Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENZE BIOMEDICHE E BIOTECNOLOGICHE

CICLO XXXII

COORDINATORE Prof. Pinton Paolo

Targeted molecular strategies for X-linked genetic disorders: the paradigmatic models of Fabry disease and Haemophilias

Settore Scientifico Disciplinare BIO/11

Dottoranda
Dott. Lombardi Silvia

(firma)

Tutore
Prof. Pinotti Mirko

(firma)

Co-tutore
Dott. Branchini Alessio

(firma)

Anni 2016/2019

Table of Contents

Abbreviations

-a	activated
aa-tRNA	aminoacyl-tRNA
AAV	adeno-associated virus
AGAL	α-galactosidase A
AON	antisense oligonucleotide
AP	activation peptide
APCC	activated prothrombin complex concentrate
aPTT	activated partial thromboplastin time
ASL	anti-codon stem loop
AT	anti-thrombin
β2m	β2-microglobulin
BP	branch point
bp	base pairs
C-terminal	carboxy-terminal
EAHAD	European Association for Haemophilia and Allied Disorders
ELISA	enzyme-linked immunosorbent assay
EMA	European Medicines Agency
ERT	enzyme replacement therapy
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
ExSpeU1	exon-specific U1 snRNA
FcRn	neonatal Fc receptor
FD	Fabry disease
FDA	Food and Drug Administration
FIX	coagulation factor IX
FV	coagulation factor V
FVII	coagulation factor VII
FVIII	coagulation factor VIII
FX	coagulation factor X
FXI	coagulation factor XI
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
Gb3	globotriaosylceramide

h-	human
HA	Haemophilia A
HB	Haemophilia B
HC	heavy chain
HEK	human embryonic kidney
HGVS	Human Genome Variation Society
HSA	human serum albumin
IgG	immunoglobulin G
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
KO	knock-out
LAMP-1	lysosomal-associated membrane protein 1
LSD	lysosomal storage disorder
LV	lentivirus
M6PR	mannose 6-phosphate receptor
MSA	mouse serum albumin
FP	fusion protein
MW	molecular weight
nd	not detectable
NMD	nonsense-mediated mRNA decay
NTC	normal termination codon
OD	optical density
OMIM	Online Mendelian Inheritance in Man
PAA	polyacrilamide
PABP	poly-(A) binding protein
PAGE	polyacrilamide gel electrophoresis
PT	prothrombin time
PTC	premature termination codon
r-	recombinant
RFU	relative fluorescence unit
rGal	Myc-tagged recombinant α-galactosidase A
RT	room temperature
RT-PCR	reverse transcriptase-PCR
SD	standard deviation

siRNA	small interfering RNA
SMA	spinal muscular atrophy
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particles
TF	tissue factor
TFPI	tissue factor pathway inhibitor
U2AF	U2 snRNP auxiliary factor
UTR	untranslated region
vWF	von Willebrand factor
WT	wild-type

Chapter 1

General Introduction

Traditionally, a broad approach has been applied to the development of treatments for genetic diseases, in an attempt to benefit the greatest possible number of patients. In several cases, these treatments have led to significant amelioration of the disease phenotypes and of patients' quality of life and lifespan. However, in many contexts major drawbacks still exist, such as the inaccessibility of specific tissues to the therapeutics, the short persistence in circulation of the infused drug and the difficulty in delivering large therapeutic transgenes. Moreover, the pathological and phenotypical heterogeneity found in patients often frustrate the attempts to apply a universal treatment.

In this context, even though a **mutation-specific therapeutic approach** can target only a subset of patients, it could still offer substantial improvements by either i) being able to reach otherwise inaccessible tissues (e.g. readthrough-inducing compounds for central nervous system, see Chapter 2), or ii) exploiting small expression cassettes compatible with adeno-associated virus (AAV)-mediated delivery (e.g. U1 snRNA-based approaches to correct splicing defects, see Chapter 4).

An additional more general approach relies on the **precise engineering** of therapeutic molecules in order to improve their functional and/or pharmacokinetic properties (e.g. exploiting human serum albumin for half-life extension purposes, see Chapter 3).

In this work, different approaches were explored for the treatment of three different X-linked disorders, chosen as paradigmatic models.

The first part of this thesis will focus on **Fabry disease**, a lysosomal storage disorder currently treated with enzyme replacement therapy. Unfortunately, the infused recombinant enzyme is not able to cross the blood-brain barrier, thus leaving untreated the central nervous system, which is one of the mainly involved organs. Given the relevant portion of patients harbouring nonsense mutations, which are associated with the most severe phenotype, a **nonsense-suppression approach** was investigated. This strategy relies on the ability of small compounds to induce the mechanism of ribosome readthrough over premature termination codons and, albeit already tested for several genetic diseases, it has never been attempted so far for Fabry disease.

The second part of the thesis will focus on replacement therapy for **Haemophilia B**. Albeit widely diffused and effective, replacement therapy still suffers from the major drawback of short half-life of the infused factor (coagulation factor IX, FIX). Several products have been developed with enhanced persistence in circulation but the research is always ongoing. In this work a double approach has been applied to improve both functional and pharmacokinetic properties of FIX. In particular, a novel **FIX-albumin fusion protein** was produced in which

a well-known gain-of-function variant of FIX has been exploited to increase the pro-coagulant features, together with a rationally engineered albumin characterised by enhanced binding to its receptor (the neonatal Fc receptor) and thus endowed with extended half-life.

Finally, in the third part of this thesis, gene therapy for **Haemophilia A** (HA) will be addressed. Recently, many progresses have been made in this field and significant results have been obtained, but the large cassette required for expression of the missing factor (coagulation factor VIII, FVIII), together with its very low expression levels, have hampered the establishment of gene therapy for HA. In this context, an alternative approach focused on mutations affecting splicing (estimated to affect 10-50% of patients) was explored, by taking advantage of **modified U1 snRNAs**. This strategy, already exploited to rescue aberrant splicing and functional protein synthesis in several diseases, can benefit from small expression cassettes compatible with adeno-associated viral delivery, with the additional advantage of maintaining the physiological regulation of target gene expression.

The aim of this work is to provide the proof-of-principle of the feasibility of each of these strategies for Fabry disease, Haemophilia B and Haemophilia A. Nevertheless, the approaches investigated here in three very specific contexts may be applied to other disease models as well.

Chapter 2

Investigation of ribosome readthrough in Fabry disease

Lombardi S, Ferrarese M, Marchi S, Pinton P, Pinotti M, Bernardi F, Branchini A.

Based on: *RNA Biol* **17**, 254–263 (2020)

2.1 Background and rationale

2.1.1 Nonsense mutations

Nonsense mutations account for ~11% of all disease-associated gene lesions and are defined as single nucleotide changes within a gene that result in the creation of an in-frame premature termination codon (PTC)[1]. A PTC reduces gene expression through two main mechanisms (Fig. 2.1). Firstly, it causes the premature termination of mRNA translation, leading to the production of a truncated protein which generally shows altered function and/or stability. Second, a PTC-bearing transcript may trigger a conserved surveillance pathway called nonsense-mediated mRNA decay (NMD), leading to the reduction in the steady-state level of cytoplasmic mRNA. Indeed, not all PTC-containing mRNAs undergo NMD as it has been demonstrated that 5-25% of transcripts normally escape degradation[2] and that PTCs located less than 50 nucleotides upstream of an exon-exon junction or in the last exon of a gene generally fail to elicit the surveillance mechanism[3]. Additionally, in a minority of cases, nonsense mutations can lead to the production of truncated proteins that can interfere with the wild-type form, thus exerting a dominant negative effect[4].

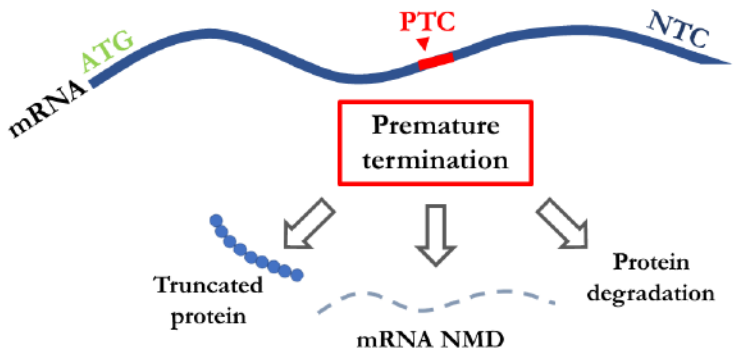

Figure 2.1. Mechanisms of gene expression reduction mediated by premature termination of translation. ATG, translation initiation codon; PTC, premature termination codon; NTC, normal termination codon; NMD, nonsense-mediated mRNA decay.

Nonsense mutations generally act by reducing the amount of full-length functional proteins, thus resulting in a loss-of-function effect, often associated to the so-called "null" phenotype. Given the relevant incidence of this type of genetic defect, in the past decades many efforts have been made to develop therapies specifically targeted to nonsense mutations suppression[5–7]. One of them relies on the natural phenomenon of termination codon suppression mediated by the ribosome readthrough mechanism.

2.1.2 Translation termination and ribosome readthrough

Eukaryotic translation comprises four major phases: initiation, elongation, termination and recycling. During translation elongation a process of aminoacyl-tRNA (aa-tRNA) sampling occurs and codons located in the acceptor (A) site of the ribosome are decoded by an aa-tRNA in a ternary complex with the elongation factor eEF1 and GTP. The accuracy of this process is strictly monitored by the ribosomal decoding centre at two steps separated by the irreversible hydrolysis of GTP[5]. Similarly, when a stop codon enters the ribosomal A site, the process of sampling still occurs. In particular, translation termination is mediated by the release factor eRF1 in complex with eRF3: the first resembles the overall shape of a tRNA and directly recognizes all stop codons (UAA, UAG, UGA), while the latter is a GTPase that binds eRF1 and assists in the termination process[8,9].

Therefore, when a stop codon enters the ribosome, a competition occurs between eRF1 and aa-tRNAs. Normally, eRF1 highly outcompetes aa-tRNAs for stop codon binding because there are no tRNAs that are cognate to one of the termination codons. However, translation termination is not a 100% efficient process and a stop codon on mRNA can be mis-recognised by an aa-tRNA, thus being suppressed by the physiological mechanism of ribosome readthrough (Fig. 2.2). During this process, eRF1 is superseded in binding the stop triplet by a near-cognate aa-tRNA (a tRNA whose anticodon is complementary to two of the three positions of a nonsense codon)[10]. As a result, instead of translation termination, an amino acid is added to the polypeptide chain and translation elongation continues until the next in-frame stop codon is reached.

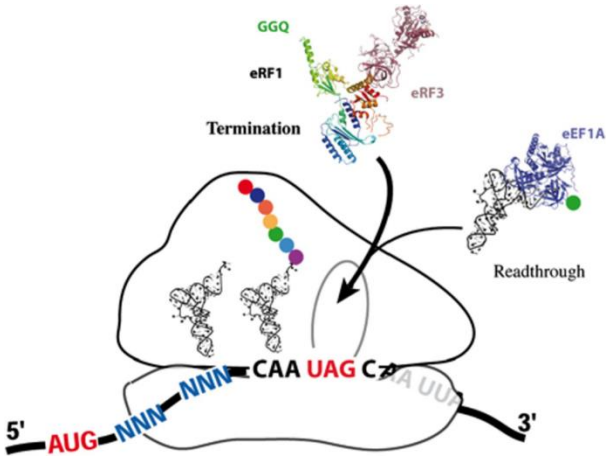

Figure 2.2. Schematic representation of the competition between termination and readthrough[11].

Ribosome readthrough was first identified as a way, for viruses, to expand their genetic information: it was observed that in *Escherichia coli* infected by RNA phage Qβ, the $tRNA^{Trp}$

stimulated readthrough over the UGA stop codon at the end of the coat protein cistron, resulting in a longer coat protein essential for the production of infective Qβ particles[12]. Since then, stop codon readthrough has been found to play important biological roles providing a mechanism to regulate gene expression through the production of different proteins from the same gene[13,14]. Nevertheless, translation termination is normally a highly efficient process and ribosome readthrough is thus a rare event.

2.1.3 Ribosome readthrough determinants

The overall frequency of readthrough can be influenced by several variables. First, the frequency of termination suppression at premature stop codons (~1%) is about 10-fold higher than at natural stop codons (<0.1%)[5,15,16]. This difference is likely due to the interaction between the termination complex and factors bound to the 3' untranslated region (UTR) of the mRNA (Fig. 2.3). Indeed, during translation, the mRNA is maintained in a closed-loop conformation whose function is both to protect the ends of the transcript from exonucleolytic degradation[17], and to enhance recycling of translational components, thus increasing the frequency of translation initiation. This closed conformation is maintained by the association of three actors: the cap-binding protein (eIF4E), which is bound to the 5'-cap structure of the mRNA, the poly(A)-binding protein (PABP), attached to the poly(A) tail, and eIF4G, that binds both eIF4E and PABP resulting in circularization of the mRNA.

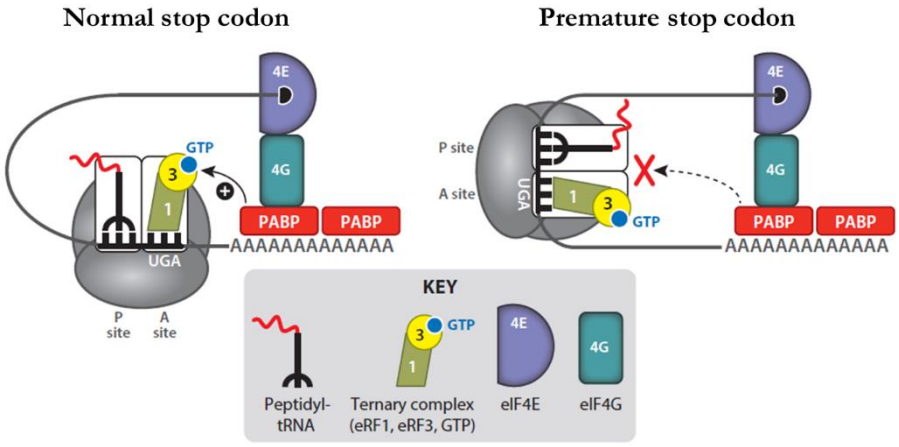

Figure 2.3. Schematic representation of normal (left) and premature (right) termination codons[5].

Along with its role in the formation of the closed-loop complex, PABP also promotes translation termination by interacting with eRF3 and stimulating polypeptide chain release[18]. In the presence of a premature stop codon, which is usually not close to the poly(A) tail, the interaction between PABP and eRF3 will likely be less efficient, leading to prolonged ribosomal

pausing and increased aa-tRNA sampling, which in turn makes the PTC more susceptible to readthrough[19].

In addition to the differences between natural stop codons and PTCs, a major determinant of readthrough frequency is the sequence context, namely the identity of the stop codon and the surrounding sequence. A first evidence came from statistical analysis of the context adjacent to the stop codons at the end of genes in several organisms, which revealed that the actual termination signal may consist of a tetranucleotide sequence[20]. Further studies on the efficiency of translation termination in yeast[21] and in mammalian translation systems[16] indicated that both upstream and downstream sequences act together with the stop codon to determine the overall efficiency of translation termination. In particular, the UGA codon was identified as the most susceptible to suppression, followed by UAG and UAA. Both studies further confirmed that also the fourth base plays a significant role in determining the occurrence of readthrough, although its effect appeared to be a function of the adjacent stop codon (Tab. 2.1). This is likely due to the network of hydrogen bonds that are formed between eRF1 and the different stop codons, together with the interaction of eRF1 with the downstream nucleotides[22].

Table 2.1. Basal readthrough levels with different stop codons and sequence contexts. Values are reported as percentage of full-length polypeptides on total protein translated *in vitro*[16].

Termination Signal	Percent Readthrough	Termination Signal	Percent Readthrough	Termination Signal	Percent Readthrough
UAA A	0.2	UAG A	0.9	UGA A	1
UAA C	0.5	UAG C	1.1	UGA C	3.8
UAA G	0.4	UAG G	0.8	UGA G	0.6
UAA U	0.2	UAG U	1.6	UGA U	0.7

However, the termination signals at the end of an open reading frame have been shown to usually promote efficient translation termination, especially among highly expressed transcripts. Moreover, the generation of carboxy(C)-terminally extended proteins following natural stop codon suppression is prevented by several cellular mechanisms. As an example, in case of stop codon readthrough, translation may proceed through the 3' UTR into the poly(A) tail leading to stalling of the ribosome. This induces a mRNA surveillance pathway called nonstop decay, which is a Ski7-dependent exosome-mediated degradation of transcripts with stalling ribosomes bound to the poly(A) tail[23]. In addition, recent studies have shown that translation into 3' UTRs may reduce C-terminally extended protein levels because 3' UTR-encoded peptides mark their resulting products for destruction, either co- or post-translationally[24].

Notably, while factors affecting readthrough occurrence have yet to be completely elucidated, several studies have identified the amino acids that are incorporated during stop codon suppression in yeast and mammalian cells[10,25,26]. The main determinant of the amino acid incorporated is the stop codon present, whereas the flanking sequences or a possible inducer have no effect on the identity of the suppressor. It has been demonstrated that UAA and UAG stop codons can be suppressed by the insertion of glutamine, tyrosine and lysine, whereas UGA codons can be suppressed by tryptophan, cysteine and arginine[10,25].

As already mentioned, the misreading of a stop codon during readthrough is mediated by a near-cognate tRNA (i.e. a tRNA with a single nonstandard Watson-Crick codon/anticodon base pair). Thus, the ability of a tRNA to decode a stop codon depends on its extent of base pairing and on the detection of that base pairing by the decoding centre of the ribosome. Recent crystallographic studies have revealed that the ribosomal A-site can tolerate nonstandard Watson-Crick base pairs and that it is the shape of the base pairs to be crucial, rather than the number or type of hydrogen bonds that are formed[27]. In addition, it has been suggested that a mechanism of "geometrical mimicry" by nonstandard pairing may be favourable for tRNA incorporation[27]. In particular, insertion of near-cognate tRNAs occurs by mispairing at either position 1 or 3 of the codon, whereas the second position is under tight selection by the ribosome and does not tolerate mismatches[10,25]. More specifically, readthrough of UAA leads to an equal distribution of position 1 and 3 mispairing, whereas UAG favours mispairings at position 1 and UGA at position 3[28]. In the first position, the nonstandard Watson-Crick U-G mispairing is predominant, due to its geometrical mimicry of a standard base pairing event. Likewise, nonstandard base pairings are allowed at the third position, with A-C favoured over G-G and A-G[28].

Interestingly, comparison of several studies indicated that, while the subset of amino acids inserted by readthrough remains the same regardless of the experimental system, their relative ratio changes when different conditions are used[10,25,26]. Indeed, readthrough occurrence is strongly influenced by pharmacological agents that bind to the ribosomal decoding centre thus altering translation termination fidelity[29].

2.1.4 Drug-induced readthrough as a nonsense suppression therapy

The first demonstration that certain low-molecular-mass drugs could stimulate ribosome readthrough of PTCs associated to genetic diseases opened the way to new therapeutic approaches based on nonsense mutations suppression. In 1996, a pioneer study showed that nonsense mutations in the *CFTR* gene could be suppressed by the aminoglycosides geneticin

and gentamicin, as measured by the appearance of full-length functional proteins in a cystic fibrosis model system[30]. Since then, readthrough induction has been explored as a therapeutic approach in more than one hundred publications for over fifty different genetic diseases[31].

The most studied readthrough-inducing compounds are aminoglycosides, a class of antibiotics characterised by a common 2-deoxystreptamine ring linked to one or more amino sugars. Aminoglycosides are clinically useful antibiotics because they inhibit prokaryotic protein synthesis at significantly lower concentrations than eukaryotic protein synthesis. In particular, it has been shown that aminoglycosides bind to specific nucleotides in the decoding site of the bacterial ribosome inducing A1492 and A1493 to flip out of the internal loop of helix 44, a conformational change similar to the transition that occurs upon cognate tRNA binding[29] (Fig. 2.4). As a result, aminoglycosides interfere with bacterial protein synthesis by facilitating amino acid misincorporation.

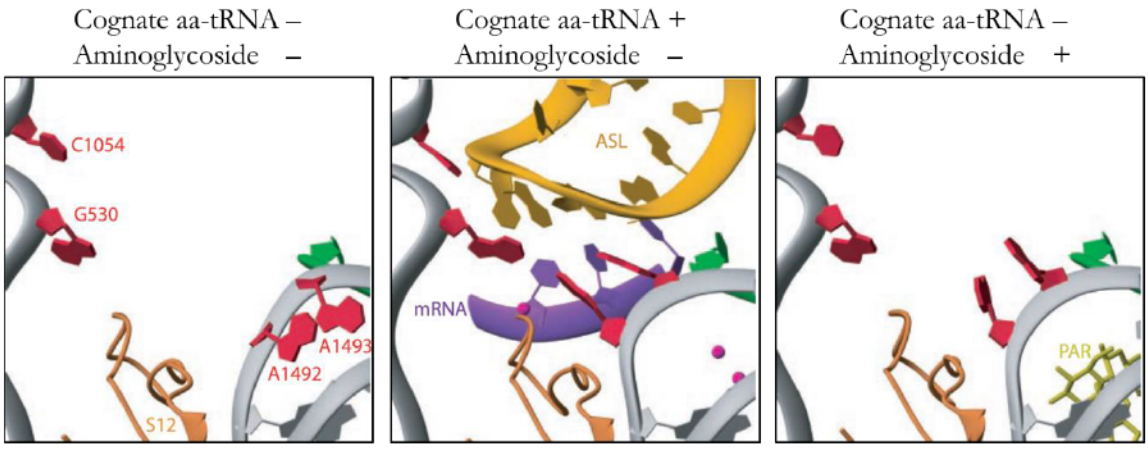

Figure 2.4. Bacterial ribosomal decoding site in different conditions. The tRNA anticodon stem-loop (yellow), the A-site mRNA codon (purple), protein S12 (brown) and important bases (red) involved in conformational changes are shown. ASL, anticodon stem-loop; PAR, paromomycin. Adapted from Ref.[29].

The major determinants of the differential aminoglycoside sensitivity observed between prokaryotes and eukaryotes are the eukaryotic 18S rRNA residues G1645 and A1754 (corresponding to *E. coli* 16S rRNA residues A1408 and G1491). The difference in these two residues cause the aminoglycoside-binding pocket to be shallower and less stable in eukaryotes, thus preventing strong binding[29,32]. Nevertheless, aminoglycosides have been shown to induce very low levels of misreading in mammalian cells, with the primary effect of increasing PTC suppression[33], thus leading to the extensive evaluation of these compounds as readthrough-inducing agents in several disease models[6,34–36].

Unfortunately, the use of aminoglycosides for lifelong nonsense suppression therapy is not feasible because they are associated with irreversible oto- and nephro-toxicity[37,38]. This damage is not directly related to the interaction of aminoglycosides with cytoplasmic ribosomes, but is rather caused by their interaction with membrane phospholipids and to their ability to alter mitochondrial ribosomes function[38–40]. To address this issue, several strategies to reduce toxicity have been tested, such as co-administration with antioxidants or liposome encapsulation[41,42], together with the rational development of aminoglycoside derivatives with enhanced nonsense suppression efficiency and reduced toxicity[43].

Extensive screening of small molecular weight compounds libraries has also led to the identification of non-aminoglycosidic molecules with readthrough-induction activity. One paradigmatic example is PTC124 (also known as ataluren or Translarna®), an orally bioavailable oxadiazole compound with PTC-suppression activity and shown to be safe, with minimal off-target side-effects and no antibacterial activity[44]. Although PTC124 mechanism of action has yet to be clarified, encouraging results have been obtained. Different diseases are currently being evaluated in clinical trials, but results are conflicting. Indeed, ataluren has been recently approved by the European Medicines Agency (EMA) for Duchenne muscular dystrophy patients with nonsense mutations[45,46], whereas the lack of conclusive results obtained in pre-clinical and clinical trials led to discontinuation of the development of ataluren for Cystic fibrosis[47].

Altogether, readthrough-inducing compounds have been tested in several models of disease. In particular, lysosomal storage disorders (LSDs) represent ideal candidates for a nonsense suppression therapeutic approach, because of the relatively low threshold of protein rescue necessary for phenotypic amelioration[31]. Among LSDs, Fabry disease has never been extensively challenged with this correction strategy.

2.1.5 Fabry disease

Anderson-Fabry disease (FD, OMIM number 301500) is an X-linked lysosomal storage disorder caused by deficiency of the lysosomal hydrolase α-galactosidase A (AGAL), which causes accumulation of undegraded substrates, mostly globotriaosylceramide (Gb3), in cells and extracellular matrix[48]. The incidence of FD has historically been estimated at one in 40,000-117,000[49], but clinical variability of different mutations, as well as variable disease severity and symptoms onset, make the disease notoriously difficult to be diagnosed. Indeed, more recent newborn screenings performed in several countries have shown a prevalence ranging from 1:1368 to 1:8882[50–52].

FD can develop into a severe classic phenotype, most often observed in patients with null or very low residual enzyme activity (1-3%), or in a generally milder nonclassical phenotype. Classic FD usually present in childhood and characteristic symptoms include neuropathic pain, hypohydrosis, angiokeratomas as well as learning and growth delay. Subsequently, ocular involvement and autonomic dysfunction appear, followed by kidney and heart failure and cerebrovascular diseases in the third decade of life. Nonclassical FD, also referred to as late-onset or atypical FD, is characterised by a more variable disease course, in which patients are generally less severely affected and disease manifestations may be limited to a single organ[48,53]. Moreover, despite the X–linked inheritance pattern of FD, most heterozygous females result clinically affected, even though the phenotype variability is higher than in males and symptoms usually occur later in life and progress more gradually[54,55]. FD diagnosis can be made by AGAL activity screening in men, whereas only *GLA* gene sequencing is reliable for confirming the diagnosis in women[48].

FD is caused by mutations in the *GLA* gene (chromosome Xq22.1), which consists of a 1290-bp coding sequence divided in 7 exons[56]. To date, over 800 different mutations have been described, the majority of which (75%) being point mutations followed by short-length rearrangements. Notably, around 10% of FD-causing variants reported to date are nonsense mutations (http://www.hgmd.cf.ac.uk/ac/index.php; http://fabry-database.org/; accessed on 16 January 2020). There is not an obvious mutation "hot spot" and most of the patients/families with FD carry unique *GLA* mutations that occurred independently[57].

The *GLA* gene encodes a polypeptide of 429 amino acids, with the first 31 residues representing a signal sequence. In particular, AGAL (EC 3.2.1.22) is a homodimeric glycoprotein with each monomer composed of two domains, a N-terminal $(\beta/\alpha)_8$ domain (amino acids 32-330), containing the active site, and a C-terminal domain (amino acids 331-429), containing eight antiparallel β-strands on two sheets in a β-sandwich. Each monomer of AGAL contains five disulphide bonds (C52–C94, C56–C63, C142–C172, C202–C223, and C378–C382) and three N-linked carbohydrate sites, two of which (N192 and N215, containing mannose 6-phospate) are responsible for lysosomal sorting through the mannose 6-phospate receptor (M6PR) pathway and for the re-captation of secreted AGAL through endocytosis[58]. The active site, which operates through a double displacement mechanism, is formed by two catalytic residues (D170 and D231) and by the side chains of other twelve residues (W47, D92, D93, Y134, C142, K168, E203, L206, Y207, R227, D266, and M267). In the dimer, thirty residues from each monomer contribute to the interface and, to date, there is no evidence for cooperativity between

the two active sites. Consistently with its lysosomal localization, AGAL operates most efficiently at low pH[59].

Deficiency of this enzyme leads to progressive storage of undegraded glycosphingolipids (mainly Gb3) in many tissues and cell types, thus resulting in cellular dysfunction, inflammation and/or fibrosis. These processes subsequently induce organ dysfunction, with the mechanism of tissue damage that is partly attributed to poor perfusion caused by storage in the vascular endothelium[48].

The gold standard for FD treatment is represented by enzyme replacement therapy (ERT), which involves administration of a recombinant form of the human α-galactosidase A and aim at reducing Gb3 accumulation. Two preparations of the enzyme are currently available: agalsidase α (Replagal®, Shire, authorized in Europe), which is produced in a modified human cell line, and agalsidase β (Fabrazyme®, Genzyme, authorized in Europe and United States), which is produced in Chinese hamster ovary cells. Endocytosis of recombinant AGAL involves a combination of multiple receptors and, besides the M6PR and mannose receptor, megalin and sortilin have also been proposed[60]. Studies with both recombinant enzymes have reported decreased cardiac mass, decreased frequency of pain crisis and clearance of Gb3 storage in skin and kidneys. In particular, the greatest benefits are observed when treatment is started at an early stage of the disease, before irreversible tissue damage takes place. Nevertheless, ERT does not completely resolve all symptoms and additional treatments are often needed. Moreover, antibodies against enzyme preparations are frequently found in serum of treated patients, but this issue has not been fully addressed to date and it is not clear if neutralising antibodies have an impact on clinical outcomes[50]. Another major drawback of ERT is the limited bioavailability of intravenously injected recombinant enzymes. Indeed, recombinant enzymes are large molecules that do not freely diffuse across membranes and are thus unable to reach therapeutic concentrations in some target tissues, particularly the brain[61].

For this reason, novel therapeutic approaches have been recently developed, namely pharmacological chaperones, substrate reduction therapy and gene therapy[50]. A recently commercialized compound is migalastat (Galafold®, Amicus Therapeutics), a small-molecule chaperone which facilitates enzyme trafficking to lysosomes by favouring native conformation and stability and thus allowing for correct trafficking. The advantage for patients is oral administration, but migalastat can only be administered to patients with amenable missense mutations, which represent around 30% of FD patients[50,62]. Another compound currently in later stages of clinical trial development is lucerastat. It is a direct inhibitor of glucosylceramide synthase, which catalyses the first step of glycosphingolipid synthesis. Therefore, by direct

inhibition of this enzyme, a lower Gb3 load is achieved[63]. Lastly, the first gene therapy clinical trials are ongoing (https://clinicaltrials.gov/; identifiers numbers: NCT02800070, NCT04040049, NCT04046224, NCT03454893). This approach is extremely promising, although several limitations remain to be addressed, such as vector neutralising antibodies, lack of long-term effects data and ethical issues of administering genome-modifying treatment to children[50,64].

In this context, drug-induced readthrough may represent a further option for the most severe patients with nonsense mutations, currently lacking alternative interventions other than replacement therapy.

2.1.6 Aim of the present work

Treatment for FD still presents many unmet needs, such as the inability of recombinant enzyme to reach the central nervous system, one of the majorly affected organs. Moreover, the only available alternative treatment, i.e. pharmacological chaperones, is not suitable for the relevant portion of severe FD patients harbouring nonsense mutations.

The aim of this work was to investigate for the first time a readthrough-mediated nonsense-suppression approach in the context of Fabry disease, for which a unique attempt on a single *GLA* nonsense variant (p.R227X) has been made to date[65]. In particular, the purpose of this project was to extensively evaluate the responsiveness of 14 rationally selected nonsense mutations to drug-induced readthrough, as well as to characterise the observed output in terms of protein stability and function. Indeed, clinically-relevant accumulation of Gb3 could be prevented by AGAL activity levels around 5-10%[66], an amount compatible with readthrough-mediated rescue.

2.2 Materials and methods

2.2.1 Nomenclature

All residues are reported according to the Human Genome Variation Society (HGVS) nomenclature[67], with numbering starting at the A (+1) nucleotide of the AUG (codon 1) translation initiation codon.

