Università degli Studi di Ferrara

Dottorato di Ricerca in

Scienze Biomediche e Biotecnologiche

Coordinatore Prof. Paolo Pinton

CICLO XXX

MODULATION AT TRANSCRIPTIONAL AND POST-TRANSLATIONAL LEVEL TO DEVELOP THERAPEUTIC APPROACHES FOR COAGULATION FACTOR DEFICIENCIES

Settore Scientifico Disciplinare BIO/11

Dottorando

Dott. Pignani Silvia

Tutori

Dott. Bovolenta Matteo

Prof. Pinotti Mirko

Co-Tutore

Dott. Branchini Alessio

2014/2017

General Introduction

Haemostasis and Coagulation

Haemostasis is a set of cellular and biochemical reactions that have a double function: to maintain the blood fluidity and to preserve the blood vessels' integrity. This process depicts a balance between pro-coagulant and anti-coagulant activities. The mechanism underlying this complex series of reactions was studied in depth for many years by researchers and this interest produced two different coagulation models through the years. The first coagulation reaction model was explained as the classical waterfall/cascade, and was described in 1964 by two independent groups[1,2]. In this model several proteins with serine-proteases activity are involved in the formation of blood clots in the presence of vessel injury. In physiological conditions these coagulation factors, that normally circulate as zymogens, are converted by cleavage into their active form and interact both together and with the pro-coagulant surfaces to produce the fibrin formation[1].

The traditional view includes two pathways: the contact intrinsic pathway and the tissue factor TF-depending extrinsic pathway. The intrinsic pathway, named because all the components involved in the reaction cascade are circulating in the blood, is composed of FXII and pre-kallikrein (PK). In this pathway, when FXII is in contact with a negatively charged or artificial surface it can be activated (FXIIa) by PK or plasmin. In the intrinsic pathway the activation of an amount of FXII leads to the generation of FXIa from FXI, which in turn, activates FIX. The association of FIXa and FVIIIa together with phospholipids and calcium activates

FX to FXa, which then converts prothrombin (FII) to thrombin (FIIa). FXII and PK zymogens have the peculiarity to be calcium independent, that allows them to "autoactivate" in a continuous manner even in presence of anticoagulant molecules. In spite of this, it has been observed that individuals with deficiencies of FXII or PK do not experience bleeding tendency in vivo, which suggests that the contact pathway may not have a fundamental role in blood coagulation, despite having an important role in thrombus formation[3,4]. In 1977, the waterfall/cascade model was improved thanks to the work of Osterud and Rapaport. Their studies showed that FVIIa-TF can activate FIX and FX, giving the first proof that the extrinsic pathway has an interaction with the intrinsic pathway during clotting[5]. While this cascade model represented a significant improvement in understanding the process and has been used for many years, more recent clinical and experimental observations[6] have shown why the waterfall/cascade hypothesis cannot completely explain the haemostasis events in vivo[7]. The model of coagulation called "cell based" was proposed in 1992, and it focused attention on the cells' role during the coagulation process.

This model is composed of an initiation phase and a propagation phase (Figure I): the initiation phase starts when the TF, also called thromboplastin, comes in contact with the blood. Tissue Factor is an integral membrane protein, and it is exposed only when damage occurs because it is not normally in contact with the blood flow. TF is expressed on the surface of several cell types, like adventitial cell, smooth muscle cells and keratinocytes. When there is a lesion of the endothelial barrier these cells are exposed to blood, consequently TF is able to bind the plasmatic serine protease FVII and generate a modest amount of thrombin. In the propagation phase, acting on the platelet surfaces, the tenase complex can be assembled with the prothrombinase complex and calcium, leading to the production of a massive production of thrombin. This "thrombin burst" occurs at the end and culminates with the conversion of soluble fibrinogen into insoluble fibrin monomers, forming the clot.

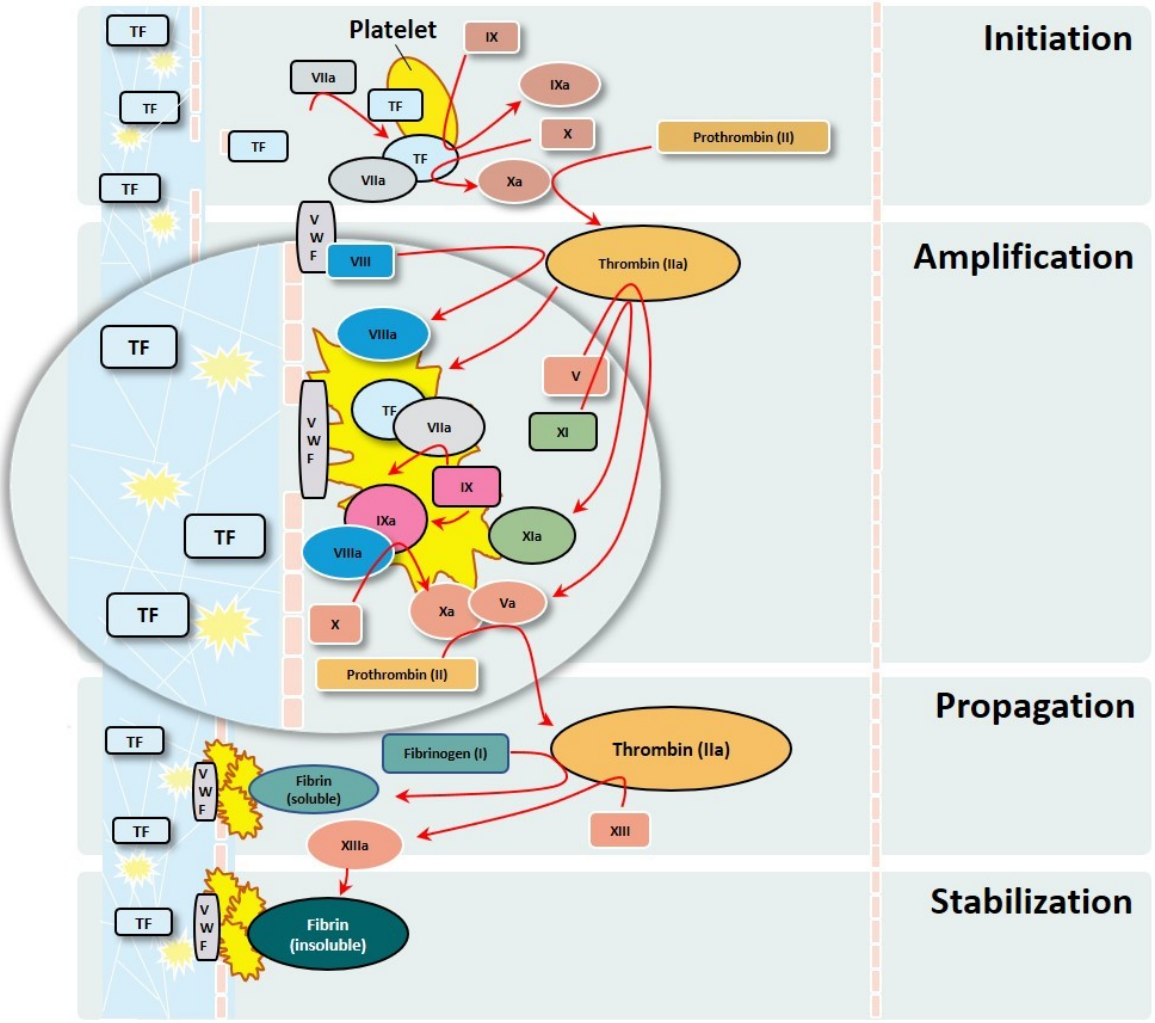

Figure I: Cell based model of coagulation (adapted from a CSL Behring image)

Role of Factor VII, VIII and IX

Factor VII is a vitamin K-dependent serine protease, it is produced in the liver and its circulating concentration is about 500 ng/ml [8,9]. It is the only coagulation factor that is present in both active and inactive forms, and it displays the shortest half-life as zymogen (approximately 5 hours). FVIIa, the active form of FVII, represents 1% of the total circulating FVII in humans[10]. The zymogen itself has negligible activity, it participates to the initiation phase after its activation to FVIIa. FVII can be activated by a number of coagulation enzymes including FIXa, FXa, FXIIa,

thrombin, plasmin or FVII-activating proteases[11]. However, FXa seems to be the most potent activator of FVII[12]. Also FIX plays an important role in this activation, which explains why humans with Hemophilia B display very low levels of FVIIa.
FVII-TF complex can be also auto-activated by FVIIa-TF complex, FVIIa has a long half-life in plasma because it cannot be inactivated by any protease inhibitor unless it is bound to TF. The FVIIa-TF complex activates FIX to FIXa and FX to FXa[13]. This initiation phase results only in a small amount of thrombin after FXa activation, because FXa and thrombin are neutralized rapidly by antithrombin (AT) and the complex FVIIa-TF-FXa by the tissue factor pathway inhibitor (TFPI).
This regulatory system, by TFPI/AT, is important because it prevents a massive thrombin generation in case of false alarm. This regulation permits the coagulation pathway to be activated only when the TF is exposed at high enough levels to overcome the inhibition by TFPI and AT. In this case, thrombin is able to enhance its own generation through the activation of FXI to FXIa and FV to FVa on the platelets surface, and also through the activation of FVIII to FVIIIa by cleaving von Willebrand factor (vWF). The recruited activated platelets provide the procoagulant phospholipids membrane and the generation of FXa is dependent on the formation of the tenase complex[14].
The activation of FX, FIX and the cofactors FV and FVIII after the generation of thrombin in the initiation phase give the beginning of the tenase complex assembly. FIXa and FVIIIa, its cofactor, binds FX and calcium on the negative charged surfaces, but mostly on platelets membrane, in order to form the tenase complex, leading to the generation of a large amount of FXa. The biggest amount of FXa is producted by the tenase complex, which is 50-fold more efficient compared to the FVIIa-TF complex. Thus, in the initiation phase the amount of FXa is insufficient to sustain hemostasis whereas during propagation a massive production of FXa is achieved as a consequence of the formation of the tenase complex. FXa participates actively to the formation of the prothrombinase complex. Together with FVa and calcium, all these events results in an explosive generation of thrombin and the subsequent clot formation. This process is tightly

controlled, in order to maintain the coagulative events at the site of injury. It is important to note that if FXa can diffuse from the site of damage to a healthy adjacent endothelium the coagulation process is inhibited by the presence of anticoagulants as AT, TFPI and thrombomodulin. It is worth noting that in the propagation phase both FVIII and FIX are required. Indeed, the absence of FVIII and FIX in the initiation phase normally occurs, the propagation phase steps is dependent from the tenase complex because lack of FVIII and FIX results in deficient fibrin clot formation and the incapacity of the hemostatic system to respond in repairing the damaged tissues[7].

The coagulation processes previously described clarify the importance of FVII. The absence of this coagulation factor is incompatible with life and it translates FVII deficiency as a rare disease, because it impairs the initiation phase of the coagulation process. On the other hand, FVIII and FIX deficiencies, named with Hemophilia A and B, are more frequent based on their role in the coagulation cascade, that falls into the propagation phase that do not affect the initial production of thrombin[14].

Coagulation factor deficiencies and mutational pattern

The impairment of the coagulation pathway might lead to bleeding disorders. This is translated in a condition with a tendency toward prolonged or excessive bleeding.

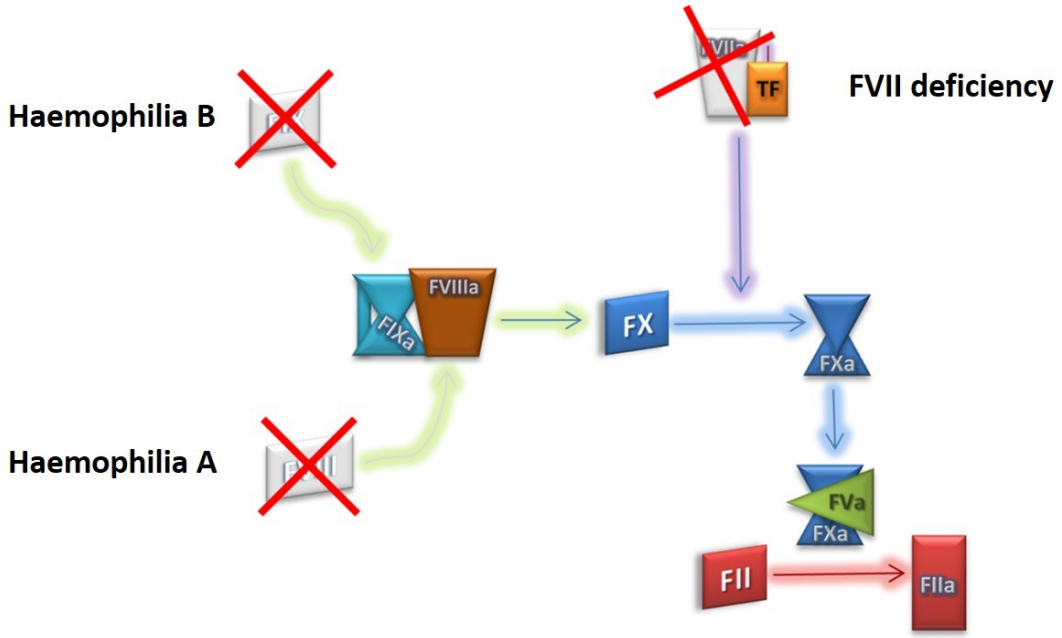

Figure II: Focus on the role of FIX, FVII and FVIII causing respectively Haemophilia B, FVII deficiency and Haemophilia A.

FVII deficiency

Factor VII deficiency, first recognized in 1951[15] , is the most common among rare autosomal recessive bleeding disorders. The incidence estimation is 1 per 300,000-500,000[16] but it can be more frequent in parts of the world where consanguinity is more frequent. It is characterized by autosomal recessive inheritance and affects females and males equally. Studies confirmed that a complete lack of FVII is not compatible with life[17], suggesting that FVII deficiency cannot be associated with complete absence of functional FVII. The research group of Triplett in 1985[18] classified the FVII Deficiency in CRM^- (activity and antigen are proportionally reduced) CRM^+ (reduced activity but normal antigen) and CRM^{red} (antigen is reduced but not proportionally with activity). This hemorrhagic disorder displays variable clinical heterogeneity, indeed the incidence can be underestimated due to the possibility of asymptomatic patients even if their FVII:C is decreased.

The most frequent symptoms, like epistaxis, that are present in males and females at the same incidence, indicate the mild nature of the disease in most cases.

However, in severe cases the symptoms are characterized by hemarthrosis, muscle hematomas and central nervous system and gastrointestinal bleeding. Other common symptoms are post-operative like excessive bleeding after dental or surgical interventions. Related to gender, women are more prevalent among bleeders due to menorrhagia. Gum bleeding and easy bruising are also more frequent in females. The life-threatening hemorrhages occur during the first six months of life and are very rare (5% of bleeds). The major risk for newborns is the trauma that causes a CNS bleeding. The minimum levels of FVII for the interaction with TF are not completely defined. Clinical phenotypes in patients with FVII deficiency do not completely correlate with the coagulant activity measured in vitro. This can be due to the presence of FVIIa that can activate the coagulation cascade also in concentration.

Haemophilia

Haemophilia A and haemophilia B are recessive, X-linked bleeding disorders characterized by deficiency or absence of coagulation factor VIII (FVIII) or IX (FIX), respectively[19–21]. Haemophilia B is much less common than haemophilia A, accounting for just 15–20% of the total haemophilic population[21,22]. From the latest global report from the World Federation of Hemophilia (WFH; compiled in 2014)[23] there are 28,775 patients with haemophilia B while patients who have haemophilia A are 143,523. The two forms of the disease were historically thought to represent the same bleeding disorder, and it was not until 1947 that haemophilia B was recognized as a separate entity[24,25].

Haemophilia A and B are clinically indistinguishable from each other, and diagnosis must be confirmed by specific factor assay. The bleeding tendency is related to the measured concentration of the factor and is classified as mild, moderate, or severe (Table I)[26].

Concentration of factor (VIIIC or IXC)	Classification	Clinical
<1% of normal (<0,01 IU/ml)	**Severe**	Spontaneous joint and muscle bleeding; bleeding after injuries accidents and surgery
1-5% of normal (0,01-0,05 IU/ml)	**Moderate**	Bleeding into joints and muscles after minor injuries; excessive bleeding after surgery and dental extractions
5-40% of normal >0,05-0,40 IU/ml)	**Mild**	Spontaneous bleeding does not occur; bleeding after surgery, dental extraction, and accidents

Table I: Classification of Haemophilia (adapted from Bolton-Maggs et al, 2003)

Haemophilia B is also called Christmas disease, and it results in deficiency of functional plasma coagulation factor IX[27]. The incidence of the disease is 1 in 30,000 live births[28] (rarer than Haemophilia A).

Haemophila A incidence is 1 in over 5000[29]. This disorder, as the other bleeding disorders previously described, determinates a prolonged bleeding after traumas, injuries and surgeries.

Patients with haemophilia experience bleeding only after serious injury, trauma or surgery and in many cases the disease is not diagnosed until an injury or tooth extraction results in prolonged bleeding. Women with mild haemophilia often experience menorrhagia, heavy menstrual periods and can hemorrhage after childbirth. People with moderate haemophilia tend to bleed after injuries, and finally, people with a severe phenotype may also have frequent spontaneous bleeding episodes.

A considerable number of mutations are involved in these types of disease. Causative mutations of F7 gene have been classified (F7 database: http://www.hgmd.cf.ac.uk/ac/index.php): missense mutations represent the most frequent portion (Figure 1.2) and occur in the 68% of cases, followed by splicing-site (13%), promoter (8%) and nonsense (6%) mutations, small insertions

and deletions (6%)[30–32]. A number of these mutations have been identified as the cause of FVII deficiency, but only a limited number of them have been characterized.

Haemophilia B has heterogenic mechanisms and involve a high number of mutations. More than 2100 mutations in the factor IX gene are recorded in the international database (www.factorix.org/). Mutations have been described in all regions of the gene and the majority of these being missense mutations (55%), followed by nonsense, deletions, splice-sites mutations, insertions and deletions.

Another database is present for the Haemophilia A mutational pattern (http://www.factorviii-db.org/). Mutations, other than intron 22 inversion that represents the 60% of Severe HA, are predominantly point mutations (85% missense, 15% nonsense) with 5% being large or small deletions and insertions (Figure II)

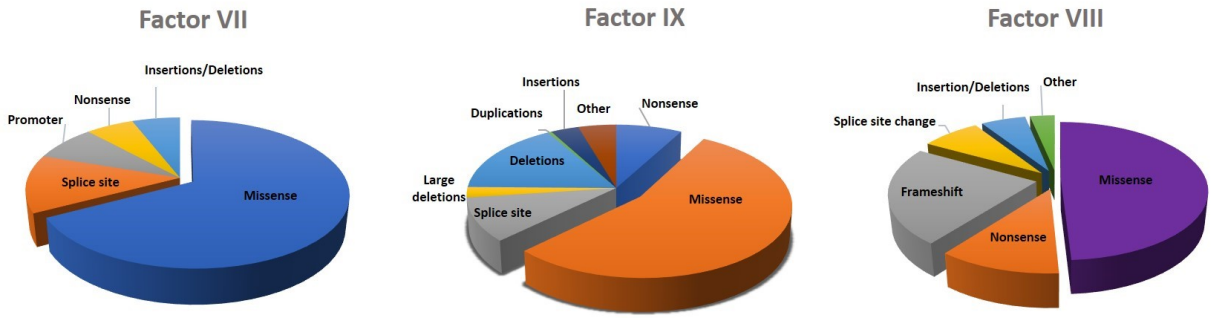

Figure III: Pie charts showing FVII, FIX and FVIII mutational pattern reported in each database (www.factorvii.org; www.factorix.org; www.factorviii-db.org).

