many of them lead to the onset of ALS (Hayashi et al., 2016; Kaur et al., 2016; Une et al., 2021). Mutations in SOD1 can cause both loss and gain of function, as well as changes in its structure. It has been determined that mutations affecting the stability of SOD1 can lead to oligomerisation and aggregate formation (Kaur et al., 2016). More

than 200 disease-related mutations have been described (https://alsod.ac.uk/), the most common being the D90A, A4V and G93A substitutions. In addition, the disease-causing mutations do not focus on a specific region of the gene, which presents a new challenge in establishing a mutation-effect relationship, as there is no apparent pattern. While not all dysfunctionalities in SOD1 are due to mutations, most of them are. Many SOD1 mutations involved in the onset and development of the disease have been identified, but how they affect the stability/activity of the protein remains a challenge. Some previous studies have demonstrated that there is indeed a relationship between mutations and the protein stability. For example, mutations in SOD1 can expose its N-terminal end and promote its binding to proteins in the endoplasmic reticulum, where it will cause stress at the ER level (Fujisawa et al., 2012). Another study (Münch & Bertolotti, 2010) showed that certain mutations increase the hydrophobicity of SOD1, making it less soluble and stimulating the formation of aggregates. Data such as these suggest that not only the genes in question are involved in the disease, but also mechanisms intrinsic to the cell itself in response to the stress provoked. There appears to be a direct relationship between these factors, with one being a consequence of the other. Against this background, the main unknown remains how and when mutations in SOD1 cause ALS, and whether they result in destabilization or aggregation.

Multiple biological models have been described for the study of SOD1 and its relationship with ALS, such as mice (Todd & Petrucelli, 2022) and yeast (Di Gregorio & Duennwald, 2018), which have been very relevant and effective models to study the role of proteins in the disease. Yeasts are eukaryotic organisms whose genome is completely sequenced. This knowledge allows us to control the cellular scenario with which we work, aiming to reduce unwanted cellular effects as much as possible and replicate the cellular context of the study in question.

In fact, many yeasts share cellular mechanisms with humans, especially cellular mechanisms of stress response or similar protein glycosylation patterns. On the other hand, yeasts also express highly conserved proteins between them and humans, such that many of them maintain partial or complete function between the two species. Although there are many more biological models such as bacteria, mice, or other eukaryotes, all these characteristics make yeast the perfect candidate for first-fundamental human studies. In the case of ALS, multiple yeast models exist to study

different parameters of the disease, and the choice of the right model is crucial for the success of the experiment. A recent study supports this idea by making a distinction between which cellular processes are most appropriate to study in yeast and which should be studied in another model (Di Gregorio & Duennwald, 2018). As demonstrated, a yeast model is ideal for studying mitochondrial functions, prion propagation and protein folding/misfolding, all of which are involved in SOD1-ALS. However, it is necessary to consider the limitations of using a single-cell model such as yeast, especially when studying neuronal communication, or the fact that SOD1 aggregation in yeast does not cause a toxic effect, as it seems to cause it in humans.

A recent novel technique has been making an impact on the study of protein mutations in recent years. Deep Mutational Scanning (DMS) (Fowler & Fields, 2014) allows simultaneous quantification of the effect of mutations in a certain protein from high-throughput DNA sequencing before and after the selection of a phenotype of interest. Thanks to this type of sequencing, it is possible to quantify how each single mutation affects a parameter of interest of a protein, such as the folding or the binding. DMS has been successfully carried out in a variety of research disciplines. In the field of virology, it has been used to determine how mutations in a SARS-CoV-2 receptor affect its binding to a cellular receptor (Starr et al., 2020). It has also provided insight into important regions of viral genomes and how mutations affect the flexibility and stability of their most highly expressed proteins (Burton & Eyre, 2021). In the field of immunology, it has been used to determine which mutations allow a therapeutic antibody to have a higher affinity for its target (Hanning et al., 2022). Recently, the technique has been used for the study of neurodegenerative diseases. An example of this is the study by (Seuma et al., 2021), in which a quantification of substitution, insertion, deletion and truncations mutations in beta amyloids was carried out to analyse the phenomenon of aggregation. Results such as those of this study demonstrate the potential use of DMS in determining the effect of mutations, and the possibility to use it for the study of mutations in SOD1 to answer one of the biggest questions about the protein's role in disease: how mutations affect its stability.

For this purpose, we propose to carry out an abundance assay (AbundancePCA) (Faure et al., 2022), by using DMS as a perturbation tool. The Abundance PCA analyses how a certain mutation affects protein folding. To quantify the results of the assay, the enzyme dihydrofolate reductase (DHFR) is used as a reporter. DHFR is

intended to be added to the expression cassette to perform a yeast culture in medium containing methotrexate (MTX). MTX is a folic acid antagonist that acts by inhibiting purine synthesis, causing cell death. Naturally, MTX also acts as an inhibitor of DHFR (Sramek et al., 2017), so it is proposed to use a modified DHFR that binds to MTX, but is not inactivated, so that it sequesters MTX from the medium and allows the cell to survive. In the assay, the enzyme will be encoded by two segments (DHFR3 and DHFR12) of the DHFR gene, so that each segment will code for a fragment of the enzyme. In the expression cassette, the DHFR3 fragment will be linked directly, via a linker, to SOD1 (in its corresponding mutated form), while the DHFR12 fragment will be located downstream of the expression cassette.

Only when the two DHFR fragments bind, a fully functional enzyme will be formed, capable of neutralizing the toxic effect of MTX, and the two DHFR fragments will bind when the mutation in question does not affect the packing (Abundance PCA) of SOD1 (**Figure 1**).

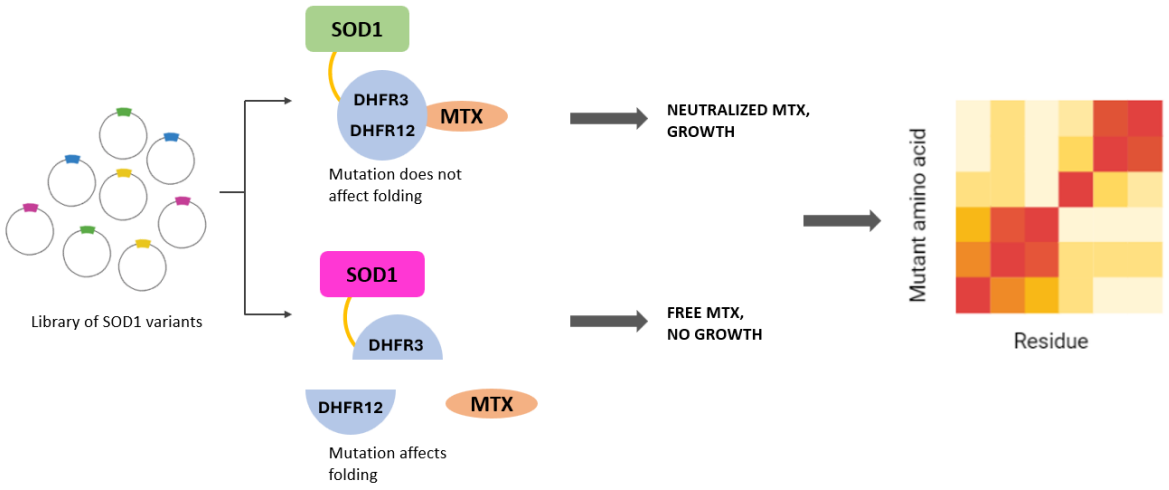


Figure 1. Plasmids encode for mutated SOD1 variants bounded to a DHFR3 fragment and to a free DHFR12 fragment. If the mutation does not affect to the SOD1 folding, DHFR3 and DHFR12 will bind and a functional DHFR will be formed, inhibiting MTX and allowing cell growth. If the mutation affects the SOD1 folding, DHFR3 will not be able to bind to DHFR12, and MTX will not be inhibited, triggering to cell death. After selection, the impact of each mutation is quantified in a heatmap.

While many genes, such as SOD1, have now been identified to cause ALS when mutated, there is still a substantial lack of knowledge regarding how mutations in these proteins can result in disease and via which mechanism, posing a great challenge when it comes to establishing the first steps towards treating or preventing disease. This is why the following project aims to carry out a study focused on SOD1, using *Saccharomyces cerevisiae* as a model. The aim of this work is to create a mutational library that allows us to model the 1007 possible mutations in SOD1 and to determine which of them affect its folding and which ones may instead result in toxicity by other means. As a result, we will create a comprehensive atlas reporting on the impact of each mutation in relation to the folding phenotype. By measuring the impact of all mutations, we will preventively interpret any of them as likely benign or pathogenic.

To carry out the abundance assay it is proposed to determine in which position with respect to DHFR (C-terminal or N-terminal) it is most effective to place the SOD1 sequence, and then design a SOD1 mutational library that include all the possible amino acid substitutions and the most known pathogenic insertions/deletions mutations. Here we present an example of the application of this method by using the SOD1 A4V mutation.

METHODS

ABUNDANCE ASSAY

First, an Abundance PCA assay was performed with SOD1 WT and the SOD1 A4V mutation to test the effectiveness of the assay on the protein in question. For this purpose, different plasmids were used with SOD1 WT and SOD1 A4V inserted. The expression cassettes of the plasmids differed according to the position at which SOD1 was inserted with respect to DHFR3 (C-terminal or N-terminal). The constructs of each plasmid are listed in **Table 1**.

Table 1*. Plasmids constructions to carry out the Abundance assay.*

Plasmid	**SOD1 position respect DHFR3**	**Linker position respect DHFR3**	**Resistance**
p045	N term	N linker	Ampicillin

p054	Empty (negative control)	-	Ampicillin
p160	C term	C linker	Ampicillin
p162	C term	N linker	Ampicillin

Primer's design

The primers to amplify the SOD1 inserts were designed in such a way that during amplification the sequences of the Nhel and Hindlll restriction targets were added to the ends of SOD1. For this purpose, the position at which SOD1 was to be inserted with respect to DHFR was considered.

Inserts and plasmid amplification and purification

SOD1 WT and SOD1 A4V inserts were amplified by PCR and purified by column. Plasmids were obtained from transformed Max Efficiency DH5 alfa bacteria and minipreps were prepared from each of them. In **Table 2** (**Appendix 1)** is shown which method to purify each sample was used**.** Purification protocols are included in **Appendix 2**

After purification, the samples were treated with 1 µL of the enzyme Dpnl to remove DNA strands methylated during replication in bacteria.

