

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Christopher William Coyle

Date

**Ancestral Sequence Reconstruction as a Lead Optimization Approach in Gene Therapy
Drug Development**

By:

Christopher William Coyle

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Molecular and Systems Pharmacology

Christopher B. Doering, Ph.D.
Advisor

H. Trent Spencer, Ph.D.
Committee Member

Renhao Li, Ph.D.
Committee Member

Baek Kim, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

Date

Ancestral Sequence Reconstruction as a Lead Optimization Approach in Gene Therapy Drug
Development

By

Christopher William Coyle

B.S., Loyola Marymount University 2016

Advisor: Christopher B. Doering, Ph.D.

An abstract of a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Science,
Molecular and Systems Pharmacology
2023

Abstract

Ancestral Sequence Reconstruction as a Lead Optimization Approach in Gene Therapy Drug Development

By Christopher William Coyle

Gene therapy is revolutionizing 21st century medicine. Currently, 27 gene therapy products are approved by the U.S. Food and Drug Administration and/or the European Medicines Agency. These initial products are predominantly focused on two disease classes, monogenic disorders and cancer. Within monogenic disorders, approved products are now available for the severe bleeding disorders, hemophilia A and B. Historically, these diseases have been treated using lifelong protein replacement therapy, which is difficult to manage from both compliance and economic standpoints. Gene therapy has the potential to change the paradigm to a one-time treatment that is effective for years instead of days. However, as a new drug class, 1st generation gene therapies are hindered by critical limitations in manufacturing efficiency, optimal and durable efficacy, and dose-related toxicities. Unlike small molecule drug optimization, relatively few methods and technologies exist to optimize gene therapies. Common methods employed include nucleic acid 'codon' optimization, vector selection and tropism engineering, and transgene product (*i.e.*, therapeutic protein) selection. The studies completed for the current dissertation were designed to test and further develop a unique strategy for gene therapy potency optimization through transgene product engineering. Improving gene therapy potency has the potential to address all three primary limitations currently hindering gene therapies. To accomplish this goal, we adopted a method historically used to study molecular evolution and applied it as a drug discovery platform.

Ancestral sequence reconstruction (ASR) utilizes bioinformatics to predict the protein sequences of ancient, extinct species, which have evolved over time to meet the ecological, physiological, and sociological demands of each species. While this approach has traditionally been applied to studies of protein and gene evolution, it also can be harnessed to facilitate the exploration and mapping of functional amino acid substitutions that confer pharmaceutically-inspired properties to protein drugs and gene therapies under development. In the current work, we present the results of ASR studies designed to identify enhanced variants of blood coagulation factors VIII (FVIII) and IX (FIX), which are associated with hemophilia A and B, respectively. As a validation of the ASR drug discovery approach, we identified ancestral variants of FVIII and FIX that display enhanced pharmaceutical properties such as specific activity, half-life and biosynthetic efficiency. Subsequently, reductionist screening strategies were successfully employed to map the amino acids responsible for the functional enhancements. In the case of FIX, an ancestral variant that displayed 10-fold higher specific activity than human FIX was identified during the initial ASR screen. Subsequently, the amino acid substitutions necessary and sufficient to confer 10-fold higher activity to human FIX were mapped and functionally characterized. Finally, by combining these 5 critical amino acid substitutions with an additional single amino acid substitution FIX variant identified in a family with elevated FIX activity levels, a final lead FIX candidate was identified and designated ET9. ET9 represents a 99% human FIX variant containing only six amino acid substitutions that confer 51-fold increased specific activity in clinical coagulation assays and 10-fold higher potency *in vivo* in hemophilia B mice compared to human FIX. ET9 is, to our knowledge, the most potent FIX variant described to date, and a premiere candidate for utilization in 2nd generation gene therapies for hemophilia B. These results further validate the ASR approach and justify its implementation in broader protein drug and gene therapy development as an effective discovery to lead optimization platform.

**Ancestral Sequence Reconstruction as a Lead Optimization Approach in Gene Therapy
Drug Development**

By

Christopher William Coyle

B.S., Loyola Marymount University 2016

Advisor: Christopher B. Doering, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Science,
Molecular and Systems Pharmacology
2023

Acknowledgments

First and foremost, I would like to thank my mentor, Dr. Chris Doering for all the support and guidance he has given me throughout my graduate training. His mentorship has given me invaluable skills that will serve me well in my future career and I will forever be grateful. I would not be the scientist I am today without having had the privilege to work under him. Dr. Trent Spencer, who served as a second mentor and a member of my dissertation committee was always available to teach me and offer guidance. Thank you to my two other committee members, Drs. Renhao Li and Baek Kim both of whom provided valuable assistance throughout the years with a complex and difficult project that involved a major change of direction late in my graduate career.

Many current and former lab members and colleagues throughout the years have assisted me greatly during this journey. Specifically, Gianna Branella, Kristopher Knight, Jordan Alexander, and Jenny Okalova from the Gene Therapy Lab, for training and assistance with critical experiments and moral support along the way. Additionally, Ernie Parker, who was always happy to help and assist, was monumentally helpful in training me in numerous challenging techniques, including protein purification. To two of my closest friends, Jamie Story and Austin Lee, thank you for your moral support and advice during these challenging years. Thank you to my parents for instilling the value of education from an early age and for their continuous love and support throughout all levels of my education.

Lastly, and most importantly, I express my deepest gratitude to my wife, Suzanne Coyle, who has stood by my side through the most challenging portion of my graduate training. Her unwavering support and the sacrifices she made helped me make it to the finish line and I could not have done it without her.

Table of Contents

Chapter 1: Introduction	1
1.1: Blood Coagulation Network.....	2
1.2 Hemophilia A and B.....	4
1.3 FVIII.....	5
1.4 FIX.....	8
1.5 Brief History of Gene Therapy.....	9
1.6 Pharmacology and Optimization Strategies of AAV Gene Therapy.....	11
1.7 Gene Therapy of hemophilia A and B using AAV vectors.....	17
1.8 Ancestral Sequence Reconstruction as a Platform for Protein Engineering.....	22
1.9 Hypothesis and objectives of this research.....	24
Chapter 2: Molecular coevolution of coagulation factor VIII and von Willebrand factor	25
2.1: Abstract.....	26
2.2: Introduction.....	26
2.3: Materials and Methods.....	27
2.4: Results.....	33
2.5: Discussion.....	45
2.6: Acknowledgements.....	48
2.7: Supplementary Figures and Tables.....	49
Chapter 3: Expression of Ancient Coagulation FVIII Molecules and Discovery of Relevant Domains	58
3.1: Abstract.....	59
3.2 Introduction.....	60
3.3 Materials and Methods.....	61
3.4 Results.....	61
3.5 Discussion.....	73
3.6 Acknowledgements.....	75
Chapter 4: Identification of coagulation factor IX variants with enhanced activity through ancestral sequence reconstruction	77
4.1: Abstract.....	78
4.2: Introduction.....	78

4.3: Materials and Methods.....	80
4.4: Results.....	86
4.5: Discussion.....	97
4.6: Acknowledgements.....	101
4.7: Supplementary Figures and Tables.....	103
Chapter 5: Humanization and functional characterization of enhanced coagulation factor IX variants identified through ancestral sequence reconstruction.....	108
5.1: Abstract.....	109
5.2: Introduction.....	110
5.3: Materials and Methods.....	112
5.4: Results.....	115
5.5: Discussion.....	129
5.6: Acknowledgements.....	133
5.7: Supplementary Figures and Tables.....	134
Chapter 6: General Discussion.....	141
6.1 Summary of Results.....	142
6.2 Implications of Findings.....	145
6.3 Limitations and Future Directions.....	147
6.4 Conclusion.....	149
References.....	150

List of Figures and Tables:

Figure 1.1: Representation of the coagulation network and the effects of hemophilia on it.....	17
Figure 1.2: Late processing and activation of FVIII.....	21
Figure 1.3: Schematic Representation of mature FIX domain structure and post-translational modifications.....	22
Figure 1.4: AAV transduction pathway.....	28
Figure 1.5: Ancestral sequence reconstruction strategy.....	38
Table 1.1: Current landscape of Gene Therapy Trials for Hemophilia B.....	33
Table 1.2: Current landscape of Gene Therapy Trials for Hemophilia A.....	36
Figure 2.1: AnVWF biosynthetic efficiency correlates with ancestral FVIII.....	48
Figure 2.2: AnVWF shows conserved multimerization and human ADAMTS13 proteolysis.....	51
Figure 2.3: AnVWF specific-activity is reciprocal to ancestral FVIII.....	53
Figure 2.4: Mammalian VWF and FVIII evolved under analogous selective pressure.....	54
Figure 2.5: AnVWF rescues murine FVIII and exhibits reduced clearance in vivo.....	58
Table 2.1: Ancestral VWF sequence identity and classification.....	49
Table 2.2: SPR steady-state K_D determinations (nM).....	56
Supplementary Figure S2.1.....	51
Supplementary Figure S2.2.....	53
Supplementary Figure S2.3.....	53
Supplementary Figure S2.4.....	54
Supplementary Figure S2.5.....	55
Supplementary Figure S2.6.....	56
Supplementary Figure S2.7.....	57
Supplementary Table S1.....	49
Supplementary Table S2.....	50
Figure 3.1: Phylogenetic tree of coagulation Factor VIII.....	62

Figure 3.2: Initial Screening of An-FVIII molecules.....	63
Figure 3.3 Expression of top AnFVIII candidates from Huh7 cells.....	64
Figure 3.4 Schematics and Expression of FVIII-An97 and FVIII-An102 hybrids.....	65
Figure 3.5 Refinement of FVIII-An97 and FVIII-An102 hybrids at single amino acid resolution.....	66
Figure 3.6 Expression of FVIII-An63 and HSQ hybrids.....	67
Figure 3.7 Expression of Initial FVIII-An84 and HSQ hybrids.....	68
Figure 3.8: Expression of FVIII-An84 A2 sub-domain hybrids.....	69
Figure 3.9 Negative single amino acid substitution screen of 84H16.....	70
Figure 3.10 Expression of FVIII-An84 A3 sub-domain hybrids.....	71
Figure 3.11: Expression of FVIII-An84 C2 sub-domain hybrids.....	82
Figure 3.12 Expression of Full Domain FVIII-An70 and HSQ hybrids.....	73
Table 3.1: Selected AnFVIII sequence identity and amino substitutions to hFVIII.....	62
Figure 4.1: AnFIX phylogeny and sequences.....	87
Figure 4.2: Recombinant AnFIX expression and characterization.....	88
Figure 4.3: Thrombin generation analysis.....	91
Figure 4.4: AAV-AnFIX gene therapy.....	93
Table 4.1: Purification of recombinant An96-Padua.....	91
Table 4.2: AAV Vector Quality Attributes.....	94
Table 4.3: AAV Vector Molar Mass and Composition.....	95
Supplementary Figure S4.1: ELISA detection of hFIX and An96-Padua.....	104
Supplementary Figure S4.2: Recombinant FIX activation by FXIa.....	104
Supplementary Figure S4.3: Thrombin generation analysis.....	105
Supplementary Figure S4.4: ED₅₀ determinations for hFIX, An96-Padua and hFIX-Padua.....	106
Supplementary Figure S4.5: TEG of whole blood obtained from control and AAV2/8-An96-Padua treated mice.....	107
Supplementary Table S4.1: AnFIX sequence variation.....	103
Supplementary Table S4.2. hFIX-Padua purification table.....	103
Supplementary Table S4.3: An96 purification table.....	103

Figure 5.1: Ancestral / human FIX hybrid schematics and expression.....	117
Figure 5.2: Negative single amino acid substitution screen of FIX-201.....	119
Figure 5.3: Humanized FIX-An96 constructs ± Padua.....	120
Figure 5.4: OSA activity curves and thrombin generation assays for recombinant FIX variants ± Padua.....	122
Figure 5.5: Tenase activity and FVIIIa binding among FIXa variants.....	124
Figure 5.6: Model of the FXase complex with ET9.....	127
Figure 5.7: Comparison of AAV2/8 FIX transgenes <i>in vivo</i>.....	129
Table 5.1: Michaelis-Menten kinetic parameters of hyperactive variants in the purified tenase assay.....	125
Supplementary Figure S5.1: FIX ELISA.....	134
Supplementary Figure S5.2: Individual Thrombograms of FIX Variants.....	135
Supplementary Figure S5.3: Amidolytic activity of recombinant FIX variants.....	136
Supplementary Figure S5.4: ED₅₀ Determination for hFIX-Padua and ET9.....	137
Supplementary Table S5.1: FIX-201 single amino acid substitutions.....	138
Supplementary Table S5.2: Minimally humanized FIX-An96 variants.....	138
Supplementary Table S5.3: Specific activity of FIX variants.....	139
Supplementary Table S4: Mutagenic Primers used to develop each single amino acid variant from Supplementary Table S5.1.....	139
Figure 6.1: Properties of AAV3-FIX-Padua.....	149

List of Abbreviations

AAP.....	assembly-activating proteins
AAV.....	adeno-associated virus
ADA.....	adenosine deaminase deficiency
AIM.....	autoinhibitory modules
An.....	ancestor/ancestral
ap.....	activation peptide
APC.....	activated protein C
aPTT.....	activated partial thromboplastin time
ASGPR.....	asialoglycoprotein receptor
ASR.....	ancestral sequence reconstruction
BDD.....	B-domain-deleted
bp.....	base pairs
co.....	codon optimized
DLS.....	dynamic light scattering
DTT.....	dithiothreitol
EF1 α	elongation-factor 1 α
EGF.....	epidermal growth factor like
ELISA.....	enzyme-linked immunosorbent assay
ER.....	endoplasmic reticulum
ETP.....	endogenous thrombin potential
FACT.....	factor assay control plasma
FIX/FIXa.....	(activated) Factor IX
FV/FVa.....	(activated) Factor V
FVII/FVIIa.....	(activated) factor VII
FVIII/FVIIIa.....	(activated) Factor VIII
FX/FXa.....	(activated) factor X
FXI/FXIa.....	(activated) factor XI
FXII/FXIIa.....	(activated) factor XII

GGCX..... gamma-glutamyl carboxylase
GP.....glycoprotein
HCC..... hepatocellular carcinoma
HMW..... high molecular weight
HSC..... hematopoietic stem cells
HSPG..... heparan-sulfate proteoglycans
HSQ.....BDD human FVIII
ITR.....inverted terminal repeat
IU.....international unit
LCO.....liver codon optimized/optimization
LMAN1.....lectin mannose binding 1
LRP1.....lipoprotein receptor-related protein-1
MAAP.....membrane associated assembly-activating proteins
MCFD2.....multiple coagulation factor deficiency 2
mFVIII.....murine FVIII
MSA.....multiple sequence alignment
MVM.....minute virus of mice
mya.....million years ago
nAb.....neutralizing antibodies
ns.....not significant
nt.....nucleotide
OSA.....one stage coagulation assay
OTC.....ornithine transcarbamylase
PAGE..... polyacrylamide gel electrophoresis
(q)PCR.....(quantative) polymerase chain reaction
PES.....polyethersulfone
PLwH.....persons living with hemophilia
pro.....protease
RU.....resonance units
sc.....self-complementary

SCID..... severe combined immunodeficiency
sd.....standard deviation
SDS..... sodium dodecyl sulfate
SEC-MALS..... size-exclusion chromatography/multiangle light scattering
sp.....signal peptide
TEAE..... treatment-emergent adverse events
TEG.....thromboelastography
TF.....Tissue Factor
TGA.....thrombin generation assay
TTR.....transthyretin
vg.....vector genomes
vp.....vector/viral particle
VWD.....von Willebrand disease
VWF.....Von Willebrand factor

Chapter 1

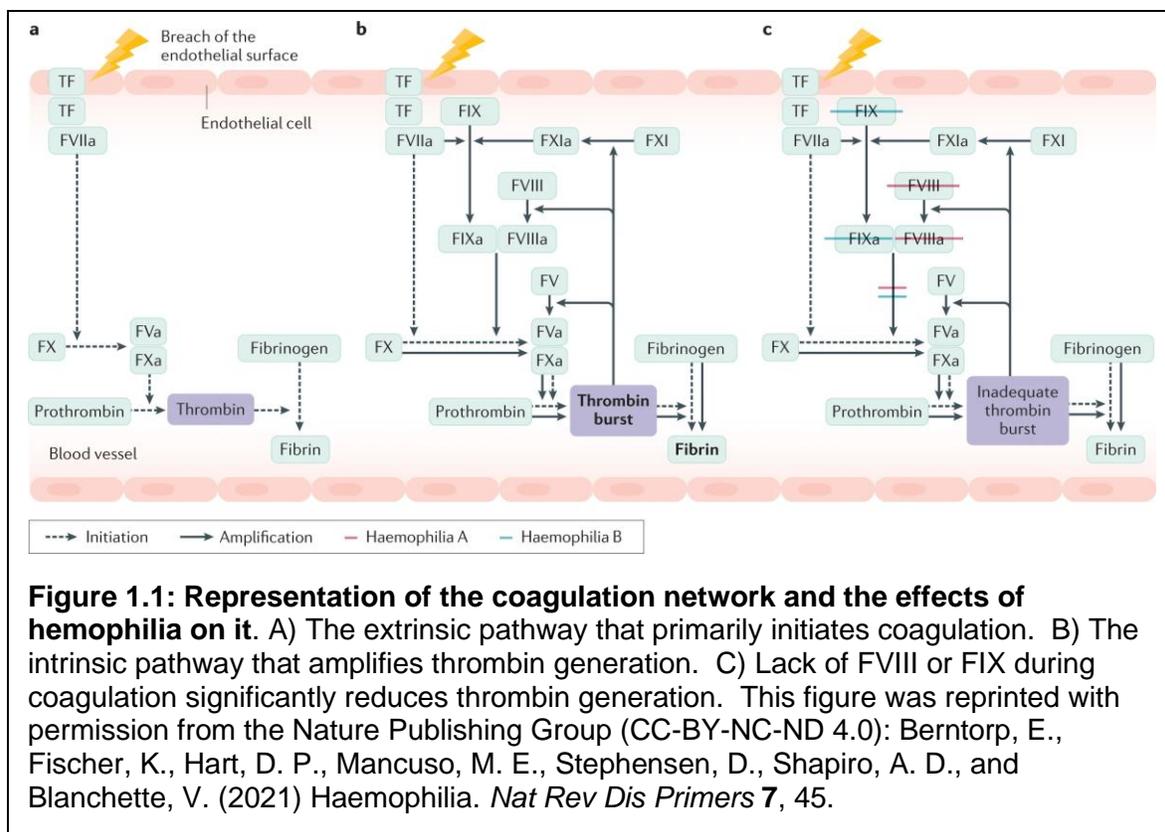
Introduction

1.1 Blood coagulation network

Hemostasis is a complex process that stops bleeding at the site of vessel injury. Hemostasis can be broken down into four stages: 1) blood vessel constriction, 2) formation of a platelet plug (primary hemostasis), 3) activation of the coagulation network (secondary hemostasis), and 4) the formation of a fibrin plug. The coagulation network, also known as secondary hemostasis, is a series of enzymatic reactions that result in the formation of fibrin (1). The coagulation network is broken down into three distinct stages: 1) the extrinsic pathway (tissue factor pathway), 2) the intrinsic pathway (contact activation pathway), and the common pathway (2). Primary hemostasis results in recruitment and adhesion of activated platelets to the site of injury. At this stage, they are ionically bound to the subendothelial collagen through several receptors: glycoprotein (GP) GPIa and GPVI directly to the collagen, and GPIb and GPIIb/IIIa through von Willebrand factor (VWF) (3). At this stage, primary hemostasis is complete.