Current and emerging therapies for coagulopathies

Effective treatments for coagulation disorders became available recently and have a profound impact on the outlook for people with severe coagulopathies. The replacement therapy is the first type of treatment for these kinds of diseases. FVIII and FIX derived from plasma became widely available in the 1970's, and were

very effective in treating bleeding episodes, and began the development of self-infusion and home therapy programs. The continuous manufacturing advances permitted high purifications of the products, resulting in less contaminants. Indeed, viral inactivation steps were introduced in 1986 to eradicate HIV transmission and hepatitis C virus. Genetically engineered factors (recombinant) concentrates became accessible from the end of the 1990's. Recombinant factors (as rFVIII, rFIX and rFVII) are the current treatments because they eliminate the risk of transmission of human and animal infectious agents. The main challenge for these replacement therapies is that the half-life of the recombinant factors is quite short. FVIIa, that is normally used as by-passing agent, has an half-life of 2-3 hours. A possible strategy was to improve the half-life and to generate longer-acting rFVIIa molecules[33]. Later, fusion protein technology was used to generate rFVII, rFVIII and rFIX fused to molecules to extend the half-life of these recombinant proteins as therapeutics: linkage to the fragment crystallizable (Fc) region of human antibodies, to polyethylene glycols (PEGs) of various sizes (PEGylation), and to albumin[34].

Another strategy that is explored for coagulation disease is represented by gene therapy. This is due to the clinical phenotype of haemophilia: a small increase of circulating coagulation factors can lead to a milder phenotype. Gene therapy is promising for coagulopathies but may not be suitable for all patients. Indeed, up to 50% of the population have neutralizing antibodies against AAV[35]. Even when anti-AAV antibodies are not present, gene therapy may be particularly challenging in young individuals where the gene expression is difficult to maintain in a growing liver. Inhibition pathways are also explored for coagulation diseases, like antibodies able target TFPI. Two of these monoclonal antibodies are currently in development[36–38]: concizumab (Novo Nordisk) and BAY 1093884 (Bayer). Another kind of inhibition therapy is based on RNA interference, in particular, ALN-AT3 (Alnylam Pharmaceuticals) is an investigational RNAi therapeutic targeting antithrombin and designed for subcutaneous prophylaxis in haemophilia A and B[39,40].

Intervention at the pre-mRNA splicing level is a promising emerging therapeutic strategy for genetic disorders. The attention has been given to the U1 small nuclear RNA (U1snRNA) that mediates the recognition of the donor splice site (5'ss), the therapeutic effect of engineered U1snRNP is to rescue aberrant splicing mutations at the 5'ss. This was widely explored for haemohilia and for FVII deficiency[41–43].

Nonsense codons result in premature termination, aberrant gene expression, and truncated proteins in inherited human disorders, including haemophilia[44]. Premature termination codon (PTC) suppressors (aminoglycoside and non-aminoglycoside compounds) have been investigated as novel treatments for haemophilia[45]. However, only in a small number of cases (and especially for hemophilia B) the cause of the disease is nonsense mutation. These treatments may offer hope for only a small portion of haemophilia population.

Emerging genome editing techniques are being protagonist of the basic research for inherited diseases in the last decade. Targeted endonucleases like zinc-finger nucleases (ZFN), transcription activator-like effectors nucleases (TALEN) and the novel CRISPR/Cas9 system have been explored also for haemophilia[46–49]. Induced pluripotent stem cells (iPS), which are derived from somatic cells, can be engineered with the CRISPR/Cas9 system to have the correct sequence for the defective factors and differentiated into functional hepatocytes that can be transplanted in patients.

List of Abbreviations

4-PBA 4-phenylbutyrate

AAV9 Adeno-associated virus 9

aPTT activated Partial Thromboplastin Time

AT antithrombin

CFTR Cystic fibrosis transmembrane conductance regulator

CHO Chinese Hamster Ovary

CNS Central Nervous system

CRISPR Cluster regularly interspaced short palindromic repeats

CRISPRa CRISPR transcription activation system

DBD DNA binding domains

dCas9 nuclease-deactivated Cas9

DMEM Dulbecco's modified Eagle's medium

EGF epidermal growth factor

ER Endoplasmic Reticulum

eTFs engineered transcription factors

F7 Factor VII gene

FBS Fetal Bovine serum

FII prothrombin

FIIa thrombin

FIX Factor IX

FVII Factor VII

FVIII Factor VIII

FXII Factor XII

gRNA guide RNA

HA Haemophilia A

HB Haemophilia B

hIAPP Human islet amyloid polypeptide

HNF4 Hepatocyte Nuclear Factor 4

iPS Induced pluripotent stem cells

KRAB Krüppel associated box

MCP MS2 coat protein

NAC N-acetylcysteine

Na-PBA Sodium phenylbutyrate

OTC Ornithine Transcarbamylase Deficiency

PDI protein disulfide isomerase

PEGs polyethylene glycols

PFA Paraformaldehyde

PK pre-kallikrein

Pnp pooled normal plasma

PTC premature termination codon

rFIX recombinant Factor IX

RVD repeat variable diresidue

SAM synergic activation mediator

SP1 Specificity protein 1

TALE Transcription Activator-like Effectors

TALEN transcription activator-like effectors nucleases

TF tissue factor

TF Transcription factor

TFPI tissue factor pathway inhibitor

TSS translation starts site

UPR unfolded protein response

UTR Untranslated region

VPR VP64, p65, Rta

vWF von Willebrand factor

ZF Zinc Fingers Proteins

ZFN zinc-finger nucleases

Chapter 1 – Chaperone Like Compounds Strategy for Haemophilia B

1.1 Factor IX: gene and protein features

The gene of human factor IX (*F9*) is localized in q27.1-q27.2 region of chromosome X. It was isolated and sequenced in 1985[50], it is approximately 34kb in length and is composed by eight exons (spanning from 25bp and 548bp) and seven introns. The transcript is 2803 bases in length and comprises a short 5′ UTR (29 bp), an open reading frame (1383 bp) and a 3′ UTR (1383 bp). Exon 1 and a part of exon 2 encode for the pre-pro leader sequence, which is removed during the protein biosynthesis. The last part of exon 2 as well as exon 3 encodes for the Gla domain containing 12 γ-carboxiglutamic acid residues. The γ-carboxilation is a post-translational modification needed for the correct folding of FIX and for its calcium binding capacity[51]. Exons 4 and 5 encode respectively for the first and the second epidermal growth factor-like domains (EGF-like 1 and 2). Exon 6 codes for the activation domain, in which factor XIa and the complex FVIIa-TF cut in order to activate FIX. The last two exons encode for the catalytic domain of the protein (Table 1.1).

Exons	Coding region	Dimension (bp)
1	Pre-pro leader sequence	88
2	Pre-pro leader, Gla domain	164
3	Gla domain	25
4	EGF-Like domain	114
5	EGF-Like domain	129
6	Activation site	203
7	Catalytic domain	115
8	Catalytic domain	571

Table 1.1: Factor IX gene exons organization

Factor IX, also called Christmas factor or antiemophilic B factor, is a vitamin K dependent serine protease glycoprotein synthetized by liver and circulating in blood as a zymogen at the concentration of 5µg/ml[5]. It features a N-terminal γ-carboxiglutamic acid-rich domain followed by two EGF-like domains and C-terminal serine protease domain. The mature protein consists in 415 aminoacids of 57 KDa molecular weight, it is synthetized as a precursor protein of 461 aminoacids with a signal pre-peptide and a pro-peptide leader of respectively 28 and 18 residues[52]. During the biosynthesis, FIX goes through different post-translational modifications as γ-carboxylation of twelve glutamic acids residues (Gla domain), glycosylation, hydroxylation, sulfatation, phosphorylation and signal peptide and pro-peptide sequence removal. The Gla residues are responsible for the high-affinity binding of the calcium ions[53,54], so conferring to FIX its biological activity.

The protein consists in a short hydrophobic segment (residues 41-46) two EGF-like domains (EGF1 residues 47-84 and EGF2 residues 85-127) an activation peptide region (146-180) and a serine protease module (181-415). The activation of FIX is in response to proteolytic cleavages at Arg145-Ala146 and Arg180-Val181 with the

release of a 35-residue activation peptide (3,40). The FIXa thus formed contain a light chain (residues 1-145) and a heavy chain (residues 181-415) held toghether by a disulfide bound. The light chain is composed by Gla, EGF1 and EGF2 domains and the heavy by the Serine Protease domain that features the catalytic triad of residues His 221 (c.57), Asp 269 (c.102) and Ser 365 (c.195)[5,55,56].

1.2 Background and Rationale

Missense mutations as the most frequent cause of Haemophilia B

Protein folding is an intermediate state of the protein formation pathway. It is important for the correct structure of proteins. Indeed, to attain functionality, proteins must fold into their three-dimensional native structure during or following translation at the ribosomes[57].

The folding process occurs into the Endoplasmic Reticulum (ER): after the polyaminoacid chain is translated from the ribosomes of the Rough Endoplasmic Reticulum it undergoes different types of post-translational modifications and it assumes a tridimensional structure[58]. Because the number of possible conformations a protein chain can adopt is very large, folding reactions are highly complex and heterogeneous, relying on the cooperation of many weak, non-covalent interactions. The newly translated protein chain collapses and the progressive increase in the number of native interactions rapidly restrict the conformational space that needs to be searched in route to the native state. The protein passes from a high energy initial state, linear and disorganized, to the tridimensional state after the different modifications.

Partially folded or misfolded states are problematic because they tend to aggregate in a concentration-dependent manner. This aggregation can be caused from the fact that these forms typically expose hydrophobic amino-acid residues and results in amorphous structures.

The partial folded state or misfolded proteins can occur in the case of the presence of a mutation, and missense mutations are the prevalent type. Missense mutations are the main cause of human genetic diseases and the most detrimental effect is exerted by amino acid changes impairing folding, thus leading to intracellular retention or degradation of the altered protein[59]. These mutations cause the variation of an amino acid that can impair the protein structure and activity based on the position in the tridimensional state of the protein, and this can be the cause of different human diseases.

The protein misfolding can follow two different pathways with different pathological consequences:

I. Protein misfolding with gain-of-function that causes toxic aggregates (Alzheimer, Parkinson's and Huntington diseases).

II. Protein misfolding loss-of-function, associated with protein degradation or protein aggregation in presence of genetic mutations[60].

The second group involves pathologies as Haemophilia B (HB), that, in the major of the cases is caused by missense mutations[28] (Figure II). These mutations involve the structure of the FIX protein and can impair its secretion, causing a severe phenotype (Figure 1). Indeed, severe phenotype of HB patients occur in the 47% of the cases in presence of missense mutations.

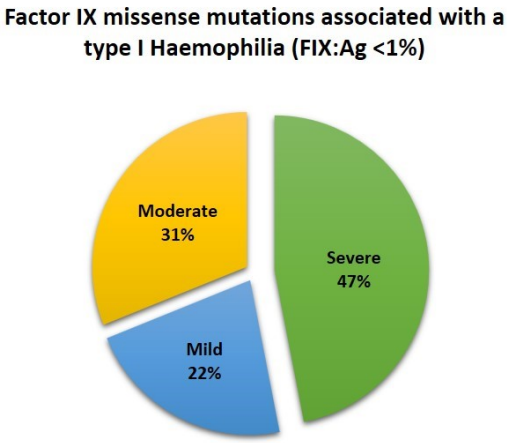

Figure 1.1: HB mutational pattern: missense mutations associated with severe type I HB.

Protein folding

The protein folding is the physical process in which a protein chain acquires its native three-dimensional structure, this process has a key role in the final conformation which permits the correct function of the new protein. The folding process take place in the endoplasmic reticulum (ER) after the translation of the aminoacid chain. The newly synthetized peptide has a linear aminoacidic conformation, and after different post-translational modifications (glycosylation, disulfide bounds) it assumes a tridimensional structure[58]. Protein folding inside cells is not generally a spontaneous process, there are many evidences indicates that many newly synthesized proteins require a complex cellular machinery of molecular chaperones and the input of metabolic energy to reach their native states efficiently[61–63]. The process occurs in a series of reactions: from the starting high energy structure, open linear and disordered, few aminoacidic residues form an initial fold. Around this first folding the rest of the protein collapses and create an intermediate folding state, that needs further modifications to reach stability.

Chaperones and intracellular processing

The correct protein folding is mediated by a class of proteins called molecular chaperones, the role of this specific type of proteins is to associate with the new peptide with hydrophobic residues or cysteines exposed[64,65].

Molecular chaperones, highly conserved in evolution, are defined as proteins able to interact, stabilize and help the nascent peptides to fold correctly and adopt their native conformation, and that are not present in the final functional protein[66].

This class of protein is involved in the assembly of macromolecular complexes, protein transport, degradation and aggregates dissociation. Molecular chaperones are highly expressed in protein synthesis districts, in the Endoplasmic reticulum lumen they form a dense matrix and they catalyze the protein folding.

Preventing aggregation, molecular chaperones have a crucial role in the correction of irregular interactions with other molecules: they establish long lasting contacts with the misfolded proteins, that are not capable of reaching a thermodynamic stable conformation. New peptides that are not able to fold correctly into the Endoplasmic reticulum are degraded from the proteasome to avoid the aggregation that can lead to a toxic effect inside the cell. The molecular chaperones are able to discriminate the misfolded proteins due to the hydrophobic residues exposed. This mediation from the molecular chaperones provide a quality check inside the cell permitting a selection on the new synthetized proteins. Therefore, only the correct folded proteins are able to translocate to the Golgi apparatus and, successively, to be secreted. On the contrary, the passage from the Endoplasmic reticulum is impaired, and the misfolded proteins are recruited from chaperones and degraded from the proteasome[67].

Chaperone like compounds as therapeutic tool

Different studies revealed that low molecular weight molecules called "Pharmacological" and "Chemical" chaperones are able to modulate protein folding, recovering the protein biosynthesis in presence of impairing mutations[60]. These molecules are able to pass the cellular membrane to act their role inside cellular organelles and help the folding after the protein synthesis[68]. Chaperone-like compounds capacity is to hold together different parts of the protein structure through electrostatic interactions, hydrogen bounds and van der Waals forces that promote the final structure of the protein[69].

This class of molecules can be divided in two main groups:

- Chemical chaperones
- Pharmacological chaperones

Chemical chaperones

These compounds are considered non-selective in capacity to stabilize the mutated proteins and help the folding, acting like molecular chaperones. This particularity is relevant if some genetic anomalies occur. Indeed, if a minor fraction of missense variant proteins undergoes a re-folding or is stabilized by the chaperones these can be relevant to restore the natural function of the protein[70].

The specific mechanism of chemical chaperones is not clearly explained in literature. The hypothesis is that these kinds of compounds can stabilize the mutant proteins, reduce the aggregation pathway, assist the mutated protein to avoid the quality control in the ER or alter molecular chaperones activity.

Pharmacological Chaperones

Pharmacological chaperones specifically and reversibly bind to misfolded target proteins. This induces conformational stabilization of the latter, consequently increasing the availability of functional protein species in the cell[71]. Pharmacological chaperone treatment stabilizes the target protein in the ER and enables its exit from the ER and trafficking through the Golgi and onward to its final destination. The peculiarity of this class of compounds is also that they promote the transition of the target from the folded state to the folded protein-pharmacological chaperone state. The interaction between the pharmacological chaperones and the mutated target protein maintains the folding state in order to rescue its biological activity[72].

1.3 Aim of the study

Hemorrhagic disorders are often caused by a reduction in circulating levels of coagulation factors, and missense mutations are the main cause. In particular, in Haemophilia B the majority of the disease-causing mutations are missense. The dramatic effect of missense mutations is the amino-acidic change that leads to an impaired folding of the protein. Misfolded proteins can aggregate in intracellular compartments or be degraded by the proteasome in the unfolded protein response (UPR) pathway.

In protein folding mechanism, a key role is covered by molecular chaperones, protein complexes able to stabilize the newly synthesized peptides in the cytosol and in the endoplasmic reticulum.

Taking into account the molecular chaperones' role, researchers developed different molecules called chemical and pharmacological chaperones. These molecules are able to modulate the protein folding and potentially to restore the biosynthesis of proteins impaired by missense mutations.

The high frequency of missense mutations associated with a reduction of secreted protein and the low therapeutic threshold, in which even a tiny increase of circulating functional protein would ameliorate the clinical phenotype of patients, makes the coagulation disorders ideal models to test a similar approach. Therefore, we tried a therapeutic model based on the treatment of cells expressing Factor IX missense mutations with chemical and pharmacological chaperones.

The aim of this work is to rescue the intracellular processing and to restore the secretion of Factor IX (FIX) in the presence of missense mutations causing severe type I Haemophilia B using chemical and pharmacological chaperones.

1.4 Materials and methods

Creation of expression vectors.

Expression vectors for recombinant FIX (rFIX) missense variants were produced by site-directed mutagenesis of the human FIX cDNA, cloned into the pCDNA3 vector, through the QuickChange® II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). All expression vectors have been validated by sequencing.

Cell culture and transfection

Human Embryonic Kidney 293 (HEK293) and Chinese Hamster Ovary (CHO) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 4 mM L-glutamine, 50 IU/ml penicillin and 50 µg/ml streptomycin (Gibco, Life Technologies, Carlsbad, CA, USA) and maintained at 37°C in a 5% CO2 atmosphere. Cells were transiently transfected with the Lipofectamine 2000 transfection reagent (Life Technologies) by using serum-free Opti-MEM (Gibco, Life Technologies) in the presence of 5 µg/ml vitamin K as medium. Selection and maintenance of stable HEK293 clones was performed by using G418 (500 µg/ml; G8168; Sigma-Aldrich, St. Louis, MO, USA) as the selection agent.

Chemical/pharmacological chaperones

4-phenylbutyrate (4-PBA; P201005), Sodium phenylbutyrate (Na-PBA; SML0309), N-acetylcysteine (NAC; A7250) and Ambroxol (A9797) were from SIGMA-Aldrich. VX-809 (Lumacaftor; S1565) was from Selleckchem (Houston, TX, USA).

Treatment of cells with chemical/pharmacological chaperones

HEK293 cells stably expressing rFIX variants were seeded in 24-well plates in DMEM medum supplemented with 5 µg/ml vitamin K without G418. The day after, fresh serum-free Opti-MEM (Life Technologies) containing 5 µg/ml vitamin K and supplemented with compounds, was added to cells and then collected 48 hours post-treatment for downstream analysis.

Cell lysates were prepared by addition of RIPA buffer directly on cells and incubation for 30 minutes at 4°C, followed by centrifugation at 3,000 x g for 5 minutes. Supernatants were then collected and stored at -80°C. Total protein content in cell lysates was determined by the BCA protein assay kit (Pierce, Thermo Fisher Scientific, Waltham, MA USA).