Enzymatic digestion

The concentration values were used to perform the calculations necessary to carry out the enzymatic digestion. The enzymatic digestion was performed with Nhel and Hindlll enzymes. An Excel template was used to calculate the required volumes of each sample, Fast Digest Buffer 10X, Nhel, Hindlll, Dep (dephosphatase) and water. The ratio 1 microgram (µg) x 1 microlitre (µL) was used to determine the required volume of inserts and plasmids.

Ligation

SOD1 inserts were ligated into the different plasmid constructs by a 1:5 (plasmid:insert)

ligation reaction. The reaction was catalysed by the T4 ligase enzyme.

Bacteria transformation

Max Efficiency DH5 alfa competent bacteria were transformed with each type of ligation and two controls were used (**Table 3**). Bacteria transformation protocol is included in **Appendix 2.**

Table 3*. Ligations used to transform bacteria.*

Plate	Plasmid	Construction
1	p045	SOD1 WT insertion in N-term
2	P045	SOD1 A4V insertion in C-term
3	p160	SOD1 WT insertion in C-term
4	p160	SOD1 A4V insertion in C-term
5	p162	SOD1 WT insertion in C-term
6	p162	SOD1 A4V insertion in C-term
Control +	p045	No SOD1, yes DHFR
Control -	p054	No SOD1, no DHFR

Plasmid purification: minipreps

For each construct, minipreps (Minipreps protocol) and Stocks were made with 800 µL of 50X Glycerol and 800 µL of cells.

Sanger sequencing

The ligation assembly reaction products were sequenced by Eurofins Genomics

(Sanger Sequencing) to check the correct assembly of the inserts with the plasmid. A total volume of 10 µL was used for sequencing: 5 µL primer at 5 µM + 5 µL plasmid. Said volume of plasmid was used as long as its concentration was within the range of 80-100 ng/ µL. For plasmids with a concentration higher than 100 ng/ µL, a dilution was carried out to achieve a concentration within the range (**Equation 1**, **Appendix 3**).

Water for molecular biology was used to achieve the required 5 µL.

For samples with a concentration of less than 80 ng/ µL, a primer volume of 2,5 µL was used at a concentration of 10 µM, which allowed pipetting a larger volume of plasmid for effective sequencing.

Yeast transformation

The yeast BY strain was used to incorporate the plasmid constructs obtained before. First, untransformed cells were grown in YPDA medium until an optical density (OD) between 0,8 and 1, corresponding to the exponential growth phase, was reached. This growth phase was used to carry out the transformation as this is when the cells reach the maximum level of competence. The linearity parameter between OD and growth was considered to adopt values within the range of 0,1-1 OD.

The absorbance of each yeast sample was measured with a spectrophotometer and, for those showing an OD value above the exponential phase range, a 1:10 dilution (900 µL of YPDA + 100 µL of cells) was made. For those where the OD value still exceeded the exponential phase range, a 1:100 dilution was made (900 µL YPDA + 100 µL cells 1:10).

Once the exponential phase was reached, the yeasts were transformed (**Appendix 2**) with the constructs on **Table 4**.

Table 4*. Plasmid constructs used to transform yeast.*

Plasmid	**Construction**
p045	SOD1 WT N-term
p045	SOD1 A4V N-term

p160	SOD1 WT C-term
p160	SOD1 A4V C-term
p162	SOD1 WT C-term
p162	SOD1 A4V C-term
p045	No SOD1, yes DHFR (control +)
p054	No SOD1, no DHFR (control -)

The table shows that not all possible plasmid-insert combinations were used. From this point on, only those constructs were used whose ligation was successful.

Yeast were grown on plate with LB medium and ampicillin and allowed to grow O/N. The following day, one colony of each construct was chopped and incubated in YPDA O/N liquid medium at 30°C and 200 rpm. The absorbance of each sample was measured and diluted to obtain an OD of 0,2, to re-establish the exponential phase to promote further growth. They were re-incubated at 30°C and 200 rpm and when they reached the end of the exponential phase, the TECAN growth test was started.

TECAN growth assay

The Tecan protocol is included in **Appendix 2**.

The absorbance of each sample was measured and a TECAN plate (96-well) was used to dilute the cells to an OD of 0,05 in 1mL of medium. The volume of cells required to achieve this concentration was calculated and pipetted into an Eppendorf. A summary of the construct cultures is shown in **Table 5**.

Table 5*. Plasmid and SOD1 constructs cultured in both +MTX and -MTX media.*

Construction	**Media**
SOD1 WT N-term (p045)	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
SOD1 A4V N-term (p045)	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
SOD1 WT C-term (p160)	-URA, -MET, -ADE

	-URA, -MET, -ADE, +MTX
SOD1 A4V C-term (p160)	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
SOD1 WT C-term (p162)	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
SOD1 A4V C-term (p162)	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
p045 no SOD1, yes DHFR	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX
p054 no SOD1, no DHFR	-URA, -MET, -ADE
	-URA, -MET, -ADE, +MTX

100 µL yeast of each construct was added to a 96-well TECAN plate, separating them according to whether they would grow on medium with or without MTX.

The TECAN was programmed to take OD readings every 10 minutes, for a period of 24 hours, at 30°C. The reading data were collected in an Excel file, from which the actual OD of each culture was calculated. The actual OD value was obtained by challenging the OD value of the medium to the OD value of the sample in question, according to the **Equation 2** (**Appendix 3**):

The values obtained were used to create a graph showing how the OD varied with respect to time (h). The graph was constructed with data up to a period of 22 hours to reflect the yeast growth curve according to the yeast constructs they carried, both in MTX and non-MTX medium.

Growth assay in bigger volume

Once the growth of yeast and Tecan had been verified, a higher volume growth trial was carried out. The aim of this assay was to test how higher volume culture conditions affected yeast growth with their constructs, and whether the same WT SOD1 - A4V SOD1 growth ratio was maintained as in Tecan. This process was necessary because, when testing growth with the actual mutational libraries, a large volume is required to

obtain results for all variants.

Yeast were transformed with the selected constructs and incubated on a plate. The next day, one colony was picked from each plate and incubated in 15 mL of sterile YPDA medium for approximately 8 hours. The OD of yeast SOD1 WT and SOD1 A4V was measured, and the exponential phase was confirmed. However, as growth was to be studied in a larger volume and only 15 mL of cells were available, it was decided to let the yeast grow for an additional hour, until a higher saturation of cells in stationary phase was achieved.

It was intended to study growth in a volume of 400 mL but given the OD reading results and the volume of cells available, the final volume was reduced to 200 mL.

From the stationary phase OD, the yeast volumes needed to reach an OD of 0,04 were calculated according to the **Equation 3** (**Appendix 3**).

We chose to dilute to 0,04 OD because in previous studies this was the concentration that gave the best results in the study.

200 mL of medium +MTX was added to 2L flasks, and the calculated volumes of yeast were transferred to 50 mL falcons. To each sample, 500 µL of YPDA was added and centrifuged for 5 min at 400 g. The supernatant was discarded and the pellet was resuspended in 300 µL of +MTX medium. The resuspended pellet was transferred to a 2L flask and incubated overnight. As a blank, yeast in YPDA medium was used. A graph was created to represent the growth curve of both WT SOD1 and A4V SOD1.

LIBRARY DESIGN

To design the SOD1 mutational libraries, its sequence was divided into three fragments of 50 amino acids each (**Figure 2**).

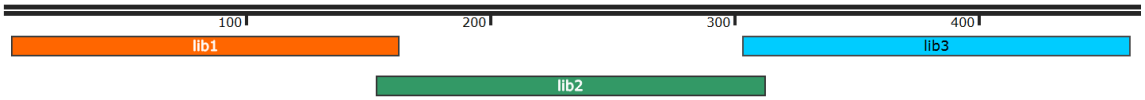


Figure 2*. SOD1 sequence divided into 3 fragments, which will be used to build the mutational libraries.*

The fragments were split with the aim of maintaining overlapping between them, in

order to carry out the mutations in each of them.

R script

To design the libraries, a script was designed in R programming language. The packages required to run the functions used are included in **Appendix 4.**

Consideration was given to the presence of overlaps between fragments, so that mutations were not repeated in these regions for each library.

The script was written according the six types of mutations that were performed:

- Single mutations

  All possible substitutions were created at each position of each fragment (53 amino acids), using the 19 possible changes (excluding substitutions that gave rise to the native amino acid and those that gave rise to stops).

  The parameters for carrying out the mutations were:

  - Mutation caused by a change of more than 2 nucleotides.
  - If the mutation could not be carried out for a change of more than 2 nucleotides, add a second synonymous mutation at another position.
  - Where possible, added the mutation following the Codon Yeast Usage. Since this study was carried out in yeast, one parameter to create the mutations was to use the most frequent codons in *S. cerevisiae* to encode each amino acid. To do this, a function was created to replace the native codons of SOD1 with the codons with the highest frequency of expression in yeast.

- Synonymous mutations

  Mutations that resulted in the same amino acid, following the same requirement that the mutation was due to a change of 2 or more nucleotides. Mutations to WT amino acid.

- Stops

  Some stop mutations were included, especially in the overlapping areas between fragments. A stop codon was added every 4 amino acids.

- Insertions

  All possible insertions were added to between each position. One requirement

was to insert amino acids after position 1 and before position 53 in each fragment. Insertions were added at random positions in the sequences, and those that resulted in amino acid duplications and stops were avoided.

- Deletions

  All possible deletions were carried out, avoiding those resulting in amino acid duplications.

LIBRARY CLONING

The 2 kb nicking plasmid pgjj191 was used as a vector for the cloning of the libraries. Each of these plasmids was designed to contain two fragments of SOD1 WT and to subsequently insert its corresponding mutational library (**Table 6**).

Table 6*. Cloning constructions to insert the SOD1 libraries.*

Construction	**SOD1 WT fragments**	**Designed to incorporate**
Empty vector 1	2-3	Library 1
Empty vector 2	1-3	Library 2
Empty vector 3	1-2	Library 3

Primer's design

Since the final step in cloning the libraries was to transfer the SOD1 expression cassette from the nicking plasmid to the abundance plasmid, the primers were designed to add the restriction target sequences of the Nhel and Hindlll enzymes at

the end fragments (1 and 3), for subsequent enzymatic digesting (**Figure 3**).