The extrinsic pathway primarily serves to initiate the coagulation network (**Figure 1.1A**). It is activated when exposed to subendothelial collagen (exposing tissue factor (TF)) encounters plasma. Factor VII circulates in the bloodstream and becomes activated Factor VII (FVIIa) when it comes into contact and complexes with TF. The FVIIa-TF complex then activates Factor X (FX) to FXa. FXa, through the common pathway, generates small amounts of thrombin, which then helps to activate the intrinsic pathway thereby promoting a positive feedback loop.

The intrinsic pathway primarily amplifies the coagulation network, resulting in large amounts of thrombin generated (**Figure 1.1B**). It begins with Factor XII (FXII), that becomes activated upon exposure to a negatively charged surface, which is often the surface of an activated platelet. FXIIa then activates Factor XI (FXI) to FXIa through proteolytic cleavage. FXIa then activates FIX to FIXa, also through proteolytic cleavage. Interestingly, the FVIIa-TF complex also activates FIX, showing interplay between the intrinsic and extrinsic pathways (4). Concurrently,



the small amount of thrombin generated from the extrinsic pathway activates FVIII. The intrinsic FXase complex is then formed when FVIIIa forms a complex with FIXa and calcium ions on a negatively charged phospholipid surface to convert FX to FXa.

The intrinsic and extrinsic pathways converge at the common pathway. Here, FXa, with its cofactor activated Factor V (FVa), converts prothrombin into thrombin. Thrombin serves many roles in the coagulation network, including activation of FVIII, FV, and FXI. Its most important function is to activate soluble fibrinogen to insoluble fibrin, which forms a gel-like framework for the blood clot. Finally, thrombin activates factor XIII; FXIIIa covalently cross-links fibrin molecules to form a more stable blood clot. As seen in figure 1.1C, when FVIII or FIX are deficient, as in the case of hemophilia A and B, respectively, there is inadequate thrombin generation. This leads to insufficient fibrin production and subsequently a weak and unstable blood clot.

1.2 Hemophilia A and B

Hemophilia A and hemophilia B are X-linked bleeding disorders caused by a deficiency in either coagulation factor VIII (FVIII) or coagulation factor IX (FIX) activity, respectively. The prevalence of hemophilia A is about 1:5,800 male births and Hemophilia B is about 1:26,000 male births (5). Hemophilia A and B are sub-classified by disease severity as: mild, moderate, or severe, based on the FVIII or FIX activity present in the patient's plasma. Severe hemophilia is characterized by < 1% factor activity; moderate hemophilia by 1% to 5% factor activity; and mild hemophilia by 5% to 40% factor activity (6). The primary symptom of hemophilia is uncontrolled bleeding. This can occur in the form of a hemorrhage, or, more commonly, into the joints, causing hemophilic arthropathy (7).

Hemophilia A and B have likely existed as long as humans have, however, the first modern doctor to describe hemophilia was John Conrad Otto, who in 1803, described “an account of an hemorrhagic disposition existing in certain families” (8). It wasn't until the 1950's that Hemophilia A and B were distinguished from each other, as they have a similar clinical presentation. In this era, hemophilia was primarily diagnosed by observation of a bleeding diathesis and when a small amount of normal blood was added to hemophilia blood, the time to clot formation was shortened. When two hemophilia patients had their blood mixed, there was no shortening of clot time, except in some rare instances. After numerous instances of this, it was determined that there are multiple forms of hemophilia and “Christmas Disease” (hemophilia B) was a recognized separate condition (9; 10).

Treatment in the 1950s and 1960s for persons living with hemophilia (PLWH) as largely limited to infusions of whole blood or fresh plasma, although these infusions seldom stopped severe bleeds (11). The first effective treatments for hemophilia were discovered in the 1970s, in the form of lyophilized plasma concentrates of coagulation factors. These concentrates could

be administered prophylactically and at the home, revolutionizing treatment and outcomes for PLWH by controlling bleeding events and reducing joint damage (12).

The current standard of care for both hemophilia A and B requires regular prophylactic intravenous infusions of either FVIII, or FIX, respectively. There are a variety of products for patients to choose from: plasma-derived, recombinant, recombinant produced from human cell lines, and for hemophilia A, a bi-specific antibody that mimics the function of FVIIIa (13-15). Additionally, there are modified recombinant products with extended half-lives that contain either polyethylene glycol moieties, F_c fusion conjugates, and albumin conjugates (16).

These treatments are efficacious; however, they require 1-3 infusions every week, which is inconvenient and can damage blood vessels over time. They are also extremely expensive with annual expenditures per patient often exceeding \$250,000 (17). Furthermore, anti-drug antibodies to the factor replacement, termed 'inhibitors', are observed in 20-30% of severe hemophilia A patients and 3% of severe hemophilia B patients (18), rendering the treatment ineffective. Due to the drawbacks of factor replacement, and inaccessibility of products in poorer countries, gene therapy is an excellent solution for the treatment or cure of hemophilia A and B.

1.3 FVIII

The coagulation factor *F8* gene is 186,000 base-pairs (bp) and located on the X-chromosome. It contains 26 exons that transcribe a 9,048 bp mRNA transcript (19). The full-length mature protein has 2,332 amino acids and has the domain structure of: A1-A2-B-ap-A3-C1-C2 (20). FVIII is primarily produced in the liver, by liver sinusoidal endothelial cells (21). Following translation, FVIII undergoes numerous post-translational modifications, beginning in the endoplasmic reticulum (ER). Here, the 19 amino acid signal peptide is removed, N-linked glycans are added to the molecule, and disulfide bonds are formed, through interactions with

chaperone proteins (22). Two additional proteins, called lectin mannose binding 1 (LMAN1) and multiple coagulation factor deficiency 2 (MCFD2) then form a complex that recruits FVIII into coat protein complex II vesicles, which facilitates FVIII's transport from the ER to the Golgi apparatus (23). In the Golgi, N-linked glycans are modified, O-linked glycans are added, certain tyrosine residues are sulfated and copper ions become bound (24; 25). Lastly, furin cleaves the FVIII twice in the B domain to create a heterodimer, and it is secreted out of the cell (26). While recombinant full-length FVIII is 2,332 amino acids, circulating FVIII displays size heterogeneity due to alternative processing of the B-domain (27).

FVIII circulates in the bloodstream at an approximate concentration of 100-200 ng/mL (~1 nM). Approximately 95% to 98% of circulating FVIII is bound to VWF (28). The bound VWF has several important functions, including: increasing the circulating half-life, increasing the stability, and protecting FVIII from FXa and activated protein C (APC) (26). FVIII has an approximate half-life of 12 hours when bound to VWF and 2 hours when unbound (29). FVIII is primarily cleared by the hepatocytes, by the CD91, low density lipoprotein related protein 1 (LRP1), and heparan-sulfate proteoglycans (HSPG) (30).

Like all other coagulation factors, FVIII requires activation before it can participate in the coagulation network. FVIII is activated by thrombin, specifically by proteolytic cleavage at R372, R740, and R1689 (31). This activation completely removes the B-domain and separates the A1 and A2 domains, creating activated FVIII (FVIIIa). FVIIIa is an unstable heterotrimer that acts as a cofactor in the intrinsic tenase complex. Illustration of the activation of FVIII can be seen in **Figure 1.2**. FVIIIa acts as a cofactor, non-covalently bound to the surface of an activated platelet at the site of injury. Here, it binds both FIXa and zymogen FX, such that FIXa can efficiently activate FX to FXa. FVIIIa can be inactivated in 2 main ways. First, by the spontaneous disassociation of the A2 domain from the heterotrimeric complex, with a half-life of

approximately 2 minutes (32) . Secondly, it is cleaved by APC, with its cofactor, Protein S at R336 and R562 (33; 34).

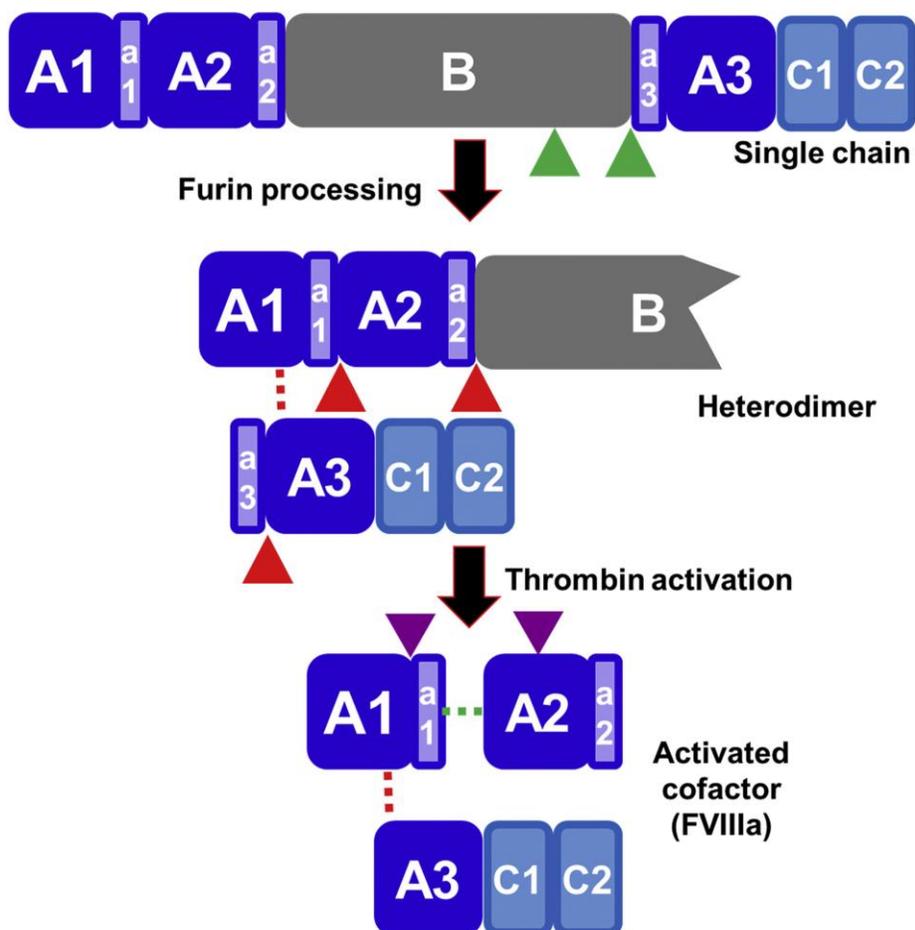
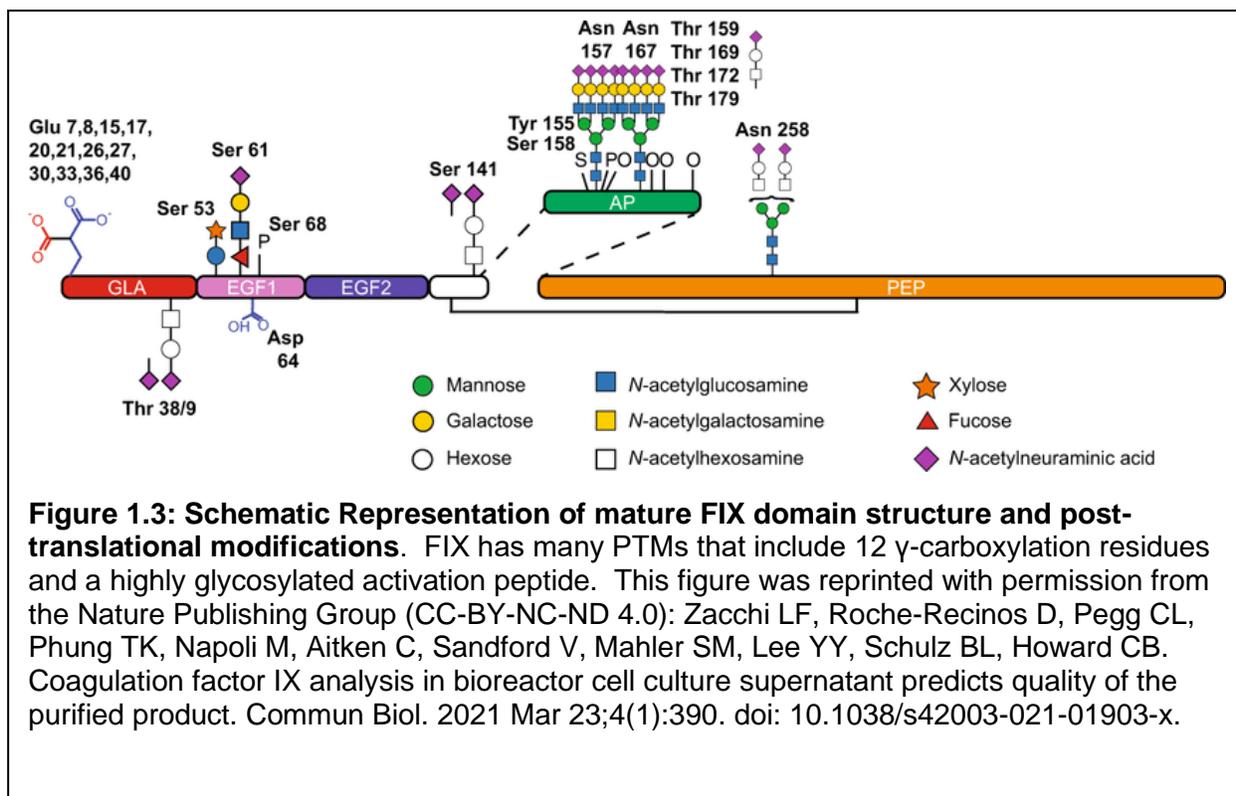


Figure 1.2: Late processing and activation of FVIII. Following translation, FVIII is post-translationally modified through a variety of methods, culminating with cleavage in the B domain by Furin. FVIII circulates primarily as a heterodimer and upon activation by thrombin, exists as an unstable ionically bound heterotrimer. This figure was reprinted with permission from the American Society of Gene and Cell Therapy (CC-BY-NC-ND 4.0): Samelson-Jones BJ, Arruda VR. Protein-Engineered Coagulation Factors for Hemophilia Gene Therapy. Mol Ther Methods Clin Dev. 2018 Dec 31;12:184-201. doi: 10.1016/j.omtm.2018.12.007.

1.4 FIX

The coagulation *F9* gene is located on the X-chromosome and spans 39,723 bp. It contains 8 exons, and is subsequently translated into a 461 amino acid protein (35). FIX is a one of the vitamin K-dependent clotting factors, which all serine proteases that share a similar domain structure (36). The domain structure of the immature protein is: signal peptide – propeptide – Gla – EGF1 – EGF2 – linker – activation peptide – protease. Just like FVIII, FIX undergoes numerous post-translational modifications. The signal peptide translocates the protein into the ER, where it is then cleaved at C28 (37). In the ER, the glutamate residues in the gla domain are gamma carboxylated. Schematic representation of mature FIX and its PTMs can be seen in **figure 1.3**. Gamma-carboxylation is critical to the function of FIX. Gamma-carboxylation is conducted by the enzyme gamma-glutamyl carboxylase (GGCX), which



requires vitamin K as a cofactor and the propeptide facilitates binding of immature FIX to GGCX. Gamma-carboxylation adds an additional carboxylic acid moiety to the glutamate, which

allows it to form a complex with calcium ions that facilitates binding to the phospholipid membrane on the surface of an activated platelet.

FIX circulates in the bloodstream at a concentration of $\sim 2 - 5 \mu\text{g/mL}$ ($\sim 70 \text{ nM}$) (38), however a majority of FIX in the body is stored in the extravascular space (39). At steady state, at least 75% of FIX is extravascular, bound to type IV collagen. It is hypothesized that the excess of extravascular FIX to plasma FIX favors necessary extravascular coagulation, while limiting pathological intravascular coagulation (39).