Evaluation of secreted protein levels

Protein levels for secreted rFIX variants in conditioned media were evaluated by a commercially ELISA kit (Affinity Biologicals, Ancaster, Ontario, Canada). Known concentrations of plasma-derived human FIX (Haematologic Technologies Inc., Essex Junction, VT, USA) were used as the reference curve.

Western blotting analysis on media and cell lysates.

Media or cell lysates were separated by SDS-PAGE in Bis-Tris precast polyacrylamide (4-12%) gels (Life Technologies) with MES-SDS as the running buffer, and then transferred to a 0.45 μm nitrocellulose membrane. Blocking of membranes was carried out by incubation with 5% w/v milk in PBS buffer over-night at 4°C. A polyclonal goat anti-human FIX (APGAFIX; Affinity Biologicals) and anti-goat HRP-conjugated (A50-101P; Bethyl Laboratories, Montgomery, TX) antibodies. Blotting images were acquired by chemiluminescence detection on the ChemiDoc MP System (Bio-Rad Laboratories, Hercules, CA, USA).

Immunofluorescence studies

Stable clones were seeded in 24-well plates on a glass slide (XXX) in DMEM medium supplemented with 5 μg/ml vitamin K. The next day cells were fixed in 4% PFA for 10 min at room temperature and permeabilized with 0.1% Triton X-100 for 10 min. After three washes with PBS, nonspecific binding sites were blocked for 1 hour with Blocking Buffer (PBS with 5% w/v milk-0.1% Tween-20). Antibodies for immunostaining, diluted in Blocking Buffer, were a polyclonal goat anti-human FIX antibody (1:200; Affinity Biologicals), polyclonal rabbit anti-PDI antibody (1:200; Abcam, Cambridge, UK) and monoclonal mouse anti-GM130 (Golga 2) (1:100; Becton Dickinson, Franklin Lakes, NJ, USA) and incubated overnight at 4°C.

After three washes with PBS (5 min each), detection was carried out with donkey anti-goat IgG (594 nm; FIX), donkey anti-rabbit IgG (488 nm; PDI) and donkey anti-mouse IgG (594 nm; Golga-2) Alexa Fluor-conjugated antibodies (Thermo Fisher Scientific). Each secondary antibody was incubated separately for 1 hour followed by three washes with PBS-0.1% Tween-20 (5 min each). Images were acquired on the inverted confocal Zeiss LSM510 microscope.

Coagulant activity of recombinant variants

Coagulant activity of the rFIX variants was assessed by activated Partial Thromboplastin Time (aPTT)-based assays. Commercially-available FIX-depleted plasma (HemosIL, Instrumentation Laboratory, Lexington, MA, USA) was supplemented with rFIX variants in conditioned medium. Coagulation times were measured on the ACLTOP700 instrument (Instrumentation Laboratory) upon addition of a contact activator (SynthASil, Hemosil) and CaCl2. Coagulation times from a standard curve obtained by serial dilutions of wild-type rFIX were used as reference.

1.5 Results

Selection of mutations

Selection of missense mutations, representing the majority of *F9* gene alterations in affected HB patients, was based on clinical phenotype and number of patients. In addition, among the missense mutations associated with severe phenotypes, a further selection led to the identification of six mutations accounting for around 50% of severe patients affected by severe type I Haemophilia B (FIX levels <1%) (Table 1.2).

Number of patients	Missense Mutation	Gene location	Nucleotide change	CpG	Protein Domain
97	p.R294Q	8	c.881G>A	Y	Serine Protease
9	p.Y115C	4	c.344A>G	N	EGF1
9	p.Y305C	8	c.915A>G	N	Serine Protease
5	p.Y161C	5	c.482A>G	N	EGF2
3	p.F424L	8	c.1270T>C	N	Serine Protease
3	p.Y450C	8	c.1349A>G	N	Serine Protease

Table 1.2 Selected mutations (FIX:Ag <1%) associated with severe HB and representing around 50% of severe patients affected by type I HB.

The six mutations selected were identified taking advantage of the Factor IX database[73], the reported mutations have shown severe clinical phenotypes in patients, with the highest frequency for the p.R294Q. All are associated with a severe reduction of circulating FIX (<1%).

Expression studies and association to Severe HB phenotype in two different models: Hek293 and CHO cells.

The vectors expressing each variant of FIX have been used to evaluate the low secretion levels associated to the disease in patients. In particular, to validate that our expression system could mimic the secretion levels found in patients, we used two different cell lines Hek293 (Human Embryonic Kidney), and CHO cells (Chinese Hamster Ovary). These two cell lines do not naturally express FIX protein, indeed they are derived from kidney and ovary tissues respectively. For this reason, these models allowed us to study the different FIX variants without background and to estimate each missense mutation secretion levels.

To demonstrate the association with severely reduced secreted FIX protein levels, and thus with severe HB, the different constructs expressing each variant have been transfected in the two different cell lines, two controls have been used: the wild-type carrying vector as a positive and the backbone as a negative control. After 48 hours from the transfection, the media have been collected and the total proteins from lysates quantified.

While the rFIX proteins were well appreciable by Western blotting analysis of cell lysates, the secreted protein levels, quantified by a polyclonal ELISA, were remarkably reduced (<1% of wild-type rFIX) for all mutations, thus recapitulating the type I HB phenotype in patients. In response to the secretion efficiency of the FIX wild-type protein that is comparable in the two cell models (Hek293 441±25 ng/ml; CHO-K1 424±170 ng/ml), the distribution of the secretion levels under the 1% threshold has been observed in both cell lines (Figure 1.2).

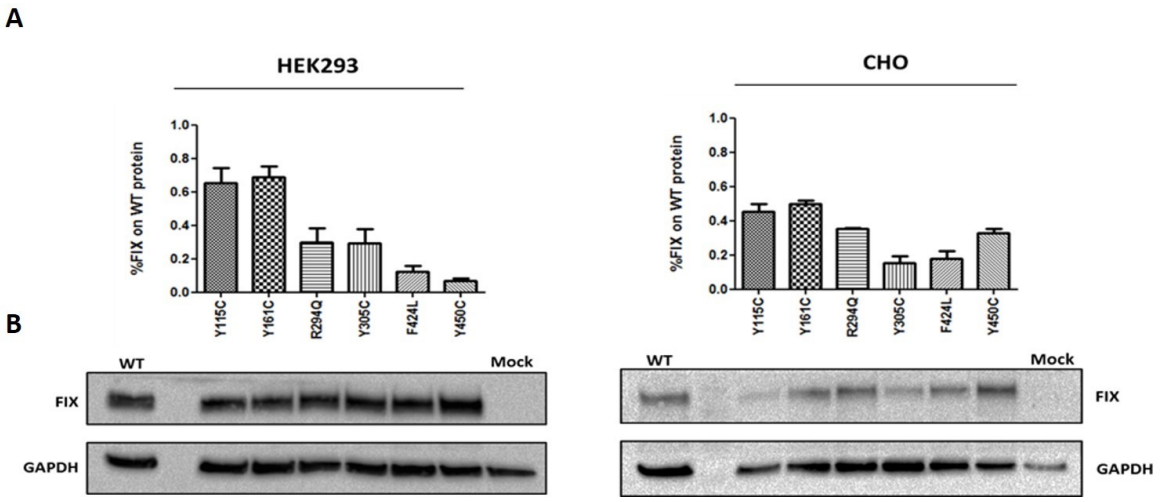

Figure 1.2 Cells transiently transfected with each FIX variant were used to confirm the clinical phenotype: A) Evaluation of secreted (ELISA) and B) intracellular (Western blotting) levels of rFIX variants in comparison with GAPDH as housekeeping in HEK293 and CHO-K1 cells.

These first results prompted us to move to stable HEK293 expression systems, ensuring a lower degree of variability on secreted levels than transient expression, to further characterize all mutations at the intracellular level and challenge all mutations with compounds.

Evaluation of Intracellular processing through immunofluorescence assays:

In the intracellular environment, as previously described, missense proteins variants can undergo two different pathways I) degradation in cytosol mediated by the proteasome or II) accumulate inside the cell compartments as the endoplasmic reticulum. For the characterization of the intracellular distribution of the selected variants we took advantage of stable clones created to express each variant and the wild type protein as a control. The co-localization studies through

immunofluorescence permitted us to evaluate if the retention hypothesis could be an explanation of the low secretion of the missense variants.

For these studies we used different antibodies targeting specific marker proteins into the cellular organelles, protein disulfide isomerase (PDI) protein for the Endoplasmic Reticulum, Golga2 (GM-130) peptide for the Golgi apparatus. The analysis of the three different fluorescent signals, involving the antibody for FIX protein, prompted us to understand the different distribution of the mutants in comparison with the wild type protein.

The analysis on the wild type protein permitted us to optimize the experimental system and verify the correct distribution in normal conditions. The confocal imaging system allowed us to observe that the stable clones expressing the wild type protein showed an elevate co-localization of the FIX signal with the Golgi apparatus, in comparison with the ER (Figure 1.3).

Interestingly, missense rFIX variants co-localized mainly in the ER and negligibly in the Golgi apparatus. This co-localization trend was evident by analysis of fluorescence intensity (expressed as Gray value) as a function of distance of the analyzed region, which indicates that the fluorescence signal corresponding to rFIX missense variants and the ER marker were nearly superimposed (Figure 1.3).

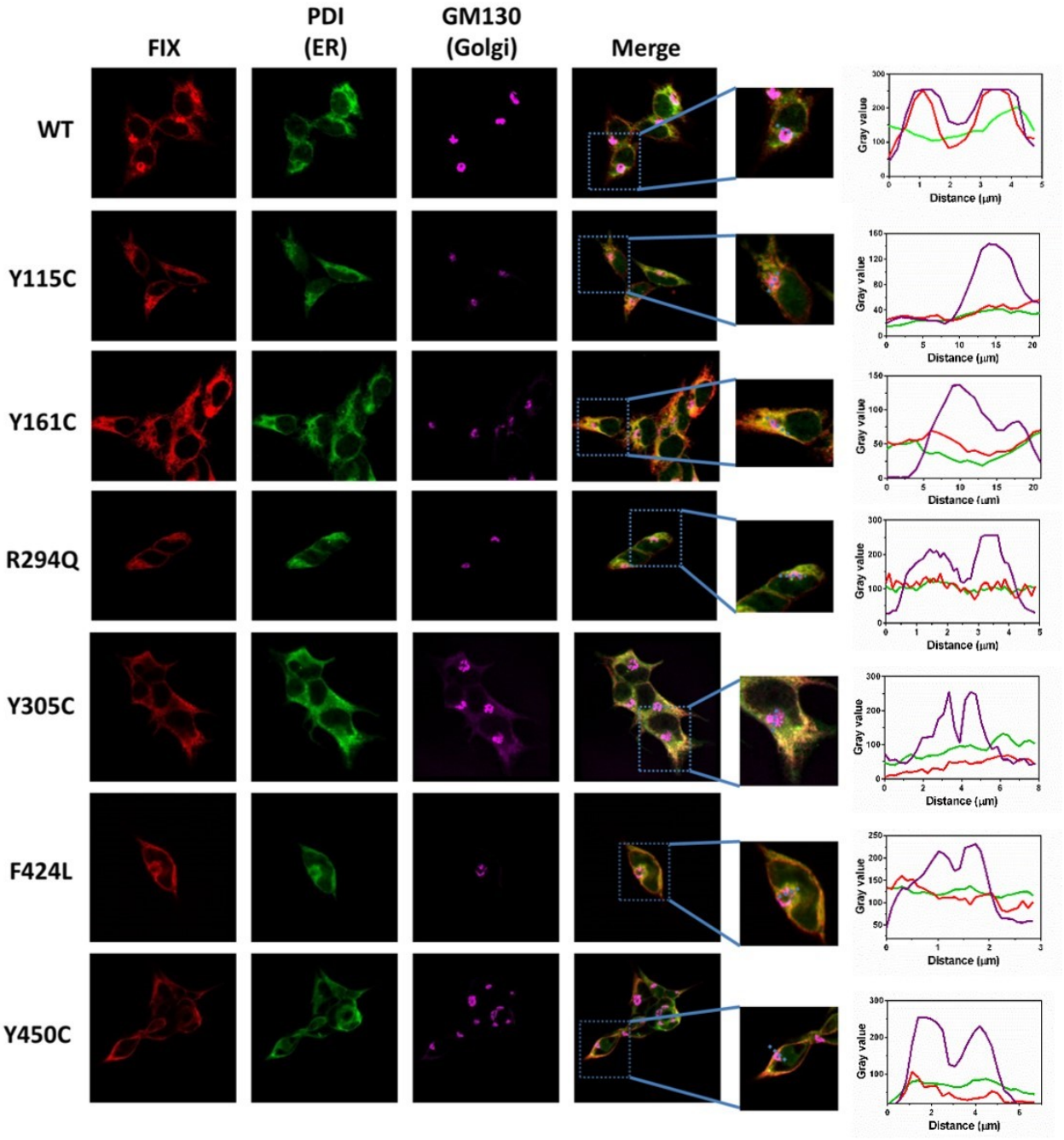

Figure1.3: Co-localization studies by immunofluorescence on HEK293 stably expressing rFIX variants. Cells were stained with polyclonal anti-FIX (red) and polyclonal anti-PDI for ER (green) or polyclonal anti-GM130 for Golgi (violet). Co-localized green and red pixels are shown in yellow color and co-localized red and purple pixels are shown in pink. To better understand the co-localization signals in the plot on the right the co-localization of the three signals are shown.

Treatment of FIX-expressing cells and evaluation of protein rescue using different chaperone-like compounds:

As chemical/pharmacological chaperones we exploited the compounds reported below:

- **4-phenylbutyrate (4-PBA)**, known to be able to restore the expression of the CFTR, α1-antitrypsin, and protein C deficiency (A267T variant);
- **Sodium phenylbutyrate (Na-PBA)**, a sodium salt variant of PBA;
- **VX-809 (Lumacaftor)**, a pharmacological chaperone in phase-3 clinical studies for CFTR mutations;
- **N-acetylcysteine (NAC)**, which has been explored for Pompe disease;
- **Ambroxol**, a chaperone candidate for Gaucher disease.

The protein levels after treatment was estimated through a ratio between the band densitometry (Western Blot) and the total protein estimation (BCA assay). After treatments, lasting 48 hours in low serum media, medium have been evaluated through Western Blotting analysis that facilitated the screening of the different compound in each of the six variants. We took advantage of total intracellular protein evaluation because the cells that we used for the experiments were stable for the FIX variants, therefore all the cells we treated expressed the variant and the result is independent from transfection efficiency.

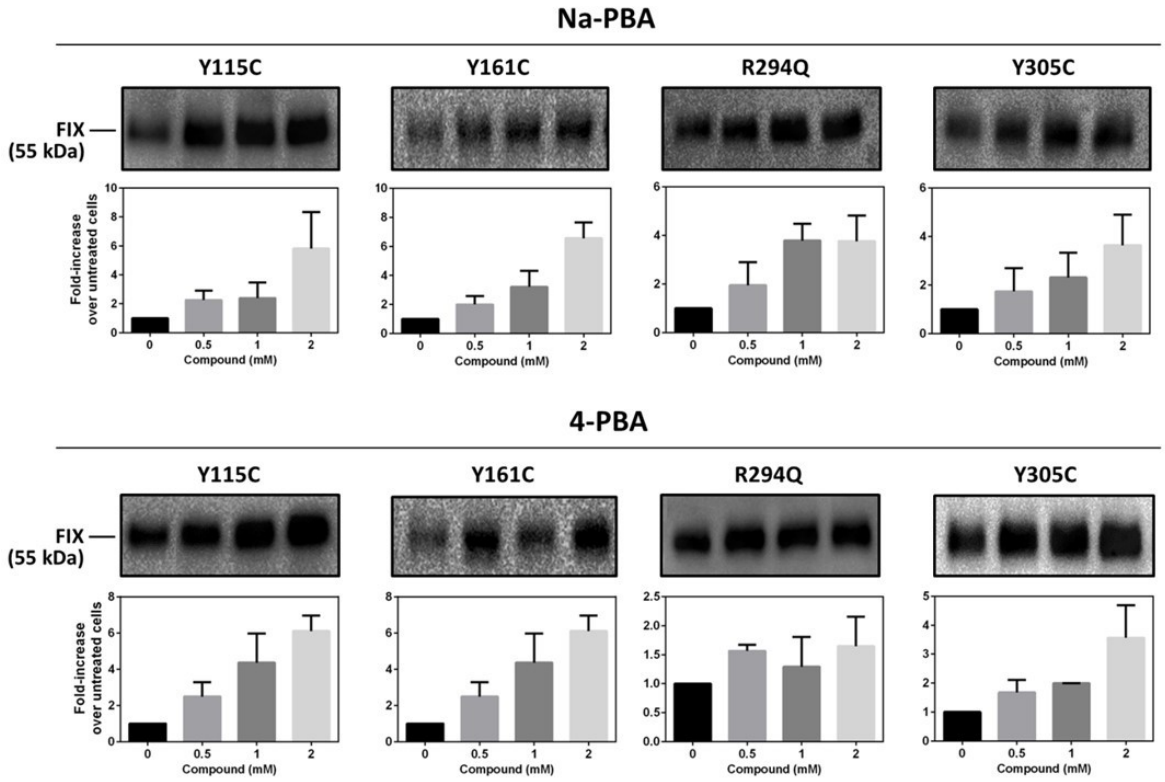

Figure 1. 4 Treatment of rFIX-expressing HEK293 clones with chemical/pharmacological chaperones. The improvement of rFIX secretion levels over the baseline values (1) from untreated cells are indicated as fold increase, which was estimated as the ratio between densitometric analysis of rFIX bands detected in Western blotting and total protein content (quantified by BCA assay).

Among the compounds tested on rFIX-expressing cells, only the 4-PBA (from 2 to 8-fold) and Na-PBA (from 3 to 7.8-fold) chemical chaperones were clearly effective in improving secretion of rFIX variants in a dose-dependent manner (Figure 4). The effect of the other compounds (VX-809, NAC and Ambroxol) was negligible. On the other hand, treatment of F424L and Y450C variants did not produce any increase in rFIX expression (Figure 1.4).

Evaluation of functional rescue over the baseline

One of the most important characteristic of coagulation factors is functionality, therefore modulation of secretion is not the unique issue when the protein as an activity as serine-proteases. Indeed, secretion rescues do not ensure that the protein function is unaltered. The observation that chemical/pharmacological chaperones were effective in improving secreted protein levels for the Y115C, Y161C, R294Q and Y305C variants prompted us to evaluate the residual pro-coagulant activity of responsive variants. After a first protein levels evaluation, the activity has been measured through aPTT assay, and interpolated, taking advantage of a standard curve using the wild type protein activity.

As shown in figure 1.5, the specific coagulant activity, referred as the ratio between coagulant activity and protein levels, was very low for the Y115C (0.012±0.002) and Y305C (0.041±0.001) variants, while that for the Y161C was not detectable. On the other hand, the observed activity for the R294Q variant exceeded the amount of rFIX protein levels, resulting in a specific coagulant activity of 1.8±0.3. Activity of the F424L and Y450C, evaluated as additional controls albeit the exclusion due to a lack of rescue after treatments, were undetectable (F424L) or similar to those previously observed by us (Y450C; Branchini et al 2013).