SOD1 fragments	Primer	Primer sequence	Annealing temperature (º C)
2-3	SOD1_2 fwd emp	catttccccgaaaagtgccACAGCAGGCTGTACCAGTGCAGGTCC	72
	HindIII SOD1_3 rev emp	ctcacatgttctttcctgcgAAGCTTTTGGGCGATCCCAATTACACCACAAG	
1	NheI-SOD1_2 fwd emp	catttccccgaaaagtgccGCTAGCATGGCGACGAAGGCCGTGTG	72
	SOD1_1 rev emp	GAGTGAGATAAACTCATGAACATGGAATCCATGCAGGCCTTC	
3	SOD1_3 fwd emp	CATGAGTTTATCTCACTCTCAGGAGACCATTGCATCATTG	72
	HindIII-SOD1_3 rev emp	ctcacatgttctttcctgcgAAGCTT TTGGGCGATCCCAATTACACCAC	
1-2	NheI-SOD1_1 fwd emp	catttccccgaaaagtgccGCTAGCATGGCGACGAAGGCCGTGTG	72
	SOD1_2 rev emo	ctcacatgttctttcctgcgTTCAATAGACACATCGGCCACACCATCTTTGTC	

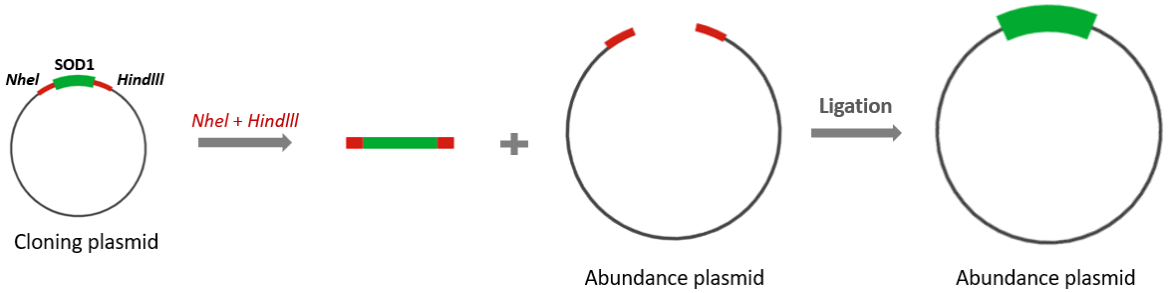


***Figure 3**. Transference of SOD1 expression cassette from the cloning plasmid to the*

abundance plasmid.

Primers were designed with four types of elements:

- Sequence homologous to the fragment under test
- NheI and HindIII restriction enzyme target sequence
- Sequence homologous to the cloning plasmid
- Sequence homologous to the binding fragment (applicable to fragments 1 and 3, which were amplified separately).

Therefore, each fragment combination was amplified with different primer designs, as shown in **Table 6**.

***Table 2**. Primer's design to amplify SOD1 fragments and the constant regions. Primers for amplification of SOD1 fragments. Red: sequence homologous to the plasmid cloning; Blue: restriction enzyme target sequence; Green: sequences homologous to fragments 1 and 3, respectively; **Black**: sequence homologous to the fragment in question. Fwd: forward; rev: reverse; emp: empty vector.*

Primers to carry out the linearisation are listed in **Table 7**. They were designed with homology to the regions adjacent and underlying the landing pad region (**Figure 4**).

***Table 3**. Primer's design to linearise the nicking vector.*

Primer	**Sequence**	**Annealing temperature (º C)**
Pgjj055 lin fwd	cgcaggaaagaacatgtgagcaaaagg	72
Pgjj191 lin rev	ggcacttttcggggaaatgtggaag	

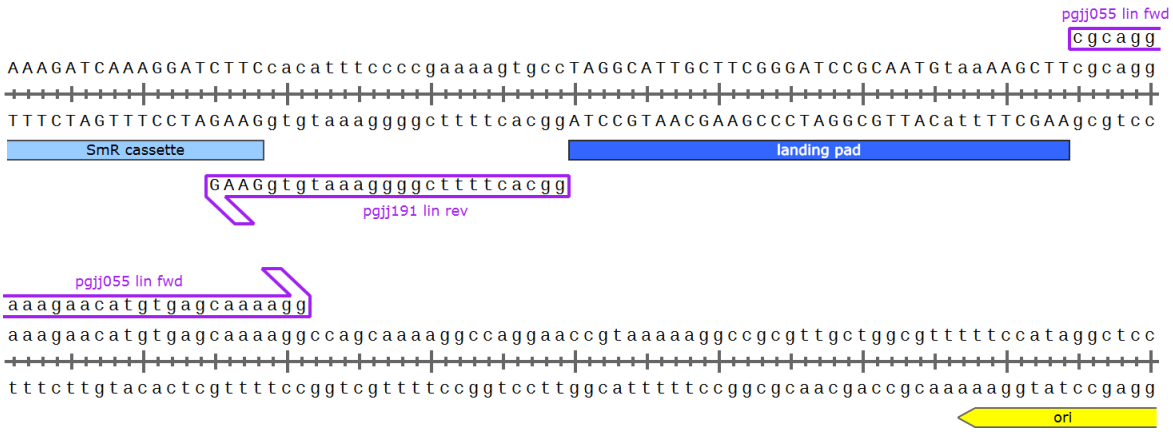


***Figure 4**. Linearisation primers joined to their sequences, to linearise the plasmid on the landing pad region.*

Plasmid linearization

The nicking vector was linearised in a non-coding region (landing pad) of 35 bp (**Figure 5**)

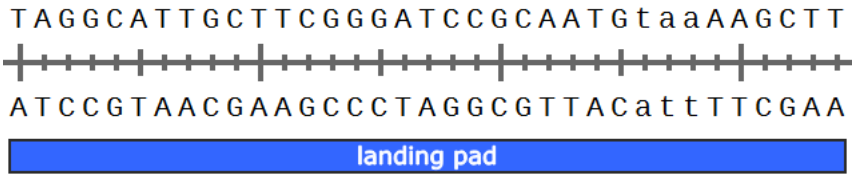


***Figure 5**. Nicking vector sequence used to linearise.*

SOD1 fragments amplification

The WT SOD1 fragments were amplified according to how they were to be inserted into their corresponding Empty Vector plasmid.

PCR protocol

Both plasmid linearisation and fragment amplification were carried out in a thermal cycler under the action of the Q5® High Fidelity DNA Polymerase enzyme. The reaction components and their respective volumes are listed in **Table 8**, **Appendix 1** The total volume of PCR reaction was 50 µL.

The programmes used for amplification and linearisation are shown **Table 8 and Table**

9 respectively**, Appendix 1**), according to the protocol established by NEW ENGLAND BioLabs®Inc.

Both correct vector linearisation and amplification of fragments were checked on a 1% agarose gel (0,5g agarose + 50 mL TAE 1X). 3 µL of DNA Ladder and 3 µL of Syber Safe were used to stain the gel. 1 µL of TriTrack DNA 6X was used to load 5 µL of each sample.

Samples were treated with Dpnl.

SOD1 fragments and plasmid purification

Linearised plasmids were purified from agarose gel band (QIAEX ll® Gel Extraction Kit). Each sample band was cut out and weighed. Considering each milligram of weight as a microlitre volume, 3 volumes of Buffer QX1 were added to 1 volume of gel and incubated at 50 °C for 10 minutes. Vortex mixed and one volume of isopropanol was added per volume of gel and vortex mixed. Samples were transferred to Mini Elute Column and centrifuged for 1 minute at 13000 rpm. The supernatant was discarded and washed with 750 µL PE Buffer and centrifuged for 1 minute at 13000 rpm. Washing was performed twice. The supernatant was discarded, and the empty column was centrifuged for drying (1 minute, 13000 rpm). Eluted with 10 µL of molecular H2O incubated at 50 ºC. Centrifuged for 1 minute at 11000 rpm.

The DNA concentration of each sample was quantified in Nanodrop.

Samples were checked on a 1% agarose gel.

Gibson assembly

The concentration values of purified plasmids and inserts were used to calculate the volumes required for a Gibson 1-10 reaction. The volumes were calculated as shown in **Gibson protocol, Appendix 3**

The corresponding reactions were prepared according to the calculated volumes and carried out at 50 °C for 3 hours.

Bacteria were then transformed, minipreps were made and the results sequenced (**Table 11, Appendix 1** shows how samples were prepared for sequencing). When analysing the results, the assembly for the Empty Vectors was considered to be effective when they met the following characteristics:

- Submit the complete inserts
- Retain the spectinomycin cassette
- No linearisation region (landing pad)
- Have the constant regions (restriction targets and homologous regions to the plasmid cloning)

Fake Libraries test

To check that the mutational libraries were going to be correctly established with their empty vectors, a test Fake Libraries was carried out. Fake Libraries are the fragments that are going to be mutated but in their wild-type form. The aim of this test was, on the one hand, to check that the assembly took place correctly and, on the other hand, to check that the amplification could be carried out with fewer cycles (10). This last requirement was necessary since the amplification of the real libraries is a process that requires as little contamination as possible.

Table 12 Appendix 2 shows the templates and the primers used to amplify the fake libraries:

After amplification, column purification was performed and the concentration of each fake library was quantified. Gibson assembly was performed with the Empty Vectors and transformed into competent bacteria. Correct assembly was checked by Sanger sequencing.

Once the effectiveness of the assembly of the SOD1 fake libraries had been checked, the Empty Vectors were linearised so that the real libraries could be inserted. The primers used to carry out the linearisation are listed in **Table 13**:

Table 13*. Primers used to linearise Empty Vectors.*

Vector	**Primer**	**Primer sequence**	**Annealing temperature (ºC)**

Empty Vector 1	Lin empty1 fwd	ACAGCAGGCTGTACCAGTGCAGG	71
	Lin empty1 rev	CAT -GCTAGC -ggcacttttcggggaaatgtgGA	
Empty Vector 2	Lin empty2 fwd	ATCTCACTCTCAGGAGACCATTGCAT	70
	Lin empty2 rev	AAACTCATGAACATGGAATCCATGCAGG	
Empty Vector 3	Lin empty3 fwd	AAGCTT - cgcaggaaagaacatgtgagcaaaagg	72
	Lin empty3 rev	TTCAATAGACACATCGGCCACACCATC	

Linearisation was checked on 1% agarose gel.

Agarose gel band purification was carried out (QIAX II protocol).

RESULTS AND DISCUSSION

LIBRARY DESIGN

From the designed script, mutations were created with the pre-established parameters in each of the libraries. The type of mutations carried out in each SOD1 library and their percentages are shown below. A total of 2000 mutations were designed and syn in each library.