Once secreted, FIX circulates as an inactive zymogen. It is primarily activated by FXIa (40), however, it is also activated to a lesser extent by FVIIa/TF complex during the initiation phase of coagulation (4). FIX is activated following cleavage at the two arginine residues on either side of the activation peptide at positions 184 and 219. When only the R184 is cleaved, it is called Factor IX α ; when only R219 is cleaved it is called FIX α , and when both are cleaved, fully activating FIX, it is called Factor IX $\alpha\beta$ (FIXa). In FIXa, the light chain (gla-EGF1-EGF2) is bonded to the heavy chain (protease domain) by a single disulfide bond (41). FIXa is an extremely poor enzyme on its own and requires FVIIIa activity to convert FX to hemostatic amounts of FXa. FIXa is inactivated by antithrombin, which irreversibly binds to FIXa (and IIa and FXa). The enzyme-inhibitor complex is then rapidly cleared from circulation (42).

1.5 Brief History of Gene Therapy

Gene therapies are a unique class of drugs, and as such, have unique considerations with respect to clinical pharmacology. The FDA defines a human gene therapy product as: “products that mediate their effects by transcription or translation of transferred genetic material, or by specifically altering host (human) genetic sequences.” This contrasts with the drug class of biologics, where even though the “effective agent” is typically a protein, that protein is made

exogenously and then administered as medicine. Essentially, with gene therapy, the patient's own body makes the medicine.

The first successful gene therapy in humans, came to be when a 4-year-old girl was suffering from adenosine deaminase deficiency (ADA), which resulted in severe combined immunodeficiency (SCID) ADA-SCID. A gamma-retroviral vector to modify patient-derived T cells. The treatment was repeated 10 more times over the next two years and was largely successful (43). As of 2003, 20% of the patient's circulating T cells expressed the retroviral gene and the patient was considered immunologically competent (44).

This early success then led to the idea of using hematopoietic stem cells (HSC) as a cellular vehicle for genetic modification and durable therapeutic benefit, as they self-renew and can provide sustained transgene expression for life. The pivotal SCID-X1 trial, based in Europe in 2000 had 20 male patients. The treatment used a gamma retroviral vector mediated transduction of HSCs and subsequent transplant of the modified cells (45). The treatment was largely successful, however, 5 of the 20 patients ended up developing leukemia (T-ALL), which was found to be due to insertional mutagenesis and activation of an endogenous proto-oncogene (46). This gave researchers pause on using integrating viral vectors but led to the development of two major advances for gene therapy: development and use of safer and self-inactivating integrating viral vectors and non-integrating viral vectors.

The first reported case of using a non-integrating viral vector for gene therapy was in 1999, to treat ornithine transcarbamylase (OTC) deficiency. Jesse Gelsinger, an 18-year-old at the time of the trial was administered intra-hepatically the escalated dose of 6×10^{13} recombinant adenoviral vector particles that encoded a functional OTC transgene. Unfortunately, within hours of the infusion, he developed a severe acute inflammatory response and died two days later (47). This caused an immediate pause on all gene therapy trials in the United States.

Adeno-associated virus (AAV) has been used for clinical gene therapy since the late 1990's, when it was used to deliver a functional CFTR gene to the maxillary sinus of cystic fibrosis patients (48). Since then, there have been over 100 clinical trials using AAV to treat primarily monogenic diseases (49). In 2012, the first AAV gene therapy was approved for hereditary lipoprotein lipase deficiency (50) in Europe. This drug, called Glybera (alipogene tiparvovec), is an AAV1 that encodes for a functional copy of the lipoprotein lipase gene with a gain of function mutation S447X that encodes a premature stop codon in exon 9 (51). Unfortunately in 2017, this drug was withdrawn from the market due to lack of sales and was never approved by the FDA (52).

Also in 2017, Luxturna (voretigene neparvovec) was approved in the United States for inherited retinal dystrophy. This AAV2 vector encoded a gene replacement for RPE65, and patients had significantly improved vision following treatment (53). In 2019, Zolgensma (Onasemnogene abeparvovec) was approved to treat spinal muscular atrophy. It utilizes an AAV9 capsid, encoding for a functional copy of the SMN1 gene. When administered during infancy, patients have rapid and significant improvements in motor function (54). 2022 and 2023 have brought on 3 additional AAV gene therapy approvals in the United States: Hemgenix (Etranacogene Dezaparvovec) for Hemophilia B, Elevidys (Delandistrogene Moxeparvovec) for Duchenne muscular dystrophy, and Roctavian (Valoctocogene Roxaparvovec) for Hemophilia A. Elevidys is an AAV rhesus isolate serotype type 74 that encodes for a shortened version of the dystrophin gene. Developed by Sarepta Therapeutics, patients receiving the therapy had confirmed transduction of muscle tissue and stabilized or improved motor function (55). Upstanza, is an AAV2 that is used to treat severe aromatic L-amino acid decarboxylase deficiency. It is currently only approved in Europe by the EMA, but is currently under review by the FDA (56).

1.6 Pharmacology and Optimization Strategies of AAV Gene Therapy

Viral vector-based gene therapies to treat monogenic diseases have benefitted from the use of AAV vectors. AAV particles were initially discovered as a contaminant in adenoviral preparations in 1965 (57). Unlike adenoviral vectors, wild-type AAV lacks essential genes needed for replication and is thus dependent on adenoviruses to complete its lifecycle. It is also far less immunogenic than adenovirus, is easily manipulated using molecular cloning techniques, and it can be produced at much higher titers than other viral vectors (58). Additionally, there are at least 13 known natural serotypes and over 100 variants, mosaic, or otherwise bioengineered serotypes. Each of the different natural serotypes have varying tissue tropism profiles; for example, AAV8 has efficient transduction to the liver and AAV9 transduces neuronal cells well and can cross the blood brain barrier (59; 60). These properties make AAV an excellent candidate for use as a non-integrating viral vector for gene therapy.

The AAV viral particle is approximately 26nm in diameter and has a single-stranded DNA genome in its wild-type form. The genome is 4.7kb and contains 2 genes and is flanked by inverted terminal repeats (ITR) (61). Through different promoters, alternative translation start sites and differential splicing, the 2 genes, *Rep* (Replication) and *Cap* (Capsid) create at least 9 different proteins. The *Rep* gene encodes 4 different proteins, termed: Rep78, Rep68, Rep52, and Rep40, which are involved in the regulation of AAV gene expression, replication of viral DNA, and viral packaging (62). The *Cap* gene encodes the 3 proteins that create the viral capsid, termed: VP1, VP2, and VP3; and 2 assembly-activating proteins (AAP), termed: AAP, and membrane associated AAP (MAAP), which assist in capsid assembly (63; 64). The ITRs serve numerous functions including acting as the viral origin of replication, serving as the packaging signal and intramolecularly recombining following transduction to form stable concatemers (65; 66).

Since AAV is replication incompetent, *in vitro* production requires either infection of a helper virus, such as adenovirus or herpes simplex virus, or transient or stable expression of

helper genes. The helper genes in this case are the *Rep* and *Cap* genes, described above. The most common method of clinical production is transfection into mammalian cells (typically HEK293). Using this method, the therapeutic transgene is inserted into an AAV transfer plasmid, between two ITRs, preceded a promoter and followed by a poly-adenylation signal and it is co-transfected with either 1 or 2 helper plasmids that contain the necessary genes for viral particle production (67). AAV particles are then purified using various techniques to be administered to patients.

AAV transduction begins with recognition by the target cell of the AAV viral particle which prompts internalization of the viral particle via clathrin-mediated endocytosis. The AAV particle is then trafficked through the cytosol and upon endosomal escape, it enters the nucleus and is uncoated, revealing the genome. Some AAV particles are degraded by the proteasome and do not enter the nucleus (68). Once the AAV genome is in the nucleus, it undergoes second strand synthesis, which is thought to be the rate-limiting step for AAV gene expression (69). AAV particles can also be “self-complementary (sc),” instead of single-stranded, where the genome is double-stranded DNA and therefore does not need to undergo second-strand synthesis. ScAAV is made by deletion of the 3’ terminal resolution site, a region within the ITR, in the transfer plasmid (70). Once the AAV genome is double-stranded, either by the host cell or upon transduction, it undergoes intramolecular circularization to form a concatemer. The concatemer then persists in the cell and transcription and subsequently, translation of the transgene ensues (68). A graphic representation of the AAV lifecycle following transduction can be seen in **figure 1.4**.

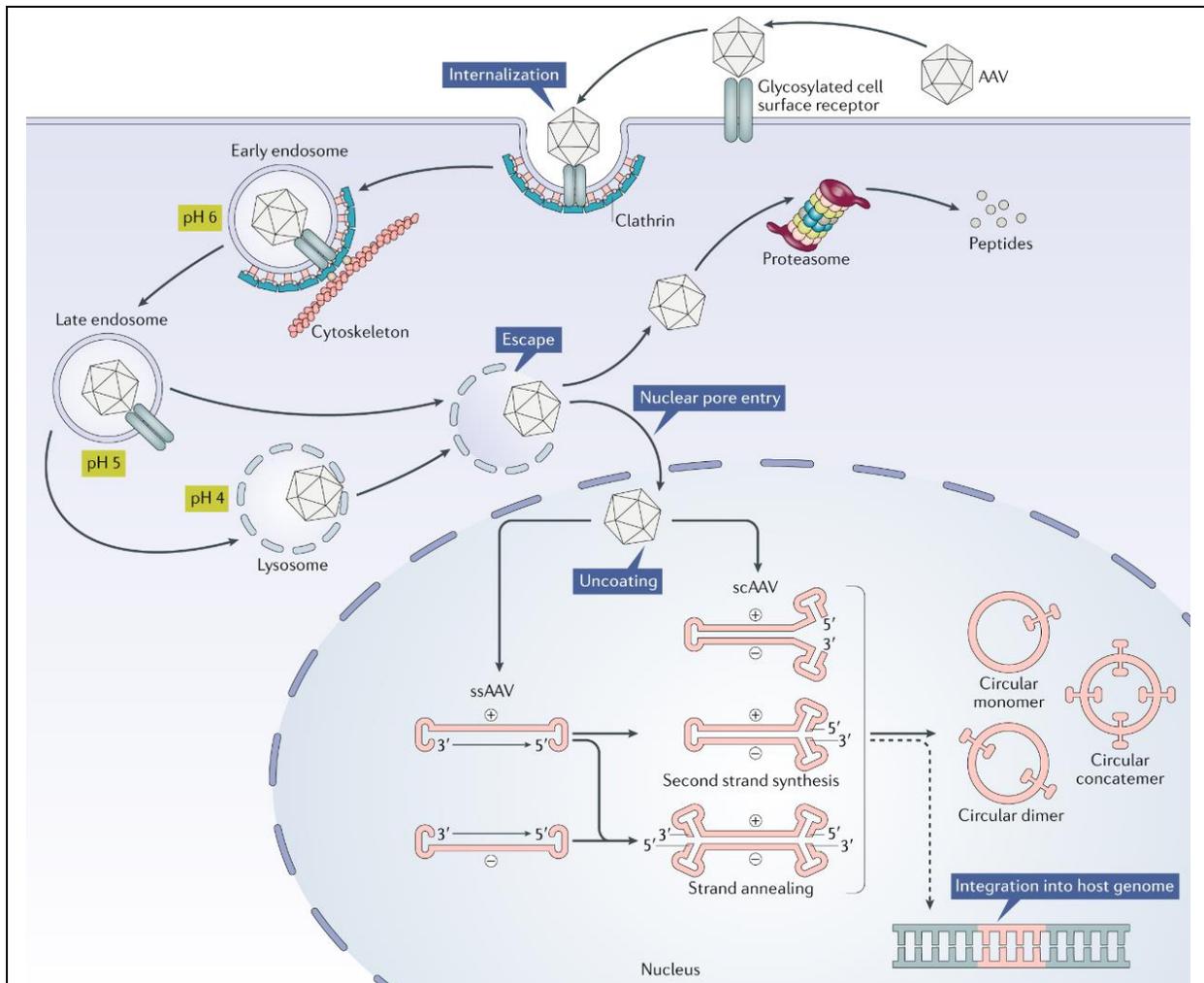


Figure 1.4: AAV transduction pathway. AAV initially enters the target cell through glycosylated cell surface receptors. It then follows the cotyskeletal pathway until it escapes the lysosome and enters the nucleus. There, it is uncoated (protein capsid removed from the genome) and second strand synthesis begins for ssAAV vectors. Once the genome is double-stranded, it circularizes to form a concatemer, where it persists episomally. Occasionally AAV is integrated into the host genome. This figure was reprinted with permission from the Nature Publishing Group (CC-BY-NC-ND 4.0): Naso MF, Tomkowicz B, Perry WL 3rd, Strohl WR. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs*. 2017 Aug;31(4):317-334. doi: 10.1007/s40259-017-0234-5. PMID: 28669112; PMCID: PMC5548848.

Depending on the clinical indication, AAV is either administered systemically into the bloodstream, which leads to broad systemic distribution (71), or directly into the target tissue or organ, which has a restrictive biodistribution, with minor leakage into systemic circulation (72).

Thorough studies of AAV biodistribution in humans are limited, as harvest of all organ tissues from patients is required for assessment. A study of Onasemnogene abeparvovec (Zolgensma), an AAV gene therapy to treat SMA was performed on 2 deceased patients and the therapy, which was administered intrathecally was widely detected in the CNS and peripheral organs, with the highest concentration of vector genomes found in the liver (73). This trend also holds true when AAV particles were administered intravenously to mice (74). Liver-directed AAV gene therapy displays dose responsiveness, where transgene protein expression is generally higher when a higher dose is administered, although the increase is not always linearly proportional to the dose (75).

The durability of liver directed AAV gene therapies remains in question to this day, as there are numerous potential mechanisms at play that control maintenance and/or decline of transgene expression. Additionally, the heterogeneous clinical response among patients further challenges our understanding of AAV gene therapy durability (76). The lifespan of hepatocytes *in vivo* is 200-400 days, therefore it is reasonable to conclude that this is a factor affecting gene therapy durability (77). Activation of the unfolded protein response by ER accumulated transgene product is another possible mechanism that can cause loss of transgene expression over time that has been shown in mice (78), but not in humans albeit with an extremely limited data set (76).

Safety is another critical aspect of clinical pharmacology. There is a clear positive correlation between AAV dose and incidence of treatment-emergent adverse events (TEAE). The primary TEAE observed following systemic AAV administration is hepatotoxicity, which manifests as elevated liver enzymes in the bloodstream. This is generally treated with a short course of corticosteroids (79). A critical safety concern that is still widely debated for liver directed AAV gene therapy is the potential for these drugs to cause hepatocellular carcinoma (HCC). Wild-type AAV can integrate into genomic DNA, through the *Rep* gene, and has been

implicated in HCC in humans (80), although since recombinant AAV lacks the *Rep* gene, this risk is mitigated. Integration by recombinant vectors occurs at a rate a very low rate (<0.01%) (81). Studies in our lab and by others, have observed HCC in liver-directed AAV treated mice, although this appears to have a correlation to enhancer-promoter strength, where stronger regulatory elements increase transactivation of proto-oncogenes and increase incidence of HCC (82). In hemophilia A dogs treated with AAV-FVIII, clonal expansion occurred, however, HCC was neither diagnosed nor reported (83). Recombinant AAV integration has not been linked to any reports of HCC in any nonrodents species.

All said, there is a catch-22 when it comes to AAV dosing: Since therapies can only be given once (due to humoral response to the AAV capsid), one wants to maximize the dosage to increase efficacy and durability, but also desires to minimize the dosage to lower the risk profile. This dichotomy is why optimization of the potency of gene therapies is critically important for the field of clinical gene therapy.

There is a large and ongoing effort in the gene therapy field to optimize viral vectors. This is due to the high cost of production and the dose limiting toxicities observed. The three main methods for optimizing viral vectors for gene therapy are: protein engineering, capsid engineering, and codon optimization.

The goal of protein engineering is to modify the primary polypeptide sequence such that the transgene product generated is more effective. The most common methods of protein engineering include: structure based and/or computational design, directed molecular evolution, random mutagenesis, post-translational modifications, and ancestral sequence reconstruction (ASR) (84).

Codon optimization is another major area of interest for gene therapy research. Briefly, codon optimization is defined as the usage of synonymous codons in the genetic payload to

maximize protein expression. Improved expression by codon optimization occurs through several potential mechanisms including: increasing translation efficiency by matching codon usage bias to the available tissue tRNA pool (85) and increasing mRNA stability (86). Additionally, most clinical AAV gene therapies remove all CpG motifs within the transgene, as this reduces immunogenicity and improves duration of transgene expression (87). Non-methylated CpG motifs are often observed in viral and bacterial DNA (and DNA from non-integrating viral vectors) and are thus detected by innate and adaptive immune cells, targeting the modified cell for apoptosis (88). Removing these motifs appears to solve this problem.

Modification of the capsid of the AAV viral particle is another method that has been used primarily to improve transduction of the target cell and evade the immune response. To improve target cell transduction, methods such as directed evolution and rational design have been utilized successfully to create novel capsid protein sequences that bind receptors on target cells with higher affinity (89). The immune response to AAV capsids comprises numerous immunological mechanisms including: 1) previously existing neutralizing antibodies (nAb) blocking transduction into target cells, 2) innate immune response to capsid following endosome breakdown, and 3) detection of transduced cells by cytotoxic T lymphocytes cells after capsids are presented on MHC class 1 molecules (89). Modifying capsid sequence to avoid ubiquitylation and proteasome-mediated capsid degradation has been shown to avoid the detection of transduced cells by cytotoxic T lymphocytes in mouse liver (90). Approximately 90% of the adult human population has been previously infected with AAV, with about half the population having pre-existing nAbs towards AAV (91). The capsid engineering methods described above in addition to shuffling different capsid regions from other serotypes to create chimeric capsids are being used to evade these nAbs (92).