The data helped understand the clinical phenotype of HB patients, in which the type I HB is associated with the virtual absence of coagulant activity, with the exclusion of the R294Q variant. Therefore, the potential functional improvement promoted by compounds would be significant only for the R294Q variant, while for the other three (Y115C, Y161C and Y305C) would be precluded by the negligible residual activity, which would also prevent the beneficial effects brought by the increase in secreted protein levels.

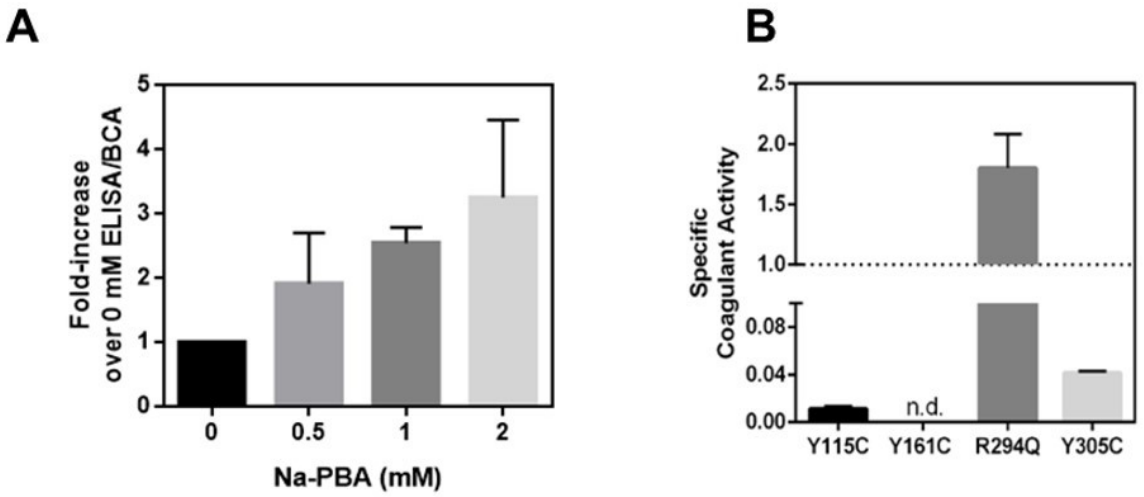

Figure 1.5 A) ELISA evaluation of increase amount of protein after treatment with Na-PBA at different concentrations. The fold-increase is related to untreated cells. B) Specific coagulant activity (activity/antigen ratio) of rFIX variants responsive to treatments.

Evaluation of R294Q variant functionality after treatment with Na-PBA.

The coagulant assay led us to identify, among the six variants studied, the p.R294Q, that showed specific activity. The small gain-of-function of this variant could be very interesting for the chaperone-like approach since the rise of protein secretion needs to be translated in an increasing of the activity. For these reasons, we decided to treat for 48 hours the p.R294Q expressing cells with the Na-PBA chemical chaperone, resulted the more effective for this variant, to collect the media for antigen and activity evaluation (Figure 1.6).

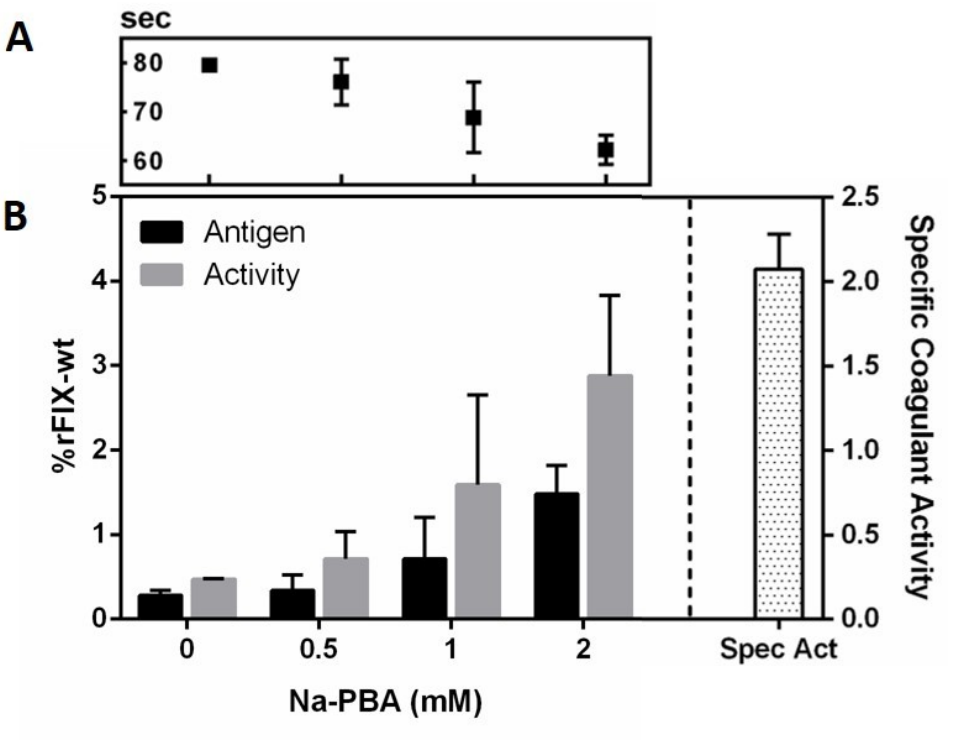

Figure 1.6: A) Reduction of aPTT (Activated Partial Thromboplastin Time) relative to the R294Q mutant treated with Na-PBA in a dose dependence manner. B) Antigen and the relative activity levels of R294Q with increasing concentrations of Na-PBA treatment, calculated on the percentage of wild type protein. The specific activity in relation with wild type is calculated on the base of the antigen levels.

To further characterize the rescue of the rFIX-R294Q variant in terms of improved secretion and function, aPTT-based assays were conducted after treatment with Na-PBA. The evaluation of secreted rFIX levels after treatment by a polyclonal ELISA revealed that protein levels increased in a dose-dependent manner to levels corresponding from 0.28±0.1% of rFIX-wt up to 0.35±0.18 (0.5 mM Na-PBA), 0.71±0.48 (1 mM) and 1.5±0.3% (2 mM) Na-PBA. Noticeably, coagulant activity measured in aPTT-based assays for the rFIX-R294Q variant from untreated cells (80±0.1 sec) significantly decreased of ~20 seconds with a similar dose-dependence after treatment with 0.5 mM (76±4.7 sec), 1 mM (68±7 sec) and 2 mM (62±3 sec) Na-PBA. Noticeably, the reliable activity/antigen ratio observed at each concentration point, resulting in a specific coagulant activity of 2.1±0.2, provides the rFIX-R294Q variant with a favorable functional output.

1.6 Discussion

In the folding pathway molecular chaperones have a big role, these protein complexes are able to stabilize the new synthetized peptides. Misfolded proteins, due to mutations disrupting the tridimensional asset of the proteins, can accumulate into the cell or be degraded by the proteasome. The result, in both cases, is a reduction of secretion and a decreased biological function. Haemophilia B is an inherited X-linked disease caused by mutation of the gene *F9* coding for Factor IX protein. In the majority of cases Haemophia B is caused by missense mutations (Figure II) that can cause an impaired protein folding.

Researchers have been studying for years molecules able to restore protein biosynthesis and modulate the folding. These compounds are called chemical and pharmacological chaperones[60]. These molecules can represent a new therapeutic strategy for diseases caused by mutations that can affect protein folding.

An interesting and relevant application of chemical chaperones was utilized for Fabry's Disease, a lysosomal accumulation disease. This approach resulted effective on rescuing protein variants with folding issues[74]. Despite the interest for the innovative potential of these compounds, there are few studies on coagulation field[75,76].

Chemical chaperones are not selective in the new peptide stabilization capacity[77]. *In vitro* studies demonstrated that chaperone-like compounds as 4-PBA (4-phenylbutirrate), a fatty acid with a low molecular weight, can restore the expression of CFTR (Cystic fibrosis transmembrane conductance regulator) in Cystic Fibrosis[78]. Furthermore, 4-PBA was successfully used to restore the secretion levels of α1-antitrypsin in presence of mutations impairing protein folding, both in cellular and animal models[79]. 4-PBA was used also to prevent and

revert the amyloid formation of Human islet amyloid polypeptide (hIAPP) associated with b-cell dysfunction and death in type 2 diabetes (T2D)[80].

To underline the potential of the chemical chaperones based approach it is worth noting that 4-PBA was approved from FDA for pharmacological use in patients affected by urea's circles diseases[77]. The 4-PBA derived compound, Na-PBA (sodium phenylbutyrate) demonstrated to prevent the ER accumulation of misfolded nephrins restoring the transportation on the cellular lumen[81]. In addition, this chemical chaperone has been applied to restore the enzymatic functionality of missense variant (as RPE65) associated with retina degeneration[82].

Pharmacological chaperones are folding mediators that are specific for protein misfolded targets. Their role is to help retrieve aberrant conformations helping them to escape from the degradation pathway[69,83]. This category of compounds has been explored for different protein aggregates pathologies, included lysosomal diseases[84]. This category involves Ambroxol chlorhydrate, a commercially available drug, that demonstrated to be effective for the treatment of Gaucher disease[85]. N-acetylcysteine (NAC) was reported as a new allosteric chaperone for Pompe disease [86]. Another example of pharmacological chaperone is the VX-809 compound, it has been explored for the correction of folding defects of CFTR[87] and it has been the object of clinical phase 3 studies[88].

Due to all these encouraging results using chemical and pharmacological chaperones, the usage of these molecules could be interesting for the development of innovative therapeutic strategies for human genetic diseases impairing protein folding. An important aspect of Haemophilia B is that also that a small increase of functional protein leads to an improvement of the patient clinical phenotype. With all these premises, we hypothesized that a chaperone-based approach could be effective to improve the intracellular processing of different protein variants.

To investigate the potential of this approach we took advantage of two expression systems, a transient model, useful to assess protein secretion levels, and a stable model, to study the different efficiency of the different treatments. In the first part

of our study we confirmed that our transient model was able to mimic the clinical phenotype observed in patients (type I Haemophilia B <1% of wild type protein) in two different cell lines. We created stable clones expressing the different missense variants, to study the intracellular trafficking and the secretion response to the different compound in order to have the same basal protein expression for each cell. The immunofluorescence studies brought us to characterize the molecular defects resulting from each missense mutation. Data analysis showed how the wild type protein colocalized mainly with the Golgi Apparatus, on the contrary missense variant resulted in a high colocalization with the Endoplasmic Reticulum and a negligible colocalization with the Golgi Apparatus (Figure 1.3), this can be due to an accumulation of the misfolded proteins in the ER explaining the low secretion levels observed (Figure 1.2A).

Screening with chaperone-like compounds revealed that in our model only two chemical chaperones were effective for our variants 4-PBA and Na-PBA and were able to increase the secretion in dose dependent manner. Treatment (1mM and 2mM) revealed an increase of protein secretion and did not result toxic in our cellular model. Interestingly, both 4-PBA and Na-PBA showed an increased secretion in the same variants (p.Y115C, p.Y161C, p.R294Q and p.Y305C). Na-PBA displayed a fold increase from 3.9 to 7 according with the different FIX variants (Figure 1.4) and 4-PBA from 2 to 6-fold compared with the untreated. The observation of the compound efficacy drove us to explore the specific coagulant activity, the aim being to verify that the secretion increment could have functional meaning. For these reasons, variant functionality was evaluated in FIX deficient plasma, that, having a very low activity (<1%) is an efficient method to mimic the haemophilic phenotype. The analysis obtained from coagulant activity and protein antigen revealed that only a variant (p.R294Q) has an encouraging activity for this type of approach. This variant displayed a specific activity of 2 compared with the wild type protein, that could explain the residue activity differences found in patients. Moreover, activated partial thromboplastin time (aPTT) in the concentrated medium after treatments with Na-PBA showed a reduction of the

time of coagulation (≈20 sec) at increasing concentrations of the compound in a dose dependent manner. From these evidences, treatments of the p.R294Q variant with the Na-PBA chemical chaperone could drive to overcome the therapeutic threshold up to 5%, converting a severe to moderate phenotype.

Interestingly, the amino-acidic change of the 294 variant, an arginine to a glutamine, seems to have a positive impact on the protein functionality comparing with the variants presenting a tyrosine to cysteine change, and also from phenylalanine to leucine. Specifically, the tyrosine to cysteine change could have worst consequences because of the disulfide bridges that this amino acid could form if it is exposed in the tridimensional protein structure.

Amino-acidic changes that can lead to a gain of function are rare and unexpected, but in the case of FIX protein it was already observed in the p.R384L, also named Padua[89]. The positive effect of this chaperone-like compound on this specific mutation is particularly relevant for the high frequency of the mutation in patients (97 as showed in Table 1.2), the highest number of patients associated with a missense mutation causing severe Haemophilia B[73]. Furthermore, the compound that demonstrated an effect for the FIX missense variant, Na-PBA, is particularly relevant in the therapeutic point of view since this compound was previously confirmed as safe and effective and dispensed to patients with urea's circles diseases, in particular for the Ornithine Transcarbamylase Deficiency (OTC)[90].

This study represents the first extensive application of chaperone-like compounds treatments on in vitro models of a human disease that involves a secreted protein with enzymatic properties. The identification of responsive variants to this kind of treatments could encourage studies on more specific compounds able to increase the protein secretion levels and the activity.

Chapter 2- Transcriptional Activation of F7 And F8 genes by CRISPR-mediated Activation

2.1 Engineered transcription factors: overview

Modulation of gene expression has been very interesting for decades. The ability to regulate expression is essential to the study of biology, from basic research to translational applications for the treatment of disease. Since the elucidation of the central dogma of molecular biology, we have been searching for ways to manipulate and perturb gene expression[91].

The first attempt to modulate gene expression was made by taking advantage of a class of proteins that naturally have this role in eukaryotic regulation: Zinc Fingers Proteins (ZF). Zinc fingers are found in 2% of all human genes, and they are by far the most abundant class of DNA binding domains (DBD) found in human transcription factors[92]. ZF proteins have a structure able to bind the DNA in a sequence specific manner. The protein is composed of a short stretch of 30 amino acids, containing two conserved cysteines and two conserved histidines[93]. The conserved residues coordinate a zinc ion that allows the finger to fold into a

compact structure containing a β-turn (which includes the conserved cysteines) and an α-helix (which includes the conserved histidines)[94,95]. C_2H_2 zinc fingers typically occur in tandem arrays, and many transcription factors have three or more fingers working in concert to recognize DNA, thereby targeting the transcription factor to its appropriate promoter. This particular structure permitted to engineer the ZF proteins to recognize specific sequences on the genomic DNA in order to have a different function depending on the effector that can be fused to the DNA-binding domain. ZF nucleases have been used for chromatin modification, DNA specific cleavage and gene activation and repression[96]. ZF TFs can contain several kinds of activation or repression domains. In many applications, ZF TFs containing the VP16 domain from herpes simplex virus or the p65 domain from the cellular transcription factor NF-kB45 are used for activating transcription. The same ZFP DNA-binding domain can be used for repressing transcription by adding the KRAB domain, that is the most common ZF repressor in human genome. Despite many achievements in using these ZF based DNA binding domains (DBD) linked to several actuators in different biological applications, there are many issues related to their engineering that have limited the widespread adoption of this platform.

The discovery of Transcription Activator-like Effectors (TALE) in *Xhantomonas* bacteria move forward the usage of DBDs proteins. Similarly to ZF proteins they show a modular DBD that can be modified to determine the specificity to a particular sequence in the genome. TALE proteins DBD is composed of a variable number of tandem repeats that differ from each other only for the amino acids in position 12 and 13 of each monomer. These residues, called repeat variable diresidue or RVD, determine the recognition with a specific nucleotide in the DNA sequence with a specific code for A, T, C and G (NI, NG, HD and NN respectively). The possibility in engineering these TALE proteins in a relatively simpler way compared to zinc fingers has determined their broad usage in last years[97].

Moreover, a third technology has emerged based on the RNA-guided DNA endonuclease Cas9 from the type II bacterial adaptive immune system CRISPR[98]. It differs from the previous ones because the targeting of specific DNA loci is due to the complementary pairing with a guide RNA (gRNA) that form a complex with Cas9 endonuclease, meaning that the system requires only to exchange the nucleotide into the gRNA expression cassette, and not to design an entire new protein.

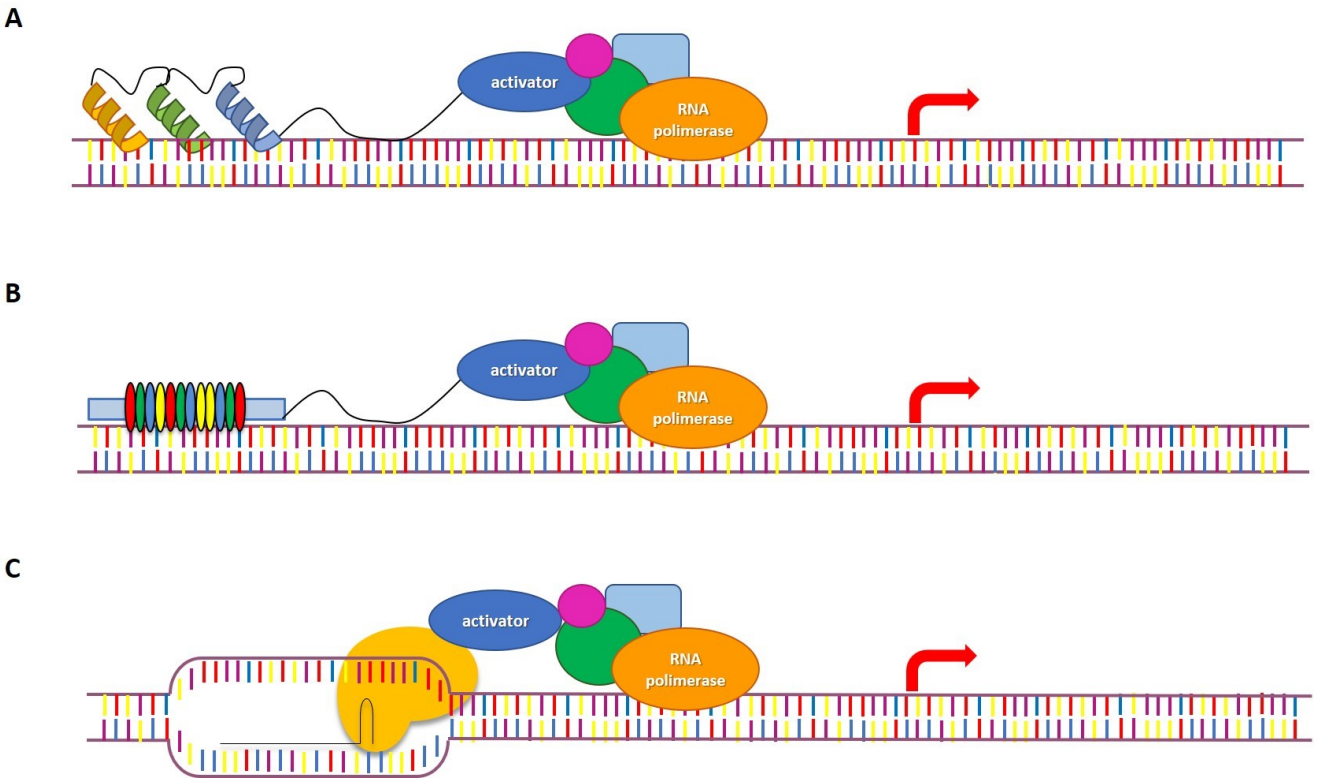

Figure 2.1 Technologies for engineering programmable DNA-binding proteins, including (A) Zinc Finger proteins, (B) TALEs and (C) CRISPR/Cas9.