SO1 Library 1

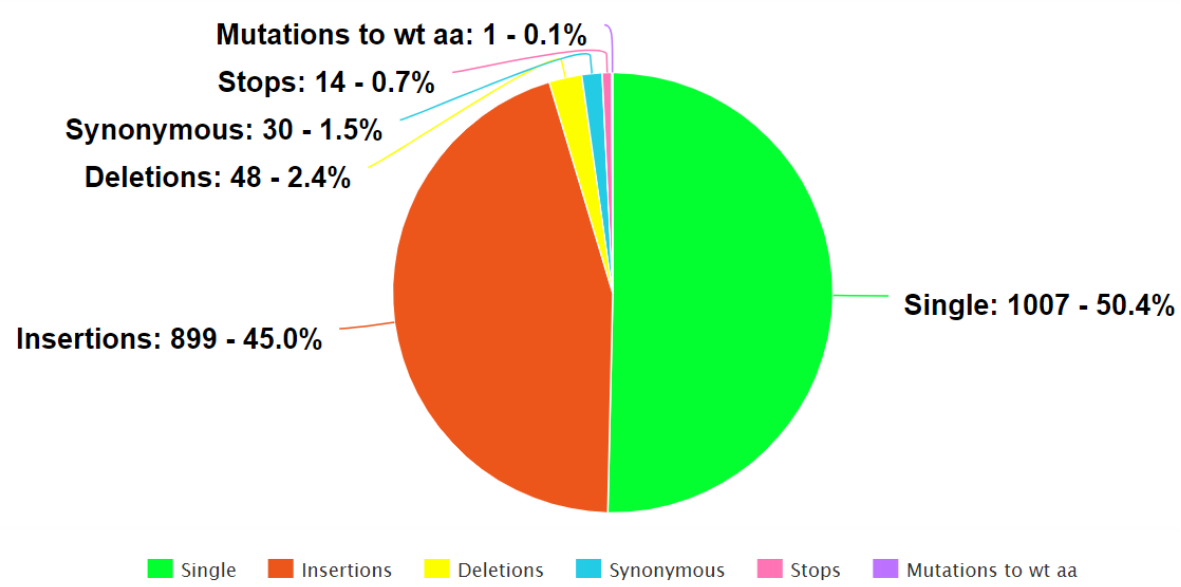


***Figure 6**. Percentage of types of mutations in SOD1 library 1.*

According to the exclusion parameters, the number of insertions in library 1 was 894 out of a possible 1009 (88%), as 115 of them resulted in duplications.

SOD1 Library 2

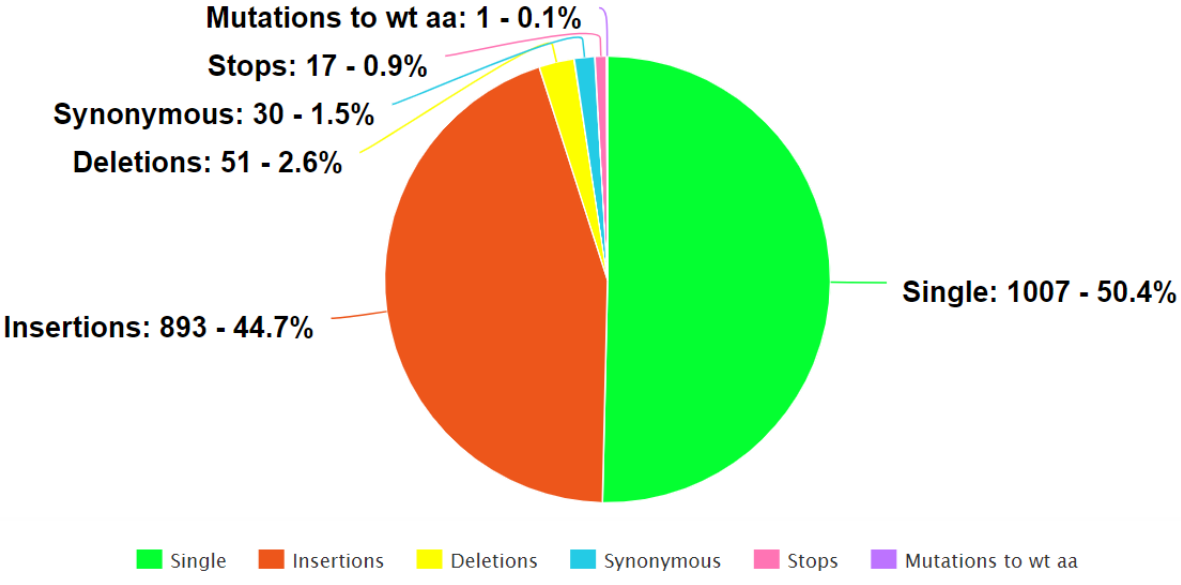


***Figure 7**. Percentage of types of mutations in SOD1 library 2.*

Applying the same principle as in library 1, library 2 insertions were 893 out of a

possible 1009 (88.6%).

SOD1 Library 3

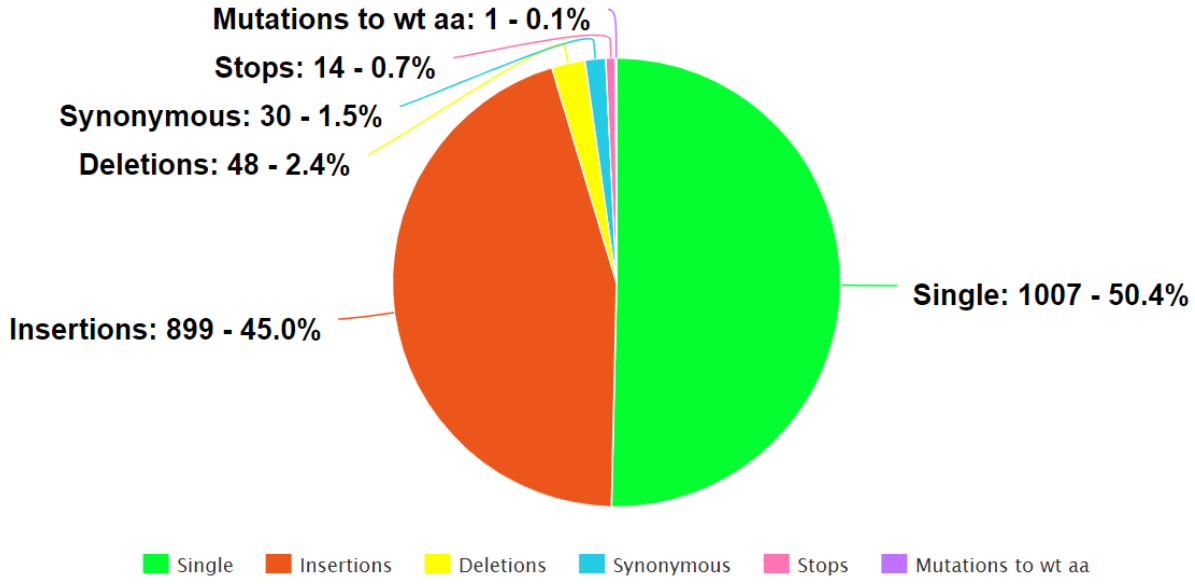


***Figure 8**. Percentage of types of mutations in SOD1 library3.*

In library 3, 899 insertions were added out of possible 1009 (89.1%).

LIBRARY CLONING

SOD1 fragments amplification and nicking plasmid linearization

A total volume of 150 µL of PCR product was purified from each insert and 90 µL of PCR product from plasmids. DNA quantifications are included in **Table 14, Appendix 1.** Here are shown the gel results for linearised plasmid and amplified SOD1 fragments.

In **Figure 9**, the first two bands (from the left) appear at the 2000 base pair height. The nicking vector has a total of 2037 bp and its landing region pad 35 bp, so when linearised, the vector size is expected to be approximately 2000 bp. The following bands correspond to SOD1 fragments 2-3, 1, 3 and 1-2, respectively.

Each fragment of S0D1 is 150 bp long, so fragments 2-3 and 1-2 will be 300 bp long (150 bp + 150 bp) and fragments 1 and 3 will be 150 bp long. As can be seen, this is as expected for each case. Although the band corresponding to fragment 1 of SOD1 has a weaker intensity than the rest of the bands, the amount of DNA could be used to correctly purify the fragment and obtain a sufficient concentration to carry out the Gibson assembly.

It is worth noting the peculiarity of the amplification of fragment 1, which in all results showed two signals, with the unwanted band being the most intense (below). One possible explanation could be that the primers used for amplification had a high homology between them, but this was not the case and, if it were the case, a less intense band would be expected. The other explanation is that the primers had homology for another region in the template, causing the amplophification of, in this case, two sequences.

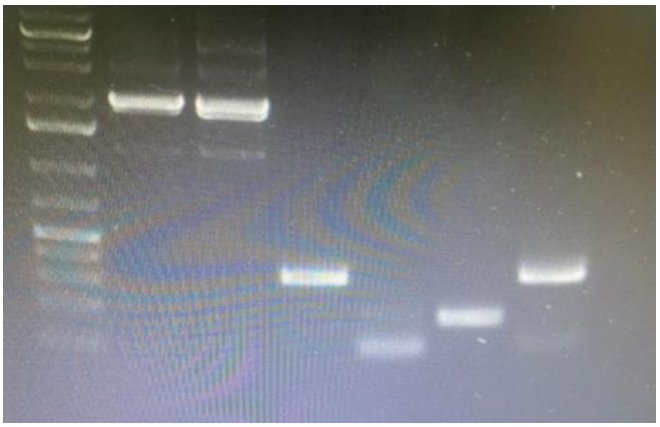


***Figure 9**. Gel result of linearised nicking vector and amplified SOD1 fragments.*

Gibson assembly reaction

Following the methodology explained in the corresponding section, the vector and inserts volumes required for the construction of the Empty Vectors were calculated (**Tables 15, 16 and 17, Appendix 1**).

Minipreps results from Gibson after bacteria transformation are included in **Table 18, Appendix 1**.

The number of colonies counted for each Gibson construction is shown in **Table 19:**

***Table 19**. Number of colonies of transformed bacteria with empty vectors.*

Empty Vector	**Colonies**
1	132
2	89
3	111

Sanger sequencing

The sequencing results of the assemblies of each Empty Vector are shown below. As discussed in the methodology section, suitable constructs had to carry the complete fragments, lack the landing pad region, carry the spectinomycin expression cassette and the constant regions added during amplification.