1.7 Gene Therapy of hemophilia A and B using AAV vectors

Hemophilia A and B are both excellent candidates for treatment with gene therapy. Despite a lower clinical incidence, trials for hemophilia B were performed first, as the gene is much smaller and therefore easier to work with prior to widespread use of modern molecular biology techniques. In 1997, mice treated with an intramuscular injection of FIX-AAV showed stable gene transfer and detectable expression of circulating FIX in Rag 1 deficient mice, which produce no mature T or B cells (93). This work was then translated to a hemophilia B dog model, where therapeutic levels of FIX were achieved by 2 different groups (94; 95). The first in-human studies for AAV-FIX were performed as intramuscular injections and resulted in subtherapeutic FIX levels (96). To directly target the liver, the next in-human AAV-FIX trials utilized an intrahepatic infusion (97). These trials were paused after detection of AAV DNA in the semen of a patient, although the DNA was found to be present only in seminal fluid, not sperm and was eventually cleared (97). At the highest dose level tested, FIX levels rose and then dropped over the subsequent 12 weeks following injection, coinciding with a rise in liver transaminases. This led researchers to theorize that cytotoxic T cells were destroying transduced hepatocytes and that this could be ablated with a short course of corticosteroids following vector administration or the administration of a lower AAV dosage (98). Since then, numerous clinical trials for hemophilia B utilizing FIX-AAV were performed and are ongoing, which finally led to the approval of Hemgenix in 2022 (**Table 1.1**).

Hemgenix, approved in 2022 is an AAV5 serotype that encodes for a liver codon optimized FIX gene with the “Padua” (p.R384L) variant, that increases specific activity ~7-8 fold (99). This pivotal approval builds on the numerous previous clinical trials for FIX gene therapy, outlined in Table 1. Hemgenix is also the most expensive drug in the world, with a price tag of \$3.5M. Despite being a major leap forward in the field, questions remain about its safety and durability. Nine out of the 54 patients in the Hemgenix trial developed transient transaminitis, which is common for liver directed AAV gene therapy. All nine cases eventually resolved but

did require the use of corticosteroids. Additionally, it is not currently known how long FIX gene therapy will last in patients. The earliest successful trials have shown FIX expression to be durable and stable at ~2-5 IU/L, depending on dose, 8 years post treatment in all 10 subjects. Additionally, FIX concentrate usage dropped by 66% and annual bleed rate has dropped by 82%. (100). For patients treated with Hemgenix, mathematical models predict they would likely have durable FIX activity levels and not require FIX replacement products for up to 25.5 years (101).

Table 1: Current landscape of Gene Therapy Trials for Hemophilia B.

Sponsor	Transgene	No. of CpG motifs in transgene	Serotype	Genome format	Method of vector delivery	Dose range (vg/kg)	Mean stable FIX activity levels	No. of patients with transaminitis	Current status
Avigen/CHOP	Wild-type FIX	19	AAV2	ssAAV	IM	2e11 to 1.8e12	Transient at a maximum level of 1.6%	0	Closed
Avigen/CHOP Coagulin-B	Wild-type FIX	19	AAV2	ssAAV	Intrahepatic artery	2e11 to 2e12	Transient with a maximum of 12%	1/2 at highest dose	Closed
St Jude/UCL	Codon-optimized FIX	0	AAV8	scAAV	Systemic	2e11 to 2e12	5.1%	4/6 patient at highest dose	Closed
Takeda (Baxalta; BAX 335)	Codon-optimized FIX + Padua mutation	99	AAV8	scAAV	Systemic	2e11 to 3e12	Transient except in 1 patient who had expression of ~20% at last report	2/6 patients treated at or above 1e12 received steroids in response to ALT elevations and 1 patient received prophylactic steroids.	Closed
Spark Therapeutics (SPK-9001, fidanacogene elaparvovec)	Codon-optimized FIX containing the Padua mutation	0	AAV-Spark100	ssAAV	Systemic	5e11	19.8% at 5 years	3/15 patients	Transitioned to phase 3 with Pfizer

uniQure (AMT-060)	Codon-optimized FIX	0	AAV5	ssAAV	Systemic	5e12-2e13	6.9% 5.2% at 5e12-vg/kg dose and 7.2% at 2e13-vg/kg dose at 5 years	2/5 at highest dose	A new program, AMT-061, that contains the FIX Padua in development with CSL Behring
CSL Behring (AMT-061, etranacogene dezaparvovec), "Hemgenix"	Codon-optimized FIX containing the Padua mutation	0	AAV5	ssAAV	Systemic	2e13	36.9%	9/54	Approved, 2022
Dimension Therapeutics (DTX101)	Codon-optimized FIX	96	AAVrh10	ssAAV	Systemic	1.6e12-5e12	6.7%	3/3 at highest dose	Closed
Freeline Therapeutics (FLT-180a, verbrinacogene setparvovec)	Codon-optimized FIX containing the Padua mutation	5	AAV-S3 synthetic capsid	ssAAV	Systemic	3.84e11 - 1.28e12	30%-279%	8/10	Clinical trials ongoing
Belief BioMed (BBM-H901)	Codon-optimized FIX containing the Padua mutation	0	AAV843 synthetic capsid	scAAV	Systemic	5e12	36.9%	2/10	Clinical trials ongoing
Sangamo Bioscience (SB-FIX)	Codon-optimized FIX	Not known	AAV6/zinc finger-mediated targeted integration into the albumin locus in hepatocytes	ssAAV	Systemic		Unknown	Unknown	Closed

Table 1: Current landscape of Gene Therapy Trials for Hemophilia B. This table was adapted from: Nathwani AC. Gene therapy for hemophilia. Hematology Am Soc Hematol Educ Program. 2022 Dec 9;2022(1):569-578.

Hemophilia A gene therapy has several key differences compared to hemophilia B gene therapy. First, the *F8* cDNA is almost twice the size of the native AAV genome and cannot fit into an AAV capsid. Even the B-domain-deleted (BDD) version of the *F8* transgene (4.5 kb) approach the size of the AAV genome and transgene expression cassettes containing a strong (and typically larger) promoter and poly A sequence are oversized for AAV packaging.

Furthermore, FVIII is more immunogenic than FIX, which presents additional concerns. Fortunately, no anti-FVIII immune responses have been observed in clinical AAV gene therapy trials to date. However, it is important to recognize that previously untreated patients (i.e., have not received FVIII containing products prior to gene therapy) and patients that have experienced an immune response to FVIII have been excluded from clinical trials and are not eligible for the approved AAV gene therapy currently. One key benefit that hemophilia A gene therapy has is that FVIII, unlike FIX, does not require any tissue-specific post-translational modifications. Therefore, functional FVIII could theoretically be produced by any cell type in the body. This led to the first in human trials where genetically modified fibroblasts, expressing FVIII, were implanted into hemophilia A patients. FVIII expression lasted at most 10 months, and laid the groundwork for later trials (102).

HFVIII has remarkably poor expression/secretion, and this property complicated not only the study of FVIII but limited the viability of early gene therapies. Porcine FVIII was determined to express at 10-14 fold higher levels (103), and it was further determined that this high expression was mediated solely by the A1 and A3 domains (104). A resulting hybrid FVIII that contained porcine A1 and ap-A3 domains with all other sequences being human was then developed (termed ET3) and has been used successfully for hemophilia A gene therapy. Due lack of need for FVIII to be expressed in the liver, gene therapies utilizing lentiviral vectors to modify HSCs *ex vivo* with ET3 were developed (105-107) and in-human clinical trials are currently underway.

Eventually, an AAV-FVIII vector was created that utilized BDD-FVIII and this vector was used in the first in-human trials for AAV-FVIII (108). Following successful Phase III clinical trial data (109) this drug, called Roctavian, was eventually approved by the FDA in 2023 (and by the EMA in 2022) and is licensed by BioMarin Pharmaceuticals (110). Roctavian is an AAV5-serotype viral vector encoding BDD FVIII with the "SQ" linker. In the pivotal phase 3 trial, 134

participants were given a single dose of 6×10^{13} vg/kg. The mean FVIII activity level was 41.9 IU/dL and participants had adverse effects like those in the Hemgenix trial, primarily transient elevation of liver transaminases. Roctavian has a list price of \$2.9 million and has the potential to keep patients free from factor concentrate usage for years. Additional clinical trials using various forms of AAV-FVIII are underway, outlined in **Table 2**.

Table 2: Current landscape of Gene Therapy Trials for Hemophilia A

Company	Name	Transgene	Serotype	Dose (vg/kg)	AAV Antibody Criteria	Clinical Trial Phase
BioMarin	valoctocogene roxaparvovec, Roctavian	ssFVIII-SQ	AAV5	6×10^{13}	Negative	Approved
Pfizer/Sangamo	giroctocogene fitelparvovec	ssFVIII-SQ	AAV6	3×10^{13}	Negative	I/II, III
Spark	SPK-8016	ssFVIII-SQ	LK03	$5-20 \times 10^{11}$	$\leq 1:5$	I/II
UCL/St.Jude		ssFVIII-V3	AAV8	$6-60 \times 10^{11}$	Negative	I/II
Bayer/Ultragenyx	DTX-201	ssFVIII-SQ	AAVhu37	$5-20 \times 10^{12}$	Negative	I/II
Takeda	TAK 754	ssFVIII-SQ	AAV8	$2-12 \times 10^{12}$	$< 1:5$	I/II

Table 2: Current landscape of Gene Therapy Trials for Hemophilia A: This table was adapted from Samelson-Jones BJ, George LA. Adeno-Associated Virus Gene Therapy for Hemophilia. *Annu Rev Med.* 2023 Jan 27;74:231-247.

1.8 Ancestral Sequence Reconstruction as a Platform for Protein Engineering

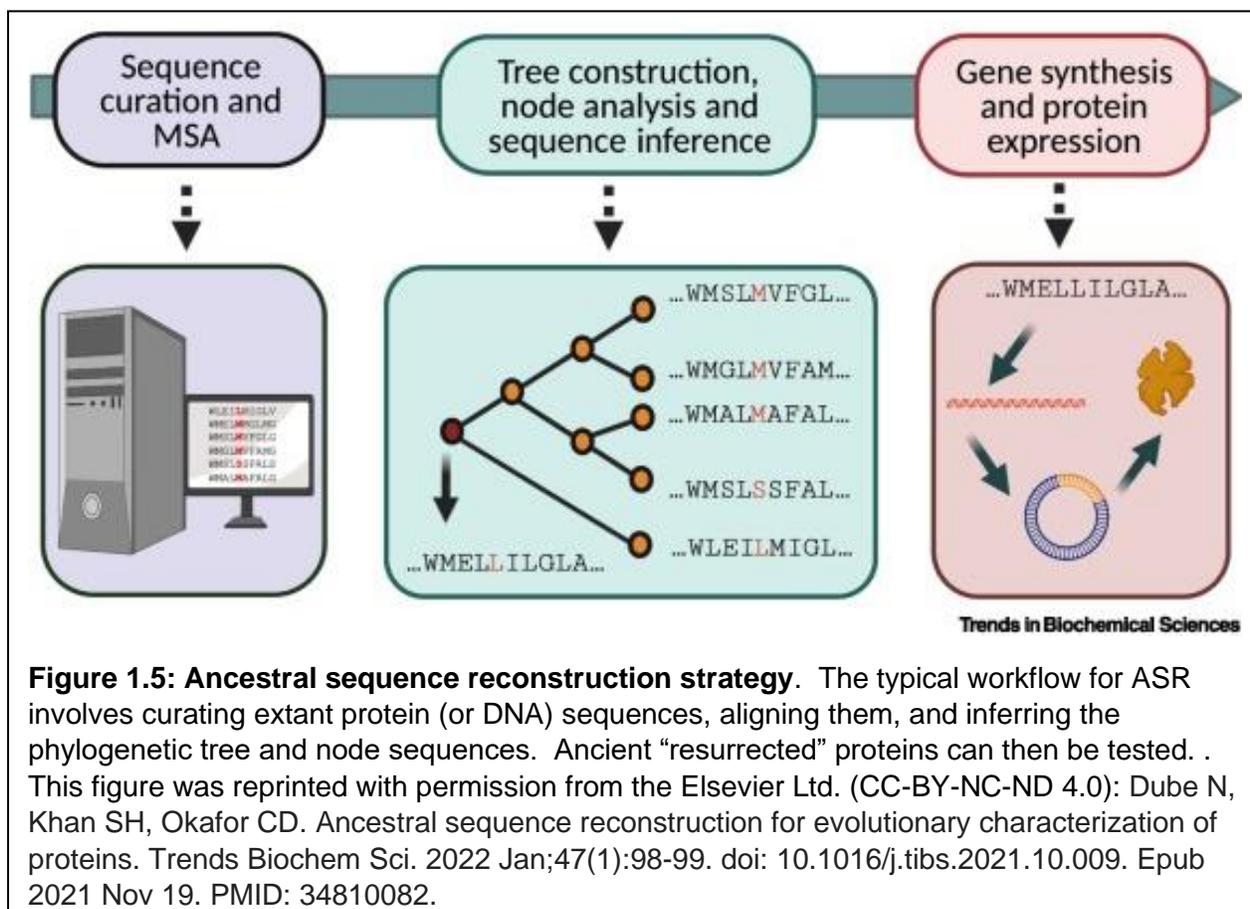
Bioengineering coagulation factors to improve their pharmaceutical properties can lead to more potent gene therapies. Previous bioengineering efforts for FVIII have primarily revolved around using the “ortholog approach,” and removal of the B domain. The B domain of FVIII is not necessary for procoagulant activity, and its removal increases mRNA concentrations 15-fold and secretion 2-fold (111). BDD FVIII also has the advantage of being able to fit within the

packaging capacity of AAV vectors. Orthologous FVIII variants that have improved properties include: porcine FVIII, which has 10-14 fold greater expression than hFVIII (103); canine FVIII, which has improved specific activity (112); ovine FVIII which has modest increases in biosynthetic efficiency, activated cofactor stability, and specific activity (113); and murine FVIII, which displays increased stability of the activated cofactor (114).

Previous bioengineering efforts for FIX have often involved alanine scanning and rational design to determine and mutate critical residues (115), however the most successful FIX engineering used clinically is the inclusion of the “Padua” variant in gene therapy for hemophilia B. This variant, discovered not with any of the aforementioned methods, but serendipitously from a family in Padua, Italy, is a single point mutation that results in a R384L substitution (116). This variant improves FIX activity 6-8-fold and has been incorporated into most newer hemophilia B gene therapies. The ortholog approach has not been as robustly explored for FIX, although canine FIX has been characterized (117).

ASR is a relatively newer protein engineering technique that our lab has successfully used to improve the potencies of FVIII and FIX gene therapies (118). ASR is a multistep process that begins with a multiple sequence alignment (MSA) of extant orthologs. From there, the phylogenetic tree is computed, and lastly, the ancestral sequences at each node of the tree are reconstructed (**Figure 1.5**). There are three main methods for performing ASR: maximum parsimony, maximum likelihood, and Bayesian inference. The maximum parsimony method was invented first and is considered the least sophisticated and least accurate. It reconstructs ancestral sequences by minimizing the number of substitutions at internal nodes without accounting for branch lengths and substitution rates (119). The maximum likelihood method treats ancestral sequences as unknowns when building the phylogenetic model and allows parameters such as branch length and substitution rates to guide the predictions (120). Both maximum parsimony and maximum likelihood methods generate a single tree, whereas

Bayesian inference methods utilize all inputs to create a collection of phylogenetic trees, each with their own calculated likelihood and variance. It is debated whether Bayesian or maximum likelihood methods produce the most accurate trees (120).



1.9 Hypothesis and objectives of this research

While treatments for hemophilia A and B are efficacious, they still have numerous downsides. Hemophilia A and B are excellent disease candidates to be treated with gene therapy. FDA approved products Hemgenix and Roctavian, while paradigm-shifting could still be significantly improved upon. Here, we sought to use ASR as a drug development platform to identify FVIII and FIX variants that can improve the potency of viral vector-based gene therapies.

Chapter 2

Molecular coevolution of coagulation factor VIII and von Willebrand factor

This research was published in *Blood Advances*.

Philip M. Zakas, **Christopher W. Coyle**, Anja Brehm, Marion Bayer, Barbara Solecka-Witulska, Caelan E. Radford, Christine Brown, Kate Nesbitt, Courtney Dwyer, Christoph Kannicht, H. Trent Spencer, Eric A. Gaucher, Christopher B. Doering, David Lillicrap

Molecular coevolution of coagulation factor VIII and von Willebrand factor

Blood Advances. 2021. 5 (3): 812–822

© 2021 by The American Society of Hematology

Philip Zakas, Chris Doering and David Lillicrap conceived the study. Philip Zakas, Chris Coyle, and Christine Brown performed the experiments. Kate Nesbitt and Courtney performed in vivo procedures. Anja Brehm, Marion Bayer, Barbara Solecka-Witulska, and Christoph Kannicht performed. Caelan Radford and Eric Gaucher performed bioinformatics and molecular evolutionary analyses. Philip Zakas, Trent Spencer, Eric Gaucher, Chris Doering, and David Lillicrap wrote the manuscript.

Note: The experimental contributions of Christopher Coyle are specific to Figure 2.3B, 2.3C, and Table 2.2, Supplementary tables S1 and S2 and Supplementary Figure S4.

The data and conclusions from this publication offer insight for the observations and conclusions reported in chapter 3.

2.1 Abstract

Ancestral sequence reconstruction provides a unique platform for investigating the molecular evolution of single gene products and recently has shown success in engineering advanced biological therapeutics. To date, the coevolution of proteins within complexes and protein-protein interactions is mostly investigated *in silico* via proteomics and/or within single-celled systems. Herein, ancestral sequence reconstruction is used to investigate the molecular evolution of 2 proteins linked not only by stabilizing association in circulation but also by their independent roles within the primary and secondary hemostatic systems of mammals. Using sequence analysis and biochemical characterization of recombinant ancestral von Willebrand factor (VWF) and coagulation factor VIII (FVIII), we investigated the evolution of the essential macromolecular FVIII/VWF complex. Our data support the hypothesis that these coagulation proteins coevolved throughout mammalian diversification, maintaining strong binding affinities while modulating independent and distinct hemostatic activities in diverse lineages.