2.2 Background and Rationale

The CRISPRa (activation) system

The Cas9 endonuclease is directed to the target sequence into the genomic DNA, this capacity is the key of the CRISPR activation system. To convert Cas9 from a DNA scissor into a gene activator, the Cas9 nucleasic domains, the RuvC and HNH, can be made inactive (nuclease-deactivated Cas9; dCas9). Disrupting these domains renders the Cas9 protein unable to cut the DNA sequence but still able to be recruited on the gRNA binding region. It is then possible to fuse transcriptional effectors directly to the dCas9, which essentially transforms the dCas9-effector fusion into an easily programmable artificial transcription factor upon being paired with a target-specific gRNA. As the RuvC and HNH domains are conserved among Cas9s from other bacterial species, this approach provides a general strategy for repurposing orthogonal Cas9s into RNA- guided DNA-binding proteins.

There are different kinds of activators that can be fused to the dCas9. The first-generation activators, in eukaryotic cells, consists of dCas9 directly fused to strong transcription activator domains of p65 or VP64, an engineered tetramer of the herpes simplex VP16 transcriptional activator domain[99] (Figure 2.2 B). Several studies demonstrated that dCas9-VP64 fusion was more effective than p65, and this fusion protein is able to activate endogenous genes and reporters[100,101]. However, the activation seen in mammalian cells is usually moderate, about 2 to 5-fold using a single gRNA. The second generation of CRISPRa are the result of several attempts to improve the direct fusion. One of the strategies is to amplify the activation by transforming the dCas9 into a scaffold that can recruit many copies of activator[102]. This strategy can be achieved fusing to dCas9 a tandem array of peptides, which recruits many copies of the VP64 effector (Figure 2.2 C),

that are fused to an engineered portion of an antibody that is able to bind the peptide fused to dCas9 called SunTag. This second-generation approach, compared with the 2-fold increase observed with dCas9-VP64 alone, allows to estimate a stronger transactivation effect. Another approach involving the recruitment of transcription activators by mediators it was described by Konermann et al[103]. This system takes advantage of a synergic activation mediator (SAM) (Figure 2.2 E) that is composed by multiple transcriptional activators that create a synergistic effect. This tool is structured with a dCas9-VP64 engineered with additional features to enhance activator recruitment, this mediation is made by one or two copies of an RNA hairpin from the MS2 bacteriophage. This hairpin interacts with an RNA binding protein called MCP (MS2 coat protein) that is fused to an activator (p65). Another strategy for CRISPR-dependent gene activation is composed by a tripartite effector fused to dCas9, the activators are VP64, p65 and Rta (Figure 2.2 D) all in tandem and are called VPR[104]. These activators are fused in order to obtain a strong gene activation. This dCas9-VPR was successfully employed in human, mouse, *Drosophila melanogaster* and *Saccharomyces cerevisiae* cells. In addition, it can upregulate endogenous expression from 5- to 300-fold at mRNA level compared to the dCas9-VP64.

Finally, another system was fused to the deactivated nuclease Cas9, this is a histone acetyltransferase called dCas9-p300. In contrast with the previously described activators, this tool, acting as an epigenome editing platform[105], and is able to promote activation at both proximal and distal enhancers of genes. In future studies, it will be interesting to fuse other epigenome modifiers to dCas9, these can be used to probe the effects that epigenetic changes could have on gene expression levels.

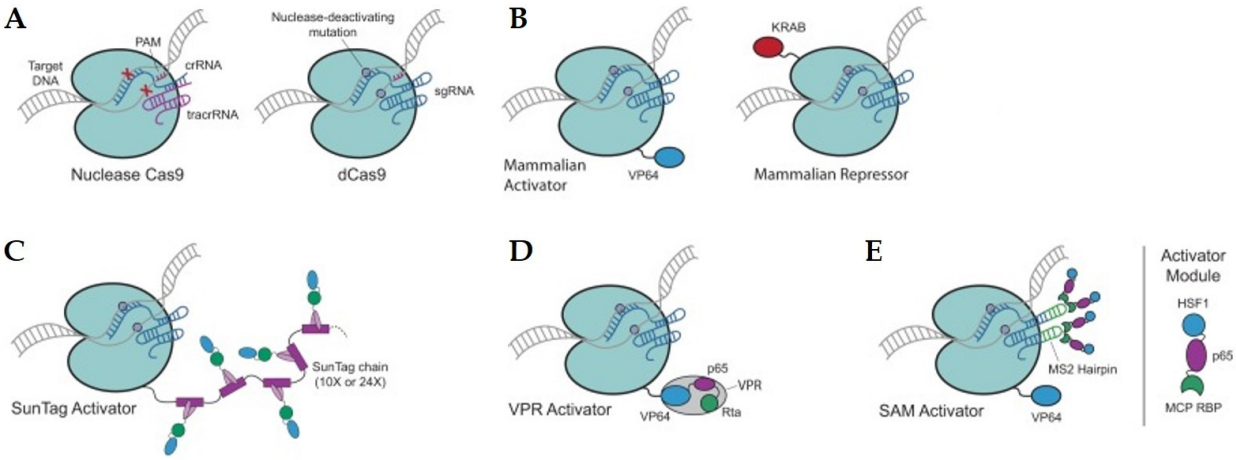

Figure 2.2 Engineered CRISPRa systems. (A) This Cas9-RNA complex is able to cleave the DNA targeted by the crRNA and which is adjacent to a PAM site (red). To adapt CRISPR for gene regulation, mutations in the nuclease domains have been introduced into Cas9 and the crRNA:tracrRNA complex convertedto a single guide RNA. (B) To turn dCas9 into an artificial transcription factor it can be fused with a VP64 activator (left) or a KRAB repressor (right). (C) The SunTag activation system (D) The VPRa ctivation system is dCas9 fused to VP64, p65, and Rta linked in tandem. (F) The SAM activation system. (Figure adapted from La Russa et al. 2015)

Application of TALE TF approach to FVII Deficiency

In recent years the transcription activation strategy has been perfected, and all the DNA binding protein exploited. The capability of regulating gene expression in the native chromosomal context may represent a way to address several genetic diseases. Transcription impairment by mutation involving promoter regions represents a small but considerable cause of severe coagulation factor defects and of all inherited human diseases. These mutations can often occur in transcriptional regulatory regions, leading to an impairment of the expression of the gene also in the presence of a unmutated coding sequence. The enhancement of transcriptional activity can effectively restore the gene expression at therapeutic levels, and most importantly in monogenic disease cases.

Driven by all these reasons our group previously explored TALE-TF proteins to restore the transcription of coagulation FVII in presence of promoter mutations. In

this work we tested different TALE proteins targeting a sequence into the FVII promoter to evaluate the efficacy in the increment of transcription. This is due to the fusion of the TALE DBD to a strong transcriptional activator: VP64. This kind of approach was previously described for the enhancement of frataxin gene, and for this disease it was demonstrated to be very effective. Moreover Chapdelaine et al moved over to express TALE proteins into AAV9 vectors in order to restore frataxin transcription[106].

Factor VII promoter mutations are about the 8% of the total mutations causing FVII Deficiency. We explored two different promoter mutations the -61 T>G and the -94 C>G impairing two natural transcription factors binding motifs for HNF4 and SP1, respectively. These mutations were found only in two patients[107,108] and caused a severe FVII deficiency phenotype. Our goal for this work was to restore the transcription of *F7* mRNA in presence of these mutations. To understand which region of the promoter was the best target for a transcription activation we designed different TALE proteins (named 1, 2, 3 and 4) targeting a wide region of the *F7* promoter using computational analysis[109]. For the first screening we took advantage of the luciferase reporter to evaluate both the wild type and the promoter activity of the two mutants, and the analysis demonstrated that the two mutations demolished the promoter activity at least of a half. Subsequently, we tested the four TALE proteins with the reporter constructs, carrying the wild type and the mutant promoters, in order to evaluate an increment of the reporter gene signal. Interestingly, only the TALE protein number 4 had an effect of the reporter expression. The selection of the most effective TALE protein prompted us to assess the transcription increase on endogenous *F7* gene in HepG2 cells. Interestingly, the transactivation of the endogenous *F7* mRNA from the analysis on total RNA was about 3.6-fold compared to the untreated cells, and we confirmed the resulted increase at the secreted protein level.

The effectiveness of the TALE protein to restore the mRNA and protein of *F7* was observed also in Hep10 cells, primary hepatocytes able to mimic the patients' hepatocytes transcription factors context. Using AAV8 we transduced the primary

hepatocytes both with the TALE-TF4 and with a negative control, AAV8-DBD. The AAV8-TALE-TF4 protein confirmed, also in this context, an increase of *F7* mRNA about 3.7-fold compared with the negative control. The protein quantification established that the increment of mRNA led to an increment of FVII protein in cell media. Being FVII a serine protease with enzymatic activity we performed FXa generation assays[110] in medium, that revealed a significant shortening of lag times between AAV8-TF4 and AAV8-DBD transduced cells, indicating an increase in the secreted levels of FVII protein able to promote the generation of FXa in plasma systems.

2.3 Aim of the study

In a previous study, we explored the Transcription Activation-like Effectors (TALE) for FVII deficiency and we confirmed that this system was able to increase the secreted protein levels.

Driven by the encouraging data obtained using TALE-TF protein for FVII deficiency we decided to explore the CRISPR activation system. Specifically, we chose to use the second generation Cas9 system fused to the tripartite activator VPR. This deactivated Cas9 protein was fused directly with three different activators: VP64, from herpes simplex (composed by four VP16 subunits) p65, also called RELA is human derived and Rta, that is an Epstein-Barr virus transcription factor. The dCas9-VPR activator demonstrate an effect higher than dCas9-VP64 (300-fold)[104]. This is the reason that drove us to explore this system for our model.

FVII deficiency is a rare coagulation disorder and as a model is effective because a small increase of protein secretion can ameliorate the clinical phenotype. We decided to apply the CRISPR activation system to the wild type and the -61 T>G mutation.

Our specific aim was to compare the CRISPR activation system with the TALE-TF system in the same context, the FVII deficiency. The CRISPRa system is versatile and easier to design compared to TALE-TF.

Further, we wanted to assess if the CRISPRa system could be effective for *F8* promoter. This tool can represent a therapeutic strategy also for Haemophia A, the most frequent X-linked bleeding disorder by maximizing the transcript levels and protein secretion.

2.4 Material and Methods

Design and assembly of F7 and F8 gRNAs

The different gRNAs targeting the *F7* and the *F8* promoter were designed according to the Benchling website tool[111]. The *F7* and *F8* promoter regions were uploaded and used as template to design the different gRNAs. Specifically, the $gRNA_{F7.5}$ was designed to recognize a sequence resembling that of the TALE-TF4. All gRNAs were assembled using the pMLM3636 vector (Addgene #43860) as a backbone with primers shown in the Table 2.1. The dCas9 – VPR expressing vector was collected from the Church Lab (Addgene #63798)[104].

$gRNA_{F7.1}$ F	ACACCGTCCCATGGGGAATGTCAAC
$gRNA_{F7.1}$ R	AAAACGTTGACATTCCCCATGGGAC
$gRNA_{F7.2}$ F	ACACCGACATTCCCCATGGGACTGA
$gRNA_{F7.2}$ R	AAAACTCAGTCCCATGGGGAATGTC
$gRNA_{F7.3}$ F	ACACCGCCTGTTGACATTCCCCA
$gRNA_{F7.3}$ R	AAAACTGGGGAATGTCAACAGGC
$gRNA_{F7.4}$ F	ACACCGGGGAATGTCAACAGGCAG
$gRNA_{F7.4}$ R	AAAACCTGCCTGTTGACATTCCCC
$gRNA_{F7.5}$ F	ACACCGCCTCCAAGGGTGACAGAG
$gRNA_{F7.5}$ R	AAAACTCTGTCACCCTTGGAGGCG
$gRNA_{hF8\ 1}$ F	AAAACTTAAGGAAAGGAGGAGAGGGC
$gRNA_{hF8\ 1}$ R	ACACCGCCCTCTCCTCCTTTCCTTAA
$gRNA_{hF8\ 2}$ F	AAAACCAGGAAGAACTGAAGTAGCAC
$gRNA_{hF8\ 2}$ R	ACACCGTGCTACTTCAGTTCTTCCTG
$gRNA_{hF8\ 3}$ F	AAAACATCGGTTACTGCTTAGTGCTC
$gRNA_{hF8\ 3}$ R	ACACCGAGCACTAAGCAGTAACCGAT
$gRNA_{hF8\ 4}$ F	AAAACAATTTCTTTGCAGAGCATTTC
$gRNA_{hF8\ 4}$ R	ACACCGAAATGCTCTGCAAAGAAATT
$gRNA_{hF8\ 5}$ F	AAAACAGGAGGGGAAAAAAGTAAAAC
$gRNA_{hF8\ 5}$ R	ACACCGTTTTACTTTTTTCCCCTCCT
$gRNA_{hF8\ 6}$ F	ACACCGTGTTCAACTGGAGAAGCAAA
$gRNA_{hF8\ 6}$ R	AAAACTTTGCTTCTCCAGTTGAACAC

Table 2.1 Primers for the creation of gRNAs target sequence.

Construction of the reporter plasmids

- pGL3 FVII wt and pGL3 -61T>C vectors were obtained by amplification of a 520bp human FVII promoter region from genomic DNA and site direct mutagenesis according to a previous work[112].
- pGL3 FVIII vector was the result of the amplification of a 1175bp sequence, the entire *F8* promoter, with primers containing XhoI and HindIII recognition sites at 5′ ends for a directional cloning of the insert.

FVIII prom F XhoI:

5′- AAA CTCGAGAAATAAATGAATAAATGCCA – 3′

FVIII prom R HindIII:

5′ – AAA AAGCTTCATGACTTATTGCTACAAAT – 3′

This region was subsequently cloned into a commercial pGL3 Basic Vector (Promega) through the two restriction sites.

Cell culture and reporter activation assay

The human hepatocyte carcinoma HepG2 and the human hepatocellular carcinoma Huh-7 cell lines were maintained under 37◦C, 5% CO2 using Dulbecco's modifed Eagle's Medium supplemented with 10% FBS, 2mM GlutaMAX (Invitrogen), 100U/ml penicillin and 100 µg/ml streptomycin. Firefly luciferase reporter activation was assessed by co-transfecting HepG2 and Huh7 cells with plasmids carrying Luciferase reporters both with dCas9-VPR and gRNAs. HepG2 cells were seeded into 12-well plates the day before transfection at densities of $0.2x10_6$ cells/well. Approximately 24 hours after initial seeding, cells were

transfected using Lipofectamine 2000 (Invitrogen). For 12-well plates we used 2µg total DNA inclusive gRNAs and dCas9-VPR plasmids and Reporters, and 100ng of Renilla Luciferase. After 48 hours from transfection, the reporter activity was established by luminometer Glomax 20/20 measurement in Dual-Luciferase assay (Promega). All fold-induction values were normalized to the activity level of Renilla Luciferase for each transfection. All experiments were performed according to the manufacturer's recommended protocol.

RT-PCR and qPCR for FVII mRNA

HepG2 cells were seeded in 12-well plates the day before transfection at density of $0.2x10^6$ cells/well. After 24 hours from seeding, cells were transfected with 2 µg total DNA of dCas9 – VPR and gRNA combination or TALE-TF. Approximately 48 hours after transfection the total RNA was extracted with RNAeasy Kit (Qiagen). The cDNA was generated by M-MLV Reverse transcriptase (Invitrogen) according to the manufacturer's protocol. The mRNAs copies were determined using pCMV-cDNA FVII at different concentration as a standard curve and using the F7 ex5 F and F7 ex7 R primers. *F7* transcript was detected by PCR with primer F7 ex5 F and F7 ex7 R. 18S and *F7* mRNAs were detected by qPCR using SsoFast EvaGreen Supermix (Bio-Rad).

Evaluation of FVII protein activity

Activated factor X (FXa) generation assays were performed as described[110]. Medium from transfected cells was added 1:1 (vol/vol) to commercial FVII-deficient plasma (BioMedical Inc., George King, USA), adjusted to a final dilution of 1:20 in a dilution buffer (20 mM HEPES, 150 mM NaCl, 0.1% PEG-8000, pH 7.4)

and sub-sampled in a 96-well microplate (Costar, Corning, NY, USA). Coagulation was triggered by the addition of a mixture containing a specific fluorogenic substrate for FXa (150 μM, American Diagnostica Inc., Greenwich, USA) and 1/100 Innovin (Siemens Healthcare, Marburg, Germany) prepared in a reaction buffer (20 mM HEPES, 150 mM NaCl, 5 mM CaCl2, 0.1% PEG-8000, pH 7.4). The generation of FXa was measured as fluorescence emission (Relative Fluorescence Units, RFU; 360 nm excitation, 465 nm emission) over time on a microplate fluorimeter (Fluoroskan Ascent FL, Thermo Fisher Scientific, Helsinki, Finland). FXa generation assays were analysed by the statistical soſtware GraphPad Prism 6 (GraphPad Soſtware, San Diego, USA). The specific parameter lag time (expressed in minutes) in FXa generation assays was extrapolated from the first derivative of relative fluorescence units (RFU) as a function of time (minutes).

2.5 Results

F7 model: comparison of TALE-TF and CRISPRa systems.

Assembly and design of FVII targeting gRNAs

We considered as our first models the wt *F7* promoter and one natural variant, the -61 T>G transversion, which falls in the binding site for the hepato-specific HNF-4 transcription factor (Figure 2.3). The proximal promoter region of *F7* gene has been shown to maintain the maximal promoter activity in HepG2 cells[113], with the HNF-4 and Sp1 binding sites included in the first 185 bp upstream of the translation starts site (TSS). We previously generated the reporters vectors for the *F7* promoter, wt and mutant, by cloning 520 bp of the human *F7* promoter region upstream of the coding sequence for firefly luciferase, thus obtaining the pGL3 FVII wt and pGL3 -61G plasmids (Figure 2.3).