<u>Empty Vector 1</u>

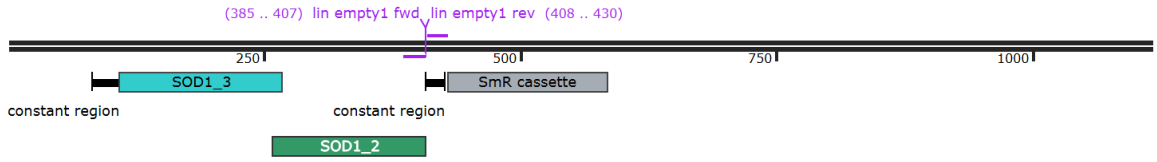


***Figure 10**. A view of how a correct Sanger sequencing result for Empty Vector 1 must be.*

It can be seen how the primers designed for linearization (lin empty1 fwd and lin empty1 rev) perfectly recognise their target sequences in the plasmid. In this case the vector is linearised at the beginning of SOD1 fragment 2, which will allow the insertion of the mutational library 1.

Empty Vector 2

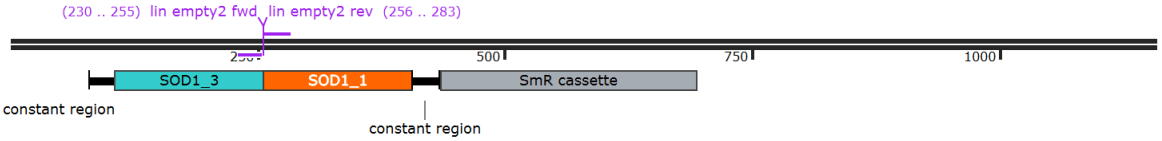


***Figure 11**. A view of how a correct Sanger sequencing result for Empty Vector 2 must be.*

In this case the linearisation primers (lin empty2 fwd and lin empty2 rev) bind between fragments 1 and 3, allowing the plasmid to be opened and the insertion of the mutational library 2.

Empty Vector 3

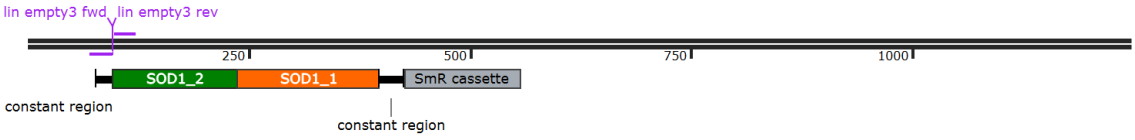


***Figure 11**. A view of how a correct Sanger sequencing result for Empty Vector 3 must be.*

Finally, the primers to linearise Empty Vector 3 (lin empty3 fwd and lin empty3 rev)

bind to the end region of fragment 2 of SOD1, which will allow the insertion of the mutational library 3.

The primers bind to one of the constant regions in each case, to carry out linearisation. The SOD1 fragments 1 and 3 constant regions correspond to the targets of the NheI and HindIII restriction enzymes. This allows enzymatic digestion of the SOD1 expression cassette and transfer to the abundance plasmid.

Fake libraries test

Concentrations of each fake library are included in Table 20. The correct amplification verification of the fake libraries is shown in **Figure 13**. Since each SOD1 fragment is 150 base pairs long, a signal was expected between bands 75 and 200 (last two bands of the Ladder). The fact that the bands are located close to 200 base pairs is due to the addition of the constant regions that have been incorporated during amplification.

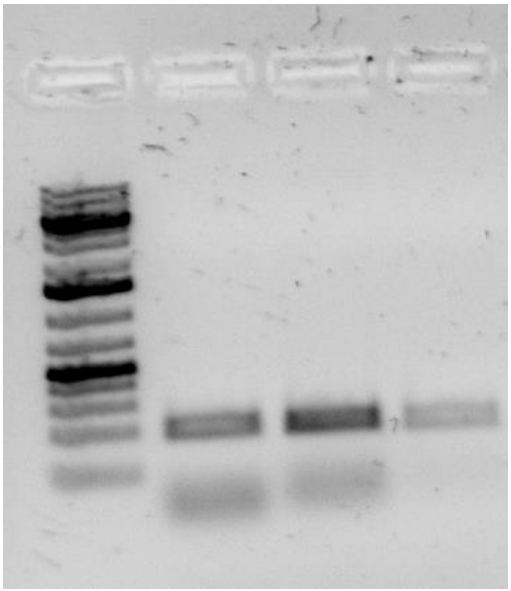


***Figure 13**. Gel of amplified SOD1 fake libraries.*

***Table 20**. Fake libraries DNA concentrations.*

Fake library	Concentration (ng/µL)
1	40
2	47

3	29

The counted colonies after transformation were: (**Table 21**):

***Table 21**. Number of colonies of transformed bacteria with fake libraries.*

Fake library	**Colonies**
1	150
2	22
3	230

ABUNDANCE ASSAY

Plasmid and SOD1 inserts concentrations

PCR-amplified SOD1 inserts and plasmid minipreps quantifications are shown in **Table 22**:

***Table 22**. SOD1 inserts and plasmid DNA quantification for abundance assay.*

Sample	**Concentration (ng/µL)**
SOD1 N-term	214,3
A4V N-term	234,8
SOD1 C-term	222,6
A4V C-term	219,3
p045	689,4
p054	292,9
p160	477,5
p162	556

Enzymatic digestion

The methodology for calculating the volumes of each reaction component has been

explained in the corresponding Methods section. The results are shown in **Table 23, Appendix 1.**

Note 1: Plasmid 164 with the N-term C-linker construct was ultimately not used in the abundance assay.

Purification

The **Table 24** shows the concentrations of plasmids and inserts after gel purification and PCR, respectively.

Table 24*. DNA quantification of digested plasmid and SOD1inserts.*

Sample	**Purificated from**	**Concentration (ng/µL)**
p045	MinElute Gel	28
p160	MinElute Gel	50
p162	MinElute Gel	49
SOD1 N-term	MinElute PCR	122
A4V N-term	MinElute PCR	211
SOD1 C-term	MinElute PCR	131
A4V N-term	MinElute PCR	181

Plasmids-inserts ligation minipreps

After cloning the plasmid-insert ligations from the transformed bacteria, the miniprep concentrations of each of the constructs were (**Table 25**):

Table 25*. DNA quantification of ligated plasmid and SOD1 inserts.*

Construction miniprep	**Concentration (ng/µL)**
p045 (WT SOD1 N-term)	106,4
p045 (A4V SOD1 N-term)	104,1
p160 (WT SOD1 C-term)	111,9
p160 (A4V SOD1 C-term)	131,5
p162 (WT SOD1 C-term)	52,9

p162 (A4V SOD1 C-term)	102,6

First yeast culture: restart exponential phase (OD = 0,2)

Yeasts were transformed with the constructs that were successful in the previous assay. The **Table 26** lists the OD readings (1:10 dilution) and the volumes of cells and medium required to achieve an OD = O.2 in a 3 mL volume:

Table 26*. OD readings at the end of the exponential phase.*

Yeast construction	**OD (600) 1:10 at the end of exp. phase**	**V_{cells} (µL)**	**V_{media} (µL)**
WT SOD1 N-term (p045)	0,883	34	2966
A4V SOD1 N-term (p045)	1,002	62,8	2932
WT SOD1 C-term (p160)	0,954	59,88	2940
A4V SOD1 C-term (p160)	0,926	64,79	2935
WT SOD1 C-term (p162)	0,857	70	2930
A4V SOD1 C-term (p162)	1,005	59,70	2941
Empty p045 (control+)	1,008	59,52	2941
Empty p054 (control-)	0,817	73,44	2927

Second yeast culture: from OD = 0,8 to OD = 0,05

The previously cultured yeasts were allowed to grow to the end of the exponential phase, OD = 0.8, and were diluted again, this time to carry out the TECAN growth

assay, OD = 0.05 in a volume of 1 mL (**Table 27**).

Table 27*. OD readings at the exponential phase.*

Yeast construction	**OD (600) in exp phase**	**V_{cells} (µL) to OD = 0,05**
WT SOD1 N-term (p045)	1,016	49,21
A4V SOD1 N-term (p045)	0,941	53,13
WT SOD1 C-term (p160)	0,960	52,1
A4V SOD1 C-term (p160)	0,942	53,1
WT SOD1 C-term (p162)	1,033	48,4
A4V SOD1 C-term (p162)	0,884	56,56
Empty p045 (control+)	0,882	56,7
Empty p054 (control-)	1,033	48,4

TECAN Growth assay

The growth test in TECAN was carried out over a period of 22 hours (**Figure 14**).