2.2.2 Mutations selection

The rational selection of the panel of nonsense mutations in the *GLA* gene (see Table 2.3) was performed using two online databases (http://www.hgmd.cf.ac.uk/ac/; http://fabry-database.org/) and on the basis of:

i. association with the classic form of FD[48];
ii. predicted responsiveness of sequence contexts to readthrough-inducing drugs[16];
iii. predicted readthrough-mediated production of full-length AGAL with relevant re-insertion of the original amino acid[25,28] and production of the wild-type enzyme.

2.2.3 Creation of expression vectors for recombinant AGAL variants

The human *GLA* cDNA (reference sequence NM_000169.3) was cloned into the pCMV6-XL4 expression plasmid through *Eco*RI-*Sac*II restriction sites. The 5′-CATTAAAAGA CTTACTTGGATCCAAAAAAAAAAAAAAAAC-3′ oligonucleotide was used to delete the natural *GLA* stop codon and to insert the *Bam*HI restriction site (underlined) through site-directed mutagenesis (QuickChange® II XL Site-Directed Mutagenesis Kit, Agilent Technologies, Santa Clara, CA, USA). The Myc-tag sequence, obtained by annealing of the forward 5′ *gatcc*GAGCAGAAACTCATCTCAGAAGAGGATCTGTAA*g* 3′ and reverse 5′ *gatcc*TTACAGATCCTCTTCTGAGATGAGTTTCTGCTC*g* 3′ oligonucleotides (*Bam*HI compatible ends in italic), was inserted downstream of the *GLA* coding sequence through cloning with *Bam*HI to obtain the pCMV6-XL4-GLA-Myc plasmid expressing the carboxy(C)-terminal Myc-tagged recombinant AGAL protein (abbreviated as rGal). Expression vectors for rGal variants were created by site-directed mutagenesis of the human *GLA* cDNA with the oligonucleotides listed in Table 2.2.

Table 2.2. List of the primers used to introduce nonsense and missense variants. The modified nucleotide (bold) and codon (underlined) are reported.

Primer	Sequence (5'→3')
GLA C52X F	GAGCGCTTCATGTG**A**AACCTTGACTGCC
GLA C56X F	GTGCAACCTTGACTG**A**CAGGAAGAGCCAG
GLA Q119X F	CCTCATGGGATTCGC**T**AGCTAGCTAATTATG
GLA W204X F	GTACTCCTGTGAGTG**A**CCTCTTTATATG
GLA W209X F	CCTCTTTATATGTG**A**CCCTTTCAAAAGC
GLA Q321X F	GTAATTGCCATCAAT**T**AGGACCCCTTGGGC
GLA Q330X F	GCAAGCAAGGGTAC**T**AGCTTAGACAGGGAG
GLA Q333X F	GGTACCAGCTTAGA**T**AGGGAGACAACTTTG
GLA W340X F	CTTTGAAGTGTG**A**GAACGACCTCTCTC
GLA R342X F	GAAGTGTGGGAA**T**GACCTCTCTCAGGCTTAG
GLA Q357X F	GCTATGATAAACCGG**T**AGGAGATTGGTGGAC
GLA Q386X F	CCTGCTTCATCACA**T**AGCTCCTCCCTGTG
GLA Y397X F	GAAGCTAGGGTTCTA**G**GAATGGACTTCAAG
GLA Q416X F	GGCACTGTTTTGCTT**T**AGCTAGAAAATACAA
GLA C56C F	GTGCAACCTTGAC**C**GCCAGGAAGAGCCAG
GLA C56W F	GTGCAACCTTGACTG**G**CAGGAAGAGCCAG
GLA Q119K F	CATGGGATTCGC**A**AGCTAGCTAATTATG
GLA Q119Y F	CATGGGATTCGC**T**A**C**CTAGCTAATTATG
GLA W209C F	GCCTCTTTATATGTG**T**CCCTTTCAAAAGCCC
GLA W209R F	GGCCTCTTTATATG**C**GGCCCTTTCAAAAG
GLA Q321K F	GCCATCAAT**A**AGGACCCCTTG
GLA Q321Y F	GCCATCAAT**T**A**C**GACCCCTTGG

2.2.4 Cell culture and transient expression of rGal variants

All experiments were conducted in human embryonic kidney (HEK) 293 cells maintained in DMEM medium (Gibco, Life Technologies, Carlsbad, CA, USA) supplemented with 10% foetal bovine serum (FBS; Gibco, Life Technologies) and 1% penicillin/streptomycin (Gibco, Life Technologies). Cells were maintained at 37°C under 5% CO_2 controlled atmosphere.

The day before transfection, $2.5x10^5$ cells/well were seeded in 12-well culture plates. Transfection was performed in serum-free medium (Opti-MEM, Gibco, Life Technologies) with the Lipofectamine 2000 reagent (Life Technologies), according to the manufacturer's protocol. Briefly, a mixture of plasmid DNA (2 μg) and Lipofectamine 2000 (2 μL) in 150 μL Opti-MEM medium was incubated at room temperature for 20 minutes and then added to cells. For co-transfection studies, cells were transfected with a mixture of 1 μg of wild-type plasmid and 1 μg of plasmid either empty or encoding missense variants. Wild-type alone (2 μg) was used as reference. Transfection medium was removed 4–6 hours after transfection and 1 mL of

fresh DMEM was added, with or without an optimized concentration (100 μg/mL) of geneticin (G418, Sigma-Aldrich, St. Louis, MO, USA)[68], depending on the experimental set-up.
Cell lysates were prepared 48 hours post-transfection as previously described[69]. Briefly, cell pellets obtained from 12-well cell culture plates were homogenised in 200 μL of water added with protease inhibitors (Halt™ Protease Inhibitor Cocktail, ThermoFisher Scientific, Waltham, MA, USA) and subjected to five freeze-thaw cycles in dry ice. The supernatant was collected after centrifugation of the homogenate at 16000 x g for 10 minutes at 4°C and stored at -20°C.

2.2.5 Evaluation of rGal protein levels

The total protein amount for each lysate was measured by BCA (Pierce BCA Protein Assay Kit; ThermoFisher Scientific), according to manufacturer's protocol.
Cell lysates, normalized for total protein amount, were separated by SDS-PAGE on Bolt 4–12% Bis-Tris Plus gels with Bolt MES SDS Running Buffer (Invitrogen, ThermoFisher Scientific) and transferred onto 0,45 μm nitrocellulose membrane. Blocking of membranes was carried out by incubation with 5% milk (w/v) in PBS buffer for 2 hours at room temperature (RT). Full-length rGal was detected by polyclonal goat anti-Myc (1:5000 in PBS-2.5% milk, incubation over-night at 4°C) and donkey anti-goat HRP-conjugated (1:8000, incubation at RT for 45 minutes) antibodies (Bethyl Laboratories, Montgomery, TX, USA). Blotting images were acquired and analysed through the Image Laboratory Software version 4.0 (Bio-Rad, Hercules, CA, USA).
Analysis of rGal dimers was performed through native PAGE and Western blot. Briefly, cell lysates were diluted in either native or SDS-containing sample buffer and separated by native PAGE on a home-made 10% PAA gel using a Tris-glycine buffer pH 8.3 as running buffer. Before transfer, gels were pre-incubated with 0.1% SDS for 15 minutes with gentle shaking. Transfer onto 0,45 μm nitrocellulose membrane was performed in ice using Tris-glycine buffer pH 8.3. Wester blotting was carried out as described above.
A double ELISA was performed with coated polyclonal goat anti-Myc antibodies (1:500, A190-104A, Bethyl Laboratories, Montgomery, Texas) followed by detection with polyclonal rabbit anti-AGAL (1:500, SAB1410536, Sigma-Aldrich) and goat anti-rabbit HRP-conjugated (1:2000, DAKO, Agilent Technologies) antibodies. Serial dilutions of cell lysates containing wild-type rGal were used as reference.

2.2.6 Evaluation of rGal activity with fluorogenic functional assays

Activity levels were evaluated according to a previously described protocol[69]. Briefly, 20 µL of cell lysates at a concentration of 50 µg/mL were incubated for 30 minutes at 37°C with 40 µL of 4-Methylumbelliferyl α-D-galactopyranoside (1 mM, Sigma-Aldrich) in 0.05 M phosphate citrate buffer (pH 5). Reactions were terminated by the addition of 200 µL of 1.0 M glycine buffer (pH 10.5). The released 4-Methylumbelliferone was determined by fluorescence measurement at 360 and 465 nm as the excitation and emission wavelengths, respectively, on a microplate fluorometer (TECAN, Salzburg, Austria). Cell lysates from HEK293 cells expressing wild-type rGal were used as reference.

The endogenous AGAL activity measured in lysates from cells transfected with the empty vector was subtracted from the activity measured in lysates from cells expressing mutant or wild-type rGal.

Specific activity was calculated as the ratio between activity and protein levels.

2.2.7 Lysosomal localization studies by immunofluorescence

Immunofluorescence studies on cells expressing the rGal variants were performed following a previously described protocol[70]. Briefly, transiently transfected cells, seeded on 12 mm glass coverslips, were fixed with 4% paraformaldehyde and permeabilized with 0.1% Triton X-100. Non-specific binding sites were blocked with PBS-5% milk (w/v)-0.1% Triton X-100 (Blocking buffer). Polyclonal goat anti-Myc (Bethyl Laboratories) and monoclonal mouse anti-lysosomal-associated membrane protein 1 (LAMP-1, clone eBioH4A3, ThermoFisher Scientific) antibodies were diluted in Blocking buffer (1:200 and 1:100, respectively) and incubated overnight at 4°C. Detection was performed by Alexa Fluor-conjugated donkey anti-goat (594 nm; Myc tag) and donkey anti-mouse (647 nm; LAMP-1) antibodies (ThermoFisher Scientific) incubated for 1 hour at RT. Images were acquired on the Zeiss LSM510 confocal microscope.

2.2.8 Statistical analysis

Data were analysed with GraphPad Prism 5 software (San Diego, CA, USA) and statistical differences were analysed by *t*-test.

2.3 Results and discussion

A total of 14 *GLA* nonsense mutations were rationally selected according to the features described above (Tab. 2.3), covering around 20% of all *GLA* nonsense mutations reported to date (http://www.hgmd.cf.ac.uk/ac/; http://fabry-database.org/; accessed on 16 January 2020). Two types of PTCs were included, namely those predicted to result in a high degree of suppression (TGA), as well as those predicted to display intermediate frequency of readthrough (TAG). In addition, all these contexts predict the insertion of the original residue.

Table 2.3. FD-associated nonsense mutations selected for the study.

Protein Variant	Nucleotide change[a]	Exon	Patients (n)	Sequence context
Cys52X	c.156C>A	1	2	TGA A
Cys56X	c.168C>A	1	2	TGA C
Gln119X	c.355C>T	2	5	TAG C
Trp204X	c.612G>A	4	8	TGA C
Trp209X	c.627G>A	4	2	TGA C
Gln321X	c.961C>T	6	3	TAG G
Gln330X	c.988C>T	6	3	TAG C
Gln333X	c.997C>T	6	2	TAG G
Trp340X	c.1020G>A	7	4	TGA G
Arg342X	c.1024C>T	7	13	TGA C
Gln357X	c.1069C>T	7	5	TAG G
Gln386X	c.1156C>T	7	2	TAG C
Tyr397X	c.1191T>G	7	1	TAG G
Gln416X	c.1246C>T	7	2	TAG C

[a]NCBI Reference Sequence: NM_000169.3.

All nonsense variants were introduced in the pCMV6-XL4-GLA-Myc plasmid and readthrough was evaluated both in basal and treated conditions. G418 (geneticin) was selected as the most potent readthrough-inducing aminoglycoside[16]. The presence of the Myc-tag at the C-terminal end of recombinant AGAL (rGal) allowed for the sole detection of recombinant full-length proteins in antibodies-based assays, excluding the confounding effect of both truncated forms of the recombinant enzyme and of the endogenous AGAL expressed by cells. A preliminary evaluation of the recombinant construct showed that the Myc-tag was compatible with AGAL enzymatic activity, as already reported by others[71].

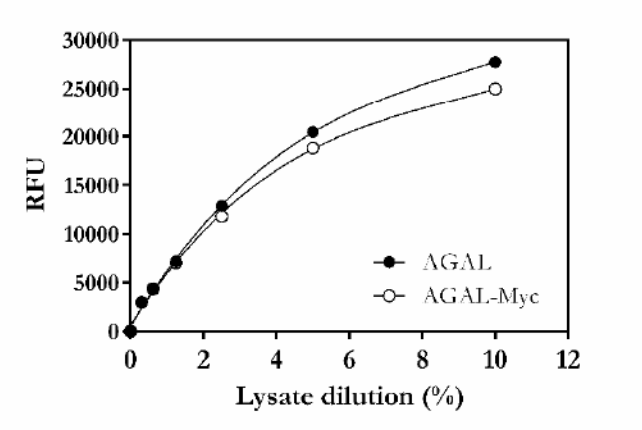

Figure 2.5. Functional evaluation of serial dilutions of recombinant untagged and Myc-tagged AGAL. RFU, relative fluorescence units.

2.3.1 Readthrough induction resulted in detectable full-length rGal proteins

Preliminary evaluation of the responsiveness of all nonsense variants to readthrough was performed by Western blotting analysis. In basal conditions (i.e. transfected cells not treated with G418) no bands corresponding to full-length rGal were detected (data not shown), an observation compatible with the very low levels of spontaneous readthrough expected. Conversely, as shown in Figure 2.6, all treated nonsense variants resulted in detectable full-length rGal and the differential intensity of bands suggested a variable degree of PTC suppression.

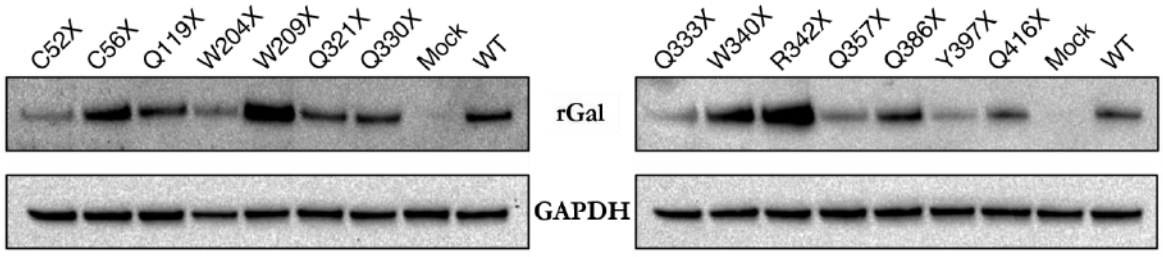

Figure 2.6. Representative western blots on full-length rGal variants resulting from treatment of transiently transfected HEK293 cells with G418. For rGal detection, wild-type lysate was further diluted (1:20) and used as reference. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as control. Mock, cells transfected with the empty vector; WT, wild-type rGal.

The amount of full-length rGal produced by readthrough was subsequently quantified by an optimized double ELISA (Fig. 2.7). This assay took advantage of an anti-Myc capture antibody and of an anti-AGAL detection antibody, thus providing an increased specificity for the recombinant full-length rGal proteins potentially arising from readthrough.

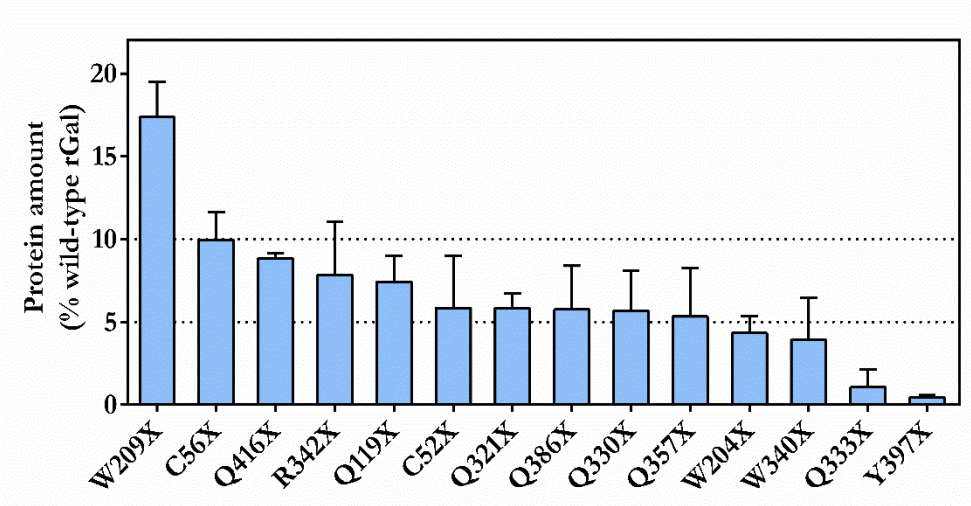

Figure 2.7. Evaluation of drug-induced readthrough over *GLA* nonsense mutations by ELISA. Results, indicated as % of wild-type rGal, are reported as mean ± standard deviation (SD) from three independent experiments. The full-length rGal arising from nonsense variants was not observed in untreated conditions.

Full-length proteins were undetectable in lysates from untreated cells, as well as in lysates from cells transfected with the empty vector (data not shown). In contrast, Figure 2.7 highlights the broad variability of responsiveness to G418-induced readthrough and allows to identify three groups of variants: the high-responders (protein levels ≥10% of wild-type rGal, n=2), the intermediate-responders (5-10%, n=8) and the low-responders (<5%, n=4). Considering the sequence context of the mutants tested, it appears that TGA PTCs are equally distributed among the three groups and that, notably, the only two variants in the high-responders group (C56X, 10.0±1.7% of wild-type rGal; W209X, 17.4±2.1%) harbour a TGA-C PTC, which has been identified as the sequence context most favourable to readthrough[16]. In comparison, variants harbouring a TAG were mainly in the intermediate-responders group (n=6), with two mutants (Q333X and Y397X) that, showing protein levels ≤1% of wild-type rGal, could be classified as very low-responders. These results are in agreement with previous knowledge on the influence of the sequence context on readthrough frequency. Moreover, they highlight the presence of other determinants in addition to the tetranucleotide stop signal, such as more distant sequence determinants[72] and the stability of the missense variants arising from readthrough. Overall, these first data showed that all nonsense variants tested in our experimental system were susceptible to G418-induced readthrough, albeit with variable efficiency, and identified 10 out of 14 variants that reached the arbitrary full-length protein threshold of 5%[66].

2.3.2 Activity of rGal variants was partially restored upon drug-induced readthrough

In the context of Fabry disease, the major goal of a correction approach such as drug-induced readthrough is represented by the rescue of enzyme activity, rather than responsiveness at the protein level. Indeed, recovery of full-length protein does not necessarily correspond to a rescue in terms of enzymatic activity, due to the readthrough-mediated synthesis of wild-type as well as missense variants with unknown features[10]. To address this issue and evaluate the functional rescue mediated by readthrough, activity assays were performed (Fig. 2.8, upper panel). Since no protein levels were detectable in the absence of induction, the functional impact of readthrough was assessed only on cell lysates collected after treatment of cells with the readthrough-inducing agent G418.

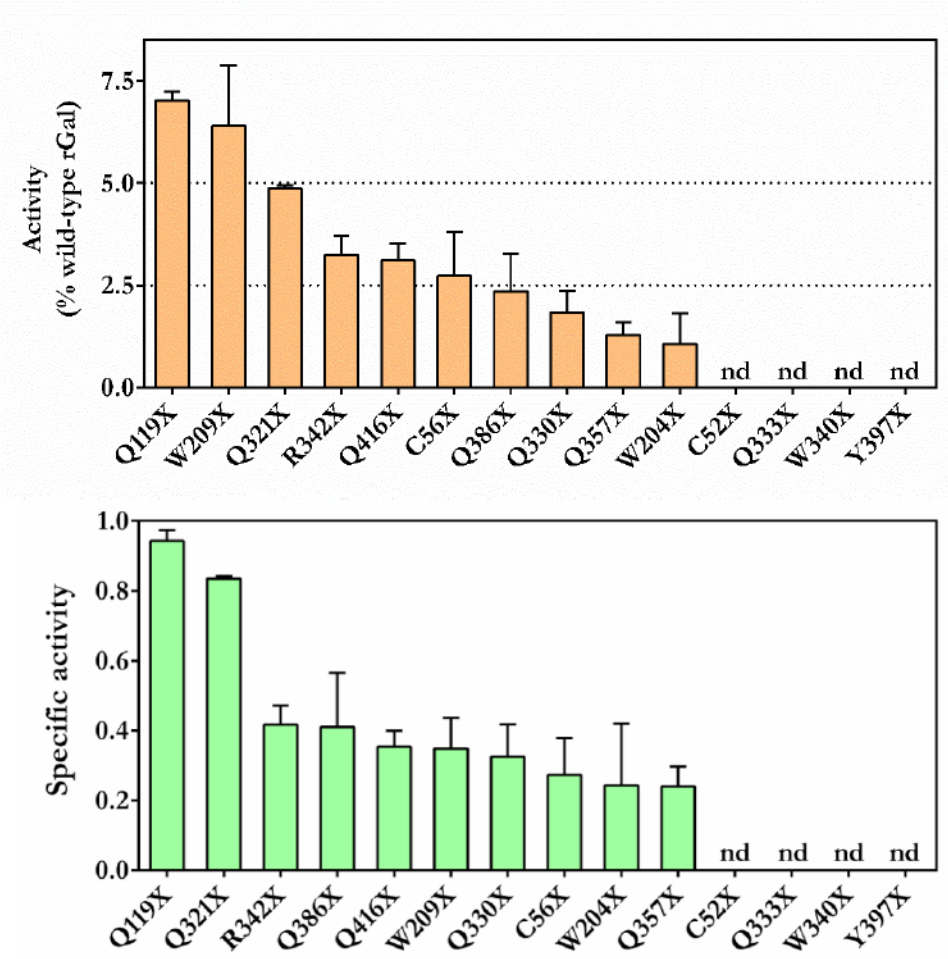

Figure 2.8. Functional evaluation of rGal variants after drug-induced translational readthrough. Activity of full-length rGal variants, produced by readthrough induction (upper panel). Specific activity, referred as the ratio between enzyme activity and protein amount (lower panel). Results, indicated as % of wild-type rGal for the activity levels, are reported as mean ± SD from three independent experiments. nd, not detectable.

As expected, evaluation of the rGal functional levels resulted in a generally lower response compared to the full-length protein levels. As shown in the upper panel of Figure 2.8, only three variants reached the relevant activity threshold of 5% (Q119X, 7.0±0.2% of wild-type rGal; W209X, 6.4±1.5%; Q321X 4.9±0.1%), whereas seven showed levels between 2-4% and activity levels were undetectable for the remaining variants. Indeed, the functional output of readthrough is not exclusively dependent on its occurrence, but rather on the identity of the amino acid inserted at the PTC.

The same considerations are further highlighted by evaluating the resulting specific activity, referred as the ratio between activity and protein levels. The lower panel of Figure 2.8 clearly shows that for only two variants (Q119X, 0.94±0.03; Q321X, 0.84±0.01) readthrough induction resulted in a specific activity similar to that of the wild-type rGal, whereas the majority of variants showed strongly reduced specific activity between 0.24 and 0.42. No specific activity was calculable for those nonsense variants with undetectable enzyme activity. It is worth to note that specific activities near 1 may derive from the production of stable and functional missense variants or from the main reinsertion of the wild-type residue thus producing mostly wild-type molecules.

These results can be interpreted in light of the different features of each residue reported in Table 2.4, namely: i) its role in AGAL structure and function, ii) evolutionary conservation, and iii) predicted impact of missense changes at a specific position. In this context, natural missense variants associated to FD are extremely informative upon the impact of amino acid changes on AGAL protein. In addition, bioinformatic tools such as SIFT (https://sift.bii.a-star.edu.sg/)[73] and PolyPhen (http://genetics.bwh.harvard.edu/pph2/)[74] may help predicting the effects of missense changes introduced by readthrough, although experimental validation is often desirable.

2.3.3 Amino acid substitutions predicted from readthrough showed differential impact on rGal protein and activity

As shown in Table 2.4, some positions are predicted to be more tolerant than others in regards of amino acid changes introduced by readthrough. For example, C52 residue is involved in formation of a disulphide bond crucial for AGAL structure and function and is completely conserved among different species. It is therefore unlikely that missense changes at this position would be tolerated in terms of protein stability and activity. Indeed, naturally occurring missense variants at this position are associated with classic FD, including two reported variants corresponding to the predicted changes introduced by readthrough (C56R and C56W). On the

Table 2.4. Specific features of the examined AGAL residues.

Original residue	Role in AGAL[a]	Evolutionary conservation[b]	Natural missense variants[c]		Missense variants arising from readthrough[d]		
			Amino acid change	Phenotype	Amino acid change	SIFT[e]	PolyPhen[f]
C52	disulphide bridge	total	C52R/G/S/Y/W	classic FD	C52R/W	damaging	probably damag.
C56	disulphide bridge	total	C56G/F/Y	classic FD	C56R/W	damaging	probably damag.
Q119	unknown	partial	none	--	Q119K/Y	tolerated	benign
W204	near active site E203	total	W204C	FD	W204C/R	damaging	probably damag.
W209	unknown	partial	none	--	W209C/R	tolerated	probably damag.(C) benign (R)
Q321	N^{E2} of Gln needed in H-bonding network	total	Q321E/R/L/H	FD	Q321K/Y	damaging	probably damag.
Q330	unknown	partial	none	--	Q330K/Y	tolerated (K) damaging (Y)	benign
Q333	unknown	partial	none	--	Q333K/Y	tolerated (K) damaging (Y)	benign (K) possibly damag. (Y)
W340	buried in hydrophobic pocket	total	W340R/S	classic FD	W340C/R	damaging	probably damag.
R342	totally buried	total	R342Q/L/P	FD	R342C/W	damaging	probably damag.
Q357	unknown	total	Q357A	FD	Q357K/Y	tolerated (K) damaging (Y)	benign
Q386	unknown	partial	Q386P	classic FD	Q386K/Y	tolerated	possibly damag. (K) probably damag. (Y)
Y397	unknown	total	none	--	Y397K/Q	tolerated	possibly damag. (K) benign (Q)
Q416	unknown	partial	Q416P	FD	Q416K/Y	tolerated	benign (K) possibly damag. (Y)

[a] Based on Ref.[59]
[b] Sequence alignment among human (NP_000160.1), chimpanzee (XP_003954083.2), rhesus macaque (XP_001093625.1), pig (NP_001171396.1), bovine (NP_001179665.1) and mouse (NP_038491.2) AGAL proteins.
[c] Reported in http://fabry-database.org/ and Ref.[59] and Ref.[69]
[d] Prediction based on Ref.[25]
[e] https://sift.bii.a-star.edu.sg/
[f] http://genetics.bwh.harvard.edu/pph2/index.shtml

contrary, Q119 residue is less conserved among different species and no FD-associated variants have been reported to date at this position. This suggests that position 119 may be more tolerant in terms of amino acid changes, as also predicted by SIFT and PolyPhen algorithms.

Nevertheless, to provide experimental evidence for the mechanism at the basis of PTC suppression output, missense changes predicted to arise from readthrough were characterised (Fig. 2.9). In particular, prediction of readthrough-deriving amino acid insertions, based on the type of PTC[10,25], prompted the expression and characterisation of the following variants as paradigmatic examples:

- C56R and C56W: the position is predicted to be not tolerant against amino acid changes and a reduced specific activity was observed upon readthrough induction;
- Q321K and Q321Y: the position is predicted to be not tolerant but specific activity was near 1;
- Q119K and Q119Y: the position is predicted to be highly tolerant and the observed specific activity upon readthrough was near 1;
- W209C and W209R: the position is predicted to be tolerant to amino acid changes introduced by readthrough but a reduced specific activity was observed.

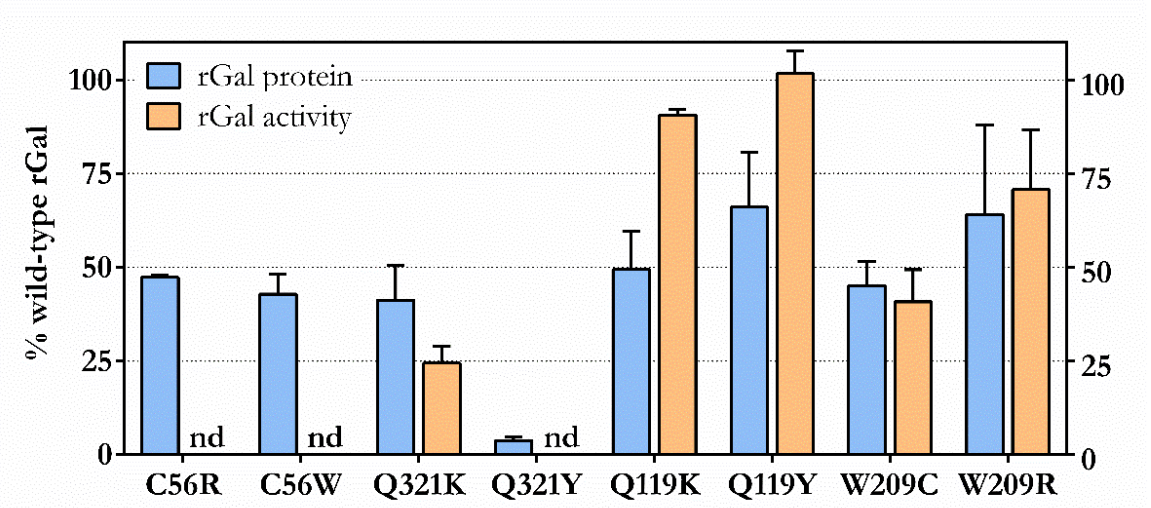

Figure 2.9. Protein amount (blue bars) and activity levels (orange bars) of rGal missense variants predicted to arise from readthrough of *GLA* PTCs. Results, indicated as % of wild-type rGal, are reported as mean ± SD from three independent experiments. nd, not detectable.

Amino acid substitutions predicted from readthrough showed differential impact on rGal protein and activity levels (Fig. 2.9).

Characterisation of missense variants at positions 56 and 321 revealed a detrimental effect of all amino acid changes introduced. In particular, protein levels <50% of wild-type (C56R 47.3±0.6% of wild-type rGal; C56W 42.9±5.3%; Q321K 41.2±9.2%; Q321Y 3.7±1.0%), as well

as not detectable activity levels for all variants except Q321K (24.4±4.5%) indicated reduced AGAL stability and activity due to the inserted amino acid changes, confirming the hypothesis of susceptibility based on positions features (Tab. 2.4). These observations suggested that readthrough over C56X leads to the synthesis of a mixture of wild-type and missense variants in which all components contribute to the protein levels but only the wild-type contributes to the activity levels, thus resulting in a low specific activity. Conversely, it can be hypothesised that the almost normal specific activity observed after readthrough over the Q321X variant was due to the prevalent re-introduction of the wild-type glutamine residue[10] rather than to the production of functional missense variants.

In contrast, characterisation of missense variants Q119K/Y and W209C/R confirmed the permissive nature of these positions. Indeed, all variants showed reduced protein levels compared to wild-type rGal (Q119K 49.4±10.3% of wild-type rGal; Q119Y 66.1±14.6%; W209C 45.1±6.4%; W209R 64.1±23.9%), but clearly detectable activity levels (Q119K 90.6±1.5% of wild-type rGal; Q119Y 101.7±6.1%; W209C 40.8±8.6%; W209R 70.8±15.9%), comparable to their respective protein amount. This resulted in specific activities similar to that of wild-type, indicating that missense changes at positions 119 and 209 had a minor impact on AGAL stability and did not impair its enzymatic function. These results suggested that readthrough over Q119X nonsense mutation leads to the synthesis of a combination of wild-type and functional missense variants, all contributing to the protein and activity levels observed. On the other hand, the reduced specific activity observed after G418 treatment of the W209X nonsense variant was not explained by the characterisation of missense variants arising from readthrough, since both substitutions were compatible with enzyme function.