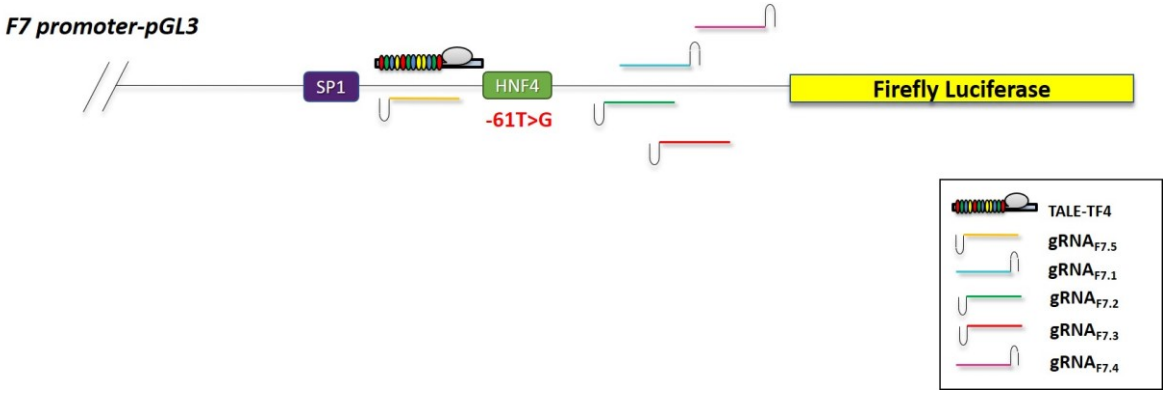

Figure 2.3: Proximal F7 promoter (520bp) reporter vector scheme and TALE-TF4 and gRNA binding sequences. The gRNAs were designed upstream the TSS and gRNA$_{F7.5}$ was designed to recognize the same sequence of TF4.

Based on computational analyses of the *F7* promoter that we performed for the previous work and of the TALE-TF number four (TF4) binding site, resulting the best to restore the transcription of *F7* transcript, we took advantage of a web tool[111] for the design of a panel of gRNAs. One of these gRNAs was designed to target almost the same region of TF4, between the SP1 and HNF-4 target sites. Indeed, the gRNA design is strictly bound to the presence of a trinucleotide sequence called PAM, without this NGG sequence the dCas9 is not able to recognize the target. The other gRNAs were designed in a region that was not studied in the previous work ranging from -52 to -22 of F7 5'- flanking region (Figure 2.3). In order to reduce off-target activation we targeted sequences from 18 to 20 bp in length spanning the region between HNF-4 binding and ATG transcription start site on FVII promoter predicted to be unique in the human genome by performing an in-silico analysis. The gRNAs were designed to target both the forward and the reverse strand on the base of the PAM sequence position, the selected sequences were cloned into a gRNA expressing vector, containing the scaffold that is able to recruit the dCas9 protein on the target site (Figure 2.4). To enhance the transcription activity, we decided to use the catalytically inactivared dCas9-VPR from *Streptococcus Pyrogens* designed by Chavez et al. Both the gRNAs and the dCas9-VPR vectors are commercially available from Addgene plasmid repository (Figure 2.4).

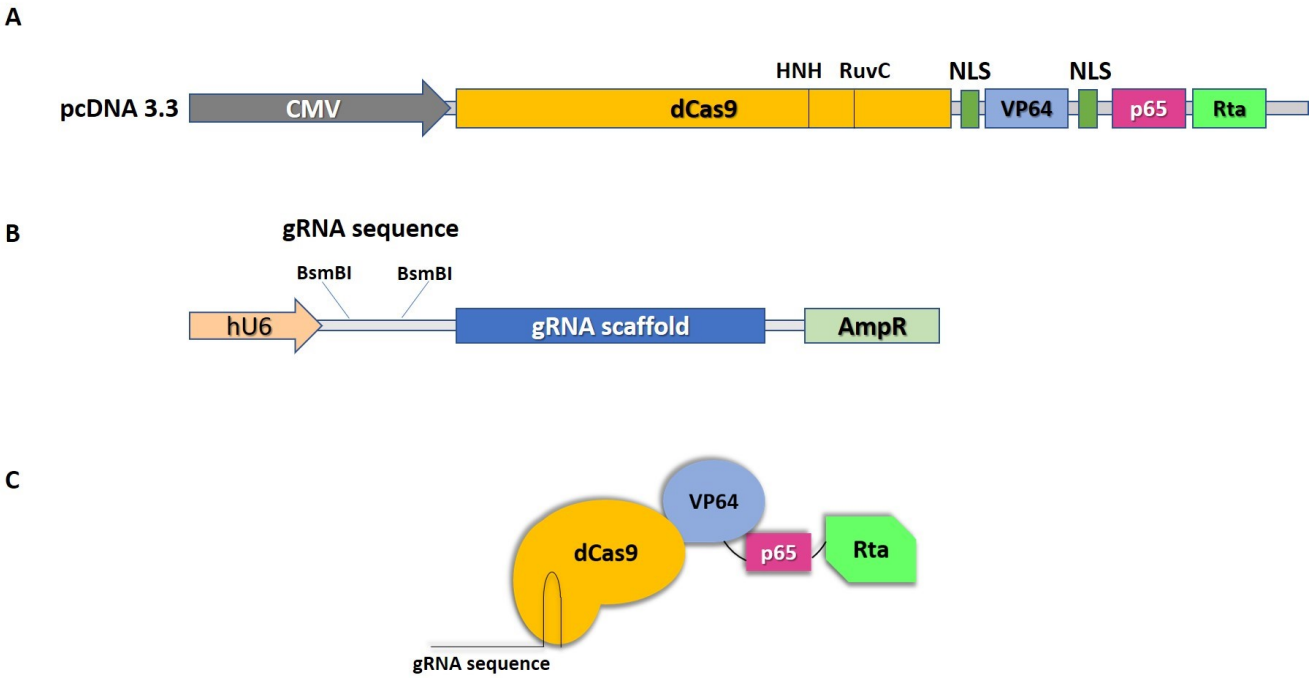

Figure 2.4: Scheme representing the spCas9-VPR expressing vector and gRNA expressing backbone. A) The dCas9-VPR protein is under the control of CMV promoter and fused to a tripartite activator domain, composed by VP64, p65 and Rta able to recruit the transcriptional machinery on a specific target drove by each gRNA. Moreover, the dCas9-VPR vector contains two NLS signals to enhance the translocation of the protein into the cellular nucleus. B) Each synthetic gRNA is expressed under the hU6 promoter control and the vector is designed to have two BsmBI sites to clone the gRNA target sequence of interest. C) Final riboprotein composed by the dCas9-VPR and the gRNA molecule.

Evaluation of promoter activity and validation of the efficacy of the CRISPRa system in two different cell lines.

The expression of the pGL3$_{F7wt}$ and pGL3$_{-61G}$ reporter plasmids both in HepG2 cell line and Huh-7 cell line confirmed the causative effect of the -61G mutation, with an abolishment of firefly luciferase transcriptional levels compared to the pGL3 FVII wt (Figure 2.5 A, D). The reporter gene under the control of the wild type promoter increased in the presence of each gRNAs, from 132±9.7 to 300±18 fold for the Huh-7 cells and from 33.1±9.1 to 36±0.5 fold for the HepG2 cells, except in the

presence of the $gRNA_{F7.2}$ which showed a minimum effect in both cell lines (Figure 2.5 B, E). Most importantly, we observed a robust increase in the transcriptional levels of the reporter gene guided by the promoter mutant co-transfecting the cells with the different gRNAs, from 580±4.3 to 927.6±34.6 in Huh-7 and from 204.7±15.1 to 467.5±192.7-fold increase in HepG2, with the same exception for the $gRNA_{F7.2}$ that showed a lower effect in both lines (16.9±0.75 and 21.8±9.5 respectively) (Figure 2.5 C, F). A negative control was represented by the dCas9-VPR vector without the gRNA and it did not show an increase of luciferase levels.

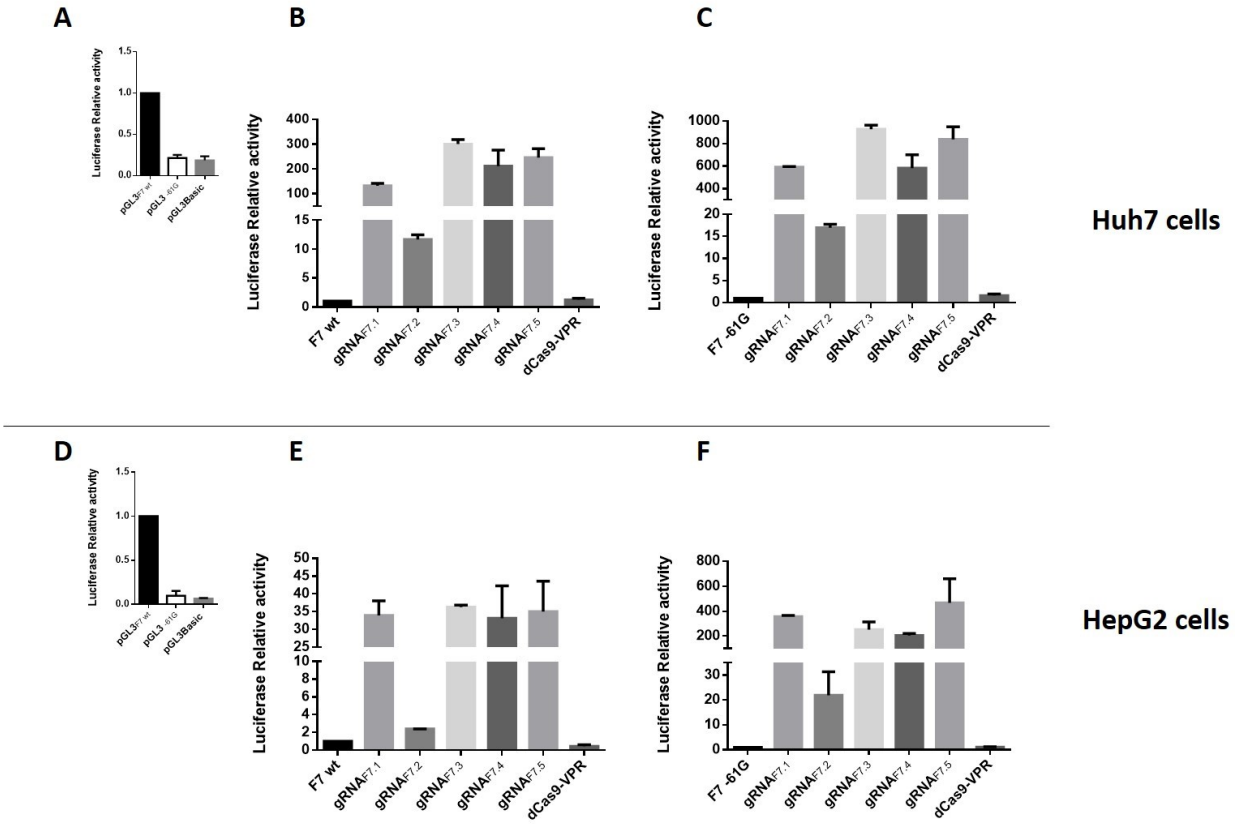

Figure 2.5 Co-expression of the reporter plasmids wt and pGL3 -61G with each gRNA designed for F7 promoter. A) Assessment of the impairment of the reporter expression in presence of -61 mutation compared with the pGL3 Basic, a negative control in Huh-7 cells. B) Reporter assay of the co-expression of the pGL3 wild type promoter in presence of the different gRNAs in Huh-7 cells, fold increases are calculated on the base of pGL3wt alone. C) Reporter assay of the -61 mutant co-expressed with the panel of gRNAs in Huh-7 cells, fold increases are calculated on pGL3 -61 alone. D)

Assessment of the impairment of the reporter expression in presence of -61 mutation in comparison with the pGL3 Basic, negative control in HepG2 cells. E) Reporter assay of the co-expression of the pGL3 wild type promoter in presence of the different gRNAs in HepG2 cells, fold increases are calculated on the base of pGL3wt alone. F) Reporter assay of the -61 mutant co-expressed with the panel of gRNAs in Huh-7 cells, fold increases are calculated on pGL3 -61 alone.

In vitro comparison of TF4 and gRNA$_{F7.5}$-dCas9-VPR: Luciferase Assay

The gRNA$_{F7.5}$ was designed to target a *F7* promoter sequence similar to that targeted by the TF4. For this reason, the comparison of the two systems was made only between TF4 and gRNA$_{F7.5}$. In order to compare the CRISPRa and TALE-TF systems we took advantage of the in vitro Luciferase assay. We tested the reporter gene increase in Huh-7 and HepG2 cell lines, transfecting both the pGL3$_{F7wt}$ and pGL3$_{-61G}$ with TF4 or gRNA$_{F7.5}$ and the dCas9-VPR expressing vector. We observed that the transactivation of the two reporter genes under the control of the different promoters increased in both cellular lines and using both systems. In Huh-7 cells we observed a strongest increase using the gRNA$_{F7.5}$ and dCas9-VPR (600-fold increase), both for the wild type and the mutant promoter. The TF4 confirmed what we observed previously, an increase of transcription of the reporter up to 80-fold (Figure 2.6 A). We compared the two systems in HepG2 to explore a different hepatic cell line, and we confirmed that the CRISPRa system is able to increase the reporter of 30-fold compared with a 15-fold of TF4 (Figure 2.6 B). Interestingly, the reporter gene increase of the pGL3$_{-61G}$ resulted higher compared with the wild type, both for TF4 and gRNA$_{F7.5}$, resulting in an increase of reporter expression of 2500-fold and 60-fold using the gRNA$_{F7.5}$ and dCas9-VPR (Figure 2.6 A and B).

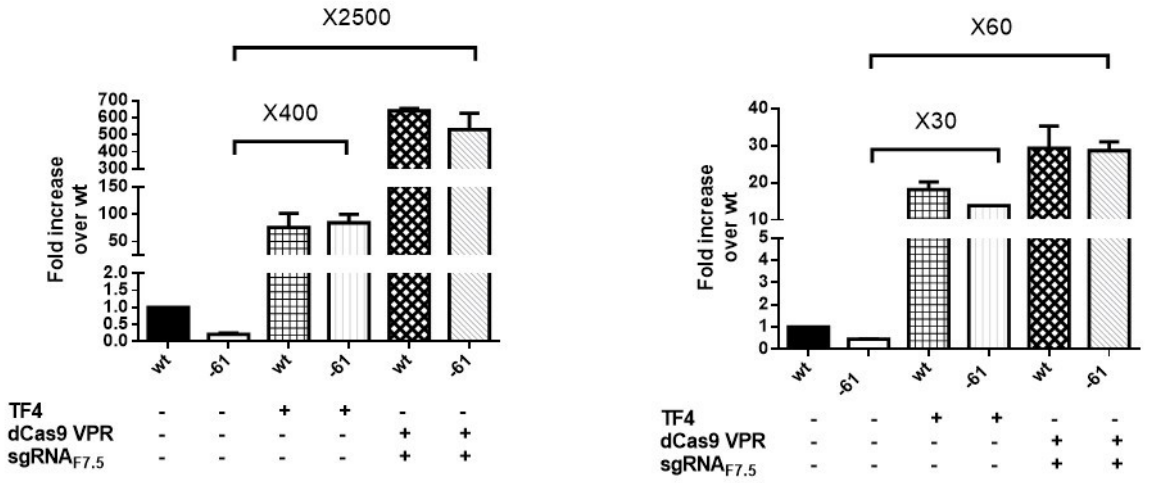

Figure 2.6 A) Huh-7 cell line: co-expression of the reporter plasmids $pGL3_{F7wt}$ and, $pGL3_{-61G}$ alone, with the synthetic TALE-TF4 and the dCas9-VPR/$gRNA_{F7.5}$. Histograms report luciferase relative activity as change fold relative to $pGL3_{F7wt}$ alone. B) HepG2 cell line: co-expression of the reporter plasmids $pGL3_{F7wt}$ and, $pGL3_{-61G}$ alone, with the synthetic TALE-TF4 and the dCas9-VPR/$gRNA_{F7.5}$. Histograms report luciferase relative activity as change fold relative to $pGL3_{F7wt}$ alone.

Effect on the endogenous context of HepG2 cells: mRNA

We chose to use the HepG2 cell line for endogenous mRNA expression, since it is a highly used liver-derived cell line, that we formerly exploited for *F7* promoter activation. To assess the capability of the $gRNA_{F7.5}$ dCas9 ribonucleoprotein complex to enhance the transcription in a chromatin context we transfected HepG2 cells with both $gRNA_{F7.5}$/Cas9 and TF4 protein in order to evaluate the best candidate able to increase transcription. Because of the low transfection levels efficiency assessed in this cell line (<3%) and the *F7* mRNA normally expressed by the hepatic cell line it was difficult to discriminate the enhancement of F7 mRNA from the background. However, the quantification of the mRNA through qPCR showed similar number of *F7* mRNA copies in all the samples, with no significative changes (Figure 2.7). This limit can be bypassed using a positive selection of transfected cells.

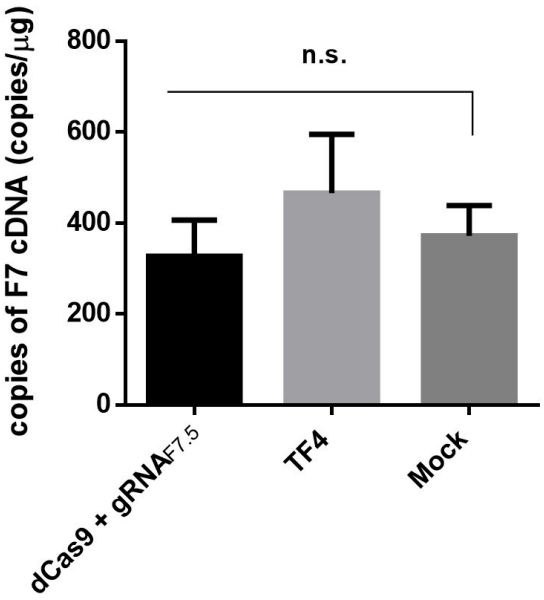

Figure 2.7: Effect of gRNAF7.5/Cas9 and TF4 on the transcriptional levels in HepG2 endogenous context. Copies of F7 cDNA transfected with CRISPRa system in comparison with the TALETF system and untransfected cells.

Correlation of TALE-TF and CRISPRa system in endogenous context: FVII protein activity.

After the promising episomal reporter data and difficulty to discriminate the mRNA increment by qPCR we decded estimate FVII protein increase in HepG2 cells. The treatments with the two systems were compared in terms of effect in protein secretion and relative activity. Being FVII a serine-protease this evaluation was made taking advantage of a FXa generation assay[110], for the standard curve we decided to use pooled normal plasma (Pnp) in order to estimate the relative

amount of FVII protein in correlation with the Pnp lag times. The fluorogenic assay showed a higher increase of FVII and relative activity, displaying shorter lag times, in cells treated with $gRNA_{F7}$/dCas9-VPR (0.11±0.008% of Pnp) than with TF4 (0.066±0.009% of Pnp) both compared with endogenous levels (0.017±0.005% of Pnp). These data showed an increase of protein that leads to a shortening lag times in presence of the CRISPRa system compared with the endogenous levels (Figure 2.8).