2.2 Introduction

The bleeding disorders von Willebrand disease (VWD) and hemophilia A arise from qualitative or quantitative deficiencies in either von Willebrand factor (VWF) or coagulation factor VIII (FVIII), respectively. Each protein is the product of a distinct gene located on different chromosomes; however, their roles in hemostasis are coordinated and concomitant. VWF has critical functions in both primary and secondary hemostasis, and FVIII is an essential cofactor for the serine protease factor IX. VWF and FVIII circulate in a tight noncovalent complex. Association with VWF protects FVIII from proteolytic degradation, prolongs the plasma half-life of the cofactor, and localizes FVIII activation at the site of injury. Impairment of this association due to mutations in the binding site of either VWF or FVIII results in the bleeding disorders VWD type 2N and nonsevere hemophilia A. Understanding the FVIII-VWF interaction is paramount to the development of novel therapies for hemophilia A and VWD; however, many of the basic

biological questions of their association have remained unanswered throughout the past 35 years of investigation.

Evolutionary studies of vertebrate coagulation have detected the earliest components of the coagulation system in lampreys and hagfish (121-124). These primitive systems generate fibrin exclusively through an extrinsic pathway of factor VIIa/tissue factor–mediated formation of the prothrombinase complex. In bony fishes, FVIII and factor IX emerged after genome duplication and subsequent diversification of the homologous coagulation proteins factor V and factor X, respectively. The emergence of additional coagulation factors allowed for increased thrombin amplification and fibrin generation and supported the development of increasingly complex vasculature systems. It remains unclear how FVIII's dependence on VWF emerged despite occurring in every species investigated.

Known species-specific differences in FVIII biosynthesis, biochemistry, and immunology can be exploited for therapeutic engineering (103; 104; 112-114; 125-128). These unique FVIII characteristics are the result of incremental amino acid replacements that can be traced throughout evolution by ancestral sequence reconstruction (ASR) (129). Although species-specific differences in VWF biology have been observed regarding platelet activation and ristocetin binding (130-132), there has been limited investigation of VWF orthologs or the FVIII-VWF interaction across species. The current study characterizes the molecular evolution of mammalian VWF and investigates the hypothesis that FVIII and VWF coevolved as critically interdependent partners. Using ancestral VWF (AnVWF) sequences spanning several mammalian lineages, the data show that VWF and FVIII evolved coordinately. To the best of our knowledge, this report provides the first phylogenetic and biochemical evidence of molecular coevolution within the hemostatic system and the first investigation of protein complex evolution within higher vertebrates.

2.3 Materials and Methods

Materials: DMEM and Opti-MEM cell culture media, HBSS, and Geneticin were purchased from Gibco Life Technologies (Waltham MA). Fetal Bovine serum was purchased from VWR (Radner, PA). Cell culture plates, flasks, and ELISA costar 96-well plates were purchased from Corning (Corning, NY) and conical tubes were purchased from Sarstedt (Nümbrecht, Germany). Restriction enzymes *NheI*, *AgeI*, *NotI*, and *BssHII* were purchased from New England Biolabs (Ipswich, MA). Heparin-Sepharose 6 Fast Flow, "VIIISelect" resin, HiTrap SPHP columns, and SFM4HEK293 media were purchased from GE Healthcare (Chicago, IL) and Poly-Prep columns were purchased from Bio-Rad (Hercules, CA). HEK cell lines were purchased from ATCC (Manassas, VA) except for Expi293™F and HEK293-H cells purchased from Thermo Fisher Scientific (Waltham, MA). Polyethylenimine, o-phenylenediamine dihydrochloride, RNAlater, bovine serum albumin, and M2 anti-FLAG antibody were purchased from Sigma-Aldrich (St. Louis, MO). TransIT-EE hydrodynamic buffer and TransIT-X2 transfection reagent were purchased from Mirus Bio (Madison, WI). DNA and RNA purification kits were purchased from Qiagen (Germantown, MD) and Omega Bio-Tek (Norcross, GA). Amicon 50K NMWL filters were purchased from EMD Millipore (Burlington, MA). Immunohistochemistry DAKO protein block, DAKO mounting medium, isotype antibody X0936, polyclonal rabbit anti-human VWF antibodies A0082 and P0226 were purchased from Agilent (Santa Clara, CA). Factor VIII chromogenic SP4 assay was purchased from Diapharma (Chester, OH). WHO standard human plasma, and the INNOVANCE VWF:Ac kit were purchased from Siemens (Munich, Germany). RNase Inhibitor RiboLock, Power SYBR™ Green PCR master mix, and Multiscribe™ reverse transcriptase were purchased from Thermo Fisher Scientific (Waltham, MA). The ExpiFectamine293™ transfection reagents and Expi293™ expression media was also purchased from Thermo Fisher Scientific. Human collagen type I was purchased from StemCell Technologies (Vancouver, Canada) and human collagen type III was purchased from Sigma-Aldrich (St. Louis, MO). Maleic anhydride activated plates were

purchased from Thermo Scientific (Waltham, MA). Recombinant B-domain deleted human FVIII was provided by Octapharma Biopharmaceuticals GmbH (Berlin, Germany).

AnVWF sequence inference: AnVWF sequence reconstruction was performed as described previously (133; 134) by using 59 extant VWF sequences, MUSCLE, and MrBayes programs. Ancestral sequences were inferred by using PAML version 4.1.

AnVWF and ancestral FVIII DNA synthesis: Ancestral amino acid sequences An101-, An88-, An84-, An70-, and An63-VWF were codon optimized (co) for human host cell expression and synthesized by using GenScript and subcloned into the pCI-Neo vector. In addition, a C-terminal glycine hinge (GGRGG), TEV protease motif (ENLYFQG), and 3x FLAG tag (DYKDDDDK)₃ with the nucleic acid sequence GGCGGCAGAGGAGGAGAGAACCTGTACTTC CAGGGCGACTATAAGGACGATGACGATAAGGATTACAAAGATGATGACGATAAGGATTAT AAAGACGATGACGATAAGTGA were added to each ancestral sequence. Ancestral FVIII complementary DNAs were generated and purified as previously described (103).

VWF enzyme-linked immunosorbent assay. Enzyme-linked immunosorbent assay (ELISA) was performed by using polyclonal antibodies A0082 and P0226 (Dako). Absorbance was interpolated to a standard curve of World Health Organization (WHO)-defined human standard plasma. VWF samples were sufficiently diluted to ensure antibody excess, and parallel absorbance standard curves were obtained for all constructs.

AnVWF expression and purification: Transient VWF plasmid transfections were performed in HEK293T and HEK116 cells. Polyclonal stable cell lines were generated in HEK293 cells. Monoclonal producer cell lines were generated from HEK116 cells using Geneticin selection. Protein was purified from serum-free Opti-MEM media (Thermo Fisher Scientific). Briefly, AnVWF was diluted 1:1 with 20 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 5 mM CaCl₂, and 0.01% Tween 80 at pH 7.2 before heparin-

sepharose affinity chromatography. Columns were equilibrated and washed with 20 mM HEPES, 5 mM CaCl₂, 50 mM NaCl, and 0.01% Tween 80 at pH 7.2. AnVWF was eluted with 650 mM NaCl. Fractions were concentrated in Amicon 50K NMWL filters (MilliporeSigma) and stored for 48 to 72 hours at 4°C. Samples were centrifuged, and VWF containing supernatant was stored at -80°C. Final concentrations of AnVWF were determined by ELISA.

Ancestral FVIII expression and purification: Production clones for An63-, 70-, 84-, and 88-FVIII were generated in HEK293-H cells. Monoclonal populations were isolated under Geneticin selection and expanded as suspension cultures in FreeStyle™ 293 media without serum. Adherent populations were cultured in triple flasks in SFM4HEK293 medium. An101-FVIII protein was collected from a polyclonal population of HEK293-H cells transduced with lentivirus encoding the An101-FVIII cDNA. Lentiviral vectors were produced as previously described (135). All ancestral FVIII was purified initially with the 'VIIISelect' resin according to the manufacturer's instructions and followed by a 5-mL SPHP cation exchange HiTrap column. FVIII was eluted with a NaCl gradient as previously described (113).

RNA isolation and reverse transcription polymerase chain reaction: RNA was purified by using the RNeasy kit (Qiagen) according to manufacturer's instructions. Cell counts, RNA collections, and VWF determinations were performed simultaneously. Transcript analysis was conducted as previously described (129) using universal VWF primers within the D3 domain (forward 5'-GAG AACGGCTACGAGTGCG, reverse 5'-CTGGAGGACAGTGTGCGTG). Transcripts per cell were calculated by using an estimated 25 pg RNA per HEK116 cell.

Multimer analysis and ADAMTS13 cleavage: Multimer analysis was performed by using 60 ng VWF as previously reported (136-138). Multimers were visualized by chemiluminescent imaging of antibody P0226 (Dako). AnVWF digestion using recombinant human ADAMTS13 was performed as previously described (139; 140). Densitometry was analyzed by using ImageJ

software (National Institutes of Health) and calculated as percentage of multimers remaining after ADAMTS13 cleavage.

In vivo AnVWF production and FVIII rescue: All animal procedures were in accordance with the Canadian Council on Animal Care guidelines and approved by the Queen's University Animal Care Committee. Linear DNA, 10 µg, in TransIT-EE buffer (Mirus Bio) was hydrodynamically injected into 8- to 12-week-old VWF^{-/-} female mice. Retro-orbital samples were collected at 24 hours' postinjection. VWF antigen was determined by ELISA and FVIII activity by chromogenic SP4 assay using a WHO-defined human standard plasma. Molar concentrations of FVIII were determined by using a specific-activity of 4000 IU/mg for murine FVIII (114).

Clearance: AnVWF (20 µg) was injected via tail vein in 8- to 12-week-old male VWF^{-/-} mice. Retro-orbital sampling was performed from alternate eyes at 2 time points. VWF determination was performed according to ELISA as described earlier.

VWF:GP1bM and VWF:RCo activity assays: The Innovance VWF:Ac GP1bM assay was performed on a BCS-XP coagulometer (Siemens) according to the manufacturer's instructions and compared with WHO-defined human standard plasma. Specific activity was defined as the ratio of GP1bM activity to VWF antigen concentration in international units per milligram. Ristocetin cofactor (VWF:RCo) activity was measured via platelet aggregometry using washed normal platelets and compared with a WHO-defined standard plasma, as previously described (141).

Surface plasmon resonance: Studies were performed at 25°C using a Biacore T200 (GE Healthcare) and CM5 sensor chips (GE Healthcare) equilibrated with immobilization running buffer (20 mM HEPES, 150 mM NaCl, 0.02% Tween 20, pH 7.4). Recombinant anti-Flag antibody (M2, #F1804; Sigma) was immobilized using the Amine Coupling Kit (GE Healthcare) according to the manufacturer's instructions to a range of ~2000 to ~7000 resonance units

(RU). VWF-Flag variants were diluted to 50 to 100 $\mu\text{g}/\text{mL}$ and injected at 2 $\mu\text{L}/\text{min}$. VWF immobilization level was between ~ 200 and ~ 500 RU. Flow cell 1 was immobilized with the anti-Flag antibody alone as a reference for systematic noise and instrument drift.

Surface plasmon resonance measurements were performed at 30 $\mu\text{L}/\text{min}$ using single-cycle kinetics with 5 sequential 120-second injections at increasing concentrations of each FVIII variant (0.55-45 nM) (142) followed by a final 300-second dissociation. The surface was regenerated with 20 mM HEPES, 600 mM NaCl, 350 mM CaCl_2 , and 0.02% Tween 20 at pH 7.4 for 2 minutes. Measurements included zero-concentration cycles of analyte before sampling for subtraction (143). All FVIII samples were measured in triplicates per run, and each VWF variant was immobilized to 3 different chips at different flow cell positions. Data were analyzed by using Biacore T200 Evaluation Software 3.0 (GE Healthcare). Sensorgrams were fitted globally by using the steady-state model after double reference subtraction.

Collagen Binding: Collagen binding assays were performed as previously reported (144; 145) using human collagen type I and III. Briefly, maleic anhydride activated plates were coated with 10 $\mu\text{g}/\text{mL}$ human collagen (95% type I, 5% type III) for 2 hours at room temperature. Plates were washed three times with 75 mM NaCl, 5 mM Tris, 0.1% Tween-20, pH 9.1 and blocked overnight in phosphate buffered saline with 5% BSA. Recombinant AnVWF protein curves were diluted in PBS with 1% BSA, 0.1% Tween-20 and applied to the collagen coated plates. VWF was detected with DAKO anti-human VWF polyclonal antibody P0226 at a dilution of 1:1000 in diluent. Absorbance of HRP mediated catalysis of OPD was conducted at 492 nm. Ancestral VWF curves were compared to human recombinant VWF, and reference human plasma was also included.

2.4 Results

Phylogenetic tree and sequence reconstruction: ASR has enabled the engineering of complex proteins when rational design and directed evolution methods have been prohibitive. Using 59 extant VWF sequences (numbered 1-59), we performed ASR as described previously (133) to generate 53 ancestral sequences (numbered 60-112) (**Figure 2.1A; Supplemental Figure S2.1A-B**). In addition, extant sequences for coagulation factors V, VII, VIII, IX, and X were applied to ASR analysis. Consistent with analyses based on the mirrortree method, (146) independent reconstruction of these protein sequences resulted in identical phylogenetic trees, providing a singular representation for the molecular evolution of multiple procoagulant proteins (118). These phylogenetic data suggest that the components of the coagulation cascade recapitulate species evolution and coevolve throughout mammalian diversification.

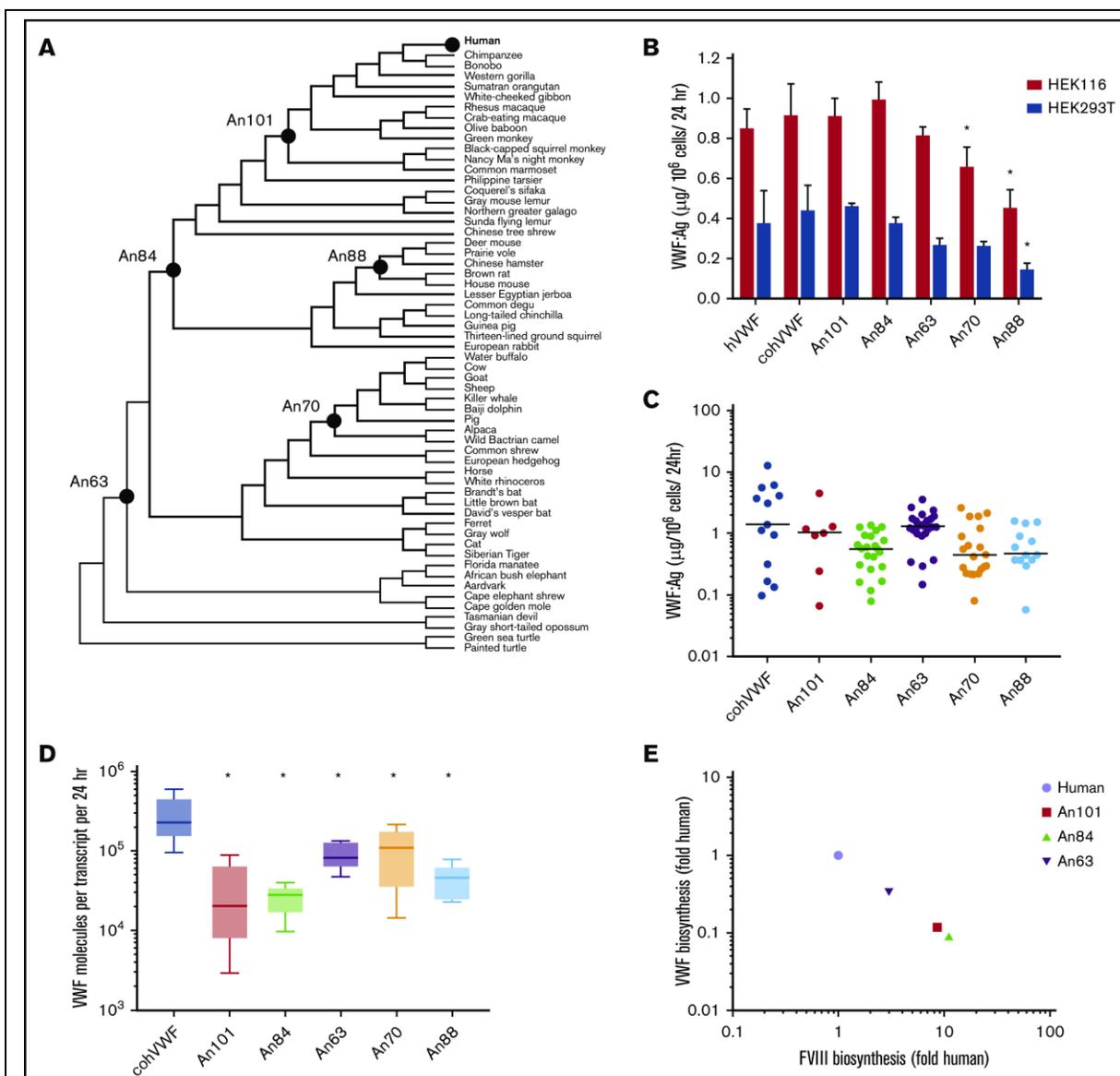


Figure 2.1: AnVWF biosynthetic efficiency correlates with ancestral FVIII. (A)

Phylogenetic tree depicting ancestral sequence reconstruction of 59 extant VWF sequences. VWF sequences selected for de novo synthesis are labeled as closed circles. (B) Production rates of AnVWF protein following transient transfection of HEK116 (red bars) and HEK293T (blue bars) cells. Error bars represent SD (n = 4). *P < .05 (Dunnett's one-way ANOVA). (C) Stable, monoclonal producer cell lines were produced from HEK116 cells, and VWF biosynthesis was measured via ELISA. Horizontal bars represent median values; n = 13, 7, 23, 29, 19, and 14 for cohVWF, An101, An84, An63, An70, and An88, respectively. No significant differences were detected compared with cohVWF (Dunn's one-way ANOVA). (D) Transcript analysis of monoclonal producer cell lines at steady state was performed by one-step reverse transcription polymerase chain reaction. VWF antigen was normalized to transcript levels to determine biosynthetic efficiency. Box plot reveals range, 25th and 75th percentiles, and median values; n = 6 for An70 and An88 and n = 5 for others. All AnVWF constructs displayed significantly reduced (*P < .05) biosynthetic efficiency compared with cohVWF (Dunnett's one-way ANOVA). (E) Correlation of VWF and FVIII biosynthetic efficiency was compared as fold change relative to human. FVIII biosynthetic efficiency was reported previously. Linear regression equation (not shown) is $y = -1.002x + 6.4 \times 10^{-5}$ with $r^2 = 1$.