Overall, expression of missense variants bearing the predicted amino acid substitutions revealed a differential impact on rGal protein and activity and provided valuable experimental evidence to interpret the results obtained from readthrough induction over *GLA* nonsense mutations.

2.3.4 The rGal W209C variant showed correct localization into lysosomes

Besides affecting protein stability and activity, missense changes could also impair AGAL trafficking to the lysosomal compartment, which is essential for the formation of a functional homodimer[71,75]. To assess if amino acid changes at position 209 affected rGal sorting to lysosomes, immunofluorescence studies were performed (Fig. 2.10). In particular, W209C variant was selected as model for both missense variants at position 209 and C56R variant was chosen as comparator because it had shown protein levels similar to the W209C variant (see Fig. 2.9).

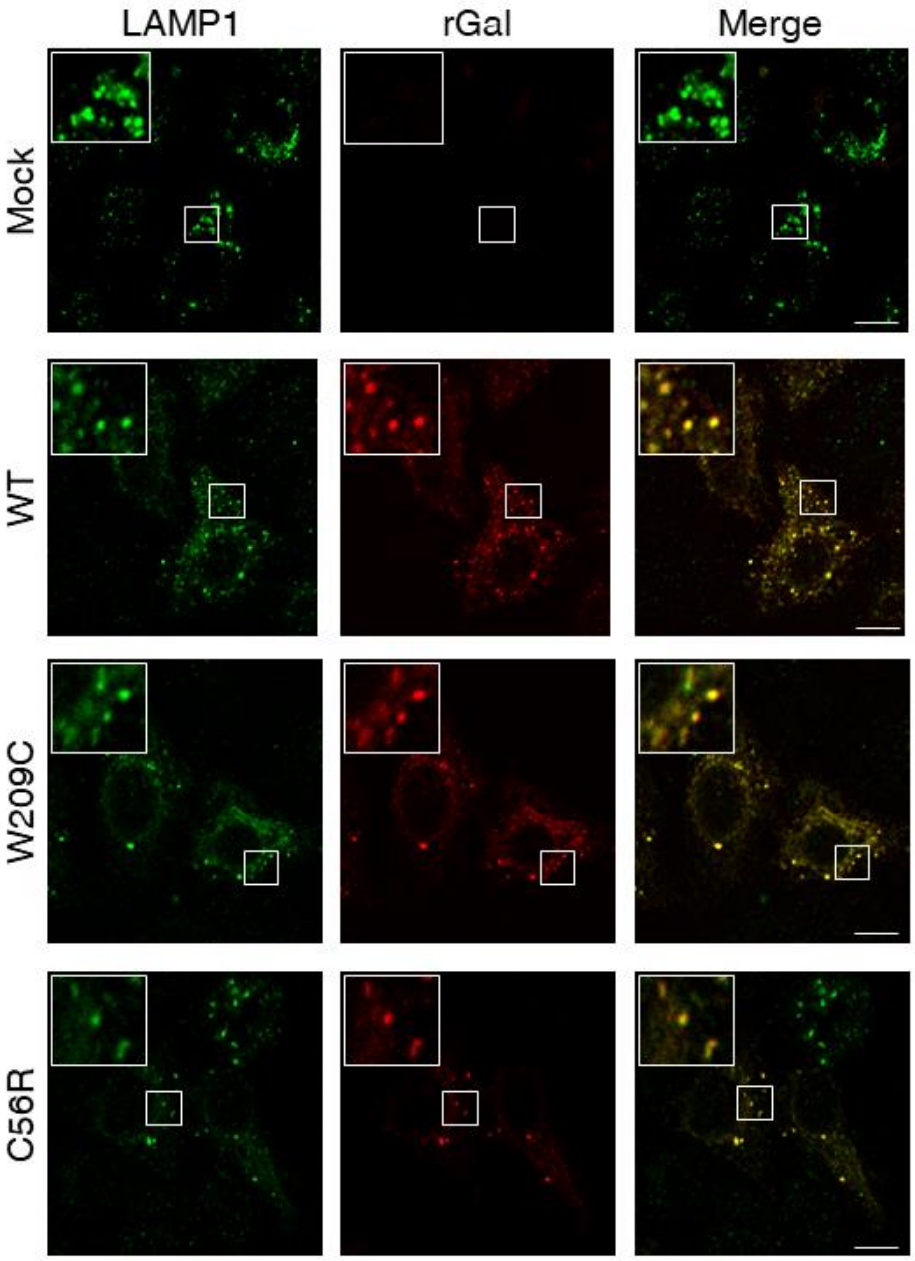

Figure 2.10. Co-localization studies with W209C and C56R variants. Fluorescently labelled secondary antibodies were used to visualize LAMP-1 (lysosomes, green) or rGal (red) as well as their co-localization (yellow). Image magnifications (white squares) for each channel are shown. Cells transfected with the empty vector (Mock) or expressing wild-type rGal (WT) were used as controls.

Both missense variants, clearly detectable by anti-rGal antibodies (Fig.2.10, red signal), were appreciably visualized into lysosomes (Fig. 2.10, green signal), as indicated by co-localization of rGal and lysosomal LAMP1 fluorescence signals, which were qualitatively comparable to that of wild-type rGal. These results indicated that the amino acid change at position 209 did not impair rGal trafficking to lysosomes and that the low specific activity observed for W209X after induction was not due to defective sorting.

2.3.5 Missense variants arising from readthrough exerted potential dominant-negative effects

The discrepancy between protein and activity levels observed for the W209X nonsense variant was not accountable for the synthesis of dysfunctional missense variants nor for a defective trafficking to lysosomes. A possible interfering effect of truncated rGal was excluded by co-expression experiments with wild-type and W209X variant in the absence of readthrough induction (Fig.2.11).

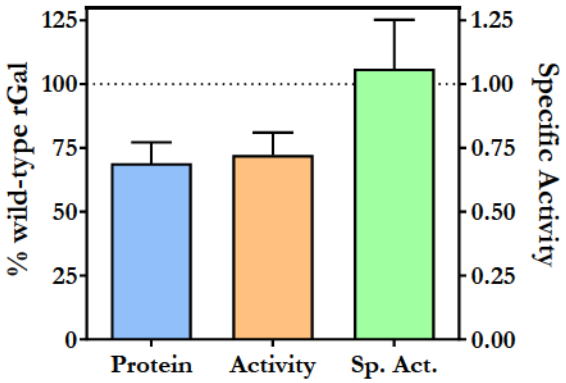

Figure 2.11. Co-expression studies of wild-type rGal with the W209X variant. Results, indicated as % of wild-type rGal, are reported as mean ± SD from five independent experiments. The dotted line indicates the specific activity of wild-type rGal expressed alone. Sp. Act., specific activity.

It was therefore hypothesized that the interaction of wild-type rGal molecules with missense monomeric variants arising from readthrough might result in the formation of dysfunctional heterodimers. To test this hypothesis, co-expression studies of wild-type and W209C/R variants were performed (Fig. 2.12).

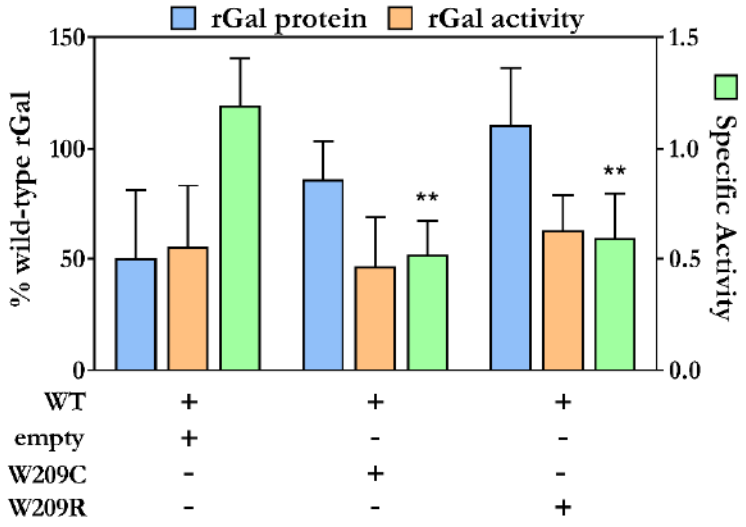

Figure 2.12. Co-expression studies of wild-type rGal with W209C/R variants. The empty vector was used to normalize the amount of plasmid during transfection. Results, indicated as % of wild-type rGal, are reported as mean ± SD from three independent experiments. **, $p < 0.05$.

Co-expression of wild-type rGal with W209C or W209R variants (Fig.2.12) resulted in increased protein levels (W209C, 86.2±17.3% of wild-type rGal; W209R, 110.7±25.5%) if compared to those obtained with wild-type rGal alone (50.3±30.8%), an observation compatible with an additive effect deriving by the synthesis of both wild-type and missense rGal variants. In contrast, activity levels (W209C, 46.9±22%; W209R, 63.0±16%) were similar to those of wild-type alone (55.6 ± 27.5%), thus resulting in a significant decrease of the specific activity of wild-type rGal only when co-expressed with W209C and W209R (W209C 0.5±0.2, p=0.0027; W209R 0.6±0.2, p=0.0061), despite both variants had shown normal specific activity when expressed alone.

These results suggested that amino acid substitutions introduced by readthrough at this position produced rGal monomers unable to form functional dimers with the wild-type, thus exerting potential dominant-negative effects on specific activity and lowering the functional rescue mediated by readthrough. This observation can be interpreted in light of the symmetric nature of the AGAL homodimer (PDB 1R46, Fig. 2.13, left panel), in which even a small distortion caused by an amino acid substitution in one of the two monomers might lead to a heterodimer with reduced catalytic efficiency (Fig. 2.13, right panel).

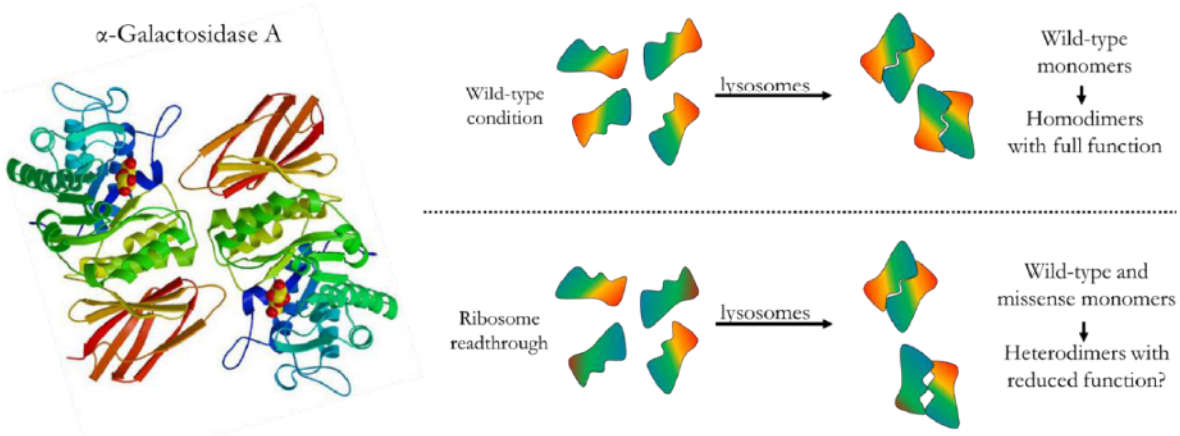

Figure 2.13. AGAL structure (left panel, PDB 1R46 from Ref.[76]) and schematic representation of the impact of readthrough-deriving amino acid insertions on AGAL biology (right panel). Once transported into lysosomes, AGAL monomers dimerize to form functional homodimers (wild-type condition) or dysfunctional heterodimers due to the presence of wild-type as well as missense monomers arising from ribosome readthrough.

To provide experimental evidence on rGal dimerization, native PAGE analysis was performed (Fig. 2.14). Lysates deriving from cells expressing wild-type rGal alone or co-expressed with W209C/R variants were evaluated with or without the presence of the strong denaturant agent (SDS).

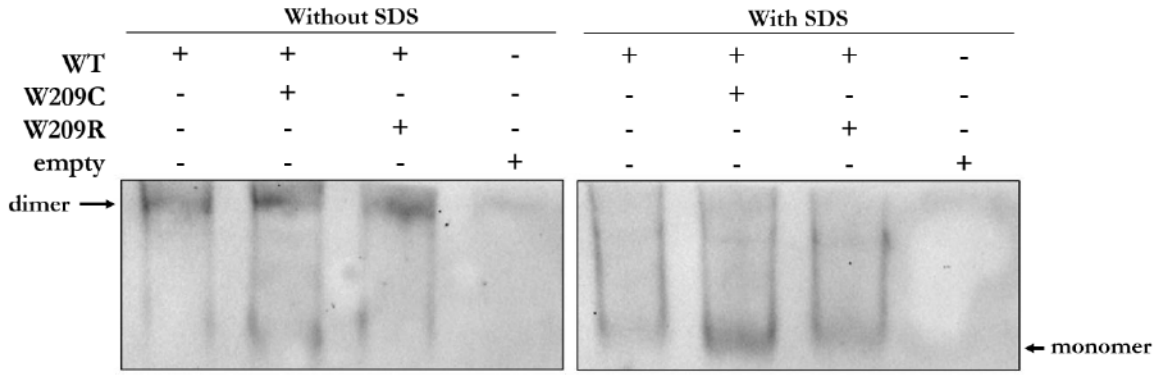

Figure 2.14. Native PAGE and Western blot on wild-type rGal expressed alone or co-expressed with W209C or W209R variants. Samples were prepared in the absence or in the presence of SDS to favour the visualization of rGal dimers or monomers, respectively. Lysates from cells transfected with the empty vector were loaded as controls.

This analysis revealed the presence of two series of bands, i) with lower electrophoretic mobility and compatible with the presence of rGal dimers, or ii) with higher electrophoretic mobility thus compatible with the monomeric rGal forms. Importantly, the differential distribution of rGal was in accordance with the absence (i) or presence (ii) of SDS. These observations provided an experimental proof for the presence in cell lysates of both rGal dimers and monomers. Unfortunately, due to a limit in the experimental set-up, the wild-type and the missense rGal variants are not distinguishable, thus preventing the actual discrimination between homodimers and heterodimers (all detected through the same Myc-tag).

The lack of direct evidence for the role of residue W209 in the AGAL structure makes the explanation for the observed negative effect upon co-expression not obvious. Interestingly, studies on the specific features of homodimers emphasized that isologous association (i.e. the binding of two identical subunits, involving identical binding domains) gives rise to "closed structures" characterised by intrinsic symmetry and stability[77,78]. Therefore, even a small distortion caused by an amino acid substitution in one of the two monomers might lead to a heterodimer with reduced stability or catalytic activity.

Overall, these findings depict a scenario in which, once transported into lysosomes, part of the amino acid changes possibly introduced by readthrough leads to the assembly of wild-type and missense variants, resulting in a fraction of functional homodimers as well as heterodimers with lower or no function (Fig. 2.13, right panel). These dominant-negative effects, already hypothesized by others in the context of AGAL enzyme[79,80], might have implications for the large proportion (60–75%) of heterozygous females presenting with signs and symptoms of FD[54,55]. Interestingly, pathological phenotypes in heterozygous females are not explained by skewed X-inactivation alone[81,82] and the mechanism hypothesized here could provide an additional explanation to interpret these observations.

2.4 Conclusions

This work identified three out of fourteen tested *GLA* nonsense mutations (i.e. Q119X, W209X and Q321X) that, due to favourable nucleotide and protein features, could be rescued by readthrough induction. Moreover, a novel dominant-negative mechanism was suggested, resulting from the interaction of different monomers and formation of dysfunctional heterodimers. To the best of our knowledge, this is the first evidence for translational readthrough on a dimeric enzyme that results in potentially dominant-negative effects related to the heterodimerization process. Hopefully, the results obtained here will foster additional studies on the effects of missense changes on altered AGAL dimerization.

The readthrough-induction approach is extremely relevant in the context of Fabry disease, because of the relative low activity threshold needed for amelioration of Gb3 deposition. In particular, residual AGAL levels around 1-10% of normal have been found to be associated with less severe phenotypes[83] and 5-10% of activity seems to be sufficient to prevent clinically-relevant accumulation of Gb3[66]. Moreover, the ability of readthrough-inducing compounds to reach central nervous system would address one of the main limits of ERT. Indeed, readthrough of nonsense mutations using small molecules is a promising strategy, but the variable outcome of preclinical and clinical studies highlights that several determinants of its efficacy need to be carefully considered.

Overall, the findings of this work further support the notion that functional readthrough is the result of a combination of benign conditions such as i) favourable nucleotide context, ii) re-insertion of the original residue, and iii) insertion of tolerated missense changes. This is critical in the context of the development of therapeutic strategy based on PTC suppression. Moreover, the discovery and development of safe and effective nonsense suppression drugs is needed to improve this therapeutic approach.

Chapter 3

Design of a novel factor IX-albumin fusion protein for the treatment of Haemophilia B

3.1 Background and rationale

3.1.1 Half-life extension technologies

Protein and peptide therapeutics have been proven to be very effective for the treatment of human diseases, and a multitude of products have reached the pharmaceutical market in the past decades[84]. Nevertheless, the therapeutic efficacy of many protein therapeutics is hampered by their short half-life, which is dictated by several factors including size, charge, proteolytic sensitivity and turnover rate of proteins they bind[85]. This has led to the development of several strategies to improve persistence in circulation, which can be divided into six general groups[85,86].

1. Genetic fusion to a naturally long-half-life protein or protein domain (e.g. Fc, transferrin or albumin fusion).
2. Genetic fusion to a recombinant polypeptide, such as XTEN (also known as recombinant PEG), HAP (homoamino acid polymer), PAS (proline-alanine-serine polymer) and ELP (elastin-like peptide).
3. Conjugation to chemical moieties, such as PEG or hyaluronic acid, in order to increase the hydrodynamic radius.
4. Polysialylation or fusion to a highly sialylated peptide (e.g. carboxy-terminal peptide, CTP) in order to increase the negative charge.
5. Non-covalent binding to long-half-life proteins, such as human serum albumin (HSA), immunoglobulin G (IgG) or transferrin.
6. Chemical conjugation to long-half-life proteins, such as human IgGs, Fc moieties or HSA.

Indeed, extended persistence in plasma provides great benefits in terms of high serum concentration of the infused drug, lower dosing frequency, and the relevant advantage of decreasing administered doses without compromising pharmacological efficacy.

3.1.2 Albumin-based half-life extension

In the context of half-life extension technologies for therapeutic proteins, genetic fusion to albumin represents a powerful option. Indeed, HSA is an ideal carrier because of its serum stability and longevity (19-21 days in humans), and genetic fusion offers the favourable opportunity of one-step synthesis, without the need for further *in vitro* processing and thus reducing manufacturing costs.

Albumin is a highly soluble and stable protein which acts as a multi-carrier of small insoluble

and hydrophobic ligands such as fatty acids, metal ions, hormones, waste products, toxins and drugs. Moreover, it is responsible for maintaining the osmotic pressure and regulating blood pH, and it is characterised by antioxidant and enzymatic properties[87–89]. Albumin is the most abundant circulating protein and is synthesized and secreted into the blood stream by hepatocytes to a concentration of 40 mg/mL in both mouse and man[87]. HSA is a non-glycosylated, heart-shaped molecule of 66.5 kDa consisting of a single polypeptide of 585 amino acids with seventeen pairs of disulfide bridges and one free-cysteine at position 34. It is formed by 67% α-helices folded in three homologous domains (DI, DII and DIII), each composed of two subdomains (A and B) connected by flexible loops[90].

The extremely long half-life of albumin, a feature shared with IgGs, is mainly due to its molecular weight above the renal threshold and to a rescue mechanism from intracellular degradation. The latter is mediated by the broadly expressed neonatal Fc receptor (FcRn), which recycles internalized albumin and IgGs back to the blood stream through a strictly pH-dependent mechanism[91] (Fig.3.1). Several other putative albumin receptors have been described, but, with the exception of the cubilin-megalin complex, they are poorly characterised and their role has not been fully addressed yet[92].

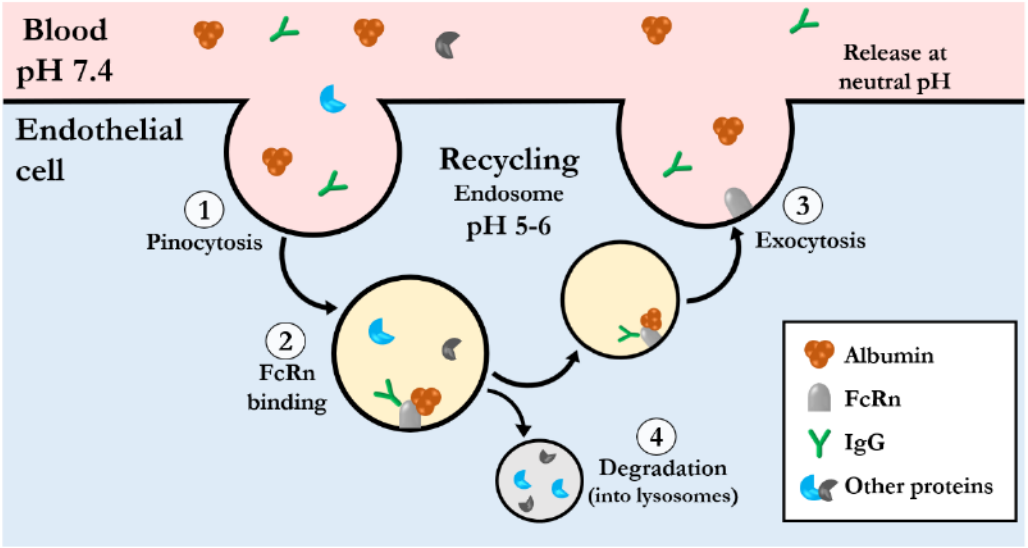

Figure 3.1. Schematic illustration of albumin and IgG rescue mechanism mediated by the pH-dependent FcRn binding. (1) IgG and albumin are taken up from the blood by pinocytosis and (2) FcRn binds them in the acidic endosomes. (3) IgG and albumin are recycled to the cell surface and released due to the increasing pH. (4) Proteins that do not bind the receptor are sorted to lysosomal degradation.

FcRn is broadly expressed in humans and it primarily resides within acidified endosomes. As shown in Figure 3.1, when IgGs and albumin are internalized by pinocytosis and reach the endosomal compartment, the low pH triggers their simultaneous binding to FcRn. The ternary complex is then recycled back to the cell surface, where the increasing pH leads to decrease of

the binding affinity and to the final release of albumin and IgGs outside of the cell. In contrast, internalized molecules that do not bind FcRn are sorted to late endosomes and are lastly degraded inside lysosomes[93–95].

FcRn is a heterodimeric glycoprotein formed by a larger 45 kDa subunit, consisting of a heavy chain (HC) related to the class I major histocompatibility complex, and by a smaller 12 kDa subunit, which is a β2-microglobulin (β2m). Specifically, the HC is formed by three extracellular domains (α1, α2 and α3), followed by a transmembrane part and a cytoplasmic tail, and it is non covalently associated to the soluble β2m subunit[96].

Investigation of FcRn-albumin interface (Fig. 3.2) revealed that the HSA C-terminal DIII domain is the primary binding site for FcRn, with three fully conserved histidine residues (H464, H510 and H535) that are crucial for the pH-dependent interaction[97]. In addition to DIII, the N-terminal DI also contributes to FcRn binding through two exposed loops including positions 80-89 and 104-114[98].

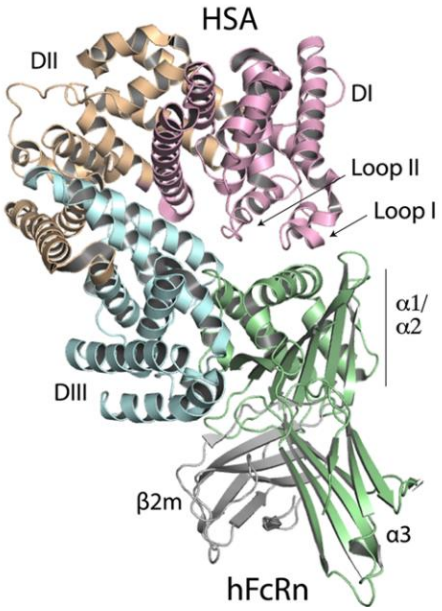

Figure 3.2. Co-crystal structure of albumin (upper part) in complex with FcRn (lower part)[99].

The increasing knowledge of FcRn-albumin binding mechanism has paved the way to the development of engineered albumin variants with improved half-life as a function of FcRn affinity. A major challenge is to improve FcRn binding without disrupting the pH dependence, which is fundamental for efficient release. A remarkable example was inspired by cross-species studies showing that mouse albumin binds human FcRn (hFcRn) more strongly than human albumin[100]. This led to the development of an HSA variant characterised by a single amino acid substitution (K573P) and which displayed a 12-fold improved affinity toward hFcRn with no detectable binding at neutral pH[101]. The improved FcRn affinity resulted in extended half-life,

as showed by pre-clinical studies in mice transgenic for hFcRn (1.4-fold longer half-life) and cynomolgus monkeys (1.6-fold). These findings, demonstrating that albumin half-life can be tailored by modulating FcRn affinity, are extremely relevant in the context of albumin-based therapeutics. Indeed, once in the blood stream, these molecules will have to compete with the abundant endogenous albumin, thus the presence of an engineered HSA variant could compensate this limitation[92,101].

Furthermore, as albumin-based therapeutics are extensively evaluated in animal models, it is crucial to understand the impact of cross-species FcRn-binding differences. Indeed, mouse albumin (MSA) has been shown to bind both human and mouse FcRn more strongly than HSA, an observation partly due to DIII proline at position 573, which is highly conserved in several species with the exception of humans[99,100]. This means that specific mouse models are required, such as mice transgenic for human FcRn. Indeed, the complex interactions between mouse and human albumin with their receptors and across the two species must be carefully considered during development and evaluation of albumin-based therapeutics in pre-clinical mouse models[99].

To date, several therapeutic proteins have been coupled to albumin to improve their pharmacokinetic properties (Tab. 3.1), showing that this approach can be successfully exploited to improve *in vivo* efficacy of a variety of drugs.

Ideal candidates for half-life extension approaches are coagulation factors, which are characterised by short half-lives of approximately 10-12 hours for FVIII, 18-34 hours for FIX and 2.3 hours for activated factor VII (FVIIa)[102]. Notably, the albumin fusion strategy has been already successfully applied to recombinant FVIIa (rFVIIa) and FIX (rFIX). In the latter case, since fusion with albumin resulted in a reduction in biological activity of FIX, a FIX-HSA fusion protein with a proteolytically cleavable linker was developed (rIX-FP, CSL Behring). Once activated, since the linker is cleaved and albumin released, this molecule behaves in the same way as endogenous activated FIX (FIXa). After preclinical studies[103] and clinical trials[104], rIX-FP has entered the market with the commercial name Idelvion® (CSL Behring), providing a 3-to-5.6-fold extended half-life molecule for Haemophilia B replacement therapy.

Table 3.1. Albumin-based therapeutics available on market or under clinical trial (adapted from Ref.[86]).

Therapeutic protein	Half-life extension technology	Product (patent name)	Half-life extension (vs native)	State
Insulin	Non-covalent binding	Levemir® (Insulin detemir)	5-7 h (4-5 min)	In market
GLP-1	Non-covalent binding	GLP-1 agonist Victoza® (Liraglutide)	13 h (2 min)	In market
GLP-1	Fusion to HSA-binding domain antibody	GSK2374697	6-10 days (2 min)	Phase 1
GLP-1	Fusion protein	Eperzan/Tanzeum® (Albiglutide)	6-8 days (2 min)	In market
FIX	Fusion protein	Idelvion® (Albutrepenonacog alfa)	102 h (18-34 h)	In market
FVIIa	Fusion protein	rVIIa-FP (CSL689)	8.5 h (2.3 h)	Phase 1
hGH	Fusion protein	hGH-HSA (TV-1106)	3-35 h (4-5 min)	Phase 1
G-CSF	Fusion protein	Egranli® (Balugrastim)	18-40 h (3.5-3.8 h)	MAA submitted (Europe)
Extendin-4	Covalent attachment	CJC-1134-PC	8 days (< 1 h)	Phase 2

GLP-1, glucagon-like peptide 1; FIX, coagulation factor IX; FVIIa, activated coagulation factor VII; FP, fusion protein; hGH, human growth hormone; G-CSF, granulocyte colony stimulating factor; MAA, marketing authorization application.

3.1.3 Coagulation and Haemophilia B

Blood coagulation is a defence mechanism that, together with inflammatory and repair responses, helps preserve the integrity of the vasculature after an injury. Coagulation involves finely tuned cellular and molecular events that culminate in the formation of a stable fibrin clot[105] (Fig. 3.3). The key step after a vessel wall injury is the exposure of sub-endothelial cell surfaces expressing tissue factor (TF) to plasma proteins, particularly to FVIIa. The TF/FVIIa complex activates small amounts of FIX to FIXa and FX to FXa. The latter is the key protease that, in complex with its cofactor (activated factor V, FVa), is directly involved in the generation of thrombin. The small amount of thrombin generated during the initiation phase boosts its own production through the activation of platelets and feedback reactions known as the amplification and propagation phases, during which a crucial feedback loop activates factor XI (FXIa) that in turn increases the amount of FIXa. The interaction of FIXa with the activated form of its cofactor (FVIIIa) strongly boosts the production of FXa that, in association with FVa, further drives the so-called thrombin burst responsible for the large-scale production of

thrombin necessary to produce a stable clot. Indeed, the great amount of generated thrombin converts fibrinogen to insoluble fibrin and activates the transglutaminase FXIII (FXIIIa), which stabilizes the clot by cross-linking the fibrin chains[106].

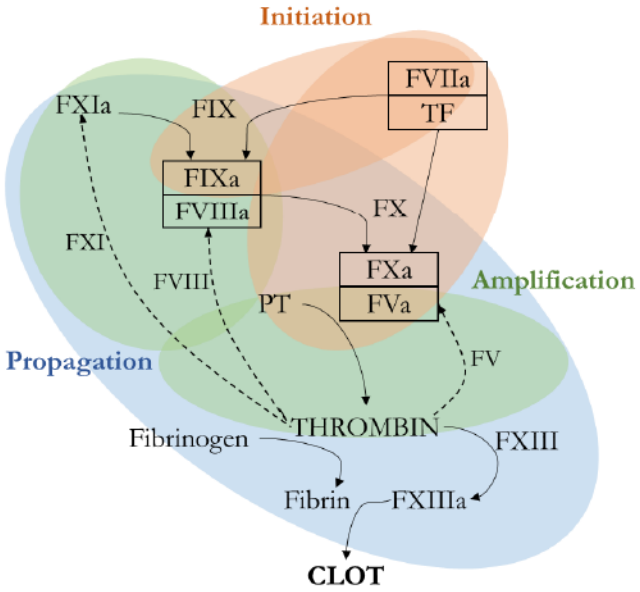

Figure 3.3. Schematic representation of the coagulation cascade. Direct (black rows) and feedback (dotted rows) reactions are indicated and can be subdivided into initiation (orange), amplification (green) and propagation (blue) phases of the coagulation process, ultimately leading to the fibrin clot. Boxed items indicate the interaction of active enzymes (FVIIa, FXa, FIXa) with their cofactors (TF, FVa, FVIIIa). PT, prothrombin. Adapted from Ref.[107].

Haemophilia B (HB; OMIM number 306900), also referred to as Christmas disease after the name of the first patient examined in detail[108], is an X-linked bleeding disorder caused by the deficiency of FIX. The estimated incidence of HB is 1 in 30000 male live births[109] and the bleeding tendency is related to the residual levels of FIX, leading to classification of the disease as severe, moderate or mild[110] (Tab. 3.2).