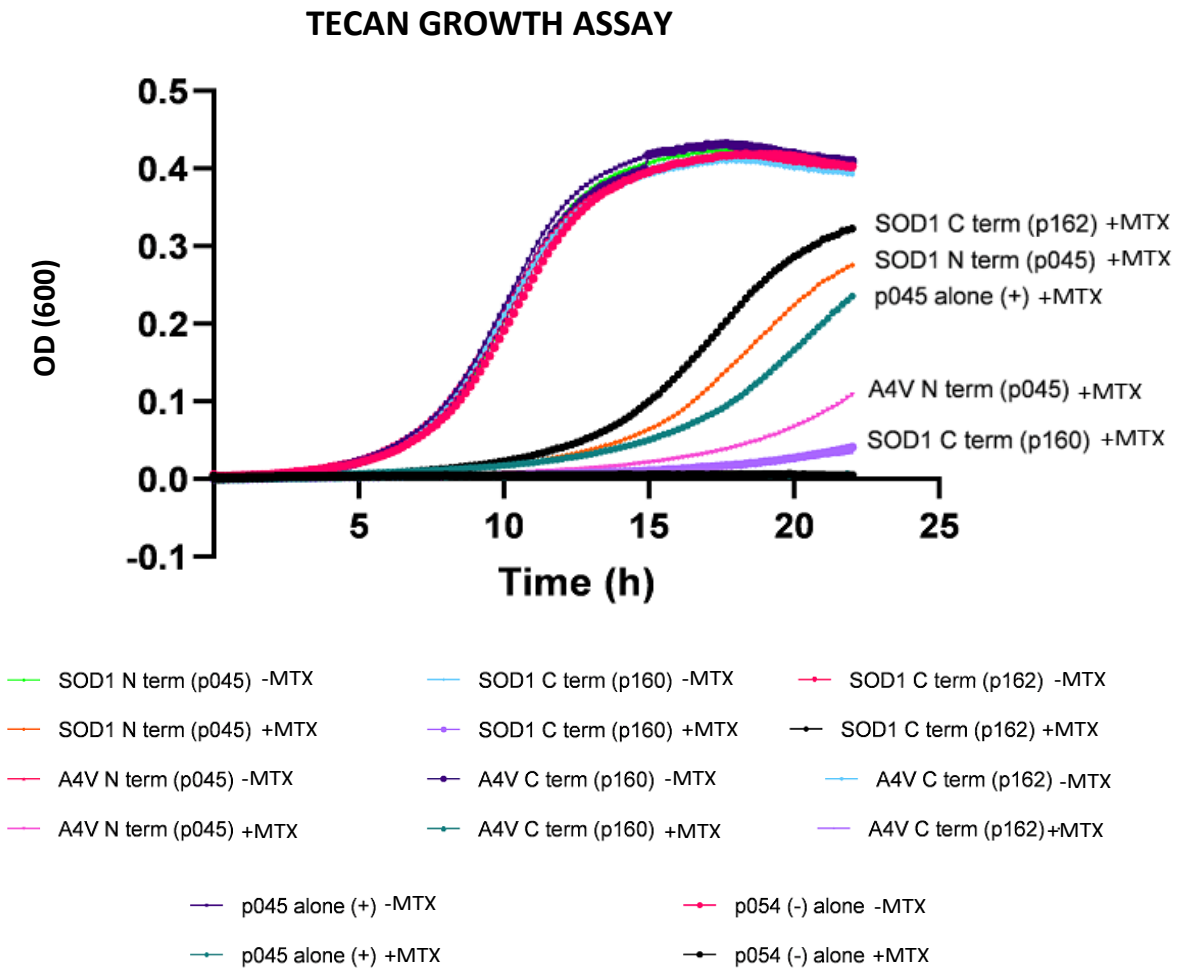


***Figure 14**. Tecan growth assay with and without MTX.*

The graph shows two types of growth curve. The more defined curve corresponds to the MTX-free culture, while the growth curves marked with their constructs are from MTX-containing cultures. Growth in MTX-free medium is evident and positive for all the constructs. This is because, although the A4V mutation may affect the stability of the protein, given the absence of MTX, the expression of the DHFR enzyme will not be essential, and therefore the yeast will be able to grow.

The situation is not the same for cultures on MTX medium. The positive control, p045 with DHFR expression, reflects the importance of the reporter enzyme in the assay. Since in this control the two DHFR fragments bind, MTX is neutralised by the enzyme, allowing the growth of yeast carrying this construct. As shown in **Table 28**, there have been some constructs that have prevented this growth: A4V C-term p160 and A4V C-

term p162. To draw conclusions as to why these constructs prevent the growth of specimens carrying it, it is necessary to consider the growth of its homologous construct, SOD1 (WT). For example, in the case of the p045 plasmid, growth is observed in both SOD1 N-term and A4V N-term, making it difficult to determine whether the reason is because the mutation in question does not affect the stability of SOD1 or because the N-term construct, with either insert, allows the correct expression of DHFR. In this case, yeast with A4V C-term p160 and A4V C-term p162 did not grow, but those carrying the same constructs with the WT insert did. The fact that they grew with one insert and not with the other allows us to determine that, in effect, the A4V mutation somehow impairs the stability of SOD1, preventing correct binding of the DHFR fragments and, consequently, allowing the toxicity of MTX.

***Table 28**. Growth results for each plasmid construction.*

Construction	**-MTX Growth**	**+MTX Growth**
SOD1 N term (p045)	Yes	Yes
A4V N term (p045)	Yes	Yes
SOD1 C term (p160)	Yes	Yes
A4V C term (p160)	Yes	No
SOD1 C term (p162)	Yes	Yes
A4V C term (p162)	Yes	No
p045 alone (+)	Yes	Yes
p054 alone (-)	No	No

From these data, plasmids with the SOD1 and A4V constructs were selected that allowed for increased growth on MTX medium (**Table 29**).

***Table 29**. Best plasmid constructions for WT SOD1 and A4V SOD1 in the Tecan growth assay.*

SOD1 insert	**Best growth**
WT	C term (p162)

A4V	N term (p045)

These results demonstrate that SOD1 can be inserted at either the C-terminal or N-terminal end of DHFR. However, placing SOD1 at the C-terminal end allows for better expression of the expression cassette in question (SOD1 + DHFR), indicating that plasmid 162 is the best candidate for carrying out the abundance assay.

Growth assay in a bigger volume

Finally, plasmids p162 C-term and p045 N-term were used to evaluate yeast growth in a volume of 200 mL, conditions in which the growth assay with the mutational libraries will be carried out.

As mentioned in the methodology, the optimum OD for working in larger volumes was 0,04, so to achieve this concentration in a 200 mL volume the volume of cells required was calculated. **Table 30** shows the OD readings of transformed yeast cultures (exponential phase) and the volume needed for each of them to decrease the concentration to OD = 0,04:

Table 30*. OD readings in exponential phase.*

Sample	**OD (exp. phase)**	**V_{cells} to OD=0,04 (mL)**
SOD1 C-term (p162)	0,660	12,12
A4V N-term (p045)	0,952	8,40

After incubating the yeasts in 200 mL of medium, OD readings were taken up to 23 hours after incubation. The OD readings are shown in **Table 31**:

Table 31*. OD readings in 200mL of media.*

Time after incubation (h)	12	14	17	20	23

OD SOD1 C-term	0,272	0,342	0,670	1,278	1,922
OD A4V N-term	0,173	0,135	0,146	0,334	0,396

The following graph shows the growth for each building.

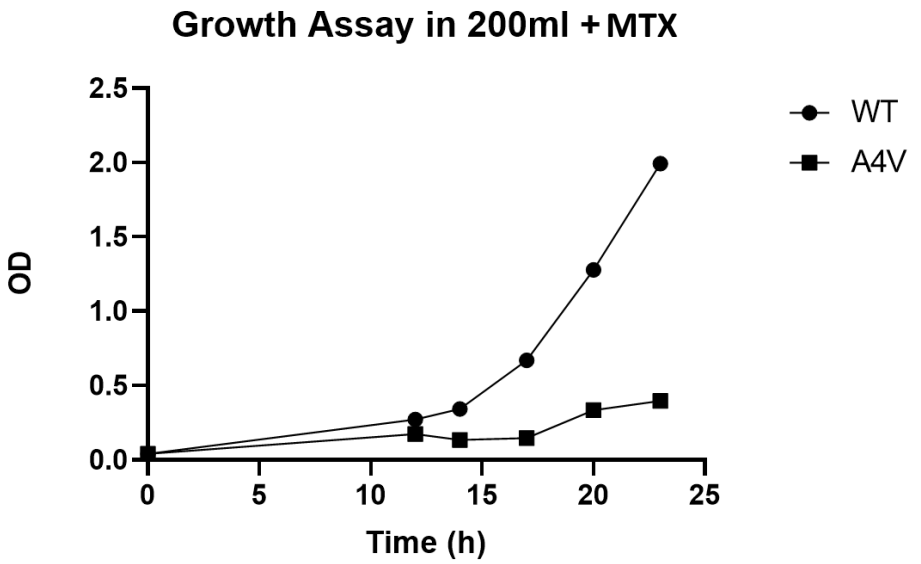


***Figure 15**. WT SOD1 vs A4V SOD1 growth assay in a 200 mL culture.*

These results demonstrate that growth correlation between WT SOD1 and A4V SOD1 is maintained at higher volume conditions. This differentiation between the two forms of SOD1 implies that it is possible to perform the abundance assay with the SOD1 mutational libraries, which must be carried out in large volumes to ensure measurement of all SOD1 variants. Although the growth curve is not fully comparable to the growth curve of the Tecan assay, the result is perfectly valid to conclude that the same growth behavior is maintained under conditions of higher volumes.

CONCLUSIONS

While mutations in the SOD1 protein are a major cause of familial and sporadic ALS, the mechanism by which they impact the protein and lead to disease onset and progression remains to be systematically elucidated. The application of Deep

Mutational Scanning in SOD1 is a promising strategy to comprehensively rationalize the impact of all possible SOD1 mutations and therefore better interpret their outcome in relation to ALS. Using a DHFR fusion as a high-throughput reporter of protein abundance, and combining this to deep sequencing, makes it possible. In this work, we performed the preliminary work required for employing DMS to quantify the effect of all mutations in SOD1 on the protein abundance and therefore, as a proxy, on its stability. This study has shown that it is relevant for the correct expression of DHFR and SOD1, and that the insertion of SOD1 at the C-terminus of DHFR allows the better discrimination between the WT protein and a known destabilizing disease mutant behavior.

Using a small plasmid to build the constructions that will incorporate the mutational libraries has proven to be an effective strategy for the correct assembly of SOD1 fragments. The successful cloning of fake libraries leads to the conclusion that the same strategy can be used for the cloning of real mutational libraries. It has been shown that a key factor determining the success of such assembly is the use of the Gibson reaction instead of enzymatic digestion, which is directly related to the size of the vector in question.

The small-scale test of the abundance assay has demonstrated the effectiveness of using Deep Mutational Scanning (DMS) to quantify the effect of the A4V mutation on SOD1 folding. However, while the method has demonstrated its ability to discriminate between one pathogenic mutation and the WT, it is true that ir has the limitation of not being able to determine at the molecular level if mutations simply destabilise the protein, lead to aggregates, or both of them.

Following up with this project and cloning the comprehensive SOD1 librares in the plasmid constructs that we generated in this work, will make it possible to run the abundance assay on thousands of SOD1 variants at the same time and therefore quantify the effect of mutations to create an atlas that will serve as a guide for the knowledge of the changes in SOD1 and its relationship with ALS.

In the continuation of this study, another possibility is to combine the method used with other high-throughput selection approaches, as the measurement of how the formation of the SOD1 dimers vares upon mutation, based on an affinity assay (Binding PCA).

From these assays, our goal is to create an atlas that discriminates between pathogenic and non-pathogenic mutations to create a breakthrough in disease

prediction. In addition, it would be useful to apply this methodology to the study of other proteins involved in ALS, such as TDP-43 and FUS, and to establish whether there are relationships between all of them, and how the destabilization aggregation of proteins of different types takes place.