We synthesized 5 AnVWF complementary DNAs for biochemical characterization, chosen based on their evolutionary position between the common mammalian ancestor and the most extensively characterized extant VWF molecules. These AnVWF sequences, termed An101-, An84-, An63-, An70-, and An88-VWF, range in total amino acid replacements from 81 to 364, representing 97% to 87% identity to human VWF (**Table 2.1**). All cysteine residues are conserved in ancestral sequences.

Table 2.1: Ancestral VWF sequence identity and classification

VWF	Amino acid replacements	% Human identity	Predicted age (mya)	Classification
cohVWF	0	100	—	Primate
An101	81	97.1	43.2	Primate: Simiiformes
An84	206	92.7	89.8	Euarchontoglires
An63	245	91.3	105.5	Eutheria
An70	307	89.1	64.2	Ungulate: Artiodactyla
An88	364	87.1	32.7	Rodentia: Muroidea

mya, million years ago.

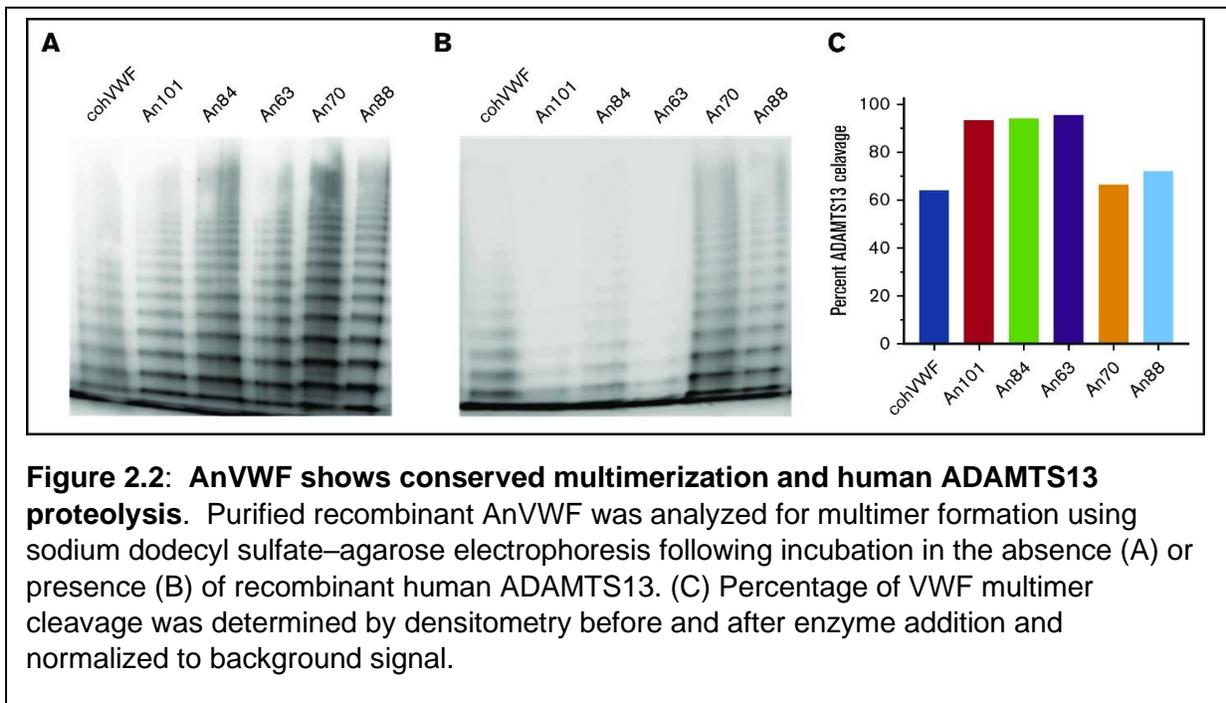
Biosynthesis, multimerization, and ADAMTS13 cleavage: Using human cell lines clinically employed in heterologous recombinant production of VWF and FVIII, we investigated the biosynthetic production of AnVWF. VWF antigen levels were determined via ELISA by using cross-reactive polyclonal anti-human VWF antibodies previously validated for measuring VWF

orthologs, including murine VWF (147). Parallel standard curves were obtained for all ancestral constructs, allowing concentration determinations from standard human plasma (**supplemental Figure S2.2A-B**). Transient transfection of either HEK116 or HEK293T cells resulted in equivalent production of the codon-optimized human (coh) VWF compared with wild type (**Figure 2.1B**). Compared with cohVWF, An70- and An88-VWF exhibited significantly reduced production in one or both cell lines ($P = .0284$ and $.0003$, respectively, for HEK116 cells; $P = .0024$ for An88-VWF in 293T cells; one-way analysis of variance [ANOVA]). Analysis of stable monoclonal cell lines at steady state, however, revealed no significant difference in VWF production despite lower median values (Dunn's one-way ANOVA) (**Figure 2.1C**). To determine if observed VWF production is transcript dependent, messenger RNA analysis was performed at the time of antigen determination. Biosynthetic efficiencies of all AnVWF molecules were significantly reduced compared with cohVWF, suggesting that translation and/or secretion are rate-limiting (Dunn's ANOVA) (**Figure 2.1D**).

Comparing biosynthesis of AnVWF vs previously reported ancestral FVIII sequences (129), we observed an inverse correlation throughout primate evolution (**Figure 2.1E**), suggesting a conserved balance of protein production. These changes are not sequential, exhibiting temporally random but reciprocal changes in the partner protein's biosynthesis.

Multimer analysis confirmed the presence of high molecular weight (HMW) multimers equivalent to cohVWF (**Figure 2.2A**). The formation of HMW multimers in the oldest ancestral sequence, An63-VWF, suggests that all mammalian VWF sequences form HMW multimers. These data are in agreement with the observation of HMW multimers in zebrafish (148) but not hagfish (121), suggesting trait emergence between 435 and 615 million years ago.

In circulation, the metalloprotease ADAMTS13 cleaves HMW multimers into smaller multimers and monomers. Throughout VWF evolution, however, ADAMTS13 proteolysis has undergone species-specific diversification abrogating cross-compatibility in some instances (131) but not others (149). Interestingly, An70- and An88-VWF within the carnivore/ungulate and rodent lineages, respectively, exhibited ADAMTS13-mediated cleavage comparable to human VWF (**Figure 2.2B-C**), suggesting that substrate diversification occurred more recently. In contrast, early ancestral and primate lineage VWF molecules (An63-, An84-, and An101-VWF) all exhibited increased and near complete cleavage of HMW multimers (**Figure 2.2C**).



AnVWF specific-activity by VWF:GPIbM and VWF:RCo: VWF activates the GPIb/IX/V complex on platelets through the VWF-A1 domain. To measure VWF activity in vitro, ristocetin is added to uncoil the mechanosensory domain of VWF, independent of shear stress. However, ristocetin does not bind all mammalian VWF molecules. To test the activity of AnVWF molecules in a ristocetin-independent manner, the GP1bM assay was used. An70- and An88-VWF displayed 1.5- to 2-fold increased specific-activity over cohVWF. The VWF molecules An101-, An84-, and An63-VWF exhibit conserved but reduced platelet activation. These differences are

likely due to amino acid replacements surrounding and within the A1 domain; however, there are no unique residue substitutions in the A1 domain of An101-VWF outside of the autoinhibitory modules (AIMs) (150). Aggregation of washed human platelets in the presence of ristocetin revealed specific-activities like GPIbM determinations, with the exception of An88-VWF. Consistent with the observation that ristocetin does not uncoil mouse or rat VWF (132), An88-VWF within the rodent lineage displayed a 4.6-fold discrepancy in activity between these assays. Ristocetin does not uncoil canine or porcine VWF (130; 151) but is able to uncoil VWF from sheep and goats (132). Considering that ristocetin activation was observed with An70-VWF (preceding pig and sheep but not dog) (**Figure 2.1A**), our data suggest that the ristocetin binding sequence may have been lost in two or more evolutionary branches within the past 78 million years.

VWF and FVIII are critical components of primary and secondary hemostasis, respectively. A recent, large epidemiologic study highlights that the predominant association between the coagulation proteins and increased risk of venous thrombotic disease is a quantitative variance of FVIII and VWF levels (152). The critical balance between hemostasis and thrombosis has likely driven a strong selective pressure on the biological activity of the coagulation proteins. Strikingly, FVIII and VWF specific-activities are inversely correlated throughout evolution (**Figure 2.3**). Although this analysis of AnVWF activity is restricted to human GPIb/platelets, our data suggest that primate evolution may have promoted primary hemostatic coverage and reduced the effect of secondary hemostatic amplification, perhaps as a modulator of thrombosis. However, a deeper analysis of associated genes within hemostasis is necessary to adequately evaluate the global effect on the hemostatic process.

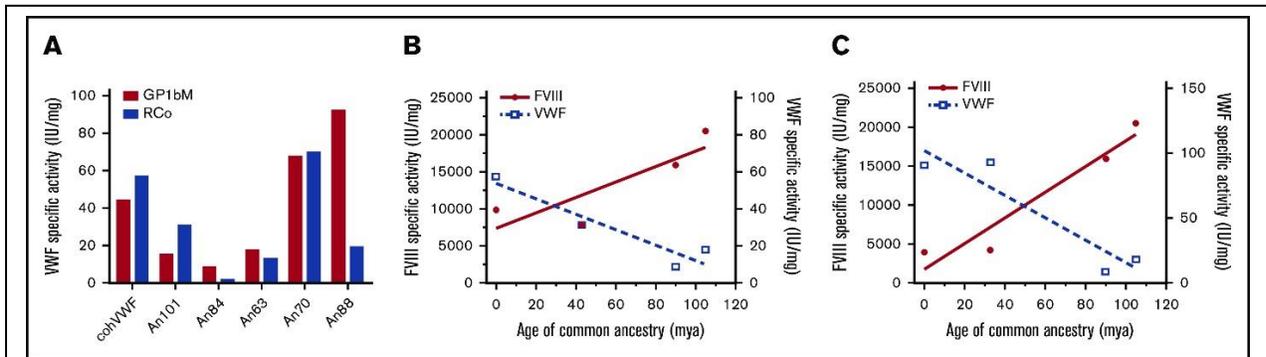


Figure 2.3: AnVWF specific activity is reciprocal to ancestral FVIII. (A) Activities of AnVWF molecules were determined by VWF:GPIbM (red bars) and VWF:RCo (blue bars) and normalized to antigen levels to generate specific-activities. VWF:RCo data are presented as the mean of 2 purified AnVWF lots. Specific-activities of VWF (open squares) and FVIII (closed circles) from the primate (B) and rodent (C) lineages were plotted in relation to their predicted age of ancestry (million years ago [mya]) as determined by TimeTree. Specific-activities of ancestral FVIII were determined by one-stage coagulation activity normalized to absorbance after extinction coefficient correction. Specific-activities of orthologous FVIII protein was determined previously.

Coordinated nonsynonymous mutation of VWF and FVIII: To identify the principal driver of coevolution, we analyzed the emergence of nonsynonymous mutations in ancestral VWF and ancestral FVIII sequences. Temporal analysis reveals a highly reciprocal pattern of mutation across all lineages (**Figures 2.4A-C**). The greatest number of substitutions in primate and ungulate VWF and FVIII was observed between 67 and 43 million years ago (**Figure 2.4A**). Rodent VWF and FVIII, however, have undergone continuous change over the past 100 million years (**Figure 2.4C**). Unlike the primate and ungulate lineages, which maintain a low

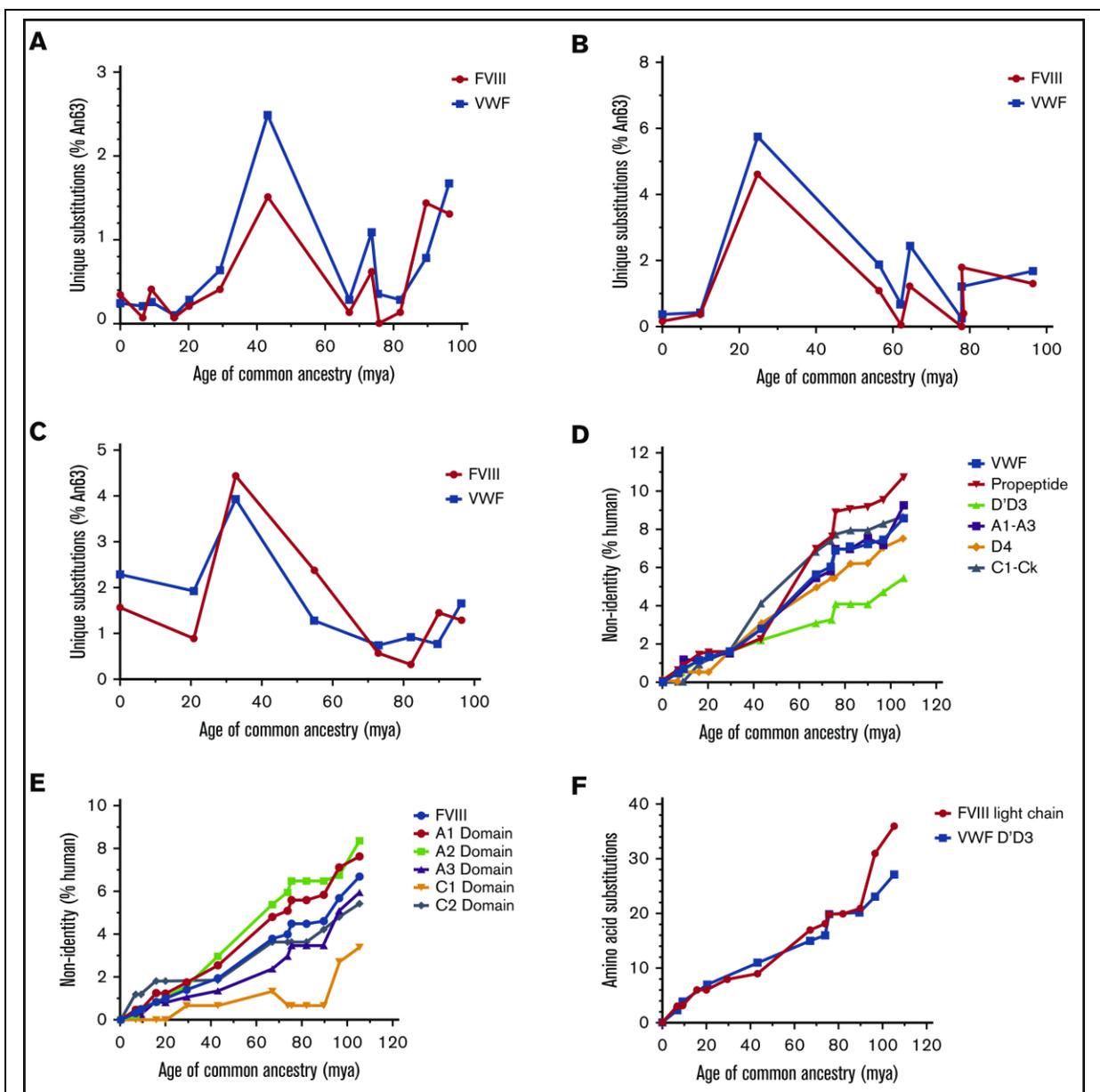


Figure 2.4: Mammalian VWF and FVIII evolved under analogous selective pressure.

Temporal emergence of unique amino acid substitutions diverging from the common ancestral sequence An63 are shown for primate (A), ungulate (B), and rodent (C) lineages. The number of unique residue differences is shown as the percentage of all amino acids. Total accumulation of mutations within specific subdomains of VWF (D) or FVIII (E) is shown as percent non-human relative to the size of each domain across primate evolution. Human VWF and FVIII define 0% nonidentity. (F) The number of amino acid substitutions within FVIII light chain or VWF D'D3 domains is shown for the primate lineage. mya, million years ago.

percentage of new mutations during the past 10 million years (<0.4%), rodent FVIII and VWF exhibit increased substitutions of >1% and 2%, respectively, suggesting that different selective

pressures and/or population sizes influence rodent coagulation factors.

Throughout primate evolution, the protein domains that experienced the greatest change are the VWF propeptide and the A1 and A2 domains of FVIII (**Figure 2.4D-E**). In the case of FVIII, this may provide an explanation for the differences in biosynthetic efficiencies and specific-activities previously observed across species. Conversely, the most conserved protein domains are the D'D3 domain of VWF and the C1 domain of FVIII, the latter of which has remained unchanged in primates for roughly 20 million years. The association of VWF and FVIII has been mapped to the D'D3 domain of VWF and the acidic a3 domain and C1/C2 domains of FVIII, respectively (153-155). Within rodent and ungulate lineages, the D'D3 VWF domains and the FVIII C1 domain are also the least disparate domains (**supplemental Figure S2.3A-D**). Notably, the rate of amino acid replacement throughout the primate VWF D'D3 domains parallels the FVIII light chain (**Figure 2.4F**).