Table 3.2. Classification of haemophilia (adapted from Ref.[110])

Concentration of factor (% of normal)	Classification	Bleeding phenotype
<1%	Severe	Spontaneous joint and muscle bleeding; bleeding after injuries, accidents and surgery
1-5%	Moderate	Bleeding into joints and muscles after minor injuries; excessive bleeding after surgery and dental extractions
5-40%	Mild	Spontaneous bleeding does not occur; bleeding after surgery, dental extractions and accidents

In both haemophilia B and A (FVIII deficiency), bleeding occurs because of a failure of the coagulation process, particularly a decreased and delayed generation of thrombin, giving rise to defects in clot formation and leading to haemorrhagic diathesis. These defects are associated with bleeding episodes affecting soft tissue, joints, and muscles, and repeated haemorrhages result in chronic arthropathy with loss of joint movement[109].

The clinical manifestations of haemophilia depend both on the severity of clotting factor deficiency and on the age of the patient[111]. The earliest and most serious complication in neonates with severe haemophilia is intracranial haemorrhage, which occurs in 1–4% of cases and can lead to permanent neurological sequelae. In severe forms, spontaneous muscle haemorrhage occurs in the lower legs, buttocks, iliopsoas muscle, and forearms, whereas bleeding after surgery is an important complication in patients affected by all severities of haemophilia. Another common and frequent manifestation of severe haemophilia is spontaneous intra-articular bleeding (i.e. haemarthrosis) and the most frequent site of bleeding is the ankle, followed by elbow and knee. Recurrent haemorrhages in the same joint cause inflammation of the synovial tissue, with progressive damage and the development of haemophilic arthropathy, whose final stage is chronic joint deformity, pain, muscle arthropathy and functional impairment. In contrast, clinical phenotypes in moderate haemophilia are heterogeneous and spontaneous bleeds are infrequent, with haemorrhages that tend to occur after injury, trauma or surgery. Similarly, mild haemophilia is not generally associated with spontaneous bleeding, with nearly all incidents caused by trauma or surgery[111]. In addition, female carriers have a tendency to bleed, but are rarely affected by severe haemophilia. The underlying genetic mechanism can be extreme Lyonisation, Turner's syndrome or carriage of a mutation by both parents[110].

Haemophilia is diagnosed either because of a known family history or after presentation of bleeding, and coagulation screening tests typically reveal a prolonged activated partial thromboplastin time (aPTT) with a normal prothrombin time (PT). The one-stage clotting assay is the established and standardised method to measure FIX activity levels in plasma and genetic analysis is recommended in all patients with haemophilia in order to establish the causative mutation[111].

HB is caused by mutations affecting the *F9* gene (chromosome Xq27), which contains eight exons and measures 33.5 kb. Causative genetic variations are scattered over the entire length of the gene and more than a thousand unique variants have been described so far (Factor IX Gene (*F9*) Variant Database; http://f9-db.eahad.org/; accessed on 16 January 2020)[112].

Missense, nonsense and splice site mutations are the most common, accounting for around

70% of variants, followed by frameshift mutations (around 17%). Large deletions in *F9* and mutations in the promoter region are relatively rare, accounting for 3% and 2% of the total, respectively. No common frequent genetic variation has been identified[111].

3.1.4 Coagulation factor IX

The *F9* gene encodes coagulation FIX (Fig. 3.4), a vitamin K-dependent protein which is synthesized in the liver and which circulates, as a zymogen, at a concentration of 5 µg/mL. FIX is synthesized as a precursor protein of 461 amino acids containing a 28-residue signal pre-peptide and an 18-residue leader pro-peptide. Mature FIX contains an N-terminal Gla domain (residues 47-86), a short hydrophobic stack (residues 87-92), two epidermal growth factor (EGF)-like domains (EGF1, residues 93-129, and EGF2, residues 134-173), followed by an activation peptide (residues 192-226) and a C-terminal serine protease domain (residues 227-461)[113].

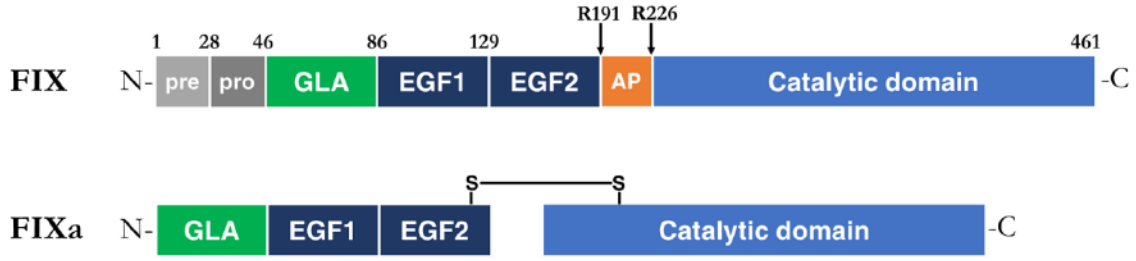

Figure 3.4. Schematic representation of FIX and FIXa domain structure. Relevant residues are reported. AP, activation peptide.

During biosynthesis, FIX undergoes several post-translational modifications[113–115], including:

- co-translational cleavage of the signal peptide, which directs endoplasmic reticulum association;
- γ-carboxylation of the first twelve Glu residues in the Gla domain, which is essential to attain a Ca^{2+}-dependent conformation and to subsequently bind phospholipid surfaces;
- glycosylation at Asn203, Asn205, Ser99, Ser107, Thr205, Thr215, Thr218 and Thr225;
- partial hydroxylation of Asp110;
- sulfation of Tyr201, which has been seen to influence the *in vivo* recovery;
- phosphorylation of Ser204;
- pro-peptide processing in the *trans*-Golgi compartment just prior to secretion from the cell, thus resulting in complete maturation of FIX.

The resulting mature protein is a single chain of 415 residues (MW 55 kDa) that contains approximately 17% carbohydrate by weight[113].

In circulation, FIX is activated to FIXa by either the TF/FVIIa/Ca^{2+} complex (initiation phase) or by FXIa/Ca^{2+} (propagation phase). This activation involves two proteolytic cleavages in FIX: the first at the Arg191-Ala192 bond and subsequently at Arg226-Val227 bond, which results in the concomitant release of a 35-residue activation peptide[116]. The FIXa thus formed contains a light chain (residues 47-191) and a heavy chain (residues 227-461) held together by a single disulfide bond (Cys178-Cys335). The light chain consists of the Gla, EGF1 and EGF2 domains, whereas the heavy chain contains the serine protease domain that features the catalytic triad of residues Ser411, His267 and Asp315[113]. The FIXa generated during the clotting process combines with its cofactor FVIIIa on platelets' surface to activate FX to FXa in the coagulation cascade. In this assembly, the protease domain and possibly the EGF2 domain of FIXa are thought to provide the primary specificity in binding to FVIIIa[117,118].

3.1.5 Haemophilia B management and novel therapies

The standard treatment for HB is replacement therapy, i.e. administration of the deficient clotting factor to achieve adequate haemostasis. The injected FIX can be either plasma-derived or recombinant, and may be administered on-demand (i.e. at the time of bleeding) or in the frame of a prophylactic regimen. For on-demand treatment, the appropriate dose, frequency and number of infusions depend on the type and severity of the bleed. Conversely, prophylaxis corresponds to the treatment with intravenous injections of factor concentrate to prevent bleeding and joint damage, with the objective of preserving normal musculoskeletal function[119]. Primary prophylaxis requires 2-to-3 infusions per week and is started at a very young age (≤2 years) before joint disease develops, whereas secondary prophylaxis begins after the onset of joint disease[111]. The evidence of success of both primary and secondary prophylaxis prompted many haemophilia patients to switch from on-demand to prophylaxis during the last decades, and the progressive enhancement of life expectancy and of quality of life made the replacement therapy of haemophilia very rewarding[120]. Moreover, cost-utility and cost-effectiveness studies showed that prophylaxis is cost-effective with respect to on-demand therapy[121]. Nevertheless, the need of multiple weekly infusions is still a significant barrier to primary prophylaxis in children, as well as a relevant concern for adult patients.

The other severe side effect of replacement therapy is the development of inhibitors, i.e. polyclonal high-affinity IgG antibodies that specifically neutralize the pro-coagulant activity of the relevant clotting factor, counteracting the management of bleeds. The development of

inhibitors is thought to be driven by both patient-related (genetics, family history) and treatment-related (intensity of exposure, product type) factors and for HB is estimated to occur in around 4-5% of patients within fifty days from the first treatment. A particular feature of anti-FIX inhibitors is the propensity for patients to develop anaphylactic reactions, thus direct medical supervision is recommended for the first infusions[111]. Patients with inhibitors can be effectively treated only with bypassing agents such as activated prothrombin complex concentrate (APCC) and/or recombinant FVIIa, which circumvent the need for FIX by generating thrombin through other mechanisms[122,123].

Recent therapeutic advancements are addressing the current issues (Tab. 3.3). In particular, half-life limitations have led to the development of enhanced half-life products. These novel drugs could simplify the prophylactic regimens for HB patients, reducing the dosage frequency and extending the protection from bleeding, thus improving adherence to treatment and rendering this therapy less distressing for patients. In addition, non-replacement products have been developed with the aim of rebalancing the haemostatic system and providing an alternative treatment for patients with and without inhibitors[124–126].

Table 3.3. Novel products for the treatment of HB (approved or in clinical trials).

Novel therapeutic (commercial name)	Mechanism of action	Status of development
Albutrepenonacog alfa (Idelvion®)	rFIX, albumin fusion	Approved by the FDA and EMA
Eftrenonacog alfa (Alprolix®)	rFIX, Fc fusion	Approved by the FDA and EMA
Nonacog beta pegol (Rebinyn/Refixia®)	rFIX, pegylation	Approved by the FDA and EMA
Fitusiran	siRNA knockdown of AT	Phase 3
Concizumab	Monoclonal antibody against TFPI	Phase 2
scAAV2/8-LP1-FIXco	Gene therapy	Phase 1/2
AskBio009 (BAX335)	Gene therapy, FIX Padua	Phase 1/2
Spk-9001	Gene therapy, FIX Padua	Phase 1/2

rFIX, recombinant FIX; FDA, Food and Drug Administration; EMA, European Medicines Agency; siRNA, small interfering RNA; AT, anti-thrombin; TFPI, tissue factor pathway inhibitor; scAAV, self-complementary adeno-associated virus; LP1, liver specific promoter; FIXco, codon-optimized FIX.

Finally, gene therapy holds the promise for a definitive cure for HB and several clinical trials have been completed or are undergoing[126]. Promising results have been obtained recently in an early-phase clinical trial using liver-directed AAV-based delivery of a naturally occurring gain-of-function FIX variant, namely FIX Padua (p.R384L)[127]. This variant was found in a thrombotic patient and showed 8-fold greater specific activity compared to wild-type FIX[128]. Notably, liver-directed gene transfer of FIX Padua did not result in thrombosis or inhibitor formation and was instead able to induce immunologic tolerance in HB dogs with pre-existing inhibitors[129–131]. However, despite the demonstration of extended therapeutic effect in clinical trials is very encouraging, substantial challenges for the development of a safe, viable and widely applicable gene therapy still remain. Up to 50% of the population have neutralizing antibodies against AAV[132] and is thus ineligible for AAV-mediated gene therapy. Moreover, even when anti-AAV antibodies are low or not present, gene therapy may be a particular challenge in young individuals because of the potential difficulty in maintaining sufficient gene expression in a growing liver. Indeed, the gene therapy landscape for HB is promising, but this approach may not be suitable for all patients[133].

3.1.6 Aim of the present work

Haemophilia B is a well-known and extensively characterised bleeding disorder. Several treatment options are currently available and novel therapeutic strategies, developed in recent years, seem to be very promising in terms of efficacy and safety. Nevertheless, current treatments still have important limitations and this scenario prompts the identification of novel optimized strategies. In particular, the short half-life of coagulation factors prompted the development of enhanced half-life products, which have the potential to improve patients' convenience and compliance. In this context, a further improvement of the biological properties of enhanced half-life products may be achieved either in terms of half-life or activity, or by the synergistic combination of these two features. Rationally-engineered and natural gain-of-function variants provide the ideal tools to develop unique molecules to be exploited for therapeutic purposes.

The aim of this project was to design, produce and characterise an innovative fusion protein between FIX and albumin, with improved features conferred by the gain-of-function FIX Padua variant and by an engineered HSA variant with strongly improved and pH-dependent FcRn binding kinetics (referred to as HSA^{QMP} for patenting reasons). The novel protein was evaluated for its pro-coagulant activity and directly compared with the commercial FIX-albumin fusion (Idelvion®) for its FcRn-binding properties and *in vivo* persistence.

3.2 Materials and methods

3.2.1 Nomenclature

All residues are reported according to the HGVS nomenclature[67], with numbering starting at the A (+1) nucleotide of the AUG (codon 1) translation initiation codon. The HGVS nomenclature differs from Legacy numbering by 46. Reference sequences: NM_000133.4; NP_000124.1.

3.2.2 Creation of expression vectors

The human FIX cDNA (including 1.4 kb of *F9* intron 1)[134,135] was PCR amplified from the pCMV5-FIXwt plasmid with the forward 5'-TCTGGTACCATGCAGCGCGTGA ACATGATC-3' (*Kpn*I restriction site underlined) and the reverse 5'-TCTCTCGAGAGTG AGCTTTGTTTTTTCCTTAATCCAG-3' (*Xho*I restriction site underlined) primers, which suppresses FIX stop codon. The PCR amplicon was then cloned into the pcDNA3 expression plasmid through *Kpn*I-*Xho*I restriction sites. Three different cleavable linkers were generated by annealing of partially complementary oligonucleotides and subsequent amplification under standard PCR conditions (the oligonucleotides' sequences are not reported for patenting reasons). The resulting linker fragments were digested with *Xho*I-*Bam*HI and cloned into the pcDNA3-FIX plasmid. Seven nucleotides including the *Xho*I site were subsequently removed by site-directed mutagenesis with primer 5´-GGAAAAAACAAAGCTCACTTC TGTGAGCCAGAC-3'. The HSA cDNA sequence corresponding to mature HSA (amino acids 25-609), either wild-type or engineered (HSA^{QMP}, sequence kindly provided by Prof. Andersen, from Oslo University Hospital), was already available in a pGEM-HSA constructs and was cloned into the pcDNA3-FIX-linker plasmid through the *Bam*HI restriction site. The FIX Padua variant (FIX p.R384L) was created by site-directed mutagenesis with 5'-CGAGCCACATGTCTTC**T**ATCTACAAAGTTCACC-3' primer (target codon underlined, changed base in bold). All the plasmids generated had the general structure pcDNA3-FIX-linker-HSA. All constructs were validated by sequencing.

3.2.3 Cell culture, transient expression and stable clones

All experiments were conducted in HEK293 cell line and cells were maintained as described in Section 2.2.4. For transient expression experiments, 24 hours before transfection $2.5x10^5$ cells/well were seeded in 12-well culture plates. Transfection was performed in serum-free

medium (Opti-MEM) with the Lipofectamine 2000 reagent, according to manufacturer's protocol. Briefly, 2 μg of plasmid DNA and 2 μL of Lipofectamine 2000 were mixed in 150 μL Opti-MEM medium and incubated at room temperature for 20 minutes, then added to the cells in a final volume of 650 μL. Transfection medium was removed 4 hours after transfection and 500 μL of fresh Opti-MEM supplemented with 5 μg/mL vitamin K was added. Media were harvested 48 hours after transfection, centrifuged for 5 min at 3000 x g and stored at -20°C.
For stable clones set up, HEK293 cells were transfected with 4 μg of plasmid DNA in 6-well plates with Lipofectamine 2000 (see above). After 24 hours, transfected cells were seeded in 100 mm plates at different dilutions (from 1:1000 to 1:24000) and cultured in DMEM supplemented with 500 μg/mL G418 (Sigma-Aldrich) as selection agent. After 14-21 days, single resistant colonies were moved to 24-well plates and cultured in DMEM supplemented with 500 μg/mL G418 and 5 μg/mL vitamin K. The best clones in terms of protein expression were selected upon a FIX-HSA-specific ELISA.
In collaboration with the group of Prof. Andersen at the Oslo University Hospital, selected clones were expanded and conditioned medium was collected. A two-step purification of FIX-HSA fusion proteins was performed through human albumin affinity chromatography followed by size exclusion chromatography.

3.2.4 Factor XIa-mediated activation and Western blotting analyses

To analyse the activation profile of FIX and FIX-HSA recombinant proteins, media from transient transfection were incubated with plasma-derived FXIa (3.6 nM final) and Innovin (1:25 final dilution, Dade® Innovin®, Siemens Healthcare, Marburg, Germany) in Reaction buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.1% PEG-8000, 5 mM $CaCl_2$) at 37°C for 4 hours. Activated proteins were then separated by SDS-PAGE on Bolt 4–12% Bis-Tris Plus gels with Bolt MES SDS Running Buffer and transferred onto 0,45 μm nitrocellulose membrane using a Tris-glycine-20% methanol running buffer. Blocking was carried out by incubation with 5% milk (w/v) in PBS buffer for 2 hours at room temperature. Recombinant proteins were detected by polyclonal goat anti-hFIX (1:2000, GAFIX-AP, Affinity Biologicals, Ancaster, ON, Canada) and donkey anti-goat HRP-conjugated (1:6000, Bethyl Laboratories, Montgomery, TX, USA) antibodies diluted in 2.5% milk (w/v) in PBS. Three washes were performed between each step with PBS-0.1% Tween-20 buffer. Western blot was developed by Clarity Western ECL Substrate (Bio-Rad, Hercules, CA, USA) and blotting images were acquired and analysed through the Image Laboratory Software version 4.0 (Bio-Rad).

3.2.5 Evaluation of secreted protein levels

Secreted rFIX levels were evaluated by ELISA, using a commercially available paired antibody set (FIX-EIA, Affinity Biologicals), according to manufacturer's protocol. Secreted FIX-HSA levels were also evaluated through a home-made ELISA. Briefly, 96-well Costar plates were coated overnight at 4°C with polyclonal anti-FIX antibody (FIX-EIA-C, Affinity Biologicals) diluted 1:100 in carbonate-bicarbonate buffer pH 9.6. Blocking was performed with 5% milk-0.1% Tween 20-PBS buffer and samples were diluted in 0.5% milk-0.1% Tween 20-PBS buffer (Sample diluent). Detection was performed through a polyclonal goat anti-HSA HRP-conjugated antibody (A80-129P, Bethyl Laboratories) diluted 1:10000 in Sample diluent. ELISA was revealed by o-phenylenediamine dihydrochloride substrate and the reaction was stopped by addition of 2.5 N H_2SO_4. A volume of 60 µL/well was added in each layer of the ELISA and the wells were washed three times with 150 µL PBS-0.1% Tween 20 after each incubation step. The absorbance was measured at 492 nm using a Sunrise spectrophotometer (TECAN).
Recombinant FIX and wild-type fusion protein were used as reference for the two ELISA.

3.2.6 Evaluation of activity levels

Activity levels were evaluated by one-stage clotting assay (aPTT), in which the sample is mixed with factor-deficient plasma and the time to clot formation is monitored after coagulation is initiated by addition of phospholipids, calcium, and a contact activation reagent. In particular, FIX-deficient plasma (HemosIL, Instrumentation Laboratory, Lexington, MA, USA) was supplemented with serial dilutions of conditioned media. Coagulation times were measured upon addition of a contact activator (SynthASil, Hemosil) and $CaCl_2$ on a ACLTOP700 instrument (Instrumentation Laboratory).
Activity levels were also assessed through a commercially available chromogenic assay (Hyphen Biomed, Neuville-sur-Oise, France), according to manufacturer's protocol. This assay uses chromogenic FXa-specific substrates to detect FX activation after sample is mixed with phospholipids, calcium, and defined amounts of purified clotting factors necessary for FXa generation.
Coagulation times (aPTT-based assay) or optical density values (chromogenic assay) from serial dilutions of an appropriate FIX protein were used as reference. The specific activity was calculated as the ratio between coagulant activity and protein levels expressed as % of the appropriate wild-type.

3.2.7 Evaluation of FcRn binding properties

FcRn binding was evaluated through an ELISA-based assay in collaboration with Prof. Andersen's group. Briefly, 96-well plates (Costar) were coated with 100 µL/well of a human IgG1 mutant variant (M252Y/S254T/T256E/H433K/N434F, 8 µg/mL, with specificity for 4-hydroxy-3-iodo-5-nitro phenylacetic acid) in PBS pH 7.4, and incubated over night at 4°C. Blocking was performed with 200 µL/well of PBS-4% skimmed milk (PBSM) and incubated for 1 hour at RT. A Hys-tagged human FcRn (10 µg/mL) in PBSM, 0.005% Tween20 (PBSTM) pH 5.5 was added and incubated for 1 hour at RT. FIX-HSA fusions were prepared in PBSTM pH 5.5, added to wells in duplicates and incubated for 1 hour at RT. Bound fusions were detected by alkaline phosphatase-conjugated polyclonal goat anti-HSA antibody (Bethyl Laboratories, Inc) diluted 1:3000 in PBSTM, and incubated for 1 hour at RT. ELISAs were developed by adding p-nitrophenyl phospate substrate (Sigma-Aldrich) diluted to 10 µg/mL in diethanolamine buffer. A volume of 100 µL/well was added in each layer of the ELISA and the wells were washed three times with 200 µL PBST pH 5.5 after each incubation step. The absorbance was measured at 405 nm using a Sunrise spectrophotometer (TECAN).

3.2.8 Mouse studies

Pharmacokinetics studies were performed in three mouse models from The Jackson Laboratory (Bar Harbor, ME): homozygous FcRn knock-out (KO) mice (B6.129X1-Fcgrt^{tm1Dcr}/Dcr), homozygous Tg32 albumin KO mice (B6.Cg-Alb^{em12Mvw} Fcgrt^{tm1Dcr} Tg(FCGRT)32Dcr/MvwJ) and hemizygous Tg32 mice (B6.Cg-Fcgrt^{tm1Dcr} Tg(FCGRT)32Dcr/DcrJ).

All mice (7-8 weeks, 3-5 mice per group) received 2 mg/kg of FIX-HSA fusions in PBS by intravenous injection. Blood collection was performed at 1, 1.5, 2, 2.5, 3, 3.5, 4 days after injection (FcRn KO mice); at 1, 1.5, 2, 2.5, 3, 4, 5, 7, 10 days after injection (hemizygous mice); at 1, 2, 3, 5, 7, 10, 11, 12, 16, 19, 23, 30, 37 days after injection (homozygous albumin KO mice). Blood (25 µL) was drawn by retro-orbital blood collection using heparinized microcapillary pipettes, mixed with 1 µL of 1% K3-EDTA and maintained on ice until centrifugation at 17000 x g for 5 min at 4°C. Plasma was isolated and diluted 1:10 in 50% glycerol/PBS solution and stored at -20°C until analysis. The studies were carried out at The Jackson Laboratory (JAX Service, Bar Harbor, ME), in accordance with guidelines and regulations approved by the Animal Care and Use Committee at The Jackson Laboratory.

The plasma concentration of FIX-HSA fusions (quantified by ELISA) is presented as percentage remaining in the circulation at different time points post-injection compared to the concentration on day 1 (100%). The β-phase half-life was calculated using the formula: t1/2 =

log 0.5/(log Ae/A0) × t, where t1/2 is the half-life of the FIX-HSA variant evaluated, Ae is the concentration remaining, A0 is the concentration on day 1 and t is the elapsed time.

3.2.9 Data analysis

Data were analysed with GraphPad Prism 5 software (San Diego, CA, USA) and Microsoft Excel 2010. Statistical differences were analysed by *t*-test.

3.3 Results and discussion

The first attempt to create a fusion protein between FIX and albumin was in 2004 and took advantage of a flexible glycine-serine linker[136], also exploited for FVIIa-HSA fusion[137]. This strategy resulted in a significant reduction of FIX activity because albumin affected FIX interaction with other coagulation factors, thereby limiting its potency. It was therefore produced a fusion protein with a proteolytically cleavable linker consisting of the amino acid sequence 182 to 200 derived from the N-terminal activation region of FIX (Idelvion®). After activation of FIX by either TF/FVIIa or FXIa, both the activation peptide and albumin were thus simultaneously cleaved, leaving FIXa free from its fusion partner[103].

In this project, it was designed, produced and characterised an improved FIX-HSA fusion protein in which the natural gain-of-function FIX Padua variant was fused through an optimized cleavable linker to an engineered albumin variant with improved affinity for FcRn.

3.3.1 A novel cleavable linker allowed optimal factor IX function

In an attempt to further optimize the fusion strategy between FIX and albumin, two novel cleavable linker sequences, together with the linker used for Idelvion®, were created and used to join the two moieties. Thus three linkers were produced, deriving from i) the amino acid sequence resembling the natural FIX Arg191-Ala192 activation site (linker a; FIX-a-HSA; equal to Idelvion®), ii) the natural FIX Arg226-Val227 activation site (linker b; FIX-b-HSA), and iii) an upstream-downstream combination of the two activation sequences (linker ab; FIX-ab-HSA). The resulting linkers sequences are not reported for patenting reasons.

The resulting three fusion proteins (FIX-a-HSA, FIX-b-HSA and FIX-ab-HSA) were transiently expressed in HEK293T cells and their activation properties were investigated by incubation with FXIa, to confirm the expected cleavage sites and the generation of fragments of the corresponding size (Fig. 3.5).

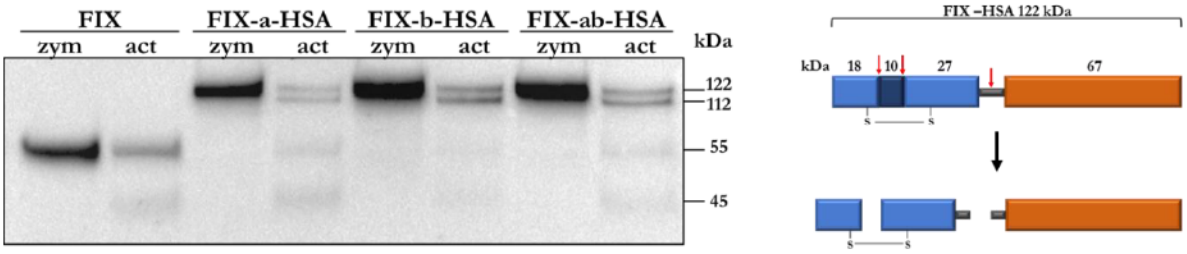

Figure 3.5. Western blotting analysis of non-activated and FXIa-activated recombinant fusion proteins (left panel). Scheme of FIX-HSA fusion protein and cleavage sites (red arrows, right panel). The molecular weight of each segment is reported. zym, zymogen; act, FXIa-activated.

Western blotting analysis with an anti-FIX antibody confirmed that all fusion proteins were effectively processed in the presence of FXIa (Fig. 3.5). In particular, cleavage of FIX-HSA proteins by FXIa resulted in partial removal of the activation peptide and/or albumin, as revealed by the presence of four different bands corresponding to i) the undigested chimaera (~122 kDa), ii) the removal of the activation peptide (~112 kDa), iii) the removal of the albumin moiety (~55 kDa), and iv) the removal of both the activation peptide and the albumin moiety (~45 kDa). As a reference, cleavage of FIX also resulted in the partial removal of the activation peptide.

To compare the secretion and functional levels of the different constructs, serial dilutions of media containing the fusion proteins were analysed (Fig. 3.6). The amount of secreted protein was quantified by a double ELISA which takes advantage of coated anti-FIX and detecting anti-HSA antibodies, thus recognising both fusion partners (Fig. 3.6, left panel). The clotting activity was assessed in FIX-deficient plasma through aPTT assay (Fig. 3.6, right panel) and, to compare the three fusion proteins, specific activity was calculated as the ratio between activity and protein levels (expressed as % of the reference fusion protein, FIX-a-HSA; inset).

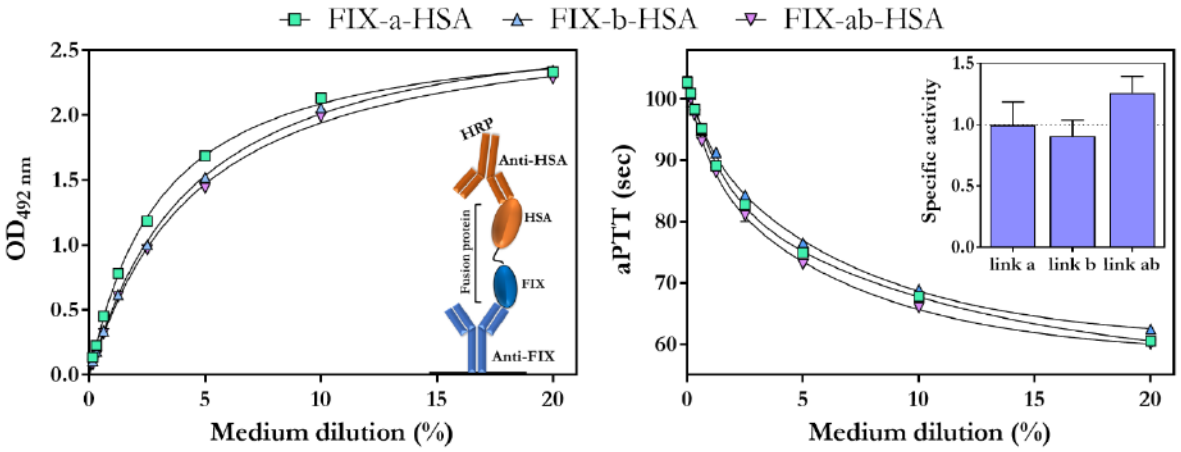

Figure 3.6. Characterisation of the best performing linker sequence for FIX-HSA fusion proteins. Secreted protein levels were measured through a home-made ELISA (left panel) and pro-coagulant activity was measured by aPTT assay (right panel). Specific activities (inset) were calculated as the ratio between activity and secreted protein levels (expressed as % of those obtained for the reference fusion, FIX-a-HSA).

The characterisation of the three fusion proteins revealed that all variants were efficiently secreted (Fig. 3.6, left panel) and that all were able to shorten coagulation times in FIX-deficient plasma in a dose-dependent manner (Fig. 3.6, right panel). As shown in Figure 3.6 (right panel, inset), this analysis also identified the linker ab as the best-working in terms of specific coagulant activity (1.26±0.13), with a significant improvement (p=0.0048) toward the reference FIX-a-HSA (1.00±0.19). These results are in agreement with the specificity preference of FXIa for an

arginine-valine instead of an arginine-alanine cleavage site, as reported in the *MEROPS* database (https://www.ebi.ac.uk/merops/index.shtml)[138]. These results prompted the selection of FIX-ab-HSA fusion protein as the scaffold for subsequent engineering steps.

3.3.2 Fusion to albumin did not affect the functional properties of factor IX Padua variant

FIX Padua (FIX^{Padua}, p.R384L) has been exploited for HB gene therapy because of its ~8-fold increased specific activity and because of its low immunogenicity[127,128,130]. Promising results were obtained and, recently, two phase 3 AAV-mediated gene therapy studies using the FIX Padua transgene have been announced[139].

These evidences prompted the adoption of FIX^{Padua} variant to improve the pro-coagulant features of the FIX-HSA fusion protein. In particular, the R384L change was introduced by site-directed mutagenesis of plasmids expressing FIX alone or the FIX-ab-HSA fusion protein, thus obtaining FIX^{Padua} and FIX^{Padua}-ab-HSA vectors, respectively. To verify the maintenance of the gain-of-function features in the presence of C-terminal albumin, fused and unfused FIX^{Padua} were expressed in HEK293T cells and specific activity was compared with the corresponding wild-type construct (Fig. 3.7).