APPENDICES

APENDIX 1

***Table 2**. Plasmids and SOD1 inserts purification method.*

Sample	**From**
SOD1 WT N-term	PCR purification
SOD1 A4V N-term	PCR purification
SOD1 WT C-term	PCR purification
SOD1 A4V C-term	PCR purification
P045	Miniprep
P054	Miniprep
P160	Miniprep
P162	Miniprep

***Table 8**. PCR compounds*

Compound	Template DNA (1 ng/µL)	dNTPSs (10 mM)	Fwd primer (10 µM)	Rev primer (10 µM)	5X Q5 Reaction Buffer	Q5 DNA Polymerase	Free water
Volume (µL)	1	1	2,5	2,5	10	0,5	32,5

***Table 9**. PCR program for amplification of SOD1 inserts.*

Step	**Temperature**	**Time**

Initial Denaturation	98 ºC	30 seconds
25 cycles	98 ºC	10 seconds
	72 ºC	20 seconds
	72 ºC	10 seconds
Final extension	72 ºC	2 minutes
Hold	4 ºC	

***Table 10**. PCR program for amplification of SOD1 inserts.*

Step	**Temperature**	**Time**
Initial Denaturation	98 ºC	30 seconds
30 cycles	98 ºC	10 seconds
	72 ºC	20 seconds
	72 ºC	10 seconds
Final extension	72 ºC	2 minutes
Hold	4 ºC	

***Table 11**. Empty Vectors Sanger sequencing preparation.*

Compound	**Volume – Concentration**	**Sequence**
Empty Vector 1	5 µL – 100 ng/µL	cgctcaagtcagaggtg
Empty Vector 2	5 µL – 100 ng/µL	cgctcaagtcagaggtg
Empty Vector 3	5 µL – 100 ng/µL	cgctcaagtcagaggtg
Primer check insert nicking rev	5 µL – 5 µM	cacctctgacttgagcg

***Table 12**. Primers used to amplify SOD1 fake libraries.*

Fake library	**Template**	**Primer**	**Primer sequence**	**Annhealing temperatura (ºC)**
1	Empty Vector 3	amp SOD1 lib1 fwd	ccgaaaagtgcc - GCTAGC - ATG	70
		amp SOD1 lib1 rev	GCACTGGTACAGCCTGCTGT	
2	SOD1	amp SOD1 lib2 fwd	GGATTCCATGTTCATGAGTTT	61
		Amp SOD1 lib2 rev	TGGTCTCCTGAGAGTGAGAT	
3	Empty Vector 1	Amp SOD1 lib3 fwd	GGCCGATGTGTCTATTGAA	63
		Amp SOD1 liib3 rev	gttctttcctgcg- AAGCTT	

***Table 14**. Amplified SOD1 fragments and linearized nicking vector quantifications.*

Sample	**Concentration (ng/µL)**
SOD1 fragments 2-3	307,9
SOD1 fragment 1	104,8
SOD1 fragment 3	105,5

SOD1 fragments 1-2	132
Linearized nicking vector	107,1

***Table 15**. Insert and vector volumes to build Empty Vector 1 by Gibson assembly.*

fragment 2-3	gibson 1-10		Concentration ng/ul	size (bp)	Concentration pmol/ul	pmol for X ng vector	300
	10	insert	307,9	150	3,157948718		
		Vector	107,1	1800	0,091538462	0,256410256	
	REACTION	VOLUMES (ul)					
	Vector	2,801120448					
	insert	0,811951932					
	Gibson Master Mix(2X)	10					
	H2O	6,386927619					

***Table 16**. Insert and vector volumes to build Empty Vector 2 by Gibson assembly.*

fragment 1 + fragment 3	gibson 1-10		Concentration ng/ul	size (bp)	Concentration pmol/ul	pmol for X ng vector	300
	10	insert 1	104,8	150	1,074871795		
		Vector	107,1	1800	0,091538462	0,256410256	
		insert 3	105,5	150	1,082051282		
	REACTION	VOLUMES (ul)					
	Vector	2,801120448					
	insert 1	2,385496183					
	insert 3	2,369668246					
	Gibson Master Mix(2X)	10					
	H2O	2,443715122					

Table 17. Insert and vector volumes to build Empty Vector 3 by Gibson assembly.

fragment 1-2	gibson 1-10		Concentration ng/ul	size (bp)	Concentration pmol/ul	pmol for X ng vector	300
	10	insert	132	150	1,353846154		
		Vector	107,1	1800	0,091538462	0,256410256	
	REACTION	VOLUMES (ul)					
	Vector	2,801120448					
	insert	1,893939394					
	Gibson Master Mix(2X)	10					
	H2O	5,304940158					

***Table 18**. Minipreps from Gibson quantifications.*

Sample	**Concentration (ng/µL)**
Empty Vector 1 miniprep	143,5

Empty Vector 2 miniprep	58,1
Empty Vector 3 miniprep	152,4

***Table 22**. Compound volumes to carry out the enzymatic digestion.*

PLASMIDS			Conc Miniprep (ng/ul)	ul x 1ug	Fast Digest Buffer 10X	NheI	HindIII	DeP	H2O
classic Nterm_Nliker	po45		689,4	1,4505367	3	1	1	1	22,5494633
neg_control	p054		292,9	3,41413452					
C-term_Clinker	p160		477,5	2,09424084	3	1	1	1	21,9057592
C-term_Nlinker	p162		556	1,79856115	3	1	1	1	22,2014388
N-term_Clinker	p164		-						
INSERTS			Conc > PCR pur (ng/ul)	ul x 1ug	Fast Digest Buffer	NheI	HindIII	DeP	H2O
WT FwdN RevN			214,3	4,66635558	2	1	1	0	11,3336444
A4V FwdN RevN			234,8	4,25894378	2	1	1	0	11,7410562
WT FwdC RevC			222,6	4,49236298	2	1	1	0	11,507637
A4V FwdC RevC			218,3	4,58085204	2	1	1	0	11,419148

APENDIX 2

PCR purification**:** MinElute® PCR Purification Kit Protocol

- 5 volumes of PB Buffer were added to 1 volume of PCR reaction and mixed.
- The volume was transferred to a MinElute column and centrifuged for 1 minute. The over-nadant was discarded and the column was returned to its tube.
- 750 µL of Buffer PE was added to the column and centrifuged for 1 minute. This step was done in duplicate. The supernatant was discarded.
- The column was transferred to a clean tube.
- 25 µL H2O was used to elute the DNA. For this, the water was incubated for 5 minutes at 50°C and, when added to the column, the column was incubated for 1 minute at 50°C. The column was centrifuged for 1 minute and the eluted DNA was recovered.
- The DNA concentration was quantified by loading 2 µL of sample into the Nanodrop.

Miniprep: NZY Miniprep Kit Protocol

- 2 mL of bacteria were centrifuged for 2 minutes at 13000 rpm.
- As much medium as possible was removed to isolate the pellet.
- The pellet was resuspended in 250 µL of Buffer A1 and mixed by Vortex.
- 250 µL of Buffer A2 was added and mixed by inverting 10 times. Incubated at room temperature for up to 4 minutes.
- 300 µL of Buffer A3 was added and mixed by inverting 10 times.
- Centrifuged 10 minutes at 13000 rpm.
- The supernatant was transferred to a Nzytech column and centrifuged 1 minute at 10800 rpm. The supernatant was discarded.
- 600 µL of Buffer A4 was added and centrifuged for 1 minute at 10800 rpm.
- The supernatant was discarded and the column was centrifuged alone for drying for 2 minutes at 10800 rpm.
- The column was transferred to a clean tube and eluted with 20 µL of water at 50°C.
- DNA concentration was quantified by loading 2 µL of sample into the Nanodrop.

Bacteria transformation: DH5-alpha Chemically Competent *E.coli* Cells Transformation Protocol

- Incubation plates and SOC medium were pre-warmed to 37 ºC.
- Mix 2 µL of DNA (PCR purification) or 4 µL of DNA (from Gibson) with 50 µL of competent bacteria (-80 ºC).
- The mixture was left on ice for 30 minutes and Heat Shock was performed at 42 ºC for 20 seconds for DH5-alpha bacteria and 45 seconds for Max Efficiency DH5-alpha bacteria.
- Bacteria were left on ice for 2 minutes.
- 450 µL of pre-warmed SOC medium was added and incubated for 1 hour at 37 °C, shaking at 200 rpm.
- Centrifuged for 2 minutes at 3000 rpm and the pellet was resuspended with 200 µL of fresh SOC medium.

- The pellet was grown on LB plate and the respective anitibiotic as a resistance marker.
- Incubated at 37°C overnight.
- The next day, a colony was chopped and grown in liquid LB medium to make minipreps.

Yeast transformation + Tecan

Day -1:

- Inoculate yeast strain from a single colony to 20 mL YPDA and grow ON at 30ºC until saturation
- Prepare Recovery Media and other reagents

Day 0:

- Morning (~10h), measure OD600 of preculture
- Inoculate 87.5 mL of warm YPDA at OD600 of 0.3
- Incubate for 4 h at 30ºC
- Prewarm medias: Recovery Media
- Harvest by centrifugating 5 min at 3,000g (two falcon tubes 50 ml)
- Resuspend in 5 mL water, transfer to a Falcon50 and complete to 25 mL.
- Centrifuge 5 min at 3,000g
- Resuspend in 5 mL SORB, transfer to new Falcon50, complete to 25 mL SORB
- Centrifuge 5 min at 3,000g.
- Resuspend in 4.3 mL SORB and incubate 30 min on a wheel.
- After incubation, split 10µL cells in PCR tubes for each independent transfo.
- OPTIONAL: freeze cells down at this stage for future transfos. The protocol can be started from this point thawing on ice frozen cell aliquots
- Dilute the Salmon Sperm (1:10 dilution)
- Boil 2 µL (for each transformation, 200 µL for 96wp) of 1 mg/mL ssDNA (10 mg/mL diluted down 10X) for 5 min and let it sit for at least 2 min on ice.
- While boiling:

    - Prewarm the plates (selection) and the recovery media at 30 degrees.
    - Go and take the yeast cells from the -80 degrees fridge
    - Take the PRC tubes. One for each transformation.
- Do the next steps UNDER THE FIRE
- Add 1.5 µL of plasmid in each corresponding tube, 2 µL of ssDNA (salmon sperm)

  and 40 µL of Plate Mixture to each well.
- Mix with the pipette and incubate 30 min at room temperature (RT).
- Add 4 µL mL of DMSO to each well.
- Heatshok at 42ºC for 20 min (Use PCR block).
- Centrifuge 5 min at 3,000 rpm and remove supernatant using pipette carefully.
- Do the next steps UNDER THE FIRE
- Remove supernatant
- Resuspend in 100 µL of Recovery Media
- Incubate for 1 h at 30ºC
- Plate 2.5 µL on Petri dishes of Plasmid Selection Media and grow at 30ºC for two days when colonies visible at the naked eye
- Put the rest of the cells in the 4 degrees fridge to use in case you need to plate again.

Day 1: Start the ON liquid culture

- Create the disposition plate excel file, several colonies for the same plasmid, always

  randomizing the positions where this fall in the 96 wp. Only required one well for the blank.
- In a deep 1mL 96 wp, add 400 µL of SC -URA/MET/ADE
- Pick the colonies of each genotype with a pipette tip and add them in each corresponding well (leave the tip for a while until you finish filling the plate). Once done remove the tips and seal it with a gas-permeable seal.