FVIII-VWF affinity determination: We next investigated the evolution of the VWF-FVIII complex of each AnVWF molecule binding to each ancestral FVIII molecule (**supplemental Figures S2.5 and S2.6; supplemental Table S2.1**). Consistent with the high sequence conservation within the binding domains, each FVIII molecule bound VWF with low nanomolar to picomolar steady-state affinities (**Table 2.2**). We hypothesized that the greatest affinity would be observed between proteins of the same phylogenetic position (e.g., human-human, An63-An63); however, this was only observed for the An88 FVIII-VWF pair. The strongest binding was observed between An70-VWF and human FVIII with a steady-state KD of 680 pM, and the weakest binding was observed between An63-VWF and human FVIII at 5.47 nM. Of the 36 protein interactions tested, An70-VWF exhibited 4 of the strongest affinity constants. An84- and An63-VWF together showed the 8 weakest affinity constants observed. An70- and An88-VWF exhibited the only significantly reduced KD values compared with the human-human KD (**supplemental Table S2.2**).

Table 2.2: SPR steady-state K_D determinations (nM)

	coh-VWF	An101-VWF	An84-VWF	An63-VWF	An70-VWF	An88-VWF
Human FVIII	1.84 ± 0.19	2.35 ± 0.31	4.05 ± 0.14 ****	5.47 ± 0.41 ****	0.68 ± 0.15 **	1.14 ± 0.06
An101-FVIII	1.68 ± 0.25	2.06 ± 0.28	3.27 ± 0.30 ****	4.25 ± 0.30 ****	0.88 ± 0.23 *	1.13 ± 0.11
An84-FVIII	1.55 ± 0.09	1.87 ± 0.21	2.56 ± 0.23	2.90 ± 0.08 **	1.08 ± 0.30	1.02 ± 0.22
An63-FVIII	2.15 ± 0.28	2.62 ± 0.39	4.49 ± 0.60 ****	3.80 ± 0.29 ****	0.95 ± 0.10 *	1.71 ± 0.10
An70-FVIII	2.17 ± 0.29	2.67 ± 0.54	5.17 ± 0.33 ****	4.03 ± 0.39 ****	0.98 ± 0.02 *	1.84 ± 0.13
An88-FVIII	0.96 ± 0.08 *	1.01 ± 0.06	1.22 ± 0.27	1.17 ± 0.30	1.28 ± 0.29	0.77 ± 0.18 **

*P < .05, **P < .005, ****P < .0001.

In addition to steady-state determinations, we performed kinetic affinity measurements but observed poor fits to a 1:1 binding model for An63-, An84-, or An101-VWF when complexed with the majority of FVIII molecules tested (**supplemental Table S2.1; supplemental Figures S2.5 and S2.6**). An88-FVIII exhibited strong 1:1 binding fits with 5 of the 6 VWF molecules tested, and both An70-VWF and An88-FVIII displayed binding characteristics that are predicted to be pharmaceutically advantageous.

AnVWF rescues FVIII in VWF^{-/-} mice: To determine if liver-derived AnVWF can functionally restore circulating levels of endogenous mouse FVIII (mFVIII) in vivo, linear DNA containing AnVWF sequences was hydrodynamically injected into VWF-deficient mice. VWF antigen levels ranged from 10% to 52% normal levels (**Figure 2.5A**). Production of An63- and An88-VWF were significantly lower than cohVWF production ($P = .0357$ and $.013$; Dunnett's ANOVA). Endogenous mFVIII activity levels were subsequently increased from 13% in mice receiving saline, to a range of 49% to 90% normal. Mice injected with An88-VWF plasmid had significantly increased rescue of mFVIII despite significantly lower VWF concentrations ($P = .0412$; Dunnett's ANOVA).

At these nonsaturating VWF concentrations, the molar ratio of cohVWF:mFVIII was calculated at roughly 29:1. An88-VWF is a rodent lineage VWF sequence and shares 94% amino acid identity to mouse VWF compared with 83% for human VWF. The concentrations of An88-VWF to mFVIII approached but did not reach equimolar levels (**Figure 2.5B**). Considering

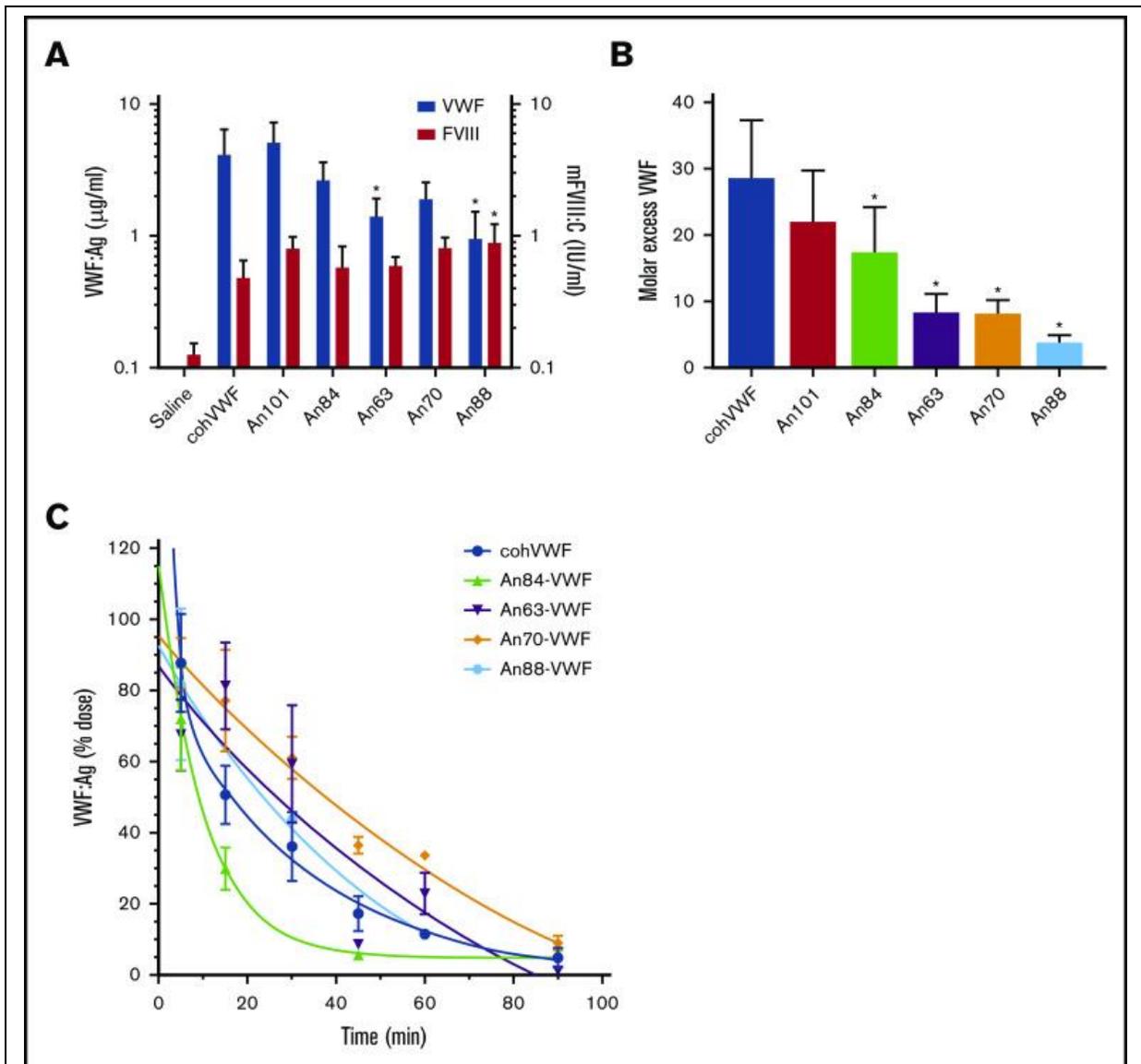


Figure 2.5: AnVWF rescues murine FVIII and exhibits reduced clearance in vivo. (A) VWF antigen (blue bars) and FVIII (red bars) activity was determined by ELISA and chromogenic assay, respectively, following hydrodynamic infusion of AnVWF plasmid DNA ($n = 4$) or saline ($n = 9$). Mice administered An88-VWF plasmid possessed significantly increased (*) FVIII activity ($P = .0412$; Dunnett's ANOVA) despite reduced VWF present in circulation ($P = .013$; Dunnett's ANOVA). (B) Molar ratios of VWF and FVIII were calculated. Molar excess of An84-, An63-, An70-, and An88-VWF was significantly reduced compared with cohVWF (* $P < .05$; Dunnett's ANOVA). (C) Male VWF-deficient mice were administered 20 μg recombinant AnVWF. VWF antigen levels were determined by ELISA normalized to the initial dose via 2-phase decay, $n = 3$ per time point.

the reduced biosynthesis and specific-activity of murine FVIII and enhanced half-life of murine

FVIIIa (114), the strong affinity of An88-VWF for An88-FVIII and 1:1 binding fit support the

hypothesis that these proteins coevolved to most efficiently maintain FVIII levels in circulation. These data support further investigation of An88-FVIII in other models of VWF deficiency to assess if the FVIII/VWF stoichiometry is consistently more favorable.

Reduced clearance of recombinant AnVWF: Clearance of VWF is mediated by the asialoglycoprotein receptor (ASGPR), lipoprotein receptor–related protein-1 (LRP1), and siglec-5 on the surface of macrophages and hepatocytes, as well as CLEC4M and stabilin-2 on sinusoidal endothelial cells. Although clearance of VWF is largely mediated by N- and O-linked glycosylation, amino acid replacements such as the R1205H Vicenza variant also affect clearance rates. These observed replacements occur throughout the D, A, and C domains of VWF, with the greatest concentration occurring in the D'D3 and A1 domains. Pharmacokinetic analysis of cohVWF and An84-, An63-, An70-, and An88-VWF in VWF-deficient mice found terminal half-lives of 23.3, 7.0, 59.8, 70.4, and 35.4 minutes, respectively (**Figure 2.5C**). Therefore, An63- and An70-VWF show a 2.6- and 3.0-fold prolonged half-life after intravenous injection.

2.5 Discussion

ASR has enabled studies of molecular evolution of numerous gene products, including hormone receptors (156; 157), metabolic enzymes (158), oxidases (159), and coagulation proteins (118; 129), as well as phenylalanine/tyrosine ammonia-lyases, for pharmaceutical development (160). Investigations of resurrected proteins have revealed insights into mechanisms of evolution and epistasis, emerging complexity within enzyme systems, and global trends of diminishing thermostability and promiscuity of enzymes (160; 161). Furthermore, this expanding field enables identification of sequence determinants responsible for novel biochemical functions for therapeutic protein engineering due to the limited number of replacements between branch points. In addition to direct identification of novel VWF sequences with therapeutically advantageous properties, this is the first known report of

coevolution of interdependent mammalian therapeutic proteins supported by both phylogenetic analysis and biochemical characterization.

Investigating the molecular evolution of VWF proteins revealed an increase in biosynthetic efficiency coupled with increased specific-activity. These data suggest that primate evolution may have promoted primary hemostatic coverage and reduced the effect of secondary hemostatic amplification, at least from the perspective of platelet activation, as a potential modulator of thrombosis. Within all lineages, the temporal emergence of novel substitutions in either protein is mirrored by substitutions in the other, with most mutations occurring outside of the VWF-FVIII binding sites. The increased emergence of substitutions coincides with the Cretaceous–Paleogene extinction ~66 million years ago. Although primate species, in contrast to ungulates, are not suggested to have greatly diversified during the Cretaceous–Paleogene extinction (162; 163), our evidence suggests that primate hemostatic proteins underwent increased diversification at this time, possibly supporting primate development during the following Eocene and Oligocene epochs.

In our investigation of the VWF-FVIII complex, we observed a heavily conserved picomolar to nanomolar affinity across all FVIII and VWF combinations. This finding is consistent with the observations that the D'D3 domain of VWF and the C1 domain of FVIII are the most conserved domains across all lineages (supplemental Figure 3A-D), suggesting a strong negative selection on mutations abrogating their association. Interestingly, early VWF molecules An63, An84, and An101 exhibited poor 1:1 kinetic binding to FVIII. Instead, these VWF molecules more appropriately fit a bivalent model of binding, perhaps suggesting changes in the FVIII-VWF multimeric complex and stoichiometry. Although each monomer of VWF contains a binding site for FVIII, the observed stoichiometry of VWF to FVIII in circulation is 50:1 (164). In our study of FVIII rescue in VWF-deficient mice (**Figure 2.5A-B**), we observed

significant decreases in the molar excess of AnVWF, supporting the presence of an additional binding site or altered stoichiometry.

The domains of greatest substitution, and likely under positive selection, are the propeptide and A domains of VWF and the A1 and A2 domains comprising the heavy chain of FVIII. Characterization of FVIII through ortholog scanning and alanine mutagenesis has identified sequences within the heavy chain responsible for biosynthesis, secretion, and engagement of the unfolded protein response (104; 165-167). Similarly, gene therapy studies codelivering the heavy and light chain of FVIII in trans show a stoichiometric imbalance favoring light chain synthesis (168). Ultimately, the key biochemical differences of FVIII across species are mediated primarily by amino acid residues within the heavy chain domains (103; 112-114).

A closer look at the VWF amino acid substitutions within A1 domain and AIMs of these ancestral sequences reveals a finite number of substitutions, which may account for increased specific-activity. Substitutions unique to An70- or An88-VWF within the A1 domain, residues 1272-1458, include E1292D, G1330A, K1332R, Q1353E, I1380V, M1393L, I1410T and D1333A, A1381V, and V1439L, respectively. Interestingly, none of these mutations is associated with VWD type 2B or any other subtype. Patients with type 2M and 2A were found to have D1302G or R1308H mutations (169); however, these substitutions are observed in An63-, An84-, An70-, and An88-VWF sequences. Similarly, V1314F or V1314D are identified in type 2A and 2B diagnoses (170), although An63-, An84-, An70-, and An88-VWF have an isoleucine substitution at this residue. These data highlight the importance of epistasis considerations in protein engineering and pathology. Within the N-terminal AIM, unique substitutions Q1238R, V1244E, A1250G, and Y1271F or P1254S and D1269N are observed in An70- or An88-VWF, respectively. Additional substitutions D1249E, L1257P, I1262T, and S1263P are observed in all ancestral sequences. The C-terminal AIM contains unique substitutions A1464V, P1470R, and D1472L in An70 and M1473V found in both An70- and An88-VWF sequences. Of interest,

D1472H, a variant that affects ristocetin-based activity measurements but not in vivo activity (171), is found in An101-VWF. At this residue, D1472Q is observed in An84-, An63-, and An88-VWF. Other residues of substitution within the CAIM include P1466A and L1469M/Q. None of these substitutions is reported in VWD pathologies but may influence the activity of in vitro VWF assays.

Within the D' and D3 domains of VWF, several substitutions are observed that may affect the FVIII-VWF complex affinity. Within the D' domain of An70- and An88-VWF, the molecules showing the strongest affinity for FVIII, these substitutions include M799V, M801T, R820K, and K833R, with only K833 residing within a structured β -sheet. Within the D3 domains, An70- and An88-VWF harbor I1001V and S1040V substitutions. As supported by our binding data, none of these mutations is associated with patient type 2N diagnoses.

As evidence of the pharmaceutical development potential of ASR, reconstruction of VWF sequences in this study resulted in proteins with enhanced and potentially advantageous biochemical characteristics. Specifically, An70-VWF exhibits increased platelet activation, increased affinity to FVIII, and a significantly prolonged half-life in mice, as well as a 3-fold improved binding affinity to human collagen (**supplemental Figure S2.7**). These characteristics are critical for pharmaceutical development of next-generation biologics and gene therapies for hemophilia A and VWD. However, despite high identity to human sequences, these proteins are feared to harbor immunogenic properties due to non-human epitopes (115; 172-174). Currently, investigations of coagulant protein immunogenicity are limited by existing animal models; however, in silico predictions of B- and T-cell epitopes are increasing in power and accuracy. An analysis of T-cell epitopes within a human-porcine FVIII chimeric protein, termed ET3', revealed no observable difference from the human FVIII epitope landscape (107), and a similar finding was observed for ancestral FVIII proteins for B-cell epitopes (129).

2.6: Acknowledgements

This work was partially supported by CIHR Foundation (grant FDN154285) (D.L.); National Institutes of Health, National Heart, Lung, and Blood Institute (grant HL137128) (C.B.D.); National Institutes of Health, National Institute of Arthritis, Musculoskeletal and Skin Diseases Research (grant R01AR069137) (E.A.G.); Department of Defense (grant MURI W911NF-16-1-0372) (E.A.G.); and Human Frontier Science Program (grant RGP0041) (E.A.G.).

Conflict-of-interest disclosure: P.M.Z., C.E.R., H.T.S., C.B.D., E.A.G., and D.L. are inventors on patent applications describing ancestral variants of coagulation factors. C.B.D. and H.T.S. are cofounders of Expression Therapeutics and own equity in the company. Expression Therapeutics licenses the intellectual property associated with ancestral coagulation factors. The terms of this arrangement have been reviewed and approved by Emory University and Georgia Institute of Technology in accordance with their conflicts of interest policies. The remaining authors declare no competing financial interests.