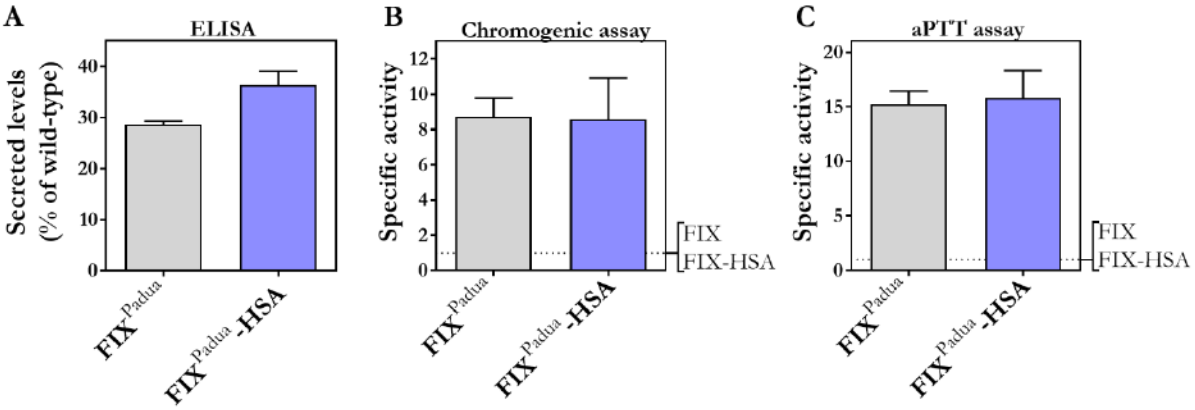

Figure 3.7. Secreted levels and activity profile of the gain-of-function FIX^{Padua} alone (grey bars) and upon HSA fusion (blue bars). (A) Secretion efficiencies are expressed as % of those of the corresponding wild-type protein. Specific activity was evaluated as the ratio between activity and secreted protein levels, as measured in chromogenic (B) and coagulant (C) assays. Results are expressed as mean ± SD from three independent experiments.

FIX^{Padua} variants showed reduced secretion compared to wild-type FIX, either alone or when fused to albumin (Fig. 3.7A). This reduction, already reported by others, is likely due to lysosomal and proteasomal intracellular degradation of the missense variant[140].

Conversely, functional evaluation of FIX^{Padua}-ab-HSA fusion protein revealed the expected improved specific activity in both chromogenic (8.6±2.3; Fig. 3.7B) and clotting (15.8±2.5; Fig. 3.7C) assays. The observed discrepancies between activity levels measured by the two assays are long-recognised and well-described for rFIX[139,141]. Indeed, the activity of rFIX measured by chromogenic assay is described to be around 70% of the clotting assay and a similar assay discrepancy was observed also in HB subjects expressing FIX Padua after AAV gene therapy[127]. Importantly, in both activity assays, the 8-to-15-fold improved specific activity of FIX^{Padua} was fully preserved upon fusion with albumin, thus supporting the use of this FIX variant as a way to boost the activity of the engineered FIX-HSA fusion protein.

A more active product for replacement therapy may allow a prolonged clinically-relevant functional profile over the therapeutic threshold even at low antigen levels, thus widening the therapeutic window and lowering the burden of treatment for patients. This may be of particular relevance because a challenge associated with current prophylaxis is the difficulty in maintaining an adequate trough level of the factor and to avoid development of joint arthropathy during patients' life[142]. Indeed, treatment is typically provided in amounts that are just sufficient to limit joint bleeds due to the high costs and limited availability, thus a hyper-functional product, requiring lower doses to achieve the same functional levels, may address these limits[131,143].

Given the high specific activity of FIX Padua, its potential for thrombogenicity needs to be carefully addressed. To date, it is unclear whether the induction of FIX activity in the high normal or above normal range could be associated with increased thrombotic risk, and all evaluations of FIX Padua safety in humans derive from gene therapy trials that are not directly translatable to a replacement therapy approach. Nevertheless, many evidences support the safety profile of FIX Padua, including the fact that FIX activity in not haemophilic people is highly variable, ranging between 50-150% of normal. In addition, a recent study on non-human primates treated with AAV5-delivered FIX Padua, showed that FIX activity up to 500% was not associated with prothrombotic state[131]. Notably, thrombosis occurred in a patient with naturally-occurring FIX-Padua and with activity levels of >700% of normal, but did not occur either in his brother nor in his mother, which had >500% and >300% FIX activity, respectively[128]. Overall, these evidences support the potential use of safe and lower doses of FIX Padua for therapeutic purposes.

3.3.3 Wild-type and engineered FIX-HSA fusion proteins were stably expressed and purified

The FIX^{Padua}-HSA fusion protein was further modified by the addition of an engineered HSA variant with enhanced half-life (designed by the group of Prof. Andersen). In particular, a novel HSA variant was exploited (referred to as HSA^{QMP} for patenting reasons), in which the K573P substitution, known to confer alone a 12-fold improved FcRn binding affinity[101], was combined with two additional amino acid changes. Notably, these three amino acid substitutions in DIII were shown to improve binding to human FcRn by 180-fold at acidic pH while barely affecting release from the receptor at neutral pH (data submitted for publication). Moreover, this variant retained an improved binding once fused with rFVIIa, thus further supporting its selection as candidate fusion partner to confer a superior FcRn binding to the proposed FIX-HSA fusion protein.

The HSA^{QMP} variant was cloned downstream of FIX^{Padua} and the previously selected linker ab, thus obtaining the FIX^{Padua}-HSA^{QMP} construct. Both engineered and wild-type FIX-HSA proteins (Fig. 3.8, left panel) were stably expressed and, in collaboration with the group of Professor Andersen at the Oslo University Hospital, both proteins were purified by affinity and size-exclusion chromatography.

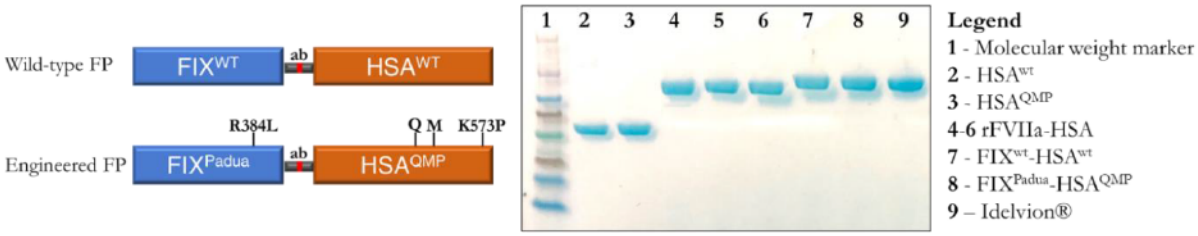

Figure 3.8. Purified fusion proteins. On the left, a schematic representation of the two fusions produced, with mutated residues indicated. On the right, SDS-PAGE of purified FIX-HSA fusion proteins together with additional controls. FP, fusion protein; ab, linker ab; WT, wild-type.

Analysis on SDS-PAGE (Fig. 3.8, right panel) showed the presence of all fusion proteins with an overall molecular weight as expected (~122 kDa). Together with the FIX-HSA proteins, the commercial product Idelvion® was loaded as control. In addition, purified rFVIIa-HSA and HSA variants alone represented further controls.

3.3.4 In vitro characterisation of engineered fusion proteins revealed improved FcRn binding

The improved FcRn-binding properties provided by the HSA^{QMP} variant were evaluated in collaboration with Prof. Andersen's group. Titrated amounts of purified fusion proteins were assessed by an ELISA-based binding assay at the optimal pH 5.8 for binding to FcRn.

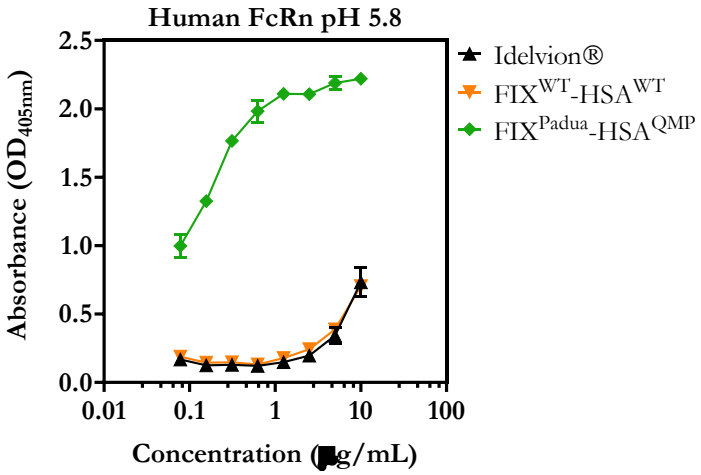

Figure 3.9. ELISA-based assay showing binding to human FcRn of titrated amounts of FIX-HSA fusions. The ELISA was performed at pH 5.8.

FIX^{Padua}-HSA^{QMP} showed stronger FcRn-binding affinity at acidic pH compared to both the wild-type fusion and Idelvion®, which, conversely, showed overlapping binding features between them (Fig. 3.9). These results demonstrated that the adopted fusion strategy does not impair the FcRn binding properties of the engineered HSA^{QMP} variant.

Besides validating its use for an improved FIX-HSA fusion protein, these data also support the use of this HSA variant as a fusion partner for other therapeutically relevant proteins. Indeed, the improved binding to FcRn has the potential to translate into an extended half-life fusion protein and into a reduced dosing frequency for HB patients.

3.3.5 In vivo studies showed enhanced half-life of the engineered fusion protein

As already mentioned, the different interspecies interactions between mouse/human albumin and mouse/human FcRn can compromise the half-life evaluation of HSA variants *in vivo*[99]. Therefore, to investigate the *in vivo* properties of FIX-HSA variants, specific mouse models were exploited (Fig. 3.10).

First, fusion proteins were challenged in hemizygous human FcRn transgenic mice expressing mouse albumin (Tg32 mice)[144]. As shown in Figure 3.10A, the engineered fusion protein was detectable in mouse plasma up to 10 days, whereas both wild-type variant and Idelvion® could

be detected only until day 5. Analysis of the clearance curves resulted in half-lives of 2.5±0.2 days for FIX^{Padua}-HSA^{QMP} variant and of 1.1±0.1 and 1.0±0.1 days for the wild-type fusion and Idelvion®, respectively, revealing a 2.5-fold improvement conferred by the engineered albumin.

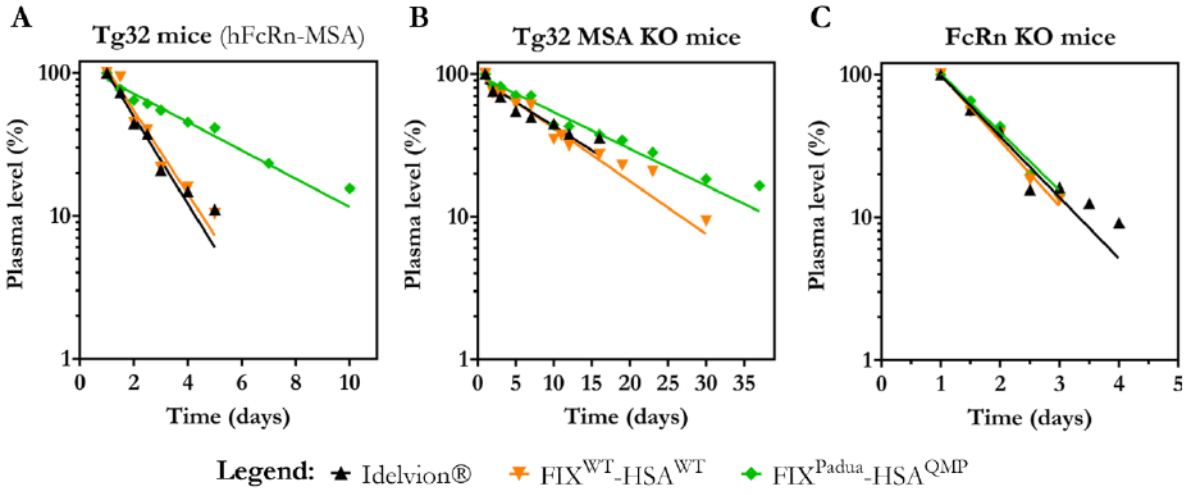

Figure 3.10. Half-life evaluation of FIX-HSA variants *in vivo*. Elimination curves of FIX-HSA in hFcRn transgenic mice (A), hFcRn and MSA KO mice (B) and in FcRn KO mice (C). The serum levels are presented as % remaining in circulation compared to that measured at day 1 after mice were intravenously injected. The values are reported as mean ± SD of five mice.

Conversely, in mice knock-out for mouse albumin (Fig. 3.10B), the differences between engineered and wild-type fusions were less pronounced with half-lives of 11.1±0.7 days for FIX^{Padua}-HSA^{QMP} and 7.8±1.4 and 7.2±0.7 days for wild-type and Idelvion®, respectively. Indeed, extended persistence in circulation was observed for all variants (detectable in plasma for up to 30-35 days). These results underline the impact of competition on half-life of albumin-based products. Indeed, the presence of a competing mouse albumin, albeit characterised by stronger affinity to human FcRn than HSA, better mimics the natural situation than the model without competition. As additional control, a mouse strain knock-out for FcRn was used. As expected, all fusion proteins were comparably and rapidly removed from circulation (Fig. 3.10C), thus confirming that the improved half-life observed in others models was dependent on FcRn-mediated rescue.

Overall, these data demonstrated that the engineered albumin is able to further extend FIX half-life in comparison with the currently available fusion protein (Idelvion®). If translated to humans, this feature would represent a great advantage in the context of a prophylactic regimen, allowing for further reduced infusion frequency and promoting patients' compliance. These findings also support the use of this engineered HSA variant for the improvement of the pharmacokinetic properties of other therapeutic proteins.

3.4 Conclusions

In this work a novel fusion protein between coagulation factor IX and human albumin was designed and produced. In particular, to obtain a molecule with improved therapeutic potential in terms of both pro-coagulant activity and half-life, the gain-of-function FIX Padua variant and an engineered albumin with improved FcRn-binding capacity were fused through a cleavable linker sequence. Evaluation of pro-coagulant properties demonstrated that the hyperfunctional features of FIX Padua were fully preserved after fusion to albumin. In addition, FcRn binding of the engineered fusion protein was strongly improved by the presence of the HSA^{QMP} variant. This resulted in 2.5-fold half-life extension in a humanized mouse model transgenic for hFcRn, with competition provided by mouse albumin and thus mimicking the natural situation. Overall, the results of this work provide a first proof-of-principle of the improved properties of a FIX^{Padua}-HSA^{QMP} fusion protein in terms of both pro-coagulant activity and half-life.

Further studies will address the functional half-life of the engineered chimaera to confirm that FIX maintains full function upon FcRn-mediated recycling within acidified endosomes. Moreover, further *in vivo* studies are ongoing, exploiting a double humanized mouse model characterised by a hybrid FcRn (human FcRn/mouse β2m heterodimer) and competition from endogenous human albumin.

If translated to HB treatment, the improved features of the novel FIX-HSA fusion designed in this work would have the potential to address many of the current limits of replacement therapy. Indeed, a further enhanced half-life factor, by widening the therapeutic window, may encourage patients to switch from on-demand to prophylaxis regimen by reducing injections frequency and the burden of treatment. It may also reduce the need for central venous lines in children and promote adherence in adolescents, improving outcomes and potentially allow for a more active lifestyle, thus ameliorating patients' quality of life. Moreover, the increased specific activity has the potential to raise trough levels with a consequent increase in effectiveness, and to possibly reduce treatment-related costs due to lower dosage needed[131,142,145].

Chapter 4

Exploring U1 snRNA-based splicing correction for Haemophilia A

4.1 Background and rationale

4.1.1 The splicing process and its regulation

In eukaryotic cells, the primary RNA transcript (pre-mRNA) is synthesized in the nucleus by RNA polymerase II, where it undergoes several modifications before being transported into the cytoplasm, in which translation takes place. A central step of mRNA maturation is the splicing process, during which the removal of non-coding introns followed by exons ligation occurs. In particular, splicing takes place in two steps involving two consecutive trans-esterification reactions (Figure 4.1, upper panel). First, an adenosine residue located within a hundred nucleotides from the 3' end of the intron (branch point sequence), carries out a nucleophilic attack on the 5' splice site (5'ss), resulting in the formation of an intron lariat intermediate. In the second step, the free 5'ss attacks the 3' splice site (3'ss), leading to the removal of the intron lariat and the formation of the spliced RNA product[146].

The machinery carrying out this process is the spliceosome, a large RNA-protein complex characterised by the presence of U1, U2, U4/U6 and U5 small nuclear RNAs (snRNAs) organized in small nuclear ribonucleoprotein particles (snRNPs)[147]. In a subset of eukaryotes there is also a less abundant minor spliceosome formed by U11/U12 and U4atac/U6atac snRNPs, which are functionally comparable to the components of the major one[148].

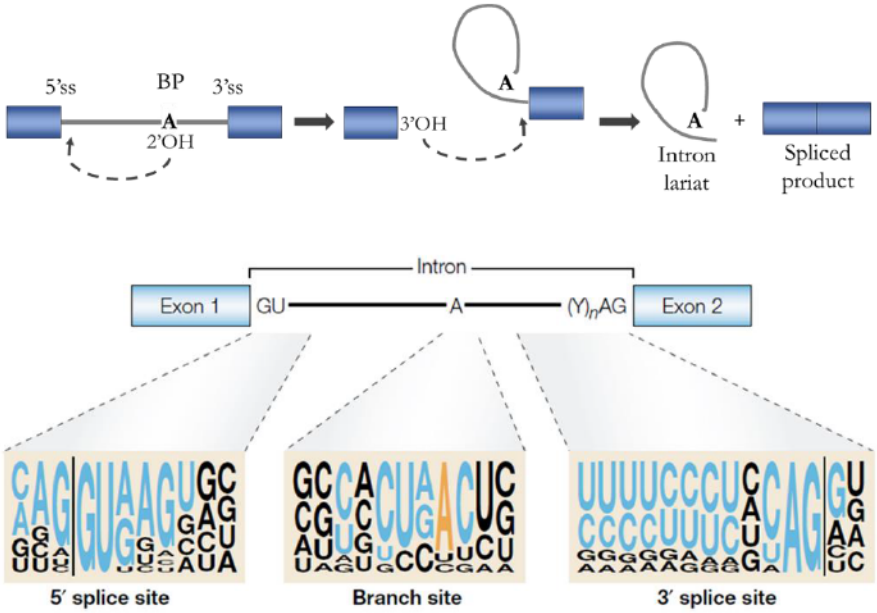

Figure 4.1. Representation of the two-step splicing reaction (upper panel, adapted from Ref.[149]) and scheme of the canonical splice sites with their consensus sequences (lower panel, from Ref.[150]). BP, branch point; $(Y)_n$, polypyrimidine tract.

Spliceosome assembly is directed by the consensus sequences that mark the exon-intron boundaries, including the 5'ss, the branch point, the polypyrimidine tract and the 3'ss (Fig. 4.1, lower panel). The 5'ss (donor site) is a degenerate 9-nucleotide motif that generally conforms to the sequence (C/A)AG - GU(A/G)AGU and is firstly bound via homology by U1 snRNA in what is known as the spliceosomal E complex. The 3' splice site (acceptor site) is characterised by three elements generally within 40 nucleotides of the intron-exon boundary: the branch point (BP), which contains the adenosine important for the first trans-esterification step of the splicing reaction, the polypyrimidine tract, a region with an high percentage of pyrimidines, and the 3'ss ((C/U)AG - G). The degenerate BP sequence ((C/U)N(C/U)U(A/G)A(C/U)) interacts with U2 snRNP, while the U2 snRNP auxiliary factor (U2AF) binds the polypyrimidine tract[151]. These signals comprise the first layer of code in splicing and are essential but not sufficient to direct splicing. Indeed, genome sequencing analyses have showed that the majority of splice sites do not precisely conform to the consensus sequence and that less than 5% of donor sites match perfectly the 9-nucleotide consensus stretch, whereas more than 25% are characterised by three or more mismatches[152].

A further level of complexity derives from the fact that pre-mRNA splicing can be constitutive or alternative. In the first case the exon is always included in the mature mRNA, while in the second multiple mRNAs can be obtained from the same pre-mRNA by differential joining of 5' and 3' splice sites. Alternative splicing is one of the main protagonists in regulation of tissue- and developmental stage-specific gene expression, by generating different protein isoforms that can function in several cellular processes. Indeed, it is estimated that more than 95% of the genes undergo alternative splicing, partially explaining the discrepancy between the 24.000 estimated protein coding genes and the about 100.000 different proteins found in the cell[153].

Due to the degenerate nature of the splice sites, other auxiliary sequences are essential in order to define exons from introns in constitutive splicing, and to modulate alternative splicing. The need for these auxiliary signals is even more significant when considering that the size of a human exon is generally comprised between 50 and 250 bp, thus much shorter compared to the thousands of residues that constitute intronic sequences, meaning that the spliceosomal machinery has to identify small exons within vast introns[147]. The cis-acting regulatory elements are classified according to their effect and localization as exonic splicing enhancers and silencers (ESE and ESS, respectively) and intronic splicing enhancers and silencers (ISE and ISS, respectively)[150](Fig. 4.2).

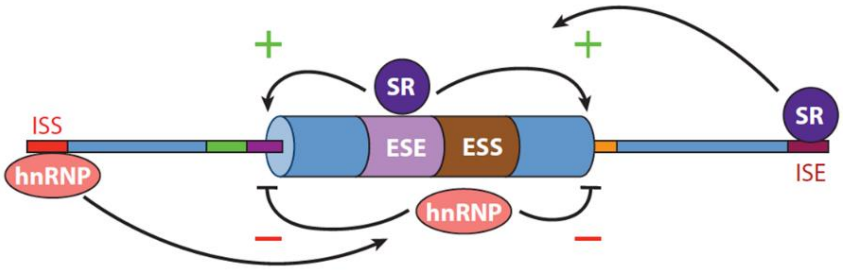

Figure 4.2. Control of pre-mRNA splicing by cis-acting intronic and exonic silencers and enhancers. A schematic exon surrounded by two introns is represented and intronic and exonic splicing enhancers (ISE, burgundy box; ESE, purple box), as well as intronic and exonic splicing silencers (ISS, red box; ESS, brown box) are indicated. Serine/arginine-rich (SR) proteins promote splicing by interacting with enhancers, whereas heterogeneous nuclear ribonucleoproteins (hnRNP) inhibit splicing by interacting with silencers[154].

The vast majority of enhancer elements contain purine-rich sequences, that are binding sites for serine/arginine-rich (SR) proteins. SR proteins have a common structural organization, characterised by the presence of one or two RNA recognition motifs at the N-terminal and the arginine/serine (RS) domain, of at least 50 amino acids and with an RS content higher than 40%, at the C-terminal. SR proteins that are bound to ESEs can promote exon definition by directly recruiting the splicing machinery through their RS domain and/or by antagonizing the action of nearby silencer elements. Notably, these two models of splicing enhancement are not necessarily mutually exclusive, as they might reflect different requirements in the context of different exons[150].

The silencer elements are principally bound by heterogeneous nuclear ribonucleoproteins (hnRNPs), which are a set of primarily nuclear proteins that bind pre-mRNA without necessarily forming stable association with other RNA-protein complexes. The vast majority of the hnRNP proteins interact with the pre-mRNA through one or more RNA binding motifs and can act through three silencing mechanisms. Direct competition occurs when the enhancer-binding site overlap with the silencer's one so that the binding of the positive and negative factors is mutually exclusive. Moreover, silencers can act by promoting the formation of protein–protein interactions that loop out the alternative exon or by nucleation and cooperative binding of additional inhibitory molecules, which polymerize along the exon over an enhancer site[147,150].

Overall, the decision of whether to include an exon reflects the interplay and the balance among many factors, including the intrinsic strength of the canonical splice sites, the combinatorial effects of positive and negative regulatory elements, as well as the effect on splicing of the RNA structure[155], transcription rate[156], chromatin remodelling and epigenetic modifications[157]. The

combination of all these elements still hampers the many attempts of effectively predicting if a given exon will be included or not in a specific biological context.

4.1.2 Splicing and disease

Given the complexity of splicing mechanism and regulation, it is not surprising that this process is extremely susceptible to the effect of mutations and thus widely involved in human diseases. In particular, it has been estimated that around 15% of point mutations that result in a human genetic disease cause defective splicing[158]. Notably, the majority of mutations considered in these surveys directly affect the canonical splice sites, whereas point mutations in the coding regions are usually classified as missense, nonsense or silent mutations. This assumption may be misleading because even exonic mutations affecting splicing regulatory elements are likely to have a strong impact on RNA processing and thus on the translated product. A paradigmatic example is represented by mutations that are classified as missense variants but also affect an ESE. In this case, the mutation could actually result in exon skipping, thus leading to a disrupted open-reading frame or a mutant protein with a large deletion, instead of the expected single amino acid change[159,160]. Indeed, misclassification of mutations leads to underestimate the frequency of splicing mutations and it is becoming evident that, depending on the complexity of the affected genes, mutations leading to splicing dysfunction may represent up to 50% of all described variations[161–163]. Importantly, the correct classification of mutations is essential for understanding structure–function relationships in the corresponding protein, for assessing the phenotypic risk in individuals with familial disease predispositions and for the design of new therapies[150].

It is well known that splicing mutations can act in different ways, depending on the affected sequence. Mutations of canonical splice sites generally lead to exon skipping or, in a minority of cases, to intron retention, with those occurring at the 5'ss being the most severe ones. This results in the synthesis of a non-functional protein or in mRNA degradation due the introduction of a premature termination codon which can trigger nonsense mediated decay. Moreover, mutations in enhancer or silencer elements can change the ratio of isoforms containing alternative exons and mutations within introns can lead to inclusion of intronic sequences through the use of a cryptic splice site[164]. To date, several diseases have been associated to splicing defects and this has led to great interest in developing therapeutic approaches to correct aberrant splicing[165].

4.1.3 Splicing-correction approaches

Overall, the fact that many human genetic diseases are caused by mutations affecting splicing has prompted the development of therapeutic approaches that directly target this process, such as antisense oligonucleotides, chemical compounds and modified U1 snRNAs.

Antisense oligonucleotides (AONs), among the most studied splicing-based therapeutics, are short (15-20 bases) synthetic oligonucleotides that bind to RNA through standard Watson-Crick base pairing, thus modulating the function of their target. In particular, AONs can be directed to canonical splicing sites or to regulatory elements where they sterically block access to the transcript by the spliceosome and splicing factors. In the past decades, many modifications have been developed to increase AONs resistance to nucleases and proteases, and to target affinity, specificity and *in vivo* stability, as well as to reduce non-specific interactions with proteins[166]. Indeed, the therapeutic application of AONs for splicing correction is promising, as highlighted by the recent approval of nusinersen (Spinraza®, Biogen), a modified AON approved for the treatment of spinal muscular atrophy[167,168]. Nevertheless, this approach lacks efficient delivery to the tissue of interest and requires repeated administrations of large amounts of costly materials, thus many studies are focusing on overcoming these obstacles[166,169].

A second splicing-correction strategy exploits small compounds that can function by directly modifying the activity of splicing factors or by indirectly altering splicing, frequently by unknown mechanisms[170]. The major advantage of this approach is that many of these compounds have already been approved and are used in clinical practice to treat diseases not related to splicing, thus proving their safety in humans. On the other hand, a major drawback of small molecule therapeutics is their lack of specificity and the poor knowledge about their exact mechanism of action, which can lead to unpredicted off-target effects[170,171].

4.1.4 Modified U1 snRNAs as splicing-targeted therapies

Another promising approach to rescue aberrant splicing is represented by the use of modified U1 snRNAs. U1 snRNA is the RNA component of the core spliceosomal U1 snRNP and is involved in 5'ss recognition through its 9-bp tail, during the first step of spliceosome assembly. It is a 164-nucleotide long molecule and has a stable and defined 4 stem-loops secondary structure that interacts with a set of U1-specific proteins, namely U1-A, U1-70K and U1-C, as well as with the Smith antigen proteins, common to all the U-rich snRNAs[172] (Fig. 4.3).

Mutations at the 5'ss can alter U1 snRNA binding, thus preventing spliceosome assembly and resulting in aberrant splicing. For this reason, modified U1 snRNAs with restored complementarity to the mutated 5'ss have been designed and have proved their efficacy as

splicing-correction molecules in several disease models[166]. However, because of the partially conserved nature of the 5'ss, modified U1 snRNAs have to deal with potential off-target effects.

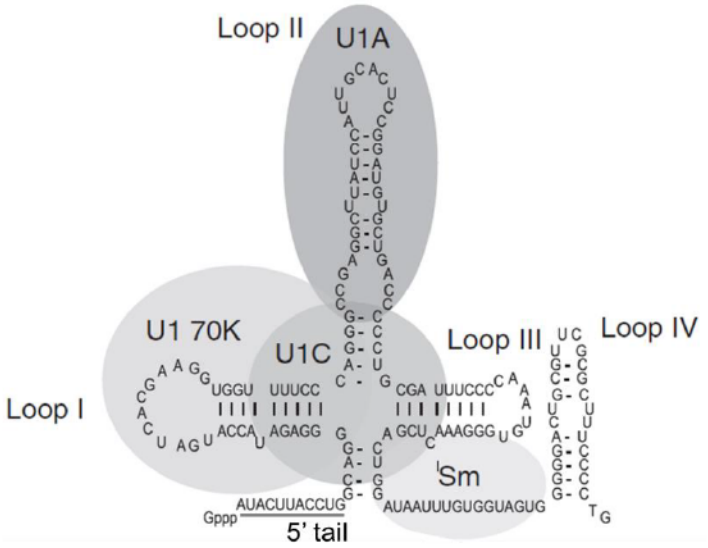

Figure 4.3. Schematic representation of U1 snRNA sequence and secondary structure and associated proteins[173]. The 5' tail, binding the donor site on the pre-mRNA, is underlined.

Interestingly, it has been shown that U1 snRNAs do not necessarily have to bind to the 5'ss to promote exon definition and that U1 snRNAs complementary to intronic sequences downstream of the 5'ss can enhance recognition of 5'ss as well as exon definition[174,175]. As a result, a second generation of correction molecules have been designed, namely exon specific U1 snRNAs (ExSpeU1s), characterised by an engineered 5' tail which binds to non-conserved intronic regions downstream of the 5'ss and allowing improved specificity and reduced potential off-target events[175]. Moreover, it has been shown that a single ExSpeU1s can be used to correct multiple exon-skipping mutations, at the 5'ss, in the polypyrimidine tract and even at exonic regulatory elements, thus extending the applicability of this molecules[175]. Recent studies revealed that ExSpeU1s splicing rescue activity is not dependent on the recruitment of endogenous U1 snRNP to the 5'ss, but on the U1-70K protein and on the loop IV structure of the U1 snRNA, thus indicating that ExSpeU1s promote exon recognition through the recruitment of splicing factors that subsequently activate the mutated 5'ss[173].

In the context of a therapeutic approach, U1 snRNAs have the major advantage of maintaining the regulated expression of the target gene and, given that their expression cassette includes promoter and regulatory sequences, to provide long-term correction. Another favourable feature of U1 snRNAs is the small cassette size, which is compatible with the delivery with the AAV vectors, one of the most successful gene therapy systems available[166]. Overall, these

features make the use of U1 snRNAs a promising novel therapeutic strategy to correct splicing defects. Indeed, both modified U1 snRNAs and ExSpeU1s have already shown their efficacy to repair different types of splicing defects in cellular and animal models of diseases such as haemophilia B[176], FVII deficiency[177], cystic fybrosis[175], spinal muscular atrophy (SMA)[178,179], Fanconi anemia[180], Netherton syndrome[181] and Familial dysautonomia[182]. Notably, studies on SMA mouse models, addressing the possible off-target effects of ExSpeU1s, showed that, out of 12414 analysed genes, only 12 had altered expression after treatment[173]. Moreover, using a cellular model with high over-expression of ExSpeU1, no relevant undesired effects on the transcriptome nor a general inhibitory effect on pre-mRNA processing was present[179]. This further indicated that ExSpeU1s may be a valid therapeutic opportunity for several splicing-related pathologies, thus prompting the investigation of this approach in other disease models.