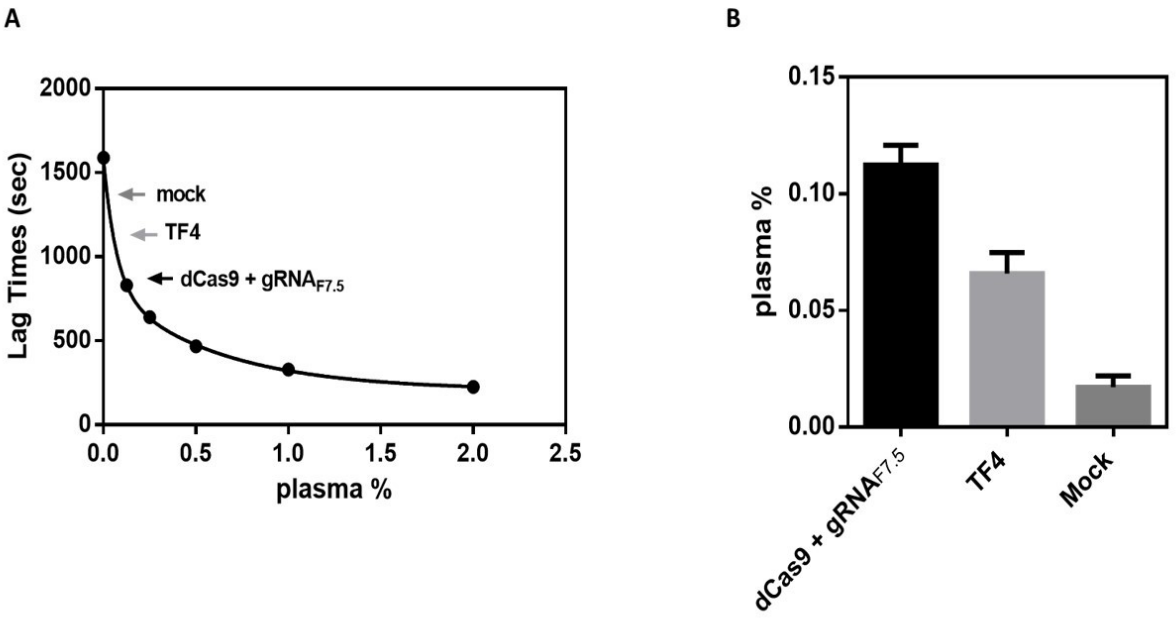

Figure 2.8: FXa generation activity in FVII-deficient plasma supplemented with medium from HepG2 cells.A) The graph reports the standard curves lag times in presence of different concentrations of Pooled normal plasma. B) The graph reports the relative amount of FVII in percentage of Pnp derived from the interpolation in the standard curve.

CRISPR activation on F8 promoter model:

Assembly of the F8 reporter plasmid and gRNA design.

Encouraged by the promising data on the *F7* promoter, we decided to design a transactivation system based on gRNAs and dCas9-VPR complex for FVIII. The *F8* promoter has not been extensively studied and it reported a weak expression. For these reasons we decided to clone the entire promoter (about 1.2Kb) into a plasmid expressing the Luciferase Reporter gene called pGL3, obtaining the $pGL3_{FVIII}$ (Figure 2.9). Based on previous studies on promoter activity and *F8* promoter mutations[114] we designed the gRNA target sequences through computational analysis. We decided to explore the promoter region from -333 to +1 which comprises two natural transcription activators binding sites, HNF1-α (from -259 to 251) and C/EBPβ (from -227 to -208). The gRNAs design was made to screen a big portion of the promoter and one gRNA was designed to bind approximately the same region of the transcription factor C/EBPβ, in order to understand if the gRNA binding sequence could result in a synergistic or in competitive effect (Figure 2.9). The selected binding sequences were cloned into a gRNA expressing vector that contains the hU6 promoter and gRNA scaffold, as showed before (Figure 2.4).

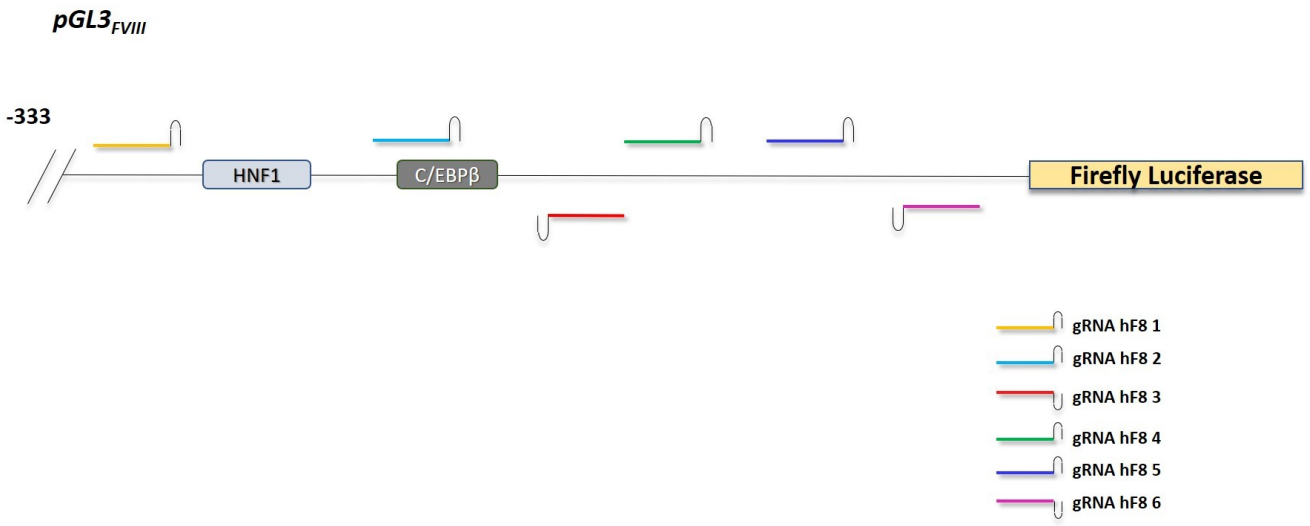

Figure 2.9: Representation of the proximal F8 promoter cloned into pGL3 vector expressing luciferase reporter gene. The gRNAs targeting this promoter were designed approximately from position -333 of the promoter to TSS.

Validation of gRNAs and dCas9-VPR effect on F8 promoter

To understand the *F8* promoter activity we transfected the Huh-7 cell line with the reporter vector expressing the wild type promoter. The expression of the reporter gene under the control of *F8* promoter resulted very low, and the levels were almost two-fold the pGL3 Basic levels, a negative control that did not contain any promoter sequence upstream of the luciferase gene (Figure 2.10 upper part). Noticeably, the combination of the different gRNAs showed an increase only using the gRNAs 1, 2, 4 and 6 (8-fold, 19-fold, 3-fold and 7-fold, respectively). To confirm that the transcription increment was directed by the ribonuclease complex gRNA/dCas9-VPR we transfected the cells only with the dCas9-VPR protein, revealing that without the gRNA binding sequence the transcriptional levels of the reporter gene remained unaltered (Figure 2.10).

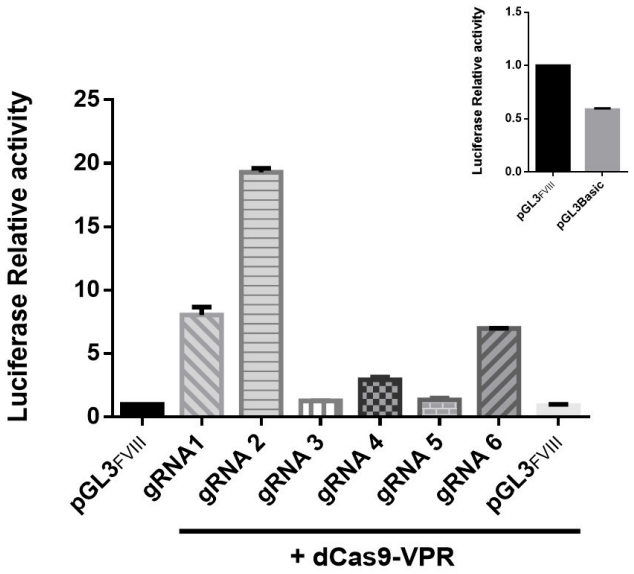

Figure 2.10: Co-expression of the FVIII reporter plasmid with dCas9-VPR and each different gRNA designed in Huh-7 cell line. Histograms reported the relative luciferase activity as change fold relative to the pGL3FVIII alone. Negative control is represented by the pGL3 Basic carrying firefly luciferase without an upstream promoter.

Combination of gRNAs to assess an improvement on transcription

We supposed that expression of multiple guide RNAs in a single cell might enable additive or even synergistic activation of endogenous gene targets as recently described for other genes activated with gRNA/dCas9-VP64 system[115] and with TALE-TFs effectors[116]. For these reasons we decided to explore the combination of the gRNAs that showed the best transactivation activities. Interestingly, the association of two gRNA in the same cells showed a maximization of the reporter gene expression, in particular using the gRNA2, that already demonstrated to be effective alone. The best combinations were gRNAs 1+2 (38.5±0.79) and gRNAs 2+4 (23.6±1.2) supporting the additive effect hypothesis. The remaining combinations showed smaller but not negligible increases (Figure 2.11).

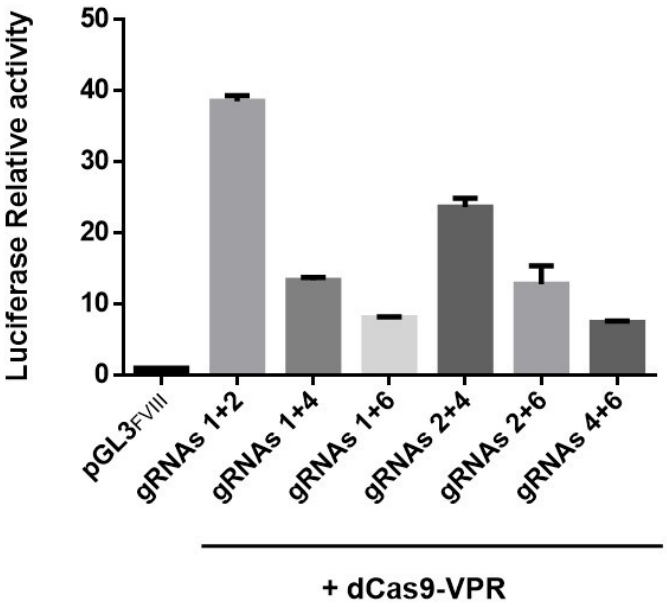

Figure 2.11: The histograms reported the relative luciferase activity detected in Huh-7 cells transfected with the reporter alone and in combination of different gRNAs.

Assessment of the specificity of gRNAs for F8 promoter

To investigate more in detail the specificity of our best candidates (gRNA1, gRNA2, gRNA6 and the combination gRNAs 1+2) we performed an experiment in Huh7 cells using a different promoter: *F7* promoter. First, we co-transfected *F7* luciferase expressing-vector with the different gRNAs and the best combination (gRNAs 1+2) and, in parallel, with the dCas9-VPR without the gRNAs and the negative control represented by pGL3Basic. In all the conditions explored, the residual *F7* promoter activity remained unaltered by the presence of the different gRNAs (Figure 2.12), thus indicating that the gRNAs are not able to increase the expression of a different promoter.

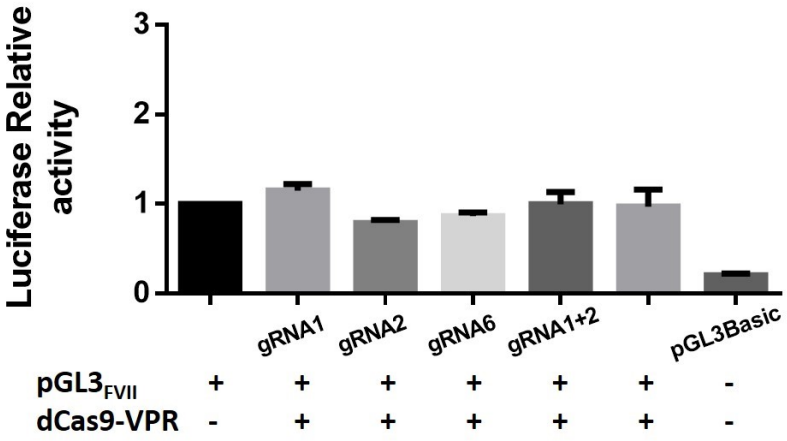

Figure 2.12: Effect of different gRNAs in Huh7 cells with $pGL3_{FVII}$ reporter plasmid., expressing firefly luciferase guided by FVII proximal promoter. The negative control is provided from pGL3 Basic. Histograms of the dual assay report firefly luciferase relative activity as change fold relative to $pGL3_{FVII}$ alone.

2.6 Discussion

FVII deficiency due to promoter mutations represents a good model to investigate the ability of engineered transcription factors (eTFs) to enhance gene transcription for therapeutic purposes. It is also worth noting that in this pathology a modest increase of functional protein levels in plasma significantly ameliorates the bleeding phenotype. In this in-vitro study we used the model of FVII deficiency that we previously created, and we exploited it in hepatic cell lines, due to the impossibility of conducting the experiments in primary hepatocytes derived from patients. We decided to use the FVII deficiency model in order to compare two different transcription activation systems in-vitro. For this study we decided to assess the transcriptional enhancement for the wild type promoter and for a promoter mutation, causative for severe FVII deficiency, the -61 T>G[107]. Taking into account that the mutation involves a binding sequence for a hepatic transcription factor, HNF-4, we confirmed that this mutation leads to a consequent decrease of functionality of the promoter and reduces the transcriptional levels, as confirmed in luciferase reporter assays (Figure 2.5; Figure 2.6). We designed different gRNAs to explore the accessibility of the *F7* promoter, and after we focused on $gRNA_{F7.5}$ because it recognizes a sequence which is similar to the TALE based TF named TF4, that previously demonstrated to be effective for this model. Indeed, in our earlier work we supposed that the position of the target sequence of the TF4, between the Sp1 and the HNF-4 binding sites, could contribute to its effectiveness. Prompted by this data we also explored another system, the CRISPR activation tool, addressed to the same sequence to compare the effects. The CRISPRa system is a new application associated to the Cas9 capability to bind DNA sequences, and for its structure this approach is easier to design and realize, consequently can be advantageous to use. We decided to use a second generation dCas9 associated to the tripartite activator VPR because in earlier works it demonstrated a low efficiency of this system in association with a unique

activator, as VP64[91]. All the gRNAs we designed for *F7* promoter resulted effective at similar levels, demonstrating that the promoter portion from -52 to -22 could be accessible for transcription activation machinery (Figure 2.5).

Nevertheless, the comparison was only possible between $gRNA_{F7.5}$ and TF4 targeting the same sequence within the promoter, which showed a higher increase of luciferase reporter in presence of the $gRNA_{F7.5}$/dCas9-VPR complex (Figure 2.6).

The data obtained from gene reporter assays showed an efficacy of the CRISPRa definitely higher than that of TF4. The ability of the $gRNA_{F7.5}$/dCas9-VPR complex to recruit the transcriptional machinery and to increase the reporter expression both in the presence and in the absence of the disease-causing mutation was also higher. However, it is worth nothing that the reporter genes are episomal systems and do not take into account the physiological context, the chromatin asset. To address this issue, we tested the efficacy on the endogenous *F7* expression treating cells with the $gRNA_{F7.5}$ and the dCas9-VPR protein in comparison with TF4. From these studies we observed that the transactivation increase is not measurable in terms of mRNA copies in HepG2 cells (Figure 2.7) because of the endogenous expression of *F7* mRNA and the low transfection efficiency of the cell line. This issue can be bypassed in further experiments aimed at the selection of only the transfected cells, thus eliminating the background, with fluorescent proteins (GFP) or antibiotic selection (Gentamycin).

However, we decided to explore the efficacy of the two transcription enhancers on the endogenous FVII protein expression of HepG2 cells. The FXa assay analyses reveled a similar situation previously seen with the usage of reporter genes. Indeed, in both cases we have an increment of FVII protein levels with relative increase of activity in comparison with the negative cells (mock). The effect on the endogen FVII protein increase is higher in the cells treated with $gRNA_{F7.5}$/dCas9-VPR than in the cells treated with TF4 (Figure 2.8), as confirmed by the shortening of the lag times.

The two systems have different DNA recognition mechanisms, and this could play a role in the transcriptional enhancement on the chromatin asset. Indeed the TALE protein structure is substantially different from the dCas9-VPR, and the protein-DNA recognition of the TALE-TF is made by weak interactions between amino acids and the major groove of the DNA molecule compared the gRNA-DNA interaction which forms a R-loop that requires an helix unwinding[117]. We can speculate that the interaction of the dCas9 and the DNA can interfere with the activation, demonstrated by the low effect using dCas9-VP64[118], and balanced with the usage of a tripartite activator.

Due to the versatility of the CRISPR activation system and the effortlessness of realization we decided to explore this technique to *F8* promoter. We already tested the potential of dCas9-VPR. However, *F8* is known to possess a weak promoter, which led us to study the entire sequence, by cloning it into a reporter plasmid (Figure 2.9). Subsequently, we designed several gRNAs able to explore a reasonable part of the promoter near the transcription start site. The reporter assays confirmed the weakness of the *F8* promoter in comparison with the expressing vector without a promoter sequence (Figure 2.10) and we were observed that able to appreciate four out of six gRNAs were able to increase the reporter gene expression. Driven by the data obtained by other groups[101,115] we tested our model with a pool of two gRNAs. Interestingly, in cells transfected with a combination of two gRNAs the reporter gene levels increase in an additive manner, and we reported that the couple of gRNAs 1+2 were able to achieve an about 40-fold increment (Figure 2.11).

An important issue that has to be considered for the correction approach with engineered transcription activators is the specificity, since there is a potential risk arising from non-specific interactions that would lead to the alteration of off-target gene expression. One way to improve the specificity is to design the gRNAs of 20 bp that, in silico analyses, predicted to be unique in the genome[119,120]. We focused in particular on the potential off-target effects that could be induced by the treatment with the the best candidates to increase the reporter levels: gRNAs 1, 2

and 6 and the gRNAs combination 1+2. The preliminary analysis was performed by co-transfecting of the gRNAs with the $pGL3_{F7wt}$ vector, carrying the firefly luciferase under the control of *F7* promoter. This experiment showed no increase in the reporter activity for all the gRNAs and for the gRNAs combination (Figure 2.12) excluding potential off-target for this reporter model. All these data require further confirmation on the endogenous context and off target analysis on the potential predicted sites.

Overall, these data suggest that the application of CRISPRa should be able to increase, in a highest way comparing with the TALE system, the FVII expression in cell models. It has also to be considered that the enhancement is referred to the wild type promoter, suggesting that a higher improvement could be achieved in the presence of promoter mutations. Moreover, the dCas9-VPR demonstrated to be more active in the presence of more than one gRNAs, as we observed for *F8* promoter, which could open a new chapter in exploring the other four gRNAs that we did not study in deep for this work. In future experiments, we are going to evaluate protein increase though ELISA quantification in cell lines treated with CRISPRa and TF4 to confirm the expression levels. Testing the CRISPRa system in primary hepatocytes would be a future approach in order to provide the rationale for subsequent experiments in vivo. While the $gRNA_{F7.5}$/dCas9-VPR seems to be the best tool to increase FVII protein expression in a cellular model, in vivo experiments are necessary to understand which of the two systems can be more efficiently customizable for the treatment of different patients.