2.7 Supplementary Figures and Tables

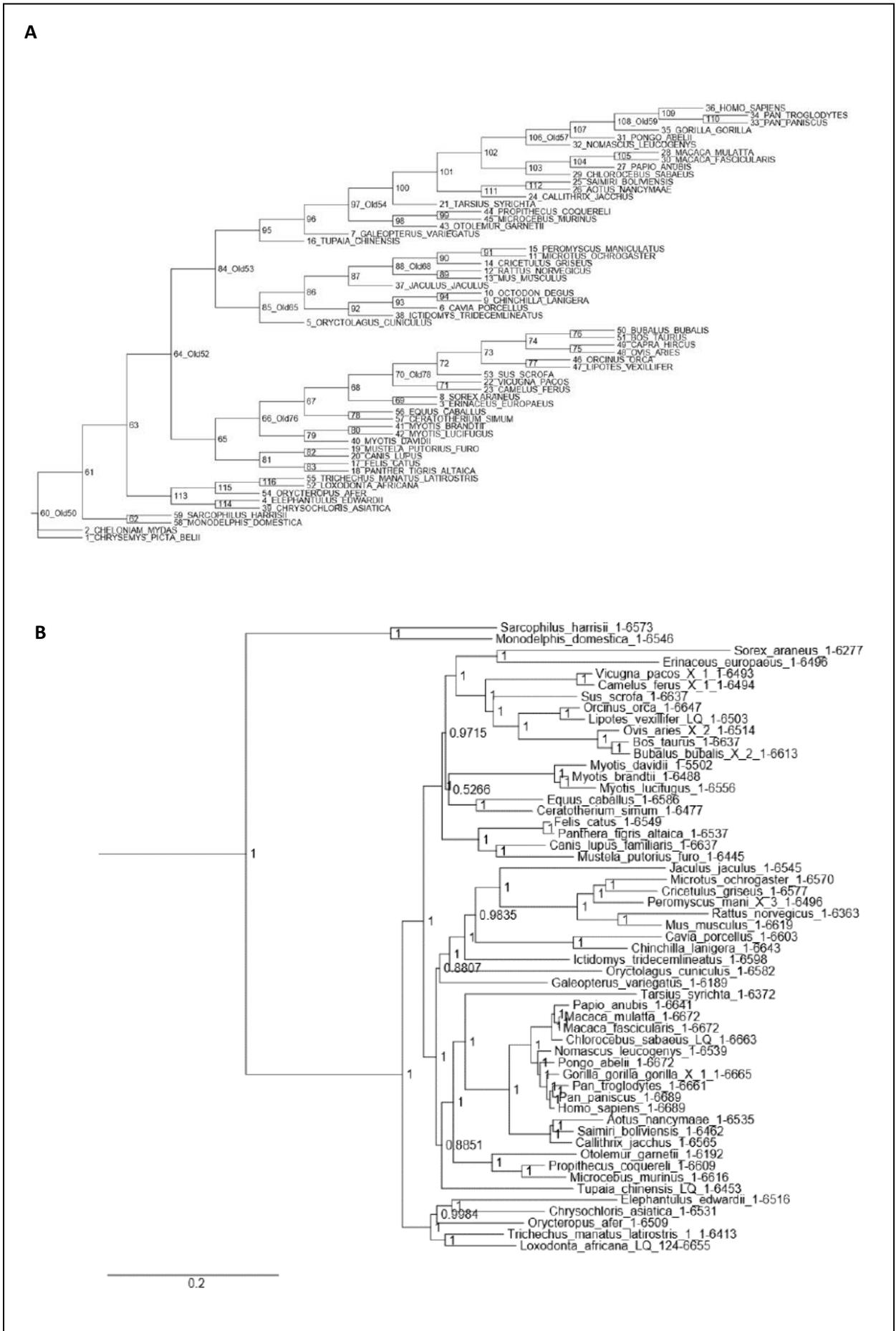
Supplemental Table 1. Kinetic SPR Affinity determinations with 1:1 binding model (nM)

K_D (nM)	coh-VWF	An101-VWF	An84-VWF	An63-VWF	An70-VWF	An88-VWF
hFVIII	0.48 (\pm 0.06)	0.45 (\pm 0.02)	1.4 (\pm 0.53)	0.88 (\pm 0.24)	0.27 (\pm 0.03)	0.42 (\pm 0.08)
An101-FVIII	0.49 (\pm 0.08)	0.49 (\pm 0.03)	1.2 (\pm 0.36)	0.83 (\pm 0.19)	0.30 (\pm 0.02)	0.44 (\pm 0.07)
An84-FVIII	0.59 (\pm 0.07)	0.56 (\pm 0.07)	1.1 (\pm 0.22)	0.60 (\pm 0.1)	0.38 (\pm 0.01)	0.53 (\pm 0.07)
An63-FVIII	0.63 (\pm 0.07)	0.57 (\pm 0.08)	1.5 (\pm 0.29)	0.88 (\pm 0.21)	0.35 (\pm 0.03)	0.64 (\pm 0.06)
An70-FVIII	0.72 (\pm 0.06)	0.52 (\pm 0.1)	2.1 (\pm 0.58)	0.85 (\pm 0.14)	0.41 (\pm 0.03)	0.66 (\pm 0.05)
An88-FVIII	0.37 (\pm 0.07)	0.54 (\pm 0.11)	0.65 (\pm 0.12)	0.57 (\pm 0.03)	0.85 (\pm 0.09)	0.61 (\pm 0.03)
	n=6	n=3	n=4	n=4	n=3	n=3

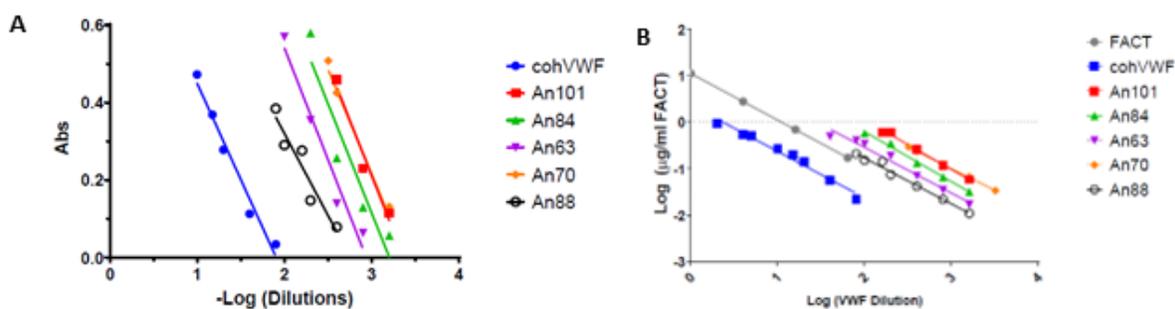
	Classification of fit quality 1; good fit
	Classification of fit quality 2; weak fit
	Classification of fit quality 3; bad fit

Supplemental Table S2. P-values of steady state affinity determinations (Dunnett's ANOVA)

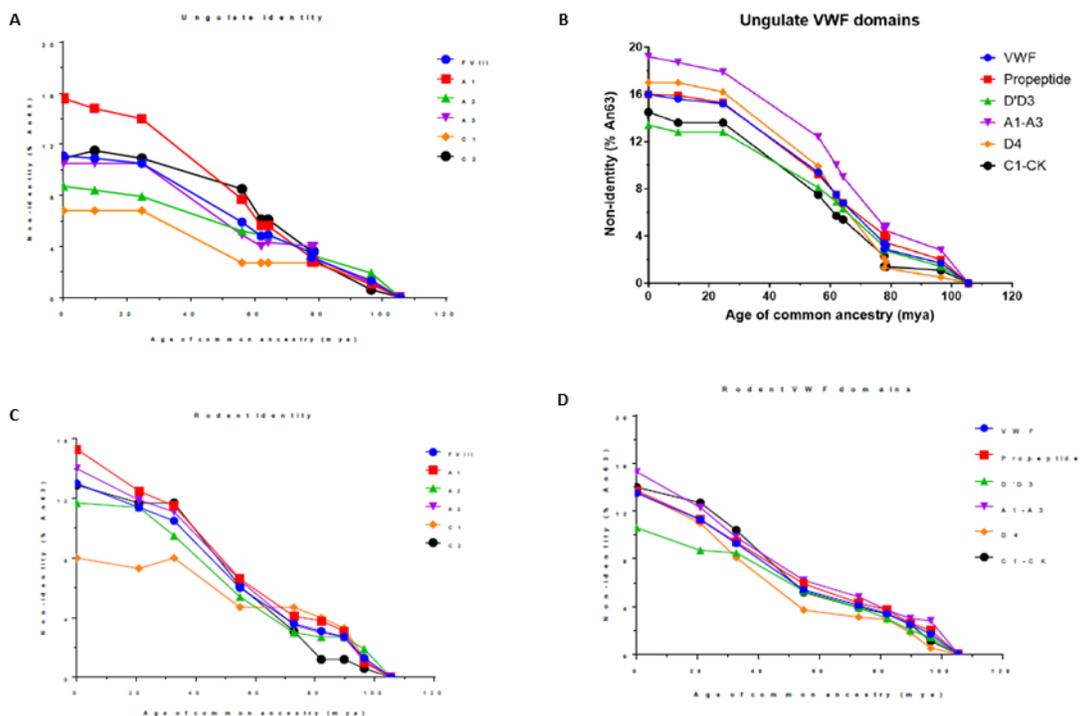
	coh-VWF	An101-VWF	An84-VWF	An63-VWF	An70-VWF	An88-VWF
hFVIII		0.5871	< 0.0001	< 0.0001	0.0013	0.1707
An101-FVIII	0.9991	0.9988	< 0.0001	< 0.0001	0.0154	0.1602
An84-FVIII	0.993	0.9999	0.1502	0.0047	0.112	0.0647
An63-FVIII	0.9855	0.0883	< 0.0001	< 0.0001	0.0331	0.9994
An70-FVIII	0.9787	0.0539	< 0.0001	< 0.0001	0.0422	>0.9999
An88-FVIII	0.0354	0.0575	0.3199	0.2186	0.4592	0.0045



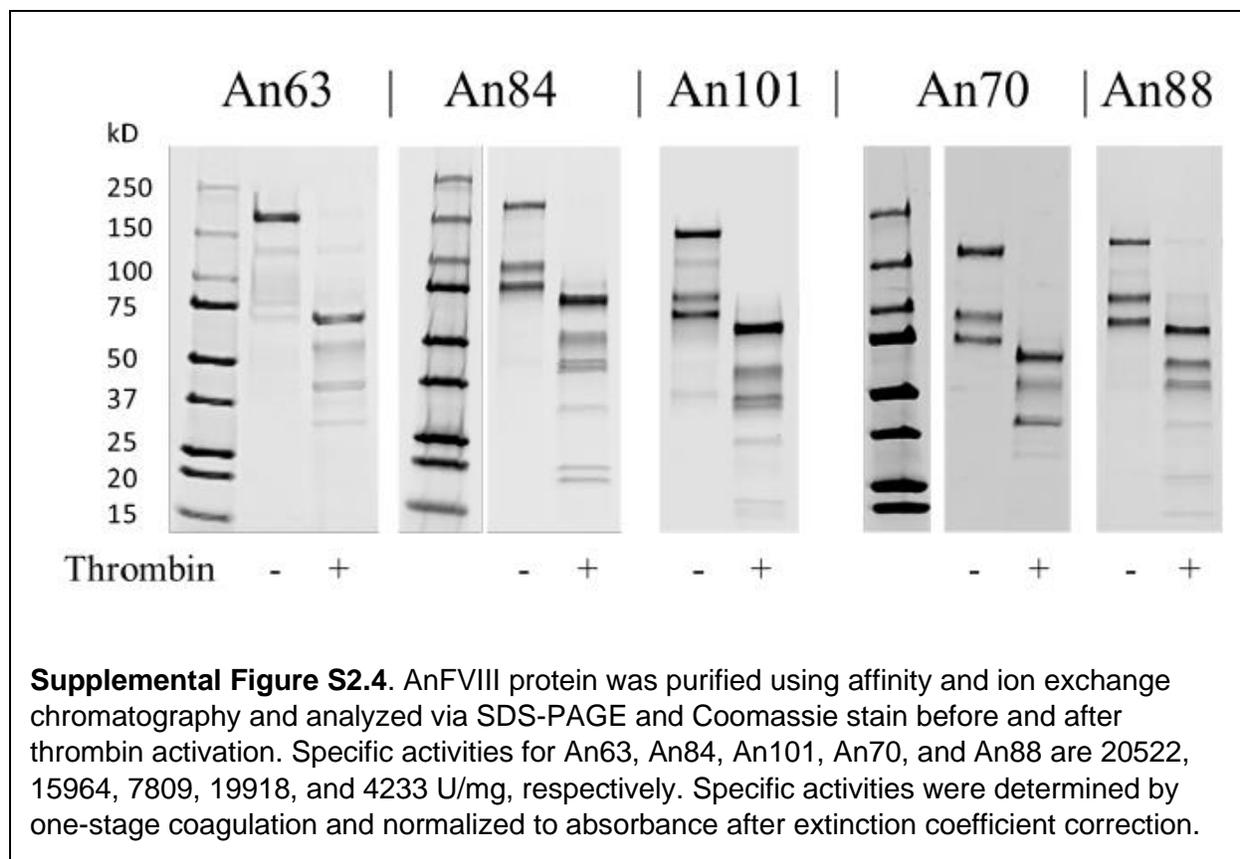
Supplemental Figure S2.1: A, Representative cladogram, and species distribution for VWF as well as coagulation factors V, VII, VIII, IX, and X. Node numbers correspond to the identities of extant (1-59) or ancestral sequences (60-112). B, Representative phylogram for VWF as well as coagulation factors V, VII, VIII, IX, and X. Scale bar represents amino acid replacements per site per unit evolutionary time. Numbers next to nodes represent the Bayesian support for that particular node.

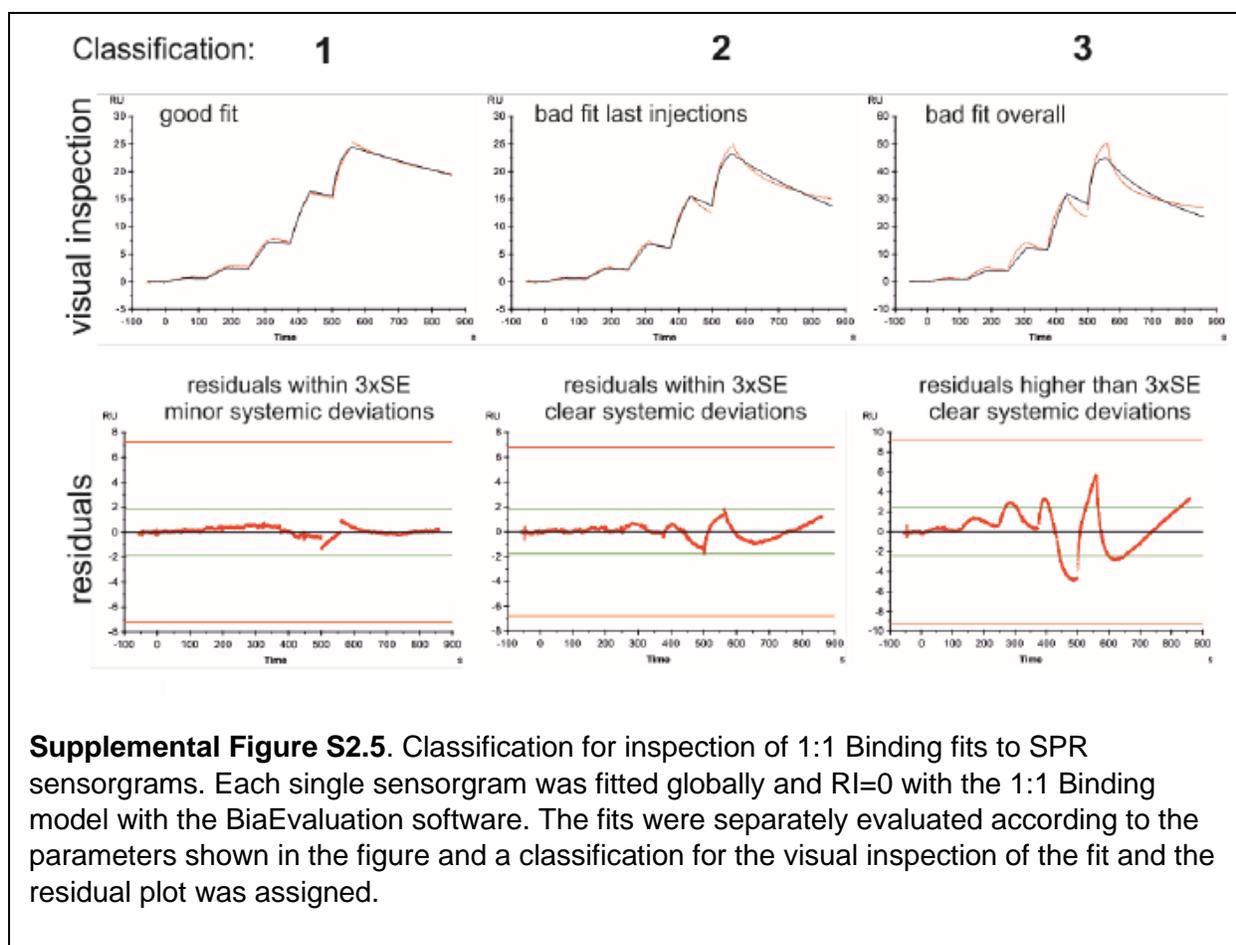


Supplemental Figure S2.2: Quantification of ancestral VWF:Ag was performed by ELISA using polyclonal antibodies across a range of dilutions to identify maximum and minimum detection limits, A. Antibody concentrations were maintained in excess and parallel OD values were observed. Using pooled normal human plasma (FACT), concentrations of VWF antigen at each dilution were elucidated, B.