4.1.5 Haemophilia A and coagulation factor VIII

Haemophilia A (HA; OMIM number 306700) is an X-linked bleeding disorder caused by deficiency of coagulation factor VIII (FVIII). Its incidence is estimated to be between 1:5000 and 1:10000 in men and the bleeding tendency is related to the residual levels of FVIII, leading to classification of the disease as severe, moderate or mild[110] (see Section 3.1.3, Tab. 3.2). Activated FVIII (FVIIIa) serves as cofactor of FIXa, which activates FX during the propagation phase of blood coagulation. HA is thus clinically indistinguishable from HB[110] and a detailed description of the clinical presentation and diagnosis of haemophilia can be found in Section 3.1.3.

HA is caused by mutations affecting *F8* gene, which is located in the distal end of the long arm of the X chromosome (Xq28), spanning 186 kb of genomic DNA. It consists of 26 exons and encodes a mature protein of 2332 amino acids[183]. At present, more than 3000 mutations within the *F8* coding and untranslated regions have been identified (Factor VIII Gene (*F8*) Variant Database; http://f8-db.eahad.org/; accessed on 16 January 2020), thus supporting the concept that HA has a very high mutational heterogeneity. Nevertheless, the most common mutation, affecting approximately half of the patients with the severe form of the disease, is the large inversion and translocation of exons 1–22, which completely disrupts the gene[183,184].

F8 gene encodes coagulation factor VIII (Fig. 4.4), which is synthesized primarily by liver sinusoidal endothelial cells[185] as a large single chain protein of around 330 kDa. FVIII has domain structure of A1-A2-B-A3-C1-C2, where both the A and C domains have internal sequence homology to the A and C domains of FV and ceruloplasmin, whereas the large B domain, representing 38% of the total sequence, is not homologous to other known structures.

Moreover, three segments of 30-40 amino acids possessing a high concentration of acidic residues (a1, a2, a3), are located downstream of A1 (residues 355-391) and A2 domains (residues 730-759), as well as upstream of the A3 domain (residues 1668-1708)[186].

In the secretory pathway, FVIII is intracellularly processed to a series of metal ion-linked heterodimers produced by cleavage at the B-A3 junction as well as at additional sites in the B-domain[187]. These cleavages generate a variably sized heavy chain (A1-A2-B; 90-200 kDa) non-covalently associated to a light chain (A3-C1-C2; 80 kDa). FVIII has two disulfide bonds in each A domain, involving four of the five Cys residues present, thus a free Cys is contained in the A1 (Cys329), A2 (Cys711) and A3 (Cys2019) domains. Post-translational modifications include phosphorylation (Thr370 and Ser1676) and N-linked glycosylation, which heavily modifies the B domain (19 out of 26 sites) as well as several sites in the A domains, and which appears dispensable for FVIII activity but critical for its efficient secretion[114,187,188]. Additionally, FVIII contains six sites of tyrosine sulfation in a1 (Tyr365), a2 (Tyr737, Tyr738 and Tyr742) and a3 (Tyr1683 and Tyr1699) domains. Sulfated tyrosine residues in the a2 and a3 acidic segments contribute to cofactor function and to von Willebrand factor (vWF) binding, respectively[187].

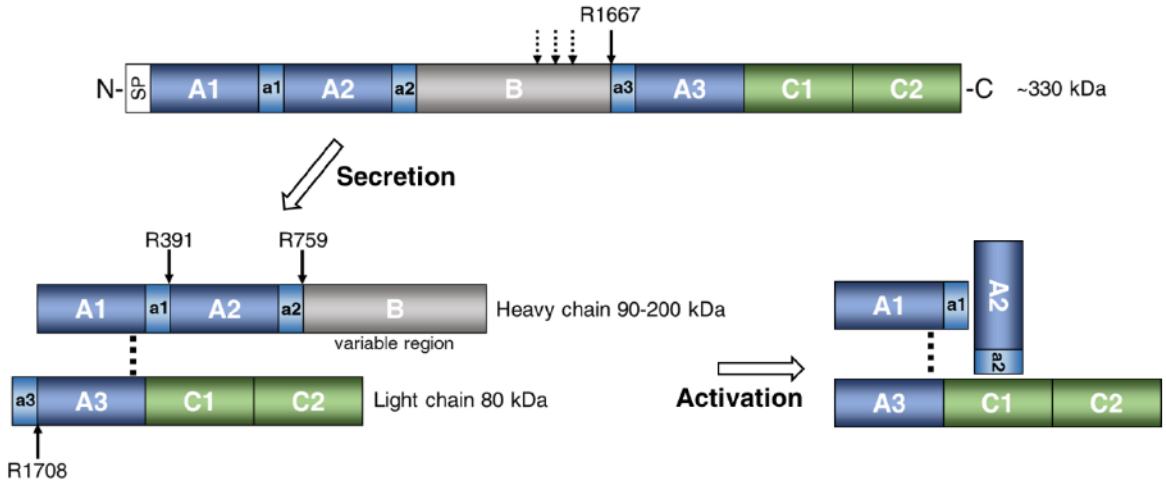

Figure 4.4. Domain structure and processing of FVIII. FVIII is secreted as a heterodimer formed by a heavy and a light chain associated through a divalent metal ion (dotted line). Activation is mediated by cleavage by thrombin and FXa at specific residues to form a heterotrimer. Cleavage sites are reported (black arrows). Adapted from Ref.[189].

In plasma, FVIII circulates as inactive precursor at a concentration of 200 ng/mL and is non-covalently bound to its carrier vWF, which is a large multimeric protein comprised of identical ~220 kDa subunits. FVIII binds vWF in a high affinity interaction (Kd ~0.3 nM) mediated by residues in the FVIII light chain, with critical sites for this interaction that have been localized

in the a3 region and the C2 domain[186]. Association with vWF stabilizes FVIII upon secretion from the cell, inhibits its binding to phospholipids and increases its half-life. The ratio of vWF to FVIII is maintained at 50:1 and an increase or decrease in the plasma vWF levels results in a corresponding change in the levels of FVIII[114].

In circulation, FVIII is activated by either thrombin or FXa, which cleave at Arg759 and subsequently at Arg391 and Arg1708. In particular, the cleavage at Arg1708 releases FVIIIa from vWF, thus enabling the interaction with negatively charged phospholipids. Upon activation, the B domain is also released and FVIIIa results as a heterotrimer composed by an A1-derived chain (50 kDa; 20-391), an A2-derived chain (43 kDa; 392-759) and a light chain fragment (A3-C1-C2; 73 kDa; 1708-2351)[114,190]. Activation of FVIII produces a transient twenty- to fifty-fold increase of its cofactor activity, which rapidly decays because of either dissociation of the A2 subunit from A1/A3-C1-C2 subunits or due to proteolytic degradation. The main physiological inactivator of FVIIIa is activated protein C which cleaves it in the A1 domain at Arg355 and in the A2 domain at Arg581. Additionally, degradation is mediated to a lesser extent by thrombin, FIXa and FXa, which cleaves FVIII at Arg1740, Arg355 and Lys55, with proteolysis at the last two sites leading to complete loss of cofactor activity[186,190].

Association of FVIIIa with FIXa constitutes the intrinsic factor Xase complex which is membrane-dependent and involves multiple inter-protein contacts. In particular, FVIIIa binds anionic phospholipid membranes via hydrophobic and electrostatic interactions mediated by the C2 domain, whereas inter-protein interactions between factors IXa and VIIIa occur over an extended interface. Indeed, the majority of the binding energy is provided by the A3-C1-C2 subunit of FVIIIa, but several studies have identified also a contribution of both A2 and A1 subunits[186].

4.1.6 Haemophilia A management and novel therapies

The management of haemophilia A is based on infusions of plasma-derived or recombinant FVIII (rFVIII) products, with the latter being either full-length or lacking the B-domain (Refacto® and Xyntha/Refacto AF®, Pfizer), found to be disposable for coagulant activity[191]. The treatment can be administered at the time of bleeding (i.e. on-demand) or as a long-term and continuous substitution therapy (prophylaxis), which is the recommended treatment in severe HA because it allows to prevent bleeding and the resultant joints damage. In addition, for mild HA patients, the synthetic drug desmopressin, which transiently increases two- to four-fold the plasma levels of FVIII and vWF, can be used to prevent or treat bleeding episodes[184].

The most serious complication of replacement therapy for HA is the occurrence of inhibitory alloantibodies against FVIII, which develop in approximately 25–30% of severely affected patients, usually within the first fifty days of treatment. The development of neutralizing inhibitors renders FVIII infusion ineffective and exposes patients to an increased risk of morbidity and mortality. Several studies have suggested the importance of both genetic and treatment-related risk factors, confirming that inhibitor formation is a complex and multifactorial process[192]. Inhibitors are characterised through the titre, which is defined as the inhibitory capacity of patient's plasma to neutralise clotting factors in normal plasma, and is quantified in Bethesda Units (BU). This characterisation is important because patients with a low titre (0.5-5 BU) and with low responding inhibitors can be treated with increased doses of the missing factor, an approach called immune tolerance induction (ITI)[193]. Conversely, patients with a high titre (>5 BU) or high responding inhibitors can be treated effectively only with bypassing agents, such as activated prothrombin complex concentrate (APCC) and recombinant FVIIa, which circumvent the need for FVIII or FIX by generating thrombin through other mechanisms[122,123,194,195].

The several disadvantages still associated to replacement therapy have prompt the development of new treatment strategies. In order to decrease the frequency of prophylactic infusions and improve patients' compliance, long-acting factor products have been developed. In particular, improved pharmacokinetic properties of rFVIII products have been obtained by site directed glycopegylation (Adynovi®, Shire) or by fusion with carrier proteins such as the Fc region of IgG (Elocta®, Biogen/Sobi). However, half-life of FVIII is highly dependent on vWF, thus the half-life extension obtained has been modest[126]. An alternative approach currently under clinical trial is represented by haemostatic rebalancing therapies. This group of nonfactor agents includes a small interfering RNA against antithrombin (Fitusiran, Alnylam Pharmaceutic) and a monoclonal antibody against tissue factor pathway inhibitor (Concizumab, Novo Nordisk), currently under phase 3 and phase 2 clinical trials, respectively. Additionally, a FVIII mimetics has recently reached the market (Emicizumab, Hemlibra®, Roche). This humanised bispecific antibody is designed to bind both FIXa and FX, thus mimicking the function of FVIIIa. Despite its demonstrated clinical efficacy, the biochemical potency of emicizumab is still limited and further improvements are ongoing[196]. Moreover, recent concerns have been raised on its potential thrombogenicity and safety[197].

Finally, gene therapy has great potential and several clinical trials are currently recruiting[125], however major challenges still exist. One is represented by the large size of FVIII expression cassette (4.4 kb for the B-domain deleted form), which hampers packaging in adeno-associated

viral vectors[126]. A further limit is represented by the low expression levels of FVIII, which are due to inefficient expression of the mRNA, significant proportion of protein misfolding with subsequent intracellular degradation and inefficient transport of the primary translation product to the Golgi[188].

Overall, recent advances have brought the widespread availability of safe and effective HA treatments, but major challenges still remain and novel approaches may help fill these gaps.

4.1.7 Aim of the present work

Although widely available, the therapeutic options for haemophilia A still present significant drawbacks. In this context, alternative and personalized therapeutic strategies, such as the modulation of pre-mRNA splicing process through modified U1 snRNAs, may be of extreme interest. Indeed, mutations affecting the splicing process account for about 8-10% of all HA patients, a largely underestimated proportion since also the abundant missense mutations can cause aberrant splicing by affecting regulatory elements. This approach, already proven to be effective in several cellular and animal models of disease[175–177,179], has never been attempted so far in HA. Notably, a correction approach based on modified U1 snRNAs would have the advantages of i) maintaining physiological gene expression regulation, ii) exploiting small expression cassette compatible with AAV-mediated delivery, and iii) being able to correct different mutations sharing the same molecular mechanism, thus broadening its applicability. Therefore, the aim of this work was to characterise splicing patterns caused by HA-causing mutations and to design and validate innovative U1 snRNA molecules able to rescue correct splicing and thus characterised by relevant therapeutic potential.

4.2 Materials and methods

4.2.1 Nomenclature

All residues are reported according to the HGVS nomenclature[67], with numbering starting at the A (+1) nucleotide of the AUG (codon 1) translation initiation codon. The HGVS nomenclature differs from Legacy numbering by 19. Intronic mutations are numbered according to their position relative to the donor site. Reference sequences: NM_000132.3; NP_000123.1.

4.2.2 Vectors construction

To create the wild-type exon 19 minigene, the genomic region of the human *F8* gene (NC_000023.11) spanning from c.5999-435 to c.6115+381 was amplified from genomic DNA of a normal subject with primers 5'-AAACATATGGTTCACCTGCAGCTTCTTGTTGCT-3' and 5'-ATACATATGGACCAGTGGCTCTCTTACTTGTTTAGG-3' (*Nde*I sites underlined) using high-fidelity *Pfu* DNA polymerase (ThermoFisher Scientific). The exon 19 amplicon was subsequently cloned into the pTB expression vector (kind gift from professor Pagani, ICGEB, Trieste, Italy) by exploiting the *Nde*I restriction site. To create minigenes with the selected variants, nucleotide changes were introduced into the wild-type minigene by site-directed mutagenesis (QuickChange II Site-Directed Mutagenesis Kit, Stratagene) with primers reported in Table 4.1.

To produce the expression vectors for the modified U7 snRNAs (pU7), a PCR containing the modified binding site of the engineered U7 snRNA has been generated by using the following primers (antisense region underlined): 5'-ACAGAGGCCTTTCCGCACTCATCCCAGCATGTAGATAATTTTTGGAG-3', 5'-ATTTAGGTGACACTATAGAA-3' (U7[a]); 5'-ACAGAGGCCTTTCCGCATGTACACCAGAAAAAGTGTGAATTTTTGGAG-3', 5'-ATTTTAGGTGACACTATAGAA-3' (U7[b]). The resulting fragments were digested with *Stu*I-*Xba*I restriction enzymes and cloned into the pSP64 plasmid (kind gift from Prof. Pagani).

The expression plasmids for the ExSpeU1s (pU1) were created from pGEM-U1wt by replacing the sequence between the *Bgl*II and *Xba*I sites with a PCR fragment generated with specific forward primer (containing the modified 5' tail of the U1 snRNA) and a constant reverse primer base pairing downstream the *Xba*I cloning site (5'-ATAGAATACAAGCTTGCATGCCTG-3'). The forward primers used (modified binding region underlined, *Bgl*II in italic) are the following: 5'-AGGCCCA*AGATCT*CATACCCACTATGCAGGGGAGATACCATGATCA-

3' ($U1^{+4}$, modified nucleotide in bold), 5'-AGGCCCA*AGATCT*CATCACATTGCTGCAG GGGAGATACCATGATCA-3' ($U1^{sh7}$), 5'-AGGCCCA*AGATCT*CATACCTCTGCCGCAG GGGAGATACCATGATCA-3' ($U1^{sh16}$), 5'-AGGCCCA*AGATCT*CATAGGTAGGGAG CAGGGGAGATACCATGATCA-3' ($U1^{sh25}$). The obtained PCR fragments were digested with *Stu*I-*Xba*I and cloned into the pU1wt plasmid.

Table 4.1. Primers used for minigene mutagenesis. Modified nucleotides (bold) and codons (underlined) are indicated.

Mutation	Variant	Forward primer sequence (5' --> 3')
c.5999G>C	p.Gly2000Ala	GGTTTTTATAAG**C**TGTTTTTGAGAC
c.6011C>G	p.Thr2004Arg	GGTGTTTTTGAG**G**AGTGGAAATGT
c.6021G>A	p.Met2007Ile	GAGACAGTGGAAA**T**ATTACCATCCAAAGC
c.6037G>A	p.Gly2013Arg	TTACCATCCAAAGCT**A**GAATTTGGCGGGTGG
c.6045G>T	p.Trp2015Cys	GCTGGAATTTGT**T**CGGGTGGAATG
c.6045G>C	p.Trp2015Cys	GCTGGAATTTG**C**CGGGTGGAATG
c.6046C>G	p.Arg2016Gly	CTGGAATTTGG**G**GGGTGGAATG
c.6046C>T	p.Arg2016Trp	CTGGAATTTGG**T**GGGTGGAATG
c.6047G>A	p.Arg2016Gln	GCTGGAATTTGGC**A**GGTGGAATG
c.6047G>T	p.Arg2016Leu	GCTGGAATTTGGC**T**GGTGGAATG
c.6049G>A	p.Val2017Met	GCTGGAATTTGGCGG**A**TGGAATGCCTTATTG
c.6053A>G	p.Glu2018Gly	GAATTTGGCGGGTGG**G**ATGCCTTATTGGCGA
c.6065G>A	p.Gly2022Asp	GAATGCCTTATTG**A**CGAGCATCTACATG
c.6082G>A	p.Gly2028Arg	CATCTACATGCT**A**GGATGAGCACAC
c.6087G>A	p.Met2029Ile	GCTGGGAT**A**AGCACAC
c.6089G>A	p.Ser2030Asn	CTACATGCTGGGATGA**A**CACACTTTTTCTGGTG
c.6092C>T	p.Thr2031Ile	CATGCTGGGATGAGCA**T**ACTTTTTCTGGTGTAC
c.6103G>A	p.Val2035Met	GCACACTTTTTCTG**A**TGTACAGCAATA
c.6104T>C	p.Val2035Ala	GCACACTTTTTCTG**C**CGTACAGCAATA
c.6107A>G	p.Tyr2036Cys	CTTTTTCTGGTGT**G**CAGCAATAGTGAG
c.6108C>T	Tyr2036Tyr	ACACTTTTTCTGGTGTA**T**AGCAATAGTGAGTAGCAA
c.6113A>G	p.Asn2038Ser	TTTTTCTGGTGTACAGCA**G**TAGTGAGTAGCAATGTG
c.6115+1G>A		CTGGTGTACAGCAATA**A**TGAGTAGCAATGTGG
c.6115+2T>C		GGTGTACAGCAATAG**C**GAGTAGCAATGTGGG
c.6115+3G>T		GTGTACAGCAATAGT**T**AGTAGCAATGTGGGC
c.6115+4A>G		GTGTACAGCAATAGTG**G**GTAGCAATGTGGGCAG
c.6115+5G>A		GTACAGCAATAGTGA**A**TAGCAATGTGGGCAG
c.6115+6T>A		GTACAGCAATAGTGAG**A**AGCAATGTGGGCAGAG
c.6115+9C>G		CAGCAATAGTGAGTAG**G**AATGTGGGCAGAGGTTC

Lentiviral vectors for FVIII expression were produced by exploiting a codon-optimized (co) human FVIII cDNA with partially deleted B domain maintaining a 226-amino acid stretch with six N-glycosylation sites (coFVIII SQ N6, kind gift from Prof. McVey, University of Surrey, UK)[188]. All variants were inserted by site-specific mutagenesis using the following primers, and modified nucleotide (bold) and codon (underlined) are indicated: 5'-CTGTACCCCG**C**CGTGTTCGAG-3' (G2000A), 5'-CCGGCATCTGG**G**GGGTGGAGTG-3' (R2016G). The coFVIII variants were subsequently cloned in the lentiviral plasmid pLNT backbone through *Xho*I-*Xba*I restriction sites, producing the pLNT-coFVIII plasmids.
All constructs were validated by sequencing.

4.2.3 Minigene-based assays

The day before transfection, HEK293T cells were seeded in 12-well plates ($2x10^5$ cells/well)[198]. Transfection was performed in Opti-MEM with the Lipofectamine 2000 reagent, according to manufacturer's protocol. Briefly, 500 ng of minigene variants were transfected alone or with a molar excess (1.5X) of the pU1/pU7 plasmids, using a 1:1 DNA:Lipofectamine ratio in a 650-µL final volume. Total RNA was isolated 24 hours post-transfection with Trizol (Life Technologies), reverse-transcribed with a mixture of oligo-dT and random examers, and finally amplified using pTB-specific primers Alfa 2,3 (5'-CAACTTCAAGCTCCTAAGCCACTGC-3') and Bra2 (5'-TAGGATCCGGTCACCAGGAAGTTGGTTAAATCA-3'). Amplified fragments were separated on 2% agarose gel. The band expected from exon 19 inclusion had a length of 356 bp, whereas exon 19 skipping was expected to produce a 239-bp band. Densitometric analysis for the quantification of correct and aberrant transcripts was performed using the ImageJ software (https://imagej.net). To verify the correct splicing, RT-PCR fragments were excised from gel, column-purified (QIAquick Gel Extraction Kit, QIAGEN, Venlo, Netherlands) and sequenced. For denaturing capillary electrophoresis analysis, the amplified fragments were labelled by using primers Alfa 2,3 and the fluorescently-labelled (FAM dye) Bra2 and run on an ABI-3100 instrument (Waltham, MA, USA)[198].

4.2.4 Lentivirus production and titration

Lentiviruses (LVs) were produced by transient co-transfection of HEK293-FT cells as previously described[188]. Briefly, the day before transfection, $2x10^7$ cells were seeded in T175 tissue culture flask. A total of 100 µg of plasmid DNA was used for transfection: 50 µg of pLNT-coFVIII, 17.5 µg of pMD.G2 (vesicular stomatitis virus glycoprotein envelope plasmid) and 32.5 µg pCMVΔ8.74 (gag-pol packaging plasmid). Plasmids were pre-complexed with 1 µL

of 10 mM poly-ethylenimine (Sigma-Aldrich), in 14 mL of OptiMEM for 20 minutes at room temperature. Transfection medium was added to the cells for 3-4 hours and then replaced by 15 mL of fresh complete DMEM medium. Viral supernatants were harvested 48 and 72 hours after transfection and cell debris was removed by centrifugation (5000 rpm for 10 minutes) and filtration (0.45 μm PVDF filter). Supernatants were subsequently concentrated by ultracentrifugation at 23000 rpm for 2 hours at +4°C. LVs were resuspended in 50 μL OptiMEM and 10-μL aliquots were stored at -80°C.

LVs titre was determined using a commercial anti-p24 ELISA kit (Lenti-X p24 Rapid Titer Kit, Takara Bio, CA, USA), according to manufacturer's instructions. The obtained quantification (expressed in pg/mL of core/capsid protein p24) was used to determine an estimated vector titre in infective units (IFU) by assuming that 1 ng of p24 is equivalent to ~1.25 x 10^7 lentiviral particles (LPs) and that, for a typical lentiviral vector, there is one IFU for every 100–1000 LPs.

4.2.5 Expression and characterisation of factor VIII missense variants

Wild-type FVIII and missense variants were expressed in both Chinese hamster ovary (CHO) DG44 and HEK293T cell lines[159,188]. HEK293T were maintained as already described (section 2.2.4). CHO cells were maintained in CD DG44 Medium (Gibco) supplemented with GlutaMAX (8 mM final concentration, Gibco) and Pluronic F-68 (1X, Gibco). Cells were maintained at 37°C with 5% CO_2 and on a shaking platform rotating at 120 rpm.

Cells were transduced at MOI 2 (as determined by p24 titre) and cells and media were harvested at 48, 72 and 96 hours post-transduction.

Genomic DNA was extracted with PureLink Genomic DNA Mini Kit (Invitrogen) and lentiviral copy number was evaluated by quantitative PCR (qPCR) using TaqMan Fast Advanced Master Mix (Applied Biosystem), following manufacturer's instructions. FVIII transgene was detected by two different sets of primers and probes, namely for BDD-FVIII and for WPRE, whereas actin β was used as normalizer.

Secreted FVIII levels were determined using a commercial ELISA kit (Asserachrom VIII:Ag, Diagnostica Stago, Asnières sur Seine, France), as per manufacturer's instructions. Briefly, all samples were tested at two different dilutions and in duplicate, loaded in pre-coated wells (monoclonal mouse anti-human FVIII antibody) and detected by a second peroxidase-coupled anti-FVIII antibody. Presence of FVIII was revealed by TMB substrate after stopping the reaction with a strong acid, and the intensity of the colour was directly proportional to the concentration of FVIII initially present in the sample. Secreted FVIII antigen levels were

expressed as % of the wild-type recombinant FVIII (rFVIII) and normalised on transduction efficiency determined by qPCR.

Cofactor activity in cell media was evaluated by Biophen Factor VIII:C Chromogenic assay (Hyphen Biomed, Neuville sur Oise, France), following manufacturer's instructions. Briefly, the assay is based on the measurement at 405 nm of released pNA, which is directly proportional to FXa activity. In presence of a constant amount of FIXa, phospholipids and calcium, FXa activity generated is directly related to the amount of FVIII in the sample. All samples were tested at two different dilutions and analysed in duplicate. A standard curve, constructed by diluting Biophen plasma calibrator (Hyphen Biomed) 1:20 and carrying out four 1:2 serial dilutions, was run in duplicate. FVIII activity levels were expressed as % of the wild-type rFVIII and normalised on transduction efficiency determined by qPCR.

4.2.6 Data analysis

All data were analysed with GraphPad Prism 5 software (San Diego, CA, USA) and Microsoft Excel 2010.

4.3 Results and discussion

This work focused on *F8* exon 19 because multiple mutations at the 5'ss, possibly rescuable with a U1 snRNA-mediated approach, were reported and associated with severe HA in the European Association for Haemophilia and Allied Disorders (EAHAD) database (http://f8-db.eahad.org/). Moreover, a previous study by this group had revealed the presence of splicing regulatory elements overlapping with the coding region of exon 19 and the presence of missense variants affecting the splicing process[159]. All selected variants are reported in Table 4.2.

Table 4.2. HA-associated point mutations selected for the study from the EAHAD *F8* variant database.

Location	Mutation[a]	Protein variant	Patients
Exon 19	c.5999G>C	p.Gly2000Ala	1
Exon 19	c.6011C>G	p.Thr2004Arg	1
Exon 19	c.6021G>A	p.Met2007Ile	2
Exon 19	c.6037G>A	p.Gly2013Arg	1
Exon 19	c.6045G>T	p.Trp2015Cys	1
Exon 19	c.6045G>C	p.Trp2015Cys	1
Exon 19	c.6046C>T	p.Arg2016Trp	61
Exon 19	c.6046C>G	p.Arg2016Gly	1
Exon 19	c.6047G>A	p.Arg2016Gln	1
Exon 19	c.6047G>T	p.Arg2016Leu	1
Exon 19	c.6049G>A	p.Val2017Met	1
Exon 19	c.6053A>G	p.Glu2018Gly	7
Exon 19	c.6065G>A	p.Gly2022Asp	2
Exon 19	c.6082G>A	p.Gly2028Arg	9
Exon 19	c.6087G>A	p.Met2029Ile	1
Exon 19	c.6089G>A	p.Ser2030Asn	33
Exon 19	c.6092C>T	p.Thr2031Ile	1
Exon 19	c.6103G>A	p.Val2035Met	5
Exon 19	c.6104T>C	p.Val2035Ala	35
Exon 19	c.6107A>G	p.Tyr2036Cys	1
Exon 19	c.6108C>T	p.Tyr2036Tyr	1
Exon 19	c.6113A>G	p.Asn2038Ser	10
Intron 19	c.6115+1G>A		1
Intron 19	c.6115+2T>C		1
Intron 19	c.6115+3G>T		1
Intron 19	c.6115+4A>G		1
Intron 19	c.6115+5G>A		2
Intron 19	c.6115+6T>A		1
Intron 19	c.6115+9C>G		1

[a]NCBI Reference Sequence: NC_000023.11

To investigate exon 19 splicing pattern in the presence of the selected mutations, the well-established minigenes approach has been exploited[161]. A minigene is a plasmid containing a hybrid and simplified version of a gene and is used to produce relatively small pre-mRNA molecules that can be processed by the spliceosome following its transfection into a cell line of interest[199]. The exon to be studied, together with part of the flanking introns, is amplified from genomic DNA and cloned into the plasmid. Once splicing has occurred, the resulting mRNA can be amplified by RT-PCR with primers specifically designed to amplify only the minigene-processed transcripts.

Minigene systems are very versatile and can be adapted to many experimental systems, and are often the preferred way if patient samples are unavailable. It has to be considered that the exon is present in a heterologous context and thus it does not necessarily recapitulate the exact splicing profile of the endogenous gene, although comparative studies have shown that minigenes do often replicate the splicing profile of the endogenous gene[161,200,201].

4.3.1 U7 snRNA-mediated masking of exon 19 resulted in loss of exon definition

To evaluate the presence of regulatory elements in *F8* exon 19, an antisense approach based on U7 snRNAs (physiologically involved in histone RNA 3' end processing) was used. Modified derivatives of the U7 snRNA (U7) have been widely used to target specific sequences on pre-mRNA and to act as antisense oligonucleotides[169,202]. In this work, two U7s were designed to mask 20-bp regions of exon 19 (Fig. 4.5) and to assess the possible presence of exonic regulatory elements in a minigene system.

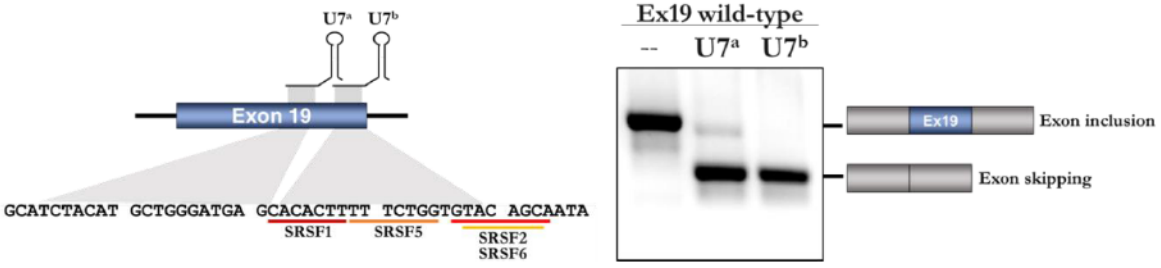

Figure 4.5. Effect of the antisense U7s on exon 19 splicing. Left panel, schematic representation of the U7 mechanism of action and predicted binding sites of SR proteins on exon 19 (ESEfinder, http://rulai.cshl.edu/). Right panel, analysis on 2% agarose gel of the splicing pattern in the absence (--) or in the presence of U7ᵃ and U7ᵇ. The RT-PCR was conducted with minigene-specific primers and the upper band corresponds to exon inclusion, whereas the lower bands correspond to exon skipping.

Co-expression with engineered U7 molecules induced major exon skipping (Fig. 4.5). This revealed the presence in those positions of splicing regulatory elements with enhancer functions, as also predicted by bioinformatic tools (ESEfinder, http://rulai.cshl.edu/)[150]. Several missense

mutations in exon 19 are located in the regions masked by the U7s and this suggests their possible influence on the splicing pattern, besides their role on protein sequence. These results are in line with those obtained by Donadon et al.[159] and further suggest that, due to the overlap of splicing and coding sequences, aberrant splicing can be caused by exonic variants affecting regulatory elements.

4.3.2 Splicing was strongly impaired by mutations at either 5' splice site and exonic regions

All selected twenty-nine variants located within *F8* exon 19 and its donor splice site were characterised for their impact on exon definition and inclusion in the final transcript. In particular, the wild-type and mutated exons were inserted in appropriate minigenes and expressed in HEK293T cells. Analysis of the splicing pattern was performed through RT-PCR with minigene-specific primers (Fig. 4.6).