Conclusions

The experimental activity on disease models such as Haemophilia B and FVII deficiency enabled us to explore two different approaches for the rescue of protein secretion and mRNA transcription in the presence of different mutations types. For the first time we provided experimental evidence that a chemical chaperone, Na-PBA, is able to restore protein levels of a coagulation serine-protease and maintain its activity. This approach might suggest a therapeutic application for diseases caused by missense mutations impairing protein folding, that are the most represented in all human diseases. It is encouraging that chemical chaperones were explored for different diseases and it is worth nothing that Na-PBA, our best candidate, it is approved by FDA as a clinical drug for urea's cycle disorders. We also compared two different transcription activator systems in a FVII deficiency model, and confirmed that engineered TFs are able to rescue the FVII protein expression. For the first time we applied the CRISPR activation system to a coagulation disorder and we achieve an increase of endogenous FVII protein determined by activity assays, and the enhancement with this new system is highest compared to TALE-TF treatment. We also explored *F8* promoter and we were able to increase its activity in episomal context. It is worth noting that this could be a good model exploitable to enhance the efficacy of other well-established correction approaches acting at the post-transcriptional levels (splicing modulation, ribosome read-through, etc.) by increasing the amount of the target mRNA.

Bibliography

1. Davie, E. W. A brief historical review of the waterfall/cascade of blood coagulation. *J. Biol. Chem.* **278,** 50819–32 (2003).

2. MACFARLANE, R. G. An Enzyme Cascade in the Blood Clotting Mechanism, and its Function as a Biochemical Amplifier. *Nature* **202,** 498–499 (1964).

3. Gailani, D. & Broze, G. J. Factor XI activation in a revised model of blood coagulation. *Science (80-.).* **253,** 909 LP-912 (1991).

4. Müller, F. & Renné, T. Novel roles for factor XII-driven plasma contact activation system. *Curr. Opin. Hematol.* **15,** (2008).

5. Osterud, B. & Rapaport, S. I. Activation of factor IX by the reaction product of tissue factor and factor VII: additional pathway for initiating blood coagulation. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5260–5264 (1977).

6. Kleinschnitz, C. *et al.* Targeting coagulation factor XII provides protection from pathological thrombosis in cerebral ischemia without interfering with hemostasis. *J. Exp. Med.* **203,** 513–518 (2006).

7. Hoffman, M. Remodeling the Blood Coagulation Cascade. *J. Thromb. Thrombolysis* **16,** 17–20 (2003).

8. Cooper, D. N., Millar, D. S., Wacey, A., Banner, D. W. & Tuddenham, E. G. Inherited factor VII deficiency: molecular genetics and pathophysiology. *Thromb. Haemost.* **78,** 151 (1997).

9. Wion, K. L., Kelly, D., Summerfield, J. A., Tuddenham, E. G. D. & Lawn, R. M. Distribution of factor VIII mRNA and antigen in human liver and other tissues. *Nature* **317,** 726–729 (1985).

10. Morrissey, J. H., Macik, B. G. & Neuenschwander, P. F. Quantitation of activated factor VII levels in plasma using a tissue factor mutant selectively deficient in promoting factor VII activation. *Blood* **81,** 734–744 (1993).

11. Wildgoose, P. & Kisiel, W. Activation of human factor VII by factors IXa and Xa on human bladder carcinoma cells. *Blood* **73,** 1888–1895 (1989).

12. Kirchhofer, D. *et al.* The tissue factor region that interacts with factor Xa in the activation of factor VII. *Biochemistry* **40,** 675–682 (2001).

13. Furie, B. & Furie, B. C. The molecular basis of blood coagulation. *Cell* **53,** 505–518 (1988).

14. Roberts, H. R., Hoffman, M. & Monroe, D. M. A cell-based model of thrombin generation. in *Seminars in thrombosis and hemostasis* **32,** 32–38 (Copyright© 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA., 2006).

15. Alexander, B. *et al.* Congenital SPCA deficiency: a hitherto unrecognized coagulation defect with hemorrhage rectified by serum and serum fractions. *J. Clin. Invest.* **30,** 596 (1951).

16. Mariani, G., Lo, C. L., Bernardi, F. & Pinotti, M. Molecular and clinical aspects of factor VII deficiency. *Blood Coagul. fibrinolysis an Int. J. Haemost. Thromb.* **9,** S83-8 (1998).

17. Rosen, E. D., Chan, J. C. Y., Idusogie, E. & Clotman, F. Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature* **390,** 290 (1997).

18. Triplett, D. A., Brandt, J. T., Batard, M. A., Dixon, J. L. & Fair, D. S. Hereditary factor VII deficiency: heterogeneity defined by combined functional and immunochemical analysis. *Blood* **66,** 1284–1287 (1985).

19. Cancio, M. I., Reiss, U. M., Nathwani, A. C., Davidoff, A. M. & Gray, J. T. Developments in the treatment of hemophilia B: focus on emerging gene

therapy. *Appl. Clin. Genet.* **6,** 91 (2013).

20. Franchini, M. & Mannucci, P. M. The history of hemophilia. in *Seminars in thrombosis and hemostasis* **40,** 571–576 (Thieme Medical Publishers, 2014).

21. Mannucci, P. M. & Tuddenham, E. G. D. The hemophilias—from royal genes to gene therapy. *N. Engl. J. Med.* **344,** 1773–1779 (2001).

22. Srivastava, A. *et al.* Guidelines for the management of hemophilia. *Haemophilia* **19,** (2013).

23. World Federation of Hemophilia. *ANNUAL GLOBAL SURVEY 2014*. (2014).

24. Chalmers, A. Hemophilia A and B: Disease Differences and the Use of Prophylactic Therapy.

25. Biggs, R. *et al.* Christmas disease. *Br. Med. J.* **2,** 1378 (1952).

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

27. Lee, C. A. *et al.* The molecular basis of haemophilia B. *Haemophilia* **4,** 350–357 (1998).

28. Giannelli, F. *et al.* Haemophilia B: database of point mutations and short additions and deletions—eighth edition. *Nucleic Acids Res.* **26,** 265–268 (1998).

29. Antonarakis, S. E. *et al.* Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* **86,** 2206–2212 (1995).

30. Mariani, G. & Bernardi, F. Factor VII deficiency. in *Seminars in thrombosis and hemostasis* **35,** 400–406 (© Thieme Medical Publishers, 2009).

31. Mariani, G. *et al.* Clinical manifestations, management, and molecular genetics in congenital factor VII deficiency: the International Registry on Congenital Factor VII Deficiency (IRF7). *Blood* **96,** 374 (2000).

32. Mariani, G. *et al.* Clinical phenotypes and factor VII genotype in congenital

factor VII deficiency. *Thromb. HAEMOSTASIS-STUTTGART-* **93,** 481 (2005).

33. Møss, J., Rosholm, A. & Lauren, A. Safety and pharmacokinetics of a glycoPEGylated recombinant activated factor VII derivative: a randomized first human dose trial in healthy subjects. *J. Thromb. Haemost.* **9,** 1368–1374 (2011).

34. Carr, M. E. & Tortella, B. J. Emerging and future therapies for hemophilia. *J. Blood Med.* **6,** 245 (2015).

35. High, K. A. The gene therapy journey for hemophilia: are we there yet? *ASH Educ. Progr. B.* **2012,** 375–381 (2012).

36. Agersø, H. *et al.* Pharmacokinetics of an anti-TFPI monoclonal antibody (concizumab) blocking the TFPI interaction with the active site of FXa in Cynomolgus monkeys after iv and sc administration. *Eur. J. Pharm. Sci.* **56,** 65–69 (2014).

37. Chowdary, P. *et al.* Safety and pharmacokinetics of anti-TFPI antibody (concizumab) in healthy volunteers and patients with hemophilia: a randomized first human dose trial. *J. Thromb. Haemost.* **13,** 743–754 (2015).

38. Gu, J. M. *et al.* Pharmacodynamics and pharmacokinetics of TFPI-neutralizing antibody (BAY1093884) in cynomolgus monkeys and prediction of human dose. *J. Thromb. Haemost.* **13,** 7 (2015).

39. Sehgal, A. *et al.* An RNAi therapeutic targeting antithrombin to rebalance the coagulation system and promote hemostasis in hemophilia. *Nat. Med.* **21,** 492–497 (2015).

40. Pasi, K. J. *et al.* A subcutaneously administered investigational RNAi therapeutic (ALN-AT3) targeting antithrombin for treatment of hemophilia: interim weekly and monthly dosing results in patients with hemophilia A or B. (2015).

41. 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).

42. Balestra, D. *et al.* Regulation of a strong F9 cryptic 5′ ss by intrinsic elements and by combination of tailored U1snRNAs with antisense oligonucleotides. *Hum. Mol. Genet.* **24,** 4809–4816 (2015).

43. 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).

44. Lee, H.-L. R. & Dougherty, J. P. Pharmaceutical therapies to recode nonsense mutations in inherited diseases. *Pharmacol. Ther.* **136,** 227–266 (2012).

45. Branchini, A. *et al.* Speci fi c factor IX mRNA and protein features favor drug-induced readthrough over recurrent nonsense mutations. **129,** 2303–2308 (2017).

46. Lee, H. J., Kweon, J., Kim, E., Kim, S. & Kim, J.-S. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* **22,** 539–548 (2012).

47. Park, C.-Y. *et al.* Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs. *Proc. Natl. Acad. Sci.* **111,** 9253–9258 (2014).

48. Zhou, Y. *et al.* Enhanced genome editing in mammalian cells with a modified dual-fluorescent surrogate system. *Cell. Mol. Life Sci.* **73,** 1–21 (2016).

49. Guan, Y. *et al.* CRISPR/Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. *EMBO Mol. Med.* **8,** 477–488 (2016).

50. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. Complete nucleotide sequences of the gene for human factor IX (antihemophilic factor B). *Biochemistry* **24,** 3736–3750 (1985).

51. Freedman, S. J., Furie, B. C., Furie, B. & Baleja, J. D. Structure of the Calcium Ion-Bound. gamma.-Carboxyglutamic Acid-Rich Domain of Factor IX. *Biochemistry* **34,** 12126–12137 (1995).

52. Schmidt, A. E. & Bajaj, S. P. Structure-function relationships in factor IX and factor IXa. *Trends Cardiovasc. Med.* **13,** 39–45 (2003).

53. Amphlett, G. W., Kisiel, W. & Castellino, F. J. The interaction of Ca2+ with human Factor IX. *Arch. Biochem. Biophys.* **208,** 576–585 (1981).

54. Bajaj, S. P. Cooperative Ca2+ binding to human factor IX. Effects of Ca2+ on the kinetic parameters of the activation of factor IX by factor XIa. *J. Biol. Chem.* **257,** 4127–4132 (1982).

55. Bajaj, S. P., Rapaport, S. I. & Russell, W. A. Redetermination of the rate-limiting step in the activation of factor IX by factor XIa and by factor VIIa/tissue factor. Explanation for different electrophoretic radioactivity profiles obtained on activation of 3H-and 125I-labeled factor IX. *Biochemistry* **22,** 4047–4053 (1983).

56. Lawson, J. H. & Mann, K. G. Cooperative activation of human factor IX by the human extrinsic pathway of blood coagulation. *J. Biol. Chem.* **266,** 11317–11327 (1991).

57. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).

58. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4,** 181 (2003).

59. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science (80-.).* **319,** 916–919 (2008).

60. Muntau, A. C., Leandro, J., Staudigl, M., Mayer, F. & Gersting, S. W. Innovative strategies to treat protein misfolding in inborn errors of

metabolism: pharmacological chaperones and proteostasis regulators. *J. Inherit. Metab. Dis.* **37,** 505–523 (2014).

61. Ellis, R. J. & Hemmingsen, S. M. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem. Sci.* **14,** 339–342 (1989).

62. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381,** 571 (1996).

63. Frydman, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* **70,** 603–647 (2001).

64. Ellgaard, L. & Helenius, A. ER quality control : towards an understanding at the molecular level. 431–437 (2001).

65. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295,** 1852–8 (2002).

66. Hartl, F. U. & Hayer-hartl, M. Converging concepts of protein folding in vitro and in vivo. **16,** 574–581 (2009).

67. Sitia, R. & Braakman, I. Quality control in the endoplasmic reticulum protein factory. *Nature* **426,** 891 (2003).

68. Arakawa, T., Ejima, D., Kita, Y. & Tsumoto, K. Small molecule pharmacological chaperones : From thermodynamic stabilization to pharmaceutical drugs. **1764,** 1677–1687 (2006).

69. Ringe, D. & Petsko, G. A. Q&A: What are pharmacological chaperones and why are they interesting? *J. Biol.* **8,** 80 (2009).

70. Cohen, F. E. & Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. *Nature* **426,** 905 (2003).

71. Leandro, P. & Gomes, C. M. Protein misfolding in conformational disorders: rescue of folding defects and chemical chaperoning. *Mini Rev. Med. Chem.* **8,** 901–911 (2008).

72. Parenti, G., Andria, G. & Valenzano, K. J. Pharmacological Chaperone Therapy : Preclinical Development , Clinical Translation , and Prospects for the Treatment of Lysosomal Storage Disorders. **23,** 1138–1148 (2015).

73. Factor IX Variant Database. Available at: www.factorix.org.

74. Rozenfeld, P. & M Neumann, P. Treatment of fabry disease: current and emerging strategies. *Curr. Pharm. Biotechnol.* **12,** 916–922 (2011).

75. Chollet, M. E., Skarpen, E., Iversen, N., Sandset, P. M. & Skretting, G. The chemical chaperone sodium 4-phenylbutyrate improves the secretion of the protein CA267T mutant in CHO-K1 cells trough the GRASP55 pathway. *Cell Biosci.* **5,** 57 (2015).

76. Roth, S. D. *et al.* Chemical chaperones improve protein secretion and rescue mutant factor VIII in mice with hemophilia A. *PLoS One* **7,** e44505 (2012).

77. Engin, F. & Hotamisligil, G. S. Restoring endoplasmic reticulum function by chemical chaperones: An emerging therapeutic approach for metabolic diseases. *Diabetes, Obes. Metab.* **12,** 108–115 (2010).

78. Lim, M., McKenzie, K., Floyd, A. D., Kwon, E. & Zeitlin, P. L. Modulation of ΔF508 cystic fibrosis transmembrane regulator trafficking and function with 4-phenylbutyrate and flavonoids. *Am. J. Respir. Cell Mol. Biol.* **31,** 351–357 (2004).

79. Burrows, J. A. J., Willis, L. K. & Perlmutter, D. H. Chemical chaperones mediate increased secretion of mutant α1-antitrypsin (α1-AT) Z: a potential pharmacological strategy for prevention of liver injury and emphysema in α1-AT deficiency. *Proc. Natl. Acad. Sci.* **97,** 1796–1801 (2000).

80. Montane, J. *et al.* Amyloid-induced β-cell dysfunction and islet inflammation are ameliorated by 4-phenylbutyrate (PBA) treatment. *FASEB J.* fj-201700236R (2017).

81. Liu, X. L. *et al.* Defective trafficking of nephrin missense mutants rescued by

a chemical chaperone. *J. Am. Soc. Nephrol.* **15,** 1731–1738 (2004).

82. Li, S. *et al.* Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense mutations in non-active sites. *J. Biol. Chem.* **289,** 18943–18956 (2014).

83. M Gomes, C. Protein misfolding in disease and small molecule therapies. *Curr. Top. Med. Chem.* **12,** 2460–2469 (2012).

84. Boyd, R. E. *et al.* Pharmacological chaperones as therapeutics for lysosomal storage diseases. *J. Med. Chem.* **56,** 2705–2725 (2013).

85. Bendikov-Bar, I., Maor, G., Filocamo, M. & Horowitz, M. Ambroxol as a pharmacological chaperone for mutant glucocerebrosidase. *Blood Cells, Mol. Dis.* **50,** 141–145 (2013).

86. Porto, C. *et al.* Pharmacological enhancement of α-glucosidase by the allosteric chaperone N-acetylcysteine. *Mol. Ther.* **20,** 2201–11 (2012).

87. Ren, H. Y. *et al.* VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol. Biol. Cell* **24,** 3016–3024 (2013).

88. Milla, C. E. *et al.* Lumacaftor/Ivacaftor in Patients Aged 6–11 Years with Cystic Fibrosis and Homozygous for F508del-CFTR. *Am. J. Respir. Crit. Care Med.* **195,** 912–920 (2017).

89. Simioni, P. *et al.* X-linked thrombophilia with a mutant factor IX (factor IX Padua). *N. Engl. J. Med.* **361,** 1671–1675 (2009).

90. Batshaw, M. L., Tuchman, M., Summar, M., Seminara, J. & Consortium, U. C. D. A longitudinal study of urea cycle disorders. *Mol. Genet. Metab.* **113,** 127–130 (2014).

91. La Russa, M. F. & Qi, L. S. The new state of the art: Cas9 for gene activation and repression. *Mol. Cell. Biol.* **35,** 3800–3809 (2015).

92. Tupler, R., Perini, G. & Green, M. R. Expressing the human genome. *Nature*

409, 832–833 (2001).

93. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **4,** 1609 (1985).

94. Berg, J. M. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci.* **85,** 99–102 (1988).

95. Lee, M. S., Gippert, G. P., Soman, K. V, Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science (80-.).* **245,** 635–637 (1989).

96. Jamieson, A. C., Miller, J. C. & Pabo, C. O. Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug Discov.* **2,** 361–368 (2003).

97. Zhang, M. *et al.* TALE: a tale of genome editing. *Prog. Biophys. Mol. Biol.* **114,** 25–32 (2014).

98. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

99. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).

100. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31,** 833–838 (2013).

101. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).

102. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160,** 339–350 (2015).

103. Konermann, S. *et al.* Genome-scale transcriptional activation by an

engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

104. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12,** 326–328 (2015).

105. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** 510–517 (2015).

106. Chapdelaine, P. *et al.* Development of an AAV9 coding for a 3XFLAG-TALEfrat# 8-VP64 able to increase in vivo the human frataxin in YG8R mice. *Gene Ther.* **23,** 606–614 (2016).

107. Arbini, Pollak, Bayeran, High, Kenneth, B. Severe factor VII deficiency due to a mutation disrupting Nuclear Factor 4 Binding Site in the factor VII promoter. *Blood* **92,** 1639–1645 (1997).

108. Carew, J. a, Pollak, E. S., High, K. a & Bauer, K. a. Severe factor VII deficiency due to a mutation disrupting an Sp1 binding site in the factor VII promoter. *Blood* **92,** 1639–1645 (1998).

109. Doyle, E. L. *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* **40,** W117–W122 (2012).

110. Branchini, A. *et al.* Natural and engineered carboxy-terminal variants: decreased secretion and gain-of-function result in asymptomatic coagulation factor VII deficiency. *Haematologica* **97,** 705–709 (2012).

111. Benchling. Available at: https://benchling.com.

112. Barbon, E. *et al.* An engineered tale-transcription factor rescues transcription of factor VII impaired by promoter mutations and enhances its endogenous expression in hepatocytes. *Sci. Rep.* **6,** 28304 (2016).

113. Pollak, E. S., Hung, H.-L., Godin, W., Overton, G. C. & High, K. A. Functional characterization of the human factor VII 5′-flanking region. *J.*

Biol. Chem. **271,** 1738–1747 (1996).

114. Nougier, C. *et al.* Characterization of four novel molecular changes in the promoter region of the factor VIII gene. *Haemophilia* **20,** 149–156 (2014).

115. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).

116. Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* **10,** 239–242 (2013).

117. Gao, X. *et al.* Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res.* **42,** (2014).

118. La Russa, M. F., Qi, L. S., Russa, M. F. La & Qi, S. The New State of the Art : Cas9 for Gene Activation and Repression. *Mol. Cell. Biol.* **35,** 3800–3809 (2015).

119. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184 (2016).

120. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827 (2013).