- Grow overnight or longer (24h max) at 30ºC, shaking in the incubator

Day 2:

- Put MTX media to prewarm
- Pick the 96 wp and add 25 µL of the “saturated” culture (make sure to mix the culture

  properly with the pipette) in a 96 wp in 75 µL water (Nuc transparent flat bottom, 4X

  dilution) and measure the OD600 culture in a Tecan machine. Pre-set Tecan machine at 30 ºC.
- Diluting cells to become OD600=1 by changing the volume of water, not the volume

  of cells. Calculate the volume of water per well required to reach an OD600=1 if we add 10 µL of the saturated cells.
- OD=1 plate with different volumes of water.
  In a new 96wp Nunc, add 95 µL of SC-URA/MET/ADE + MTX per well (using

  multichannel and reservoir). This will go into the Tecan, don’t touch the ceiling or bottom of the plate with your fingers. Add 5 µL of OD=1 cells into the MTX media plate (keep the same order columns/rows always).

SORB

100 mM LiOAc, 10 mM Tris pH 8.0, 1 mM EDTA, 1M sorbitol

1L SORB = 10g LiOAc, 182 f sorbitol, 10 mL Tris 1M (1000X), 2 mL EDTA 0,5M (500X)

Plate mixture

100 mM LiOAc, 10 mM Tris-HCl pH 8 (from 1M stock), 1 mM EDTA/NaOH (from 0,5 M stock), pH 8, 40% PEG3350

1L plate mixture = 10g LiOAc, 400 g PEG3350, 10 Ml Tris, 2 mL EDTA

Recovery media

YPDA + Sorbitol 0,5M

1L Recovery media = 10g Yeast Extract, 20 g Peptone, 20g Glucose, 91g Sorbitol, up

to 1L with water

APPENDIX 3

Equation 1

The ratio C1 x V1 = C2 x V2 was used, where:

- C1 = plasmid concentration (greater than 100 ng/µL)
- C2 = 100 ng/µL
- V1 = plasmid volumen required to achieve the dilution concentration
- V2 = 5 µL

By isolating the value of V1, it is possible to know how much plasmid needs to be pipetted to dilute to 100 ng/µL:

$$V1 = \frac{C2\ x\ V2}{C1}$$

Equation 2

$$Real\ OD600\ =\ (Well\_x\ -\ blank)\ x\ dilution\ factor\ x\ 10.$$

Equation 3

$$C1xV1 = C2xV2$$

where C1 = OD 0.05 (Tecan); C2 = OD 0.04; V1 = Vyeast; V2 = 200 mL.

Gibson protocol

The DNA concentration (ng/µL) and the length of each sample (bp) were used to obtain a concentration in picmoles/µL. For this, the following ratio was used, considering that the approximate weight of a base pair is 650 Daltons:

$$DNA\ concentration\ \left(\frac{pmol}{ul}\right) = \frac{DNA\ concentration\ \left(\frac{nmol}{ul}\right) x\ 1000}{DNA\ size\ (bp) x\ 650}$$

A quantity of 300 ng of vector was defined to carry out the reaction, and this value was

used to calculate the amount of vector in picomoles, according to the following expression:

$$pmol\ x\ 300\ ng\ of\ vector = \frac{300\ ng\ of\ vector}{Vector\ concentration\ (\frac{ng}{ul})}\ x\ Vector\ concentration\ (\frac{pmol}{ul})$$

Vector and insert volumes were calculated according to:

$$Vvector = \frac{300\ ng\ of\ vector}{Vector\ concentration\ (\frac{ng}{ul})}$$

$$Vinsert = \frac{pmol\ x\ 300\ nf\ of\ vector}{Insert\ concentration\ (\frac{pmol}{ul})}$$

10 µL of Gibson Master Mix 2X was used, so the total reaction volume was 20 µL. Water was used to achieve the required 10 µL of vector and insert.

Compound	**Volume**
Gibson Master Mix 2X	10 µL
Vector	X µL
Insert	Y µL
Water	10 – (X+Y) µL

APPENDIX 4

R script required packages

- ggplot2
- seqinr
- RGenetics
- openxlsx

REFERENCES

Andersen, P. M. (2006). Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. *Current Neurology and Neuroscience Reports*, *6*(1), 37–46. https://doi.org/10.1007/S11910-996-0008-9

Di Gregorio, S. E., & Duennwald, M. L. (2018). ALS yeast models—Past success stories and new opportunities. In *Frontiers in Molecular Neuroscience* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fnmol.2018.00394

Faure, A. J., Domingo, J., Schmiedel, J. M., Hidalgo-Carcedo, C., Diss, G., & Lehner, B. (2022). Mapping the energetic and allosteric landscapes of protein binding domains. *Nature*, *604*(7904), 175–183. https://doi.org/10.1038/s41586-022-04586-4

Fujisawa, T., Homma, K., Yamaguchi, N., Kadowaki, H., Tsuburaya, N., Naguro, I., Matsuzawa, A., Takeda, K., Takahashi, Y., Goto, J., Tsuji, S., Nishitoh, H., & Ichijo, H. (2012). A novel monoclonal antibody reveals a conformational alteration shared by amyotrophic lateral sclerosis-linked SOD1 mutants. *Annals of Neurology*, *72*(5), 739–749. https://doi.org/10.1002/ANA.23668

Hayashi, Y., Homma, K., & Ichijo, H. (2016). SOD1 in neurotoxicity and its controversial roles in SOD1 mutation-negative ALS. In *Advances in Biological Regulation* (Vol. 60, pp. 95–104). Elsevier Ltd. https://doi.org/10.1016/j.jbior.2015.10.006

Hilton, J. B., White, A. R., & Crouch, P. J. (2016). Endogenous Cu in the central nervous system fails to satiate the elevated requirement for Cu in a mutant SOD1 mouse model of ALS. *Metallomics : Integrated Biometal Science*, *8*(9), 1002–1011. https://doi.org/10.1039/C6MT00099A

Kaur, S. J., McKeown, S. R., & Rashid, S. (2016a). Mutant SOD1 mediated pathogenesis of Amyotrophic Lateral Sclerosis. In *Gene* (Vol. 577, Issue 2, pp. 109–118). Elsevier B.V. https://doi.org/10.1016/j.gene.2015.11.049

Kaur, S. J., McKeown, S. R., & Rashid, S. (2016b). Mutant SOD1 mediated pathogenesis of Amyotrophic Lateral Sclerosis. *Gene*, *577*(2), 109–118. https://doi.org/10.1016/J.GENE.2015.11.049

Masrori, P., & Van Damme, P. (2020). Amyotrophic lateral sclerosis: a clinical review. In *European Journal of Neurology* (Vol. 27, Issue 10, pp. 1918–1929). Blackwell Publishing Ltd. https://doi.org/10.1111/ene.14393

Mejzini, R., Flynn, L. L., Pitout, I. L., Fletcher, S., Wilton, S. D., & Akkari, P. A. (2019). ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Frontiers in Neuroscience*, *13*, 1310. https://doi.org/10.3389/FNINS.2019.01310/BIBTEX

Münch, C., & Bertolotti, A. (2010). Exposure of Hydrophobic Surfaces Initiates Aggregation of Diverse ALS-Causing Superoxide Dismutase-1 Mutants. *Journal of Molecular Biology*, *399*(3), 512–525. https://doi.org/10.1016/J.JMB.2010.04.019

Peggion, C., Scalcon, V., Massimino, M. L., Nies, K., Lopreiato, R., Rigobello, M. P., & Bertoli, A. (2022). SOD1 in ALS: Taking Stock in Pathogenic Mechanisms and the Role of Glial and Muscle Cells. In *Antioxidants* (Vol. 11, Issue 4). MDPI. https://doi.org/10.3390/antiox11040614

Seuma, M., Faure, A., Badia, M., Lehner, B., & Bolognesi, B. (2021). The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial alzheimer's disease mutations. *ELife*, *10*, 1–49. https://doi.org/10.7554/ELIFE.63364

Seuma, M., Lehner, B., & Bolognesi, B. (2022). An atlas of amyloid aggregation: the impact of substitutions, insertions, deletions and truncations on amyloid beta fibril nucleation.

Nature Communications 2022 13:1, *13*(1), 1–13. https://doi.org/10.1038/s41467-022-

34742-3

Sheng, Y., Chattopadhyay, M., Whitelegge, J., & Selverstone Valentine, J. (2013). SOD1 Aggregation and ALS: Role of Metallation States and Disulfide Status. *Current Topics in Medicinal Chemistry*, *12*(22), 2560–2572. https://doi.org/10.2174/1568026611212220010

Sramek, M., Neradil, J., & Veselska, R. (2017). Much more than you expected: The non-DHFR-mediated effects of methotrexate. *Biochimica et Biophysica*

Acta (BBA) - General Subjects, *1861*(3), 499–503. https://doi.org/10.1016/J.BBAGEN.2016.12.014

Todd, T. W., & Petrucelli, L. (2022). Modelling amyotrophic lateral sclerosis in rodents. *Nature Reviews Neuroscience 2022 23:4*, *23*(4), 231–251. https://doi.org/10.1038/s41583-022- 00564-x

Une, M., Yamakawa, M., Watanabe, Y., Uchino, K., Honda, N., Adachi, M., Nakanishi, M., Umezawa, A., Kawata, Y., Nakashima, K., & Hanajima, R. (2021). SOD1-interacting proteins: Roles of aggregation cores and protein degradation systems. *Neuroscience Research*, *170*, 295–305. https://doi.org/10.1016/J.NEURES.2020.07.010

Workman, A. (2020). Nucleation and kinetics of SOD1 aggregation in human cells for ALS1.

Molecular and Cellular Biochemistry, *466*(1–2), 117–128. https://doi.org/10.1007/S11010-020-03693-Y/METRICS

SELF-EVALUATION

Being part of the Protein Phase Transitios in Health and Disease research group at IBEC has been an enriching and exciting experience. Although before I started my internship I had no experience beyond what I had gained at university, I was taught and trained in the group to carry out the study almost independently. I was able to learn and practice basic molecular biology, biochemistry and chemistry skills, and to discover the potential of bioinformatics in research.

The dynamics of working in the lab were very well established. The independence of being able to organise my time to carry out my experiments allowed me to order my ideas and understand the study I was conducting much better. In addition, the weekly meetings with another research group at IBEC, whose approach is based on the biophysics of protein dynamics, allowed me to take a completely different perspective on my study, improve my understanding of articles and my way of asking questions.

I have learned a lot from the study I have done and I intend to continue to be part of the group for the next year in my master's degree.