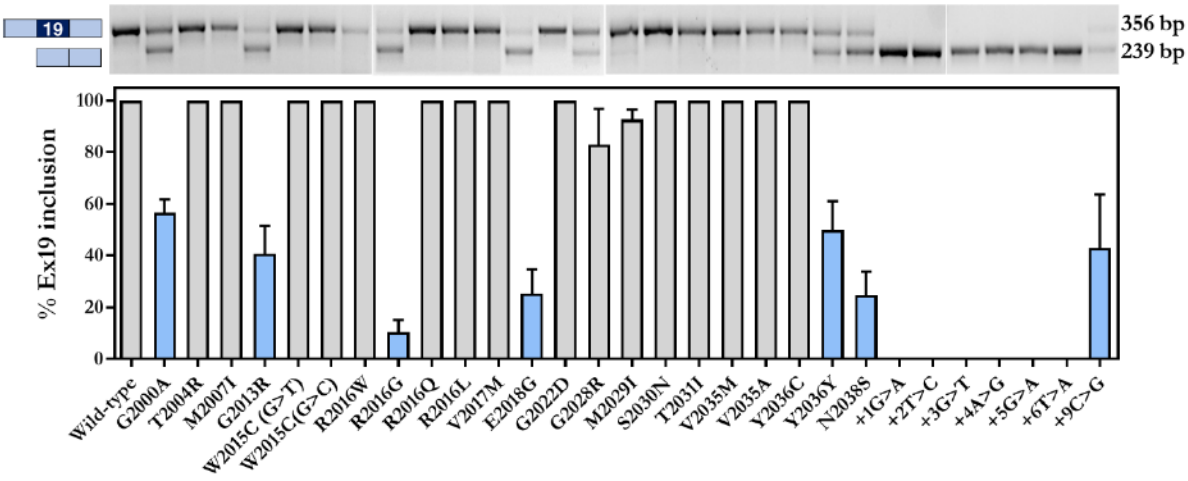

Figure 4.6. Analysis of exon 19 splicing pattern. RT-PCR amplified products from transfected HEK293T cells were resolved on 2% agarose gel (upper panel, representative images) and the amount of exon 19 inclusion was calculated as percentage of the total transcript by densitometry on the obtained bands (lower panel). Blue bars indicate exon inclusion below 60%. Results are reported as mean ± SD from three independent experiments. Ex19, exon 19.

As shown in Figure 4.6, five out of six mutations at the donor splice site caused complete exon skipping, whereas the +9C>G variant reduced correct transcript to 43.0±20.7% of the total. Indeed, the first six intronic positions of the 5'ss are part of the consensus sequence, thus nucleotide changes in this region are more likely to result in exon skipping. Conversely, +9 position is less conserved and this may explain the reduced but still detectable presence of the correct transcript.

In comparison, the majority of the exonic variants did not cause significant exon skipping, whereas six out of twenty-two reduced the correct transcript to levels below 60% of the total (Fig. 4.6, blue bars), thus indicating a role of these positions in exon 19 definition. Notably, p.G2000A variant is caused by a G>C change in the first position of the exon and p.N2038S variant by a A>G change at the third position from the end of the exon. These sites are part of the 3'ss and 5'ss consensus sequences, respectively, and are thus likely to be critical for exon definition. Conversely, the mechanism underlying aberrant splicing caused by the remaining four variants is more difficult to infer. To help the interpretation of the results obtained by the minigenes analysis, a bioinformatic tool that predicts the effects of exonic variants on splicing regulatory elements was employed. HOT-SKIP (https://hot-skip.img.cas.cz/)[203] is an open-source tool that systematically examines all possible substitutions in each exonic position and calculates the total number of ESSs, ESEs and their ratio. As reported in Table 4.3, three of the analysed changes (p.R2016G, p.E2018G and p.Y2036Y) led to an increased ESS/ESE ratio, thus supporting the observed decrease in exon definition. Differently, the p.G2013R variant was not predicted to result in a relevant imbalance, suggesting the presence of additional, still not identified regulatory elements.

Table 4.3. *In silico* analysis of the effects of identified exonic variants on ESS/ESE ratio, using HOT-SKIP algorithm. The original nucleotide is the first reported, whereas the second corresponds to the mutant.

Variant	Nucleotide change	ESSs	ESEs	ESS/ESE ratio
G2013R	G	2	14	**0.14**
	A	1	8	**0.13**
R2016G	C	2	1	**2.00**
	G	24	0	**24.00**
E2018G	A	2	10	**0.20**
	G	10	1	**10.00**
Y2036Y	C	0	4	**0.00**
	T	7	0	**7.00**

Even though *in silico* prediction tools should be used with great care and experimental analysis is preferred, the combinatorial use of these resources still represents a useful option for the interpretation of the effects of exonic variants on the process of splicing[147]. Altogether, these results showed that impaired exon 19 splicing was caused by mutations affecting both 3' and 5' splice sites, as well as exonic regulatory elements.

4.3.3 Missense variants showed differential impact on factor VIII protein biology

Following the identification of six exonic mutations with strong influence on splicing (p.G2000A, p.G2013R, p.R2016G, p.E2018G, p.Y2036Y and p.N2038S), further characterisation of their pleiotropic effects was provided by expressing the resulting FVIII missense variants (Fig. 4.7). Since the very low secretion efficiency of recombinant full-length FVIII hampers the evaluation of missense variants with possible low secretion and/or activity levels, lentiviral-mediated delivery of a codon-optimized B-domainless FVIII cDNA was exploited[188]. This cassette includes the strong and ubiquitous spleen focus forming virus (sffv) promoter and had shown expression levels ~30-fold higher than the full-length FVIII with no codon optimization[188]. Three of the selected variants (p.G2013R, p.E2018G and p.N2038S), had already been characterised by Donadon et al.[159] and in this study the remaining were analysed.

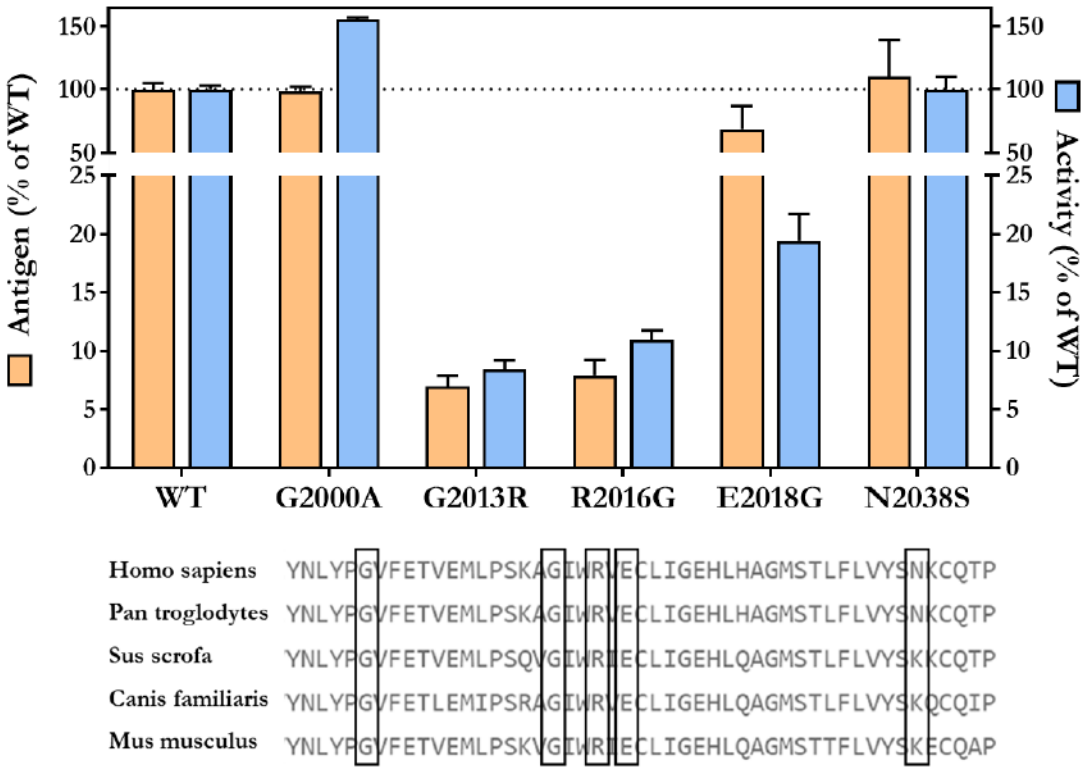

Figure 4.7. Characterisation of FVIII missense variants. Upper panel, secreted FVIII antigen (orange bars) and cofactor activity (blue bars) levels expressed as % of wild-type rFVIII and normalized on virus copy number per cell determined by qPCR. Results are reported as mean ± SD from three independent experiments. Lower panel, alignment of FVIII sequences (Homo sapiens NP_000123.1; Pan troglodytes XP_003317837.1; Sus scrofa NP_999332.2; Canis familiaris NP_001003212.1; Mus musculus NP_001154845.1). Investigated residues are reported (rectangles). WT, wild-type recombinant FVIII.

The expressed missense variants showed differential impact on FVIII secretion and function (Fig. 4.7).

The Gly-to-Arg and Arg-to-Gly changes at positions 2013 and 2016, respectively, resulted in strong reduction of FVIII secretion levels, below 10% of the wild-type. Correspondingly, functional levels were also strongly reduced and comparable to the protein levels. This resulted in specific activities of 1.20±0.04 for p.G2013R and 1.42±0.31 for p.R2016G variants, thus indicating that both these amino acid changes are likely to primarily affect FVIII secretion, rather than FVIII function. Conversely, p.E2018G variant showed secreted levels similar to the wild-type but strongly impaired cofactor activity (19.4±2.3% of wild-type), thus indicating a major functional defect[159]. Interestingly, the two remaining missense variants p.G2000A and p.N2038S revealed no deleterious effects on either FVIII secretion or activity, thus suggesting a major pathological role for the splicing defect rather than for the altered protein sequence. Indeed, p.G2000A variant, albeit at a highly conserved position, is characterised by a rather conservative amino acid change. Moreover, crystallographic structure of FVIII revealed that p.G2000 is an exposed residue located in a random coil area of the A3 domain, further supporting the tolerant nature of this position. Similarly, N2038 is an exposed residue and it is poorly conserved among different species, suggesting its tolerance in terms of amino acid changes[159].

Overall, of the six exonic variants found to be associated with aberrant splicing, i) three had also a detrimental impact on either FVIII secretion or activity (p.G2013R, p.R2016G, p.E2018G), ii) two were located in a tolerant position and did not alter protein biology (p.G2000A, p.N2038S), and iii) one was a synonymous variant, thus, by definition, not modifying the amino acid sequence of FVIII (p.Y2036Y).

The knowledge of the impact of missense changes on FVIII protein biology is of extreme interest in the context of a splicing-correction approach. Indeed, the ideal candidates for ExSpeU1-mediated rescue are those missense variants that do not strongly alter FVIII secretion nor function, as well as synonymous mutations.

4.3.4 An optimized ExSpeU1 rescued exon 19 inclusion with high efficiency

To reduce possible off-target effects, the second generation of exon-specific U1 snRNAs (ExSpeU1) was exploited, targeting non-conserved intronic sequences downstream of the 5'ss. In particular, three different "shifted" ExSpeU1s were designed ($U1^{sh7}$, $U1^{sh16}$, $U1^{sh25}$), whose 5' tails were modified to base pair at 7, 16 or 25 positions downstream of the donor site of exon 19.

To preliminary identify the most effective molecule, all ExSpeU1s were screened upon the same minigene, harbouring the +4A>G variant (Fig. 4.8). In addition, a modified U1 snRNA with a modified 5' tail that base pair exactly to the mutant donor site was also tested ($U1^{+4}$).

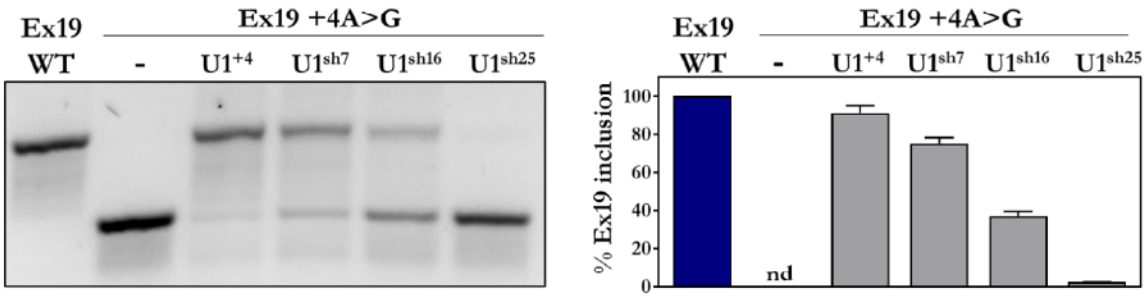

Figure 4.8. Analysis of exon 19 splicing pattern after co-expression of U1-based correction molecules. RT-PCR amplified products were resolved on 2% agarose gel (left panel, representative image). Percentage of correct transcript was evaluated by densitometry (right panel). Results are reported as mean ± SD from three independent experiments. WT, minigene with wild-type exon 19; nd, not detectable.

Minigene analysis showed that the co-expression of +4A>G mutant minigene with modified U1 snRNAs produced variable splicing rescue (Fig. 4.8, lanes 3-6), whereas the mutant RNA had the expected aberrant splicing pattern caused by the intronic point mutation (i.e. complete exon skipping, lane 2) and the wild-type minigene was correctly processed (lane 1). The most efficient correction was mediated by the complementary $U1^{+4}$, showing almost complete inclusion of exon 19 (91.2±3.9% of the total transcript). However, the correction effect of this molecule is restricted to the +4A>G specific mutation and its potential complementarity to other normal 5'ss may determine unwanted off-target effects[175]. Conversely, ExSpeU1s were designed to bind to non-conserved intronic regions, thus potentially improving their specificity. Notably, the ExSpeU1 base pairing at 7 positions downstream of 5'ss ($U1^{sh7}$) induced relevant splicing rescue (75.2±3.1% of the total transcript), and was therefore chosen as the optimal candidate for the correction approach. The others ExSpeU1s ($U1^{sh16}$, $U1^{sh25}$) showed modest correction ability, apparently dependent upon their distance from the 5'ss, and were not pursued in the study.

Overall, screening of different modified U1 snRNAs identified a unique ExSpeU1 ($U1^{sh7}$) with strong correction activity and expected to have reduced off-target effects due to its complementarity with a non-conserved intronic region.

4.3.5 A unique ExSpeU1 was able to rescue correct splicing of 5'ss and exonic variants

Once identified the molecule with the optimal splicing correction efficiency ($U1^{sh7}$), the approach was extended to the entire panel of exon 19 mutations that had been previously shown to cause aberrant splicing (Fig. 4.9). The three missense variants found to be detrimental for FVIII biology, namely p.G2013R, p.R2016G and p.E2018G, were not challenged with the ExSpeU1-based approach because rescue of correct splicing would nevertheless translate into the synthesis of a strongly impaired protein.

Splicing pattern was evaluated by RT-PCR and agarose gel electrophoresis, as well as by sequencing of RT-PCR products. Some of the variants were further analysed by capillary electrophoresis.

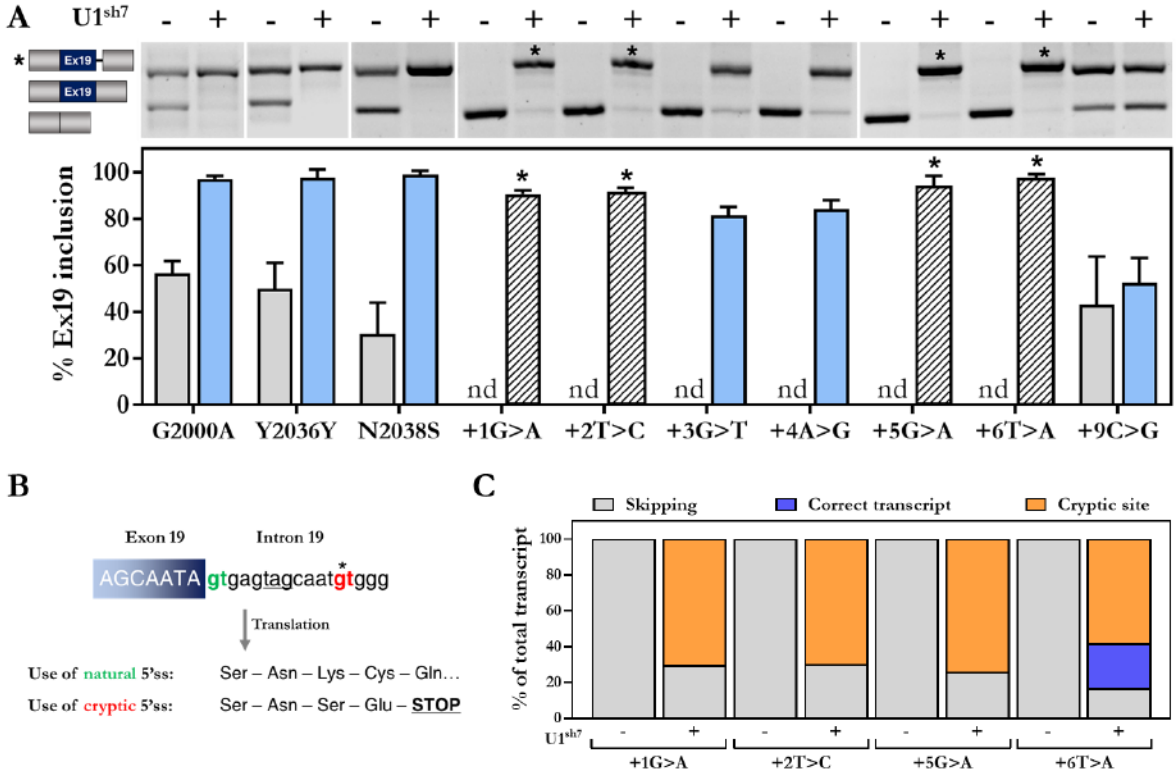

Figure 4.9. Analysis of exon 19 splicing pattern after co-expression with $U1^{sh7}$. (A) Minigene-based analysis of the splicing pattern in the absence (grey bars) or in the presence (blue bars) of $U1^{sh7}$. RT-PCR amplified products were resolved on 2% agarose gel (upper panel, representative images) and percentage of correct transcript, evaluated by densitometry, is reported as mean ± SD from three independent experiments. Transcripts deriving from use of a cryptic 5'ss are indicated by asterisks and striped bars. (B) Partial sequence of exon 19 donor site. Both natural (green) and cryptic (red, asterisk) splice sites are indicated. The deriving amino acid sequences are indicated (lower panel) as well as the new stop codon introduced by the use of cryptic site (underlined). (C) Analysis of transcripts by capillary electrophoresis. Results are presented as percentage of the total transcript in the absence (-) or in the presence (+) of $U1^{sh7}$.

Analysis of the splicing pattern of mutant minigenes after co-expression with U1^{sh7} revealed a variable response in terms of production of the correct transcript (Fig. 4.9, panel A).

The three exonic variants (p.G2000A, p.Y2036Y and p.N2038S) showed complete rescue after co-expression with the U1^{sh7}, as confirmed by sequencing of the RT-PCR products. Similarly, the intronic variants +3G>T and +4A>G, which had shown complete exon skipping without treatment, were almost completely rescued after co-expression with the U1^{sh7}, resulting in correct transcript levels of 81.4±3.7% and 84.2±3.8% of the total transcript, respectively. On the contrary, treatment of +9C>G variant resulted in a non-relevant rescue of exon 19 inclusion (from 43.0±20.7% to 52.2±10.8% of the total), probably due to the decreased complementarity of the U1^{sh7} to the mutated intronic region caused by the +9C>G variant itself. Finally, co-expression of U1^{sh7} with +1G>A, +2T>C, +5G>A and +6T>A variants induced the usage of a cryptic splice site located 12 nucleotides downstream of the natural 5'ss, resulting in retention of part of intron 19 in the mature transcript (Fig. 4.9, panel B). Further analysis through capillary electrophoresis (Fig. 4.9, panel C) confirmed that, after U1^{sh7} co-expression, the major part of the transcript resulted from the cryptic site usage (+1G>A, 70.4% of the total; +2T>C, 69.9%; +5G>A, 74.0%; +6T>C, 58.2%), while the correct transcript was produced in a minor part and only for the +6T>A variant after treatment (25.1% of total transcript). Cryptic donor site usage resulted in the presence of part of intron 19 inside of the mature mRNA, also including an in-frame UAG premature termination codon (Fig. 4.9, panel B, underlined). The presence of a PTC in the mature transcript may result in mRNA degradation by nonsense-mediated decay and/or in the synthesis of a truncated FVIII lacking around 300 residues at its C-terminal end. Notably, this portion corresponds to the entire C1 and C2 domains which are essential for FVIII stability and function. Therefore, cryptic site usage would result in the production of a truncated, unstable and non-functional protein.

Overall, these results showed that, in this experimental system, aberrant splicing caused by two of the intronic variants as well as by all tested exonic variants was completely rescued by a unique ExSpeU1. It is worth to note that, if translated at the protein level, correction of the intronic (+3G>T and +4A>G) and synonymous (pY2036Y) variants would result in the production of a wild-type FVIII, and correction of the two missense variants (p.G2000A and p.N2038S) would result in the synthesis of efficiently secreted and fully functional FVIII missense variants. Altogether, this study identified a unique ExSpeU1 able to correct five different splicing mutations, located at either the 5'ss or in exonic sequences, thus providing the first evidence of the feasibility of this approach for a panel of HA-causing variants.

Notably, in the context of HA, even low increases of FVIII activity levels, compatible with a splicing correction approach, would result in a significant amelioration of the patient's phenotype. Indeed, FVIII levels around 1-5% are associated with a moderate phenotype and levels >5% with a mild one, with consequent eradication of spontaneous bleedings.

It is worth noting that therapeutic strategies based on binding to complementary target sequences may have off-target effects that need to be carefully addressed. To date, however, promising results have been obtained on ExSpeU1 efficacy and specificity. In a recent study by Donadon and colleagues, an AAV9-delivered ExspeU1 was able to efficiently recover severe SMA mice[179]. They demonstrated that even very low levels of ExSpeU1 expression (0.1–0.3% of the wild-type U1) were sufficient for significant correction of aberrant splicing and pathologic phenotype. In addition, over-expression studies proved ExSpeU1's high specificity and ability to promote robust splicing correction without inducing unintended changes in gene expression or alternative splicing[173,179].

Overall, the available data on the ExSpeU1-based approach are very promising and this strategy presents several advantages. First, the short length of the ExSpeU1 cassette (~500 bp) represents a concrete advantage in the context of HA, because the large dimensions of *F8* gene still represent a limiting step for its insertion in AAV vectors and gene therapy applications. Moreover, the splicing-targeted correction maintains the regulation of gene expression in the correct cell-specific chromosomal context. Finally, binding of the ExSpeU1s to not conserved intronic sequences, beside enhancing specificity, has the potential to be effective on multiple mutations, thus expanding the therapeutic potential of these molecules.

Altogether, the data presented here, albeit preliminary and with some limitations, represent the first proof-of-principle of the ExSpeU1s-mediated correction approach in the context of Haemophilia A.

4.4 Conclusions

This work provided the proof-of-principle of a correction strategy based on ExSpeU1 to rescue different types of HA-causing mutations (splicing, missense, synonymous) sharing the same pathological molecular mechanism.

In particular, the presence in *F8* exon 19 of positive regulatory elements, already suggested by others[159], was further confirmed by a U7 snRNA-mediated antisense approach. Subsequently, among all point mutations affecting exon 19 (http://f8-db.eahad.org/), thirteen variants associated to aberrant splicing were identified. The further characterisation of exonic variants for their impact on FVIII protein biology revealed that two out of five missense variants affecting splicing did not affect FVIII secretion nor function thus representing, together with a synonymous mutation, the ideal candidates for a splicing-correction strategy. Finally, ten variants were challenged with a selected ExSpeU1. Here splicing pattern analysis showed that a unique molecule was able to rescue five variants, either located at the donor splice site or in the exon. Overall, this work provides a preliminary evidence of the potential of ExSpeU1 to correct multiple variants affecting different splicing elements, thus expanding its therapeutic potential.

Further studies will evaluate the ExSpeU1-mediated splicing correction at the protein level, to assess whether the treatment results in a consistent rescue of FVIII biosynthesis and function. This phase will take advantage of splicing-competent constructs, in which exon 19 (either wild-type or mutated) and its flanking introns will be inserted in a FVIII-expressing vector[176]. Moreover, to overcome the lack of HA mice harbouring splicing mutations, a HA mouse model will be created by lentiviral-mediated delivery of splicing-competent FVIII cassettes. In this context, the AAV-mediated delivery of ExSpeU1 will allow the evaluation of the therapeutic potential of this molecule and the investigation of possible off-target effects.

Chapter 5

General conclusions

Many efforts have been done to develop a definitive and lifelong cure for genetic diseases, and promising results have been recently obtained thanks to major advances in gene therapy approaches[64,127,204]. Nevertheless, important drawbacks still exist, thus supporting investigation of alternative treatment strategies.
This thesis explored three different approaches in different genetic diseases and concluding considerations are presented here for each of them.

In Chapter 2, investigation of **drug-induced readthrough** in the context of **Fabry disease** identified three nonsense variants that, due to favourable nucleotide and protein features, could be rescued by readthrough induction. Nevertheless, as highlighted also by the variable outcome of preclinical and clinical studies, the functional rescue achievable by a nonsense suppression approach strongly depends on the combination of several conditions that need to be carefully considered in every specific context (i.e. favourable sequence context and re-inserted amino acids). This work revealed an additional level of complexity for readthrough-mediated rescue, namely the homodimeric nature of the α-galactosidase A enzyme. This finding is extremely relevant in the context of nonsense suppression approaches for FD and for other disorders involving dimeric or multimeric enzymes. Moreover, if occurring in heterozygous females affected by FD, an impaired heterodimerization process could help interpret the pathological phenotypes observed.

In Chapter 3, the design and characterisation of a novel **factor IX-albumin fusion protein** was performed in order to produce an improved molecule for replacement therapy of **Haemophilia B**. Particularly, a gain-of-function FIX endowed of 8-to-15-fold improved pro-coagulant activity was fused to a rationally engineered HSA variant with increased FcRn binding affinity, resulting in a 2.5-fold extended half-life in a panel of mouse models with different FcRn/albumin settings. Direct comparison with the commercial FIX-HSA fusion protein (Idelvion®) demonstrated the concrete improvement provided by this novel chimaera in terms of half-life. The rational selection of optimal fusion partners performed here can be extremely relevant also for the improvement of other therapeutic proteins. Further studies will evaluate half-life and pro-coagulant activity of the engineered fusion protein in a double transgenic mouse model, expressing a hybrid human FcRn and human albumin, thus providing an additional evaluation of physiological competition.

In Chapter 4, a U1 snRNA-based **splicing correction** approach was explored in the context of **Haemophilia A**. The extensive analysis of all reported variants on *F8* exon 19 identified thirteen point mutations with detrimental impact on the splicing process, disrupting either the donor splice site or exonic regulatory elements. Three exonic variants were also found to be

associated with virtually normal FVIII secretion and cofactor activity levels, thus supporting the potential benefit conferred by splicing correction of these mutations. A selected ExSpeU1 was subsequently challenged over the whole panel of splicing-affecting variants and completely rescued the correct transcript for three exonic and two intronic variants. Future studies will address this splicing correction approach at the protein level through an *in vitro* expression system, as well as at the phenotypic level through the AAV-mediated delivery of ExSpeU1 in a Haemophilia A mouse model.

Overall, this thesis provided the first proof-of-principle of:

i. the feasibility of a nonsense suppression approach for Fabry disease, also highlighting the need for careful preliminary consideration of the several determinants of functional readthrough, not least the possible impact on quaternary protein structure;
ii. the improved pro-coagulant and pharmacokinetic features of a novel FIX-HSA fusion protein conferred by the gain-of-function FIX Padua variant and the rationally engineered HSA^{QMP} variant designed to bind FcRn with higher affinity;
iii. the therapeutic potential of a single ExSpeU1 to correct splicing defects caused by multiple *F8* gene mutations, thus supporting further studies on the efficacy and specificity of this promising approach.

Beside hopefully contributing to the development of novel and more effective treatments for human diseases, the study of alternative therapeutic approaches has the meaningful potential to **drive basic research** as well, thus expanding our knowledge and providing new directions for future investigations.

References

1. Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.* **29**, 1037–1047 (2008).
2. Isken, O. & Maquat, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Amp Dev.* **21**, 1833–3856 (2007).
3. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).
4. Inoue, K. *et al.* Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat. Genet.* **36**, 361–369 (2004).
5. Keeling, K. M., Xue, X., Gunn, G. & Bedwell, D. M. Therapeutics Based on Stop Codon Readthrough. *Annu. Rev. Genomics Hum. Genet.* **15**, 371–394 (2014).
6. Lee, H.-L. R. & Dougherty, J. P. Pharmaceutical therapies to recode nonsense mutations in inherited diseases. *Pharmacol. Ther.* **136**, 227–266 (2012).
7. Nagel-Wolfrum, K., Möller, F., Penner, I., Baasov, T. & Wolfrum, U. Targeting Nonsense Mutations in Diseases with Translational Read-Through-Inducing Drugs (TRIDs). *BioDrugs* **30**, 49–74 (2016).
8. Salas-Marco, J. & Bedwell, D. M. GTP Hydrolysis by eRF3 Facilitates Stop Codon Decoding during Eukaryotic Translation Termination. *Mol. Cell. Biol.* **24**, 7769–7778 (2004).
9. Alkalaeva, E. Z., Pisarev, A. V., Frolova, L. Y., Kisselev, L. L. & Pestova, T. V. In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. *Cell* **125**, 1125–1136 (2006).
10. Roy, B. *et al.* Ataluren stimulates ribosomal selection of near-cognate tRNAs to promote nonsense suppression. *Proc. Natl. Acad. Sci.* **113**, 12508–12513 (2016).
11. Bidou, L., Rousset, J.-P. & Namy, O. Translational errors: from yeast to new therapeutic targets: Translational fidelity. *FEMS Yeast Res.* **10**, 1070–1082 (2010).
12. Hirsh, D. Tryptophan transfer RNA as the UGA suppressor. *J. Mol. Biol.* **58**, 439–458 (1971).
13. Beznosková, P., Wagner, S., Jansen, M. E., von der Haar, T. & Valášek, L. S. Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Res.* **43**, 5099–5111 (2015).
14. Eswarappa, S. M. *et al.* Programmed Translational Readthrough Generates Antiangiogenic VEGF-Ax. *Cell* **157**, 1605–1618 (2014).
15. McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J. & Tate, W. P. Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc. Natl. Acad. Sci.* **92**, 5431–5435 (1995).

16. Manuvakhova, M., Keeling, K. & Bedwell, D. M. Aminoglycoside antibiotics mediate context-dependent suppression of termination codons in a mammalian translation system. *RNA* **6**, 1044–1055 (2000).

17. Chen, C.-Y. A. & Shyu, A.-B. Mechanisms of deadenylation-dependent decay: Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2**, 167–183 (2011).

18. Cosson, B. *et al.* Poly(A)-Binding Protein Acts in Translation Termination via Eukaryotic Release Factor 3 Interaction and Does Not Influence [PSI+] Propagation. *Mol. Cell. Biol.* **22**, 3301–3315 (2002).

19. Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112–8 (2004).

20. Brown, C. M., Stockwell, P. A., Trotman, C. N. A. & Tate, W. P. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res.* **18**, 6339–6345 (1990).

21. Bonetti, B., Fu, L., Moon, J. & Bedwell, D. M. The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences inSaccharomyces cerevisiae. *J. Mol. Biol.* **251**, 334–345 (1995).

22. Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524**, 493–496 (2015).

23. Klauer, A. A. & van Hoof, A. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress: Degradation of mRNAs that lack a stop codon. *Wiley Interdiscip. Rev. RNA* **3**, 649–660 (2012).

24. Arribere, J. A. *et al.* Translation readthrough mitigation. *Nature* **534**, 719–723 (2016).

25. Blanchet, S., Cornu, D., Argentini, M. & Namy, O. New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae. *Nucleic Acids Res.* **42**, 10061–10072 (2014).

26. Pranke, I. *et al.* Factors influencing readthrough therapy for frequent cystic fibrosis premature termination codons. *ERJ Open Res.* **4**, 00080–02017 (2018).

27. Westhof, E., Yusupov, M. & Yusupova, G. Recognition of Watson-Crick base pairs: constraints and limits due to geometric selection and tautomerism. *F1000Prime Rep.* **6**, (2014).

28. Roy, B., Leszyk, J. D., Mangus, D. A. & Jacobson, A. Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. *Proc. Natl. Acad. Sci.* **112**, 3038–3043 (2015).

29. Ogle, J. M. Recognition of Cognate Transfer RNA by the 30S Ribosomal Subunit. *Science* **292**, 897–902 (2001).

30. Howard M, Frizzell RA, Bedwell DM. Aminoglycoside antibiotics restore CFTR function by overcoming premature stop mutations. *Nat Med* **2**, 467–9 (1996).

31. Keeling, K. Nonsense Suppression as an Approach to Treat Lysosomal Storage Diseases. *Diseases* **4**, 32 (2016).

32. Fan-Minogue, H. & Bedwell, D. M. Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. *RNA* **14**, 148–157 (2007).

33. Wilhelm, J. M., Jessop, J. J. & Pettitt, S. E. Aminoglycoside antibiotics and eukaryotic protein synthesis: stimulation of errors in the translation of natural messengers in extracts of cultured human cells. *Biochemistry* **17**, 1149–1153 (1978).

34. Pinotti, M. *et al.* Intracellular readthrough of nonsense mutations by aminoglycosides in coagulation factor VII. *J. Thromb. Haemost.* **4**, 1308–1314 (2006).

35. Clancy, J. P. *et al.* Evidence that Systemic Gentamicin Suppresses Premature Stop Mutations in Patients with Cystic Fibrosis. *Am. J. Respir. Crit. Care Med.* **163**, 1683–1692 (2001).

36. Wagner, K. R. *et al.* Gentamicin treatment of Duchenne and Becker muscular dystrophy due to nonsense mutations. *Ann. Neurol.* **49**, 706–711 (2001).

37. Guthrie, O. W. Aminoglycoside induced ototoxicity. *Toxicology* **249**, 91–96 (2008).

38. Lopez-Novoa, J. M., Quiros, Y., Vicente, L., Morales, A. I. & Lopez-Hernandez, F. J. New insights into the mechanism of aminoglycoside nephrotoxicity: an integrative point of view. *Kidney Int.* **79**, 33–45 (2011).

39. Dehne, N., Rauen, U., de Groot, H. & Lautermann, J. Involvement of the mitochondrial permeability transition in gentamicin ototoxicity. *Hear. Res.* **169**, 47–55 (2002).

40. Hobbie, S. N. *et al.* Genetic analysis of interactions with eukaryotic rRNA identify the mitoribosome as target in aminoglycoside ototoxicity. *Proc. Natl. Acad. Sci.* **105**, 20888–20893 (2008).

41. Campbell, K. C. M. *et al.* Prevention of noise- and drug-induced hearing loss with d-methionine. *Hear. Res.* **226**, 92–103 (2007).

42. Schiffelers, R. Liposome-encapsulated aminoglycosides in pre-clinical and clinical studies. *J. Antimicrob. Chemother.* **48**, 333–344 (2001).

43. Nudelman, I. *et al.* Redesign of aminoglycosides for treatment of human genetic diseases caused by premature stop mutations. *Bioorg. Med. Chem. Lett.* **16**, 6310–6315 (2006).

44. Welch, E. M. *et al.* PTC124 targets genetic disorders caused by nonsense mutations. *Nature* **447**, 87–91 (2007).

45. Haas, M. *et al.* European Medicines Agency review of ataluren for the treatment of ambulant patients aged 5 years and older with Duchenne muscular dystrophy resulting from a nonsense mutation in the dystrophin gene. *Neuromuscul. Disord.* **25**, 5–13 (2015).

46. McDonald, C. M. *et al.* Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet* **390**, 1489–1498 (2017).

47. Kerem, E. *et al.* Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Respir. Med.* **2**, 539–547 (2014).

48. Zarate, Y. A. & Hopkin, R. J. Lysosomal Storage Disease 3 Fabry's disease. **372**, 9 (2008).

49. Mehta, A. *et al.* Fabry disease defined: baseline clinical manifestations of 366 patients in the Fabry Outcome Survey. *Eur. J. Clin. Invest.* **34**, 236–242 (2004).

50. Cairns, T. *et al.* Hot topics in Fabry disease. *Postgrad. Med. J.* **94**, 709–713 (2018).

51. Burlina, A. B. *et al.* Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. *J. Inherit. Metab. Dis.* **41**, 209–219 (2018).

52. Mechtler, T. P. *et al.* Neonatal screening for lysosomal storage disorders: feasibility and incidence from a nationwide study in Austria. *The Lancet* **379**, 335–341 (2012).

53. Arends, M. *et al.* Characterization of Classical and Nonclassical Fabry Disease: A Multicenter Study. *J. Am. Soc. Nephrol.* **28**, 1631–1641 (2017).

54. Wang, R. Y., Lelis, A., Mirocha, J. & Wilcox, W. R. Heterozygous Fabry women are not just carriers, but have a significant burden of disease and impaired quality of life. *Genet. Med.* **9**, 34–45 (2007).

55. Deegan, P. B. Natural history of Fabry disease in females in the Fabry Outcome Survey. *J. Med. Genet.* **43**, 347–352 (2005).

56. Desnick RJ, Astrin KH, Bishop DF. Fabry disease: molecular genetics of the inherited nephropathy. *Adv Nephrol Necker Hosp* **18**, 113–27 (1989).

57. Elstein, D., Altarescu, G. & Beck, M. *Fabry Disease*. (Springer Science+Business Media B.V., 2010).

58. Medin, J. A. & Murray, G. J. Correction in trans for Fabry disease: Expression, secretion, and uptake of a-galactosidase A in patient-derived cells driven by a high-titer recombinant retroviral vector. *Proc Natl Acad Sci USA* 6 (1996).

59. Garman, S. C. & Garboczi, D. N. The Molecular Defect Leading to Fabry Disease: Structure of Human α-Galactosidase. *J. Mol. Biol.* **337**, 319–335 (2004).

60. Meng, X.-L. *et al.* Molecular basis for globotriaosylceramide regulation and enzyme uptake in immortalized aortic endothelial cells from Fabry mice. *J. Inherit. Metab. Dis.* **39**, 447–455 (2016).

61. Parenti, G., Andria, G. & Ballabio, A. Lysosomal Storage Diseases: From Pathophysiology to Therapy. *Annu. Rev. Med.* **66**, 471–486 (2015).

62. Hughes, D. A. *et al.* Oral pharmacological chaperone migalastat compared with enzyme replacement therapy in Fabry disease: 18-month results from the randomised phase III ATTRACT study. *J. Med. Genet.* **54**, 288–296 (2017).

63. Guérard, N. *et al.* Lucerastat, an Iminosugar for Substrate Reduction Therapy: Tolerability, Pharmacodynamics, and Pharmacokinetics in Patients With Fabry Disease on Enzyme Replacement. *Clin. Pharmacol. Ther.* **103**, 703–711 (2018).

64. Tuttolomondo, A., Simonetta, I. & Pinto, A. Gene Therapy of Anderson-Fabry Disease. *Curr. Gene Ther.* **19**, 3–5 (2019).

65. Matalonga, L. *et al.* Effect of Readthrough Treatment in Fibroblasts of Patients Affected by Lysosomal Diseases Caused by Premature Termination Codons. *Neurotherapeutics* **12**, 874–886 (2015).

66. Clarke, J. T. R., West, M. L., Bultas, J. & Schiffmann, R. The pharmacology of multiple regimens of agalsidase alfa enzyme replacement therapy for Fabry disease. *Genet. Med.* **9**, 504–509 (2007).

67. den Dunnen, J. T. *et al.* HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* **37**, 564–569 (2016).

68. Branchini, A. *et al.* Differential functional readthrough over homozygous nonsense mutations contributes to the bleeding phenotype in coagulation factor VII deficiency. *J. Thromb. Haemost.* **14**, 1994–2000 (2016).

69. Lukas, J. *et al.* Functional Characterisation of Alpha-Galactosidase A Mutations as a Basis for a New Classification System in Fabry Disease. *PLoS Genet.* **9**, e1003632 (2013).

70. Pignani, S. *et al.* The chaperone-like sodium phenylbutyrate improves factor IX intracellular trafficking and activity impaired by the frequent p.R294Q mutation. *J. Thromb. Haemost.* **16**, 2035–2043 (2018).

71. Siekierska, A. *et al.* α-Galactosidase Aggregation Is a Determinant of Pharmacological Chaperone Efficacy on Fabry Disease Mutants. *J. Biol. Chem.* **287**, 28386–28397 (2012).

72. Dabrowski, M., Bukowy-Bieryllo, Z. & Zietkiewicz, E. Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biol.* **12**, 950–958 (2015).

73. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).

74. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1-7.20.41 (2013).

75. Lemansky, P., Bishop, D. F., Desnick, R. J., Hasilik, A. & von Figura, K. Synthesis and processing of alpha-galactosidase A in human fibroblasts. Evidence for different mutations in Fabry disease. *J. Biol. Chem.* **262**, 2062–2065 (1987).

76. Garman, S. C. Structure-function relationships in α-galactosidase A: Structure-function relationships in α-galactosidase A. *Acta Paediatr.* **96**, 6–16 (2007).

77. Mei, G., Di Venere, A., Rosato, N. & Finazzi-Agrò, A. The importance of being dimeric: The importance of being dimeric. *FEBS J.* **272**, 16–27 (2004).

78. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. 31.

79. Miyamura, N. *et al.* A carboxy-terminal truncation of human alpha-galactosidase A in a heterozygous female with Fabry disease and modification of the enzymatic activity by the

carboxy-terminal domain. Increased, reduced, or absent enzyme activity depending on number of amino acid residues deleted. *J. Clin. Invest.* **98**, 1809–1817 (1996).

80. Keshav S. Gastrointestinal manifestations of Fabry disease. in *Fabry Disease: Perspectives from 5 Years of FOS* Chapter 28 (2006).

81. Juchniewicz, P. *et al.* Female Fabry disease patients and X-chromosome inactivation. *Gene* **641**, 259–264 (2018).

82. Maier EM, Osterrieder S, Whybra C, Ries M, Gal A, Beck M, Roscher AA, Muntau & AC. Disease manifestations and X inactivation in heterozygous females with Fabry disease. *Acta Paediatr Suppl* **95**, 30–8 (2006).

83. Ishii, S. *et al.* Mutant α-galactosidase A enzymes identified in Fabry disease patients with residual enzyme activity: biochemical characterization and restoration of normal intracellular processing by 1-deoxygalactonojirimycin. *Biochem. J.* **406**, 285–295 (2007).

84. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136–1145 (2018).

85. Strohl, W. R. Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. *BioDrugs* **29**, 215–239 (2015).

86. Zaman, R. *et al.* Current strategies in extending half-lives of therapeutic proteins. *J. Controlled Release* **301**, 176–189 (2019).

87. Peters T. *All about albumin: biochemistry, genetics, and medical applications*. (Academic Press, 1996).

88. Kragh-Hansen, U., Chuang, V. T. G. & Otagiri, M. Practical Aspects of the Ligand-Binding and Enzymatic Properties of Human Serum Albumin. *Biol. Pharm. Bull.* **25**, 695–704 (2002).

89. Fasano, M. *et al.* The extraordinary ligand binding properties of human serum albumin. *IUBMB Life Int. Union Biochem. Mol. Biol. Life* **57**, 787–796 (2005).

90. Dockal, M., Carter, D. C. & Rüker, F. The Three Recombinant Domains of Human Serum Albumin: Structural Characterization and Ligand Binding Properties. *J. Biol. Chem.* **274**, 29303–29310 (1999).

91. Chaudhury, C. *et al.* The Major Histocompatibility Complex–related Fc Receptor for IgG (FcRn) Binds Albumin and Prolongs Its Lifespan. *J. Exp. Med.* **197**, 315–322 (2003).

92. Bern, M., Sand, K. M. K., Nilsen, J., Sandlie, I. & Andersen, J. T. The role of albumin receptors in regulation of albumin homeostasis: Implications for drug delivery. *J. Controlled Release* **211**, 144–162 (2015).

93. Ober, R. J., Martinez, C., Vaccaro, C., Zhou, J. & Ward, E. S. Visualizing the Site and Dynamics of IgG Salvage by the MHC Class I-Related Receptor, FcRn. *J. Immunol.* **172**, 2021–2029 (2004).

94. Andersen, J. T., Dee Qian, J. & Sandlie, I. The conserved histidine 166 residue of the human neonatal Fc receptor heavy chain is critical for the pH-dependent binding to albumin. *Eur. J. Immunol.* **36**, 3044–3051 (2006).

95. Chaudhury, C., Brooks, C. L., Carter, D. C., Robinson, J. M. & Anderson, C. L. Albumin Binding to FcRn: Distinct from the FcRn−IgG Interaction †. *Biochemistry* **45**, 4983–4990 (2006).

96. Simister NE, Mostov KE. An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184–7 (1989).

97. Andersen, J. T. *et al.* Structure-based mutagenesis reveals the albumin-binding site of the neonatal Fc receptor. *Nat. Commun.* **3**, (2012).

98. Sand, K. M. K. *et al.* Unraveling the Interaction between FcRn and Albumin: Opportunities for Design of Albumin-Based Therapeutics. *Front. Immunol.* **5**, (2015).

99. Nilsen, J. *et al.* Human and mouse albumin bind their respective neonatal Fc receptors differently. *Sci. Rep.* **8**, (2018).

100. Terje Andersen, J., Bekele Daba, M., Berntzen, G., Michaelsen, T. E. & Sandlie, I. Cross-species Binding Analyses of Mouse and Human Neonatal Fc Receptor Show Dramatic Differences in Immunoglobulin G and Albumin Binding. *J. Biol. Chem.* **285**, 4826–4836 (2010).

101. Andersen, J. T. *et al.* Extending Serum Half-life of Albumin by Engineering Neonatal Fc Receptor (FcRn) Binding. *J. Biol. Chem.* **289**, 13492–13502 (2014).

102. Mannuccio, P. M. Half-life extension technologies for haemostatic agents. *Thromb. Haemost.* **113**, 165–176 (2015).

103. Metzner, H. J., Weimer, T., Kronthaler, U., Lang, W. & Schulte, S. Genetic fusion to albumin improves the pharmacokinetic properties of factor IX. *Thromb. Haemost.* **102**, 634–644 (2009).

104. Santagostino, E. *et al.* Long-acting recombinant coagulation factor IX albumin fusion protein (rIX-FP) in hemophilia B: results of a phase 3 trial. *Blood* **127**, 1761–1769 (2016).

105. Furie, B. & Furie, B. C. Molecular and cellular biology of blood coagulation. *N. Engl. J. Med.* **326**, 800–806 (1992).

106. Monroe, D. M. & Hoffman, M. What Does It Take to Make the Perfect Clot? *Arterioscler. Thromb. Vasc. Biol.* **26**, 41–48 (2006).

107. Balestra, D. & Branchini, A. Molecular Mechanisms and Determinants of Innovative Correction Approaches in Coagulation Factor Deficiencies. *Int. J. Mol. Sci.* **20**, 3036 (2019).

108. Biggs, R. *et al.* Christmas disease: a condition previously mistaken for haemophilia. *Br Med J* **2**, 1378 (1952).

109. Mannucci, P. M. & Tuddenham, E. G. D. The Hemophilias — From Royal Genes to Gene Therapy. *N. Engl. J. Med.* **344** (23), 1773–1779 (2001).

110. Bolton-Maggs, P. H. & Pasi, K. J. Haemophilias A and B. *The Lancet* **361**, 1801–1809 (2003).

111. Peyvandi, F., Garagiola, I. & Young, G. The past and future of haemophilia: diagnosis, treatments, and its complications. *The Lancet* **388**, 187–197 (2016).

112. Rallapalli, P. M., Kemball-Cook, G., Tuddenham, E. G., Gomez, K. & Perkins, S. J. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J. Thromb. Haemost.* **11**, 1329–1340 (2013).

113. Schmidt, A. Structure–Function Relationships in Factor IX and Factor IXa. *Trends Cardiovasc. Med.* **13**, 39–45 (2003).

114. Kaufman, R. Post-translational Modifications Required for Coagulation Factor Secretion and Function. *Thromb. Haemost.* **79**, 1068–1079 (1998).

115. Hansson, K. & Stenflo, J. Post-translational modifications in proteins involved in blood coagulation. *J. Thromb. Haemost.* **3**, 2633–2648 (2005).

116. Di Scipio, R. G., Kurachi, K. & Davie, E. W. Activation of human factor IX (Christmas factor). *J. Clin. Invest.* **61**, 1528–1538 (1978).

117. Bajaj, S. P. *et al.* Factor IXa:Factor VIIIa Interaction: helix 330–338 of Factor IXa interacts with residues 558–565 and spatially adjacent regions of the a2 subunit of Factor VIIIa. *J. Biol. Chem.* **276**, 16302–16309 (2001).

118. Stoilova-McPhie, S., Villoutreix, B. O., Mertens, K., Kemball-Cook, G. & Holzenburg, A. 3-Dimensional structure of membrane-bound coagulation factor VIII: modeling of the factor VIII heterodimer within a 3-dimensional density map derived by electron crystallography. *Blood* **99**, 1215–1223 (2002).

119. Nilsson, I. M., Berntorp, E., Löfqvist, T. & Pettersson, H. Twenty-five years' experience of prophylactic treatment in severe haemophilia A and B. *J. Intern. Med.* **232**, 25–32 (1992).

120. Morfini, M. Pharmacokinetic drug evaluation of albutrepenonacog alfa (CSL654) for the treatment of hemophilia. *Expert Opin. Drug Metab. Toxicol.* **12**, 1359–1365 (2016).

121. Farrugia, A. *et al.* Issues in assessing products for the treatment of hemophilia – the intersection between efficacy, economics, and ethics. *J. Blood Med.* 185 (2015) doi:10.2147/JBM.S79091.

122. Sjamsoedin LJ, Heijnen L, Mauser-Bunschoten EP, van Geijlswijk JL, van & Houwelingen H, van Asten P, Sixma JJ. The effect of activated prothrombin-complex concentrate (FEIBA) on joint and muscle bleeding in patients with hemophilia A and antibodies to factor VIII. A double-blind clinical trial. *N Engl J Med* **305**, 717–21 (1981).

123. Key, N. S., Aledort LM, Beardsley D & Cooper HA. Home treatment of mild to moderate bleeding episodes using recombinant factor VIIa (Novoseven) in haemophiliacs with inhibitors. *Thromb Haemost* **80**, 912–8 (1998).

124. Morfini, M. & Zanon, E. Emerging drugs for the treatment of hemophilia A and B. *Expert Opin. Emerg. Drugs* **21**, 301–313 (2016).

125. Weyand, A. C. & Pipe, S. W. New therapies for hemophilia. *Blood* **133** (5), 389-39810 (2019).

126. Ling, G., Nathwani, A. C. & Tuddenham, E. G. D. Recent advances in developing specific therapies for haemophilia. *Br. J. Haematol.* **181**, 161–172 (2018).

127. George, L. A. *et al.* Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N. Engl. J. Med.* **377**, 2215–2227 (2017).

128. Simioni, P., Gavasso, S., Finn, J. D. & Arruda, V. R. X-Linked Thrombophilia with a Mutant Factor IX (Factor IX Padua). *N. Engl. J. Med.* 5 (2009).

129. Finn, J. D. *et al.* The efficacy and the risk of immunogenicity of FIX Padua (R338L) in hemophilia B dogs treated by AAV muscle gene therapy. *Blood* **120**, 4521–4523 (2012).

130. Crudele, J. M. *et al.* AAV liver expression of FIX-Padua prevents and eradicates FIX inhibitor without increasing thrombogenicity in hemophilia B dogs and mice. *Blood* **125**, 1553–1561 (2015).

131. Spronck, E. A. *et al.* Enhanced Factor IX Activity following Administration of AAV5-R338L "Padua" Factor IX versus AAV5 WT Human Factor IX in NHPs. *Mol. Ther. - Methods Clin. Dev.* **15**, 221–231 (2019).

132. Calcedo, R., Vandenberghe, L. H., Gao, G., Lin, J. & Wilson, J. M. Worldwide Epidemiology of Neutralizing Antibodies to Adeno-Associated Viruses. *J. Infect. Dis.* **199**, 381–390 (2009).

133. Dolan, G. *et al.* Haemophilia B: Where are we now and what does the future hold? *Blood Rev.* **32**, 52–60 (2018).

134. Kurachi S et al. Role of intron I in expression of the human factor IX gene. *J. Biol. Chem.* **270**, 5276–5281 (1996).

135. Miao, C. H. *et al.* Inclusion of the Hepatic Locus Control Region, an Intron, and Untranslated Region Increases and Stabilizes Hepatic Factor IX Gene Expression in Vivo but Not in Vitro. *Mol. Ther.* **1**, 522–532 (2000).

136. Sheffield, W. P. *et al.* Effects of genetic fusion of factor IX to albumin on in vivo clearance in mice and rabbits. *Br. J. Haematol.* **126**, 565–573 (2004).

137. Schulte, S. Use of albumin fusion technology to prolong the half-life of recombinant factor VIIa. *Thromb. Res.* **122**, S14–S19 (2008).

138. Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).

139. Samelson-Jones, B. J. & Arruda, V. R. Protein-Engineered Coagulation Factors for Hemophilia Gene Therapy. *Mol. Ther. - Methods Clin. Dev.* **12**, 184–201 (2019).

140. Le Quellec, S. *et al.* Recombinant Adeno-Associated Viral Vectors Expressing Human Coagulation FIX-E456H Variant in Hemophilia B Mice. *Thromb. Haemost.* **119** (12), 1956-1967 (2019).

141. Wilmot, H. V., Hogwood, J. & Gray, E. Recombinant factor IX: discrepancies between one-stage clotting and chromogenic assays. *Haemophilia* **20**, 891–897 (2014).

142. Lambert, T. *et al.* Practical aspects of extended half-life products for the treatment of haemophilia. *Ther. Adv. Hematol.* **9**, 295–308 (2018).

143. Skinner, M. W. *et al.* Achieving the unimaginable: Health equity in haemophilia. *Haemophilia* [published online ahead of print, 2019 Nov 13].

144. Roopenian, D. C. *et al.* Albumin-deficient mouse models for studying metabolism of human albumin and pharmacokinetics of albumin-based drugs. *mAbs* **7**, 344–351 (2015).

145. Mancuso, M. & Santagostino, E. Outcome of Clinical Trials with New Extended Half-Life FVIII/IX Concentrates. *J. Clin. Med.* **6**, 39 (2017).

146. Baralle, D. Splicing in action: assessing disease causing sequence changes. *J. Med. Genet.* **42**, 737–748 (2005).

147. Baralle, M. & Baralle, F. E. The splicing code. *Biosystems* **164**, 39–48 (2018).

148. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**, 5–14 (2003).

149. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).

150. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285–298 (2002).

151. Will, C. L. & Luhrmann, R. Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707–a003707 (2011).

152. Zhang, X. H.-F. Sequence Information for the Splicing of Human Pre-mRNA Identified by Support Vector Machine Classification. *Genome Res.* **13**, 2637–2650 (2003).

153. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).

154. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).

155. Buratti, E. & Baralle, F. E. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol. Cell. Biol.* **24**, 10505–10514 (2004).

156. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).

157. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in Alternative Pre-mRNA Splicing. *Cell* **144**, 16–26 (2011).

158. Krawczak, M. *et al.* Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* **28**, 150–158 (2007).

159. Donadon, I. *et al.* Clustered *F8* missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. *Haematologica* **103**, 344–350 (2018).

160. Tajnik, M. *et al.* Molecular Basis and Therapeutic Strategies to Rescue Factor IX Variants That Affect Splicing and Protein Function. *PLOS Genet.* **12**, e1006082 (2016).

161. Baralle, D. & Buratti, E. RNA splicing in human disease and in the clinic. *Clin. Sci.* **131**, 355–368 (2017).

162. Teraoka, S. N. *et al.* Splicing Defects in the Ataxia-Telangiectasia Gene, ATM: Underlying Mutations and Consequences. *Am. J. Hum. Genet.* **64**, 1617–1631 (1999).

163. Ars, E. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.* **9**, 237–247 (2000).

164. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* **18**, 472–482 (2012).

165. Chabot, B. & Shkreta, L. Defective control of pre–messenger RNA splicing in human disease. *J. Cell Biol.* **212**, 13–27 (2016).

166. Coutinho, M. F., Matos, L., Santos, J. I. & Alves, S. RNA Therapeutics: How Far Have We Gone? in *The mRNA Metabolism in Human Disease* (ed. Romão, L.) vol. 1157 133–177 (Springer International Publishing, 2019).

167. Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense Masking of an hnRNP A1/A2 Intronic Splicing Silencer Corrects SMN2 Splicing in Transgenic Mice. *Am. J. Hum. Genet.* **82**, 834–848 (2008).

168. Mercuri, E. *et al.* Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* **378**, 625–635 (2018).

169. Meyer, K. *et al.* Rescue of a severe mouse model for spinal muscular atrophy by U7 snRNA-mediated splicing modulation. *Hum. Mol. Genet.* **18**, 546–555 (2009).

170. Havens, M. A., Duelli, D. M. & Hastings, M. L. Targeting RNA splicing for disease therapy: RNA splicing for disease therapy. *Wiley Interdiscip. Rev. RNA* **4**, 247–266 (2013).

171. Stoilov, P., Lin, C.-H., Damoiseaux, R., Nikolic, J. & Black, D. L. A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc. Natl. Acad. Sci.* **105**, 11218–11223 (2008).

172. Kondo, Y., Oubridge, C., van Roon, A.-M. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* **4**, (2015).

173. Rogalska, M. E. *et al.* Therapeutic activity of modified U1 core spliceosomal particles. *Nat. Commun.* **7**, (2016).

174. Roca, X. & Krainer, A. R. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat. Struct. Mol. Biol.* **16**, 176–182 (2009).

175. Fernandez Alanis, E. *et al.* An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* **21**, 2389–2398 (2012).

176. Balestra, D. *et al.* An Exon-Specific U1snRNA Induces a Robust Factor IX Activity in Mice Expressing Multiple Human FIX Splicing Mutants. *Mol. Ther. - Nucleic Acids* **5**, e370 (2016).

177. Balestra, D. *et al.* An engineered U1 small nuclear RNA rescues splicing-defective coagulation *F7* gene expression in mice. *J. Thromb. Haemost.* **12**, 177–185 (2014).

178. Dal Mas, A., Rogalska, M. E., Bussani, E. & Pagani, F. Improvement of SMN2 Pre-mRNA Processing Mediated by Exon-Specific U1 Small Nuclear RNA. *Am. J. Hum. Genet.* **96**, 93–103 (2015).

179. Donadon, I. *et al.* Rescue of spinal muscular atrophy mouse models with AAV9-Exon-specific U1 snRNA. *Nucleic Acids Res.* **47**(14), 7618–7632 (2019).

180. Mattioli, C. *et al.* Unusual splice site mutations disrupt FANCA exon 8 definition. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1842**, 1052–1058 (2014).

181. Dal Mas, A. *et al.* Exon-Specific U1s Correct SPINK *5* Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. *Hum. Mutat.* **36**, 504–512 (2015).

182. Donadon, I. *et al.* Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. *Hum. Mol. Genet.* **27**, 2466–2476 (2018).

183. Margaglione, M. *et al.* The Italian AICE-Genetics hemophilia A database: results and correlation with clinical phenotype. *Haematologica* **93**, 722–728 (2008).

184. Franchini, M. & Mannucci, P. M. Hemophilia A in the third millennium. *Blood Rev.* **27**, 179–184 (2013).

185. Do, H., Healey, J. F., Waller, E. K. & Lollar, P. Expression of Factor VIII by Murine Liver Sinusoidal Endothelial Cells. *J. Biol. Chem.* **274**, 19587–19592 (1999).

186. Fay, P. Factor VIII Structure and Function. *Int. J. Hematol.* **83**, 103–108 (2006).

187. Fay, P. J. Activation of factor VIII and mechanisms of cofactor action. *Blood Rev.* **18**, 1–15 (2004).

188. Ward, N. J. *et al.* Codon optimization of human factor VIII cDNAs leads to high-level expression. *Blood* **117**, 798–807 (2011).

189. Pipe, S. W. Functional roles of the factor VIII B domain. *Haemophilia* **15**, 1187–1196 (2009).

190. Camire, R. M. & Bos, M. H. A. The molecular basis of factor V and VIII procofactor activation. *J. Thromb. Haemost.* **7**, 1951–1961 (2009).

191. Toole, J. J. *et al.* A large region (=95 kDa) of human factor VIII is dispensable for in vitro procoagulant activity. *Proc Natl Acad Sci USA* 4 (1986).

192. Astermark, J. FVIII inhibitors: pathogenesis and avoidance. *Blood* **125**, 2045–2051 (2015).

193. Hay, C. R. M. & DiMichele, D. M. The principal results of the International Immune Tolerance Study: a randomized dose comparison. *Blood* **119**, 1335–1344 (2012).

194. Franchini, M. & Mannucci, P. M. Inhibitors of propagation of coagulation (factors VIII, IX and XI): a review of current therapeutic practice: Inhibitors of factors VIII, IX and XI. *Br. J. Clin. Pharmacol.* **72**, 553–562 (2011).

195. Franchini, M. & Lippi, G. Immune tolerance induction for patients with severe hemophilia A: a critical literature review. *J. Thromb. Thrombolysis* **32**, 439–447 (2011).

196. Muto, A. *et al.* Anti-factor IXa/X bispecific antibody (ACE910): hemostatic potency against ongoing bleeds in a hemophilia A model and the possibility of routine supplementation. *J. Thromb. Haemost.* **12**, 206–213 (2014).

197. Aledort, L. M. Deaths Associated with Emicizumab in Patients with Hemophilia A. *N. Engl. J. Med.* **381**, 1878–1879 (2019).

198. Balestra, D. *et al.* An Altered Splicing Registry Explains the Differential ExSpeU1-Mediated Rescue of Splicing Mutations Causing Haemophilia A. *Front. Genet.* **10**, (2019).

199. Vibe-Pedersen, K., Kornblihtt, A. R. & Baralle, F. E. Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. *EMBO J.* **3**, 2511–2516 (1984).

200. Tournier, I. *et al.* A large fraction of unclassified variants of the mismatch repair genes *MLH1* and *MSH2* is associated with splicing defects. *Hum. Mutat.* **29**, 1412–1424 (2008).

201. Théry, J. C. *et al.* Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur. J. Hum. Genet.* **19**, 1052–1058 (2011).

202. Schümperli, D. & Pillai, R. S. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell. Mol. Life Sci.* **61**, 2560–2570 (2004).

203. Raponi, M. *et al.* Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in *BRCA1* exon 6. *Hum. Mutat.* **32**, 436–444 (2011).

204. Anguela, X. M. & High, K. A. Entering the Modern Era of Gene Therapy. *Annu. Rev. Med.* **70**, 273–288 (2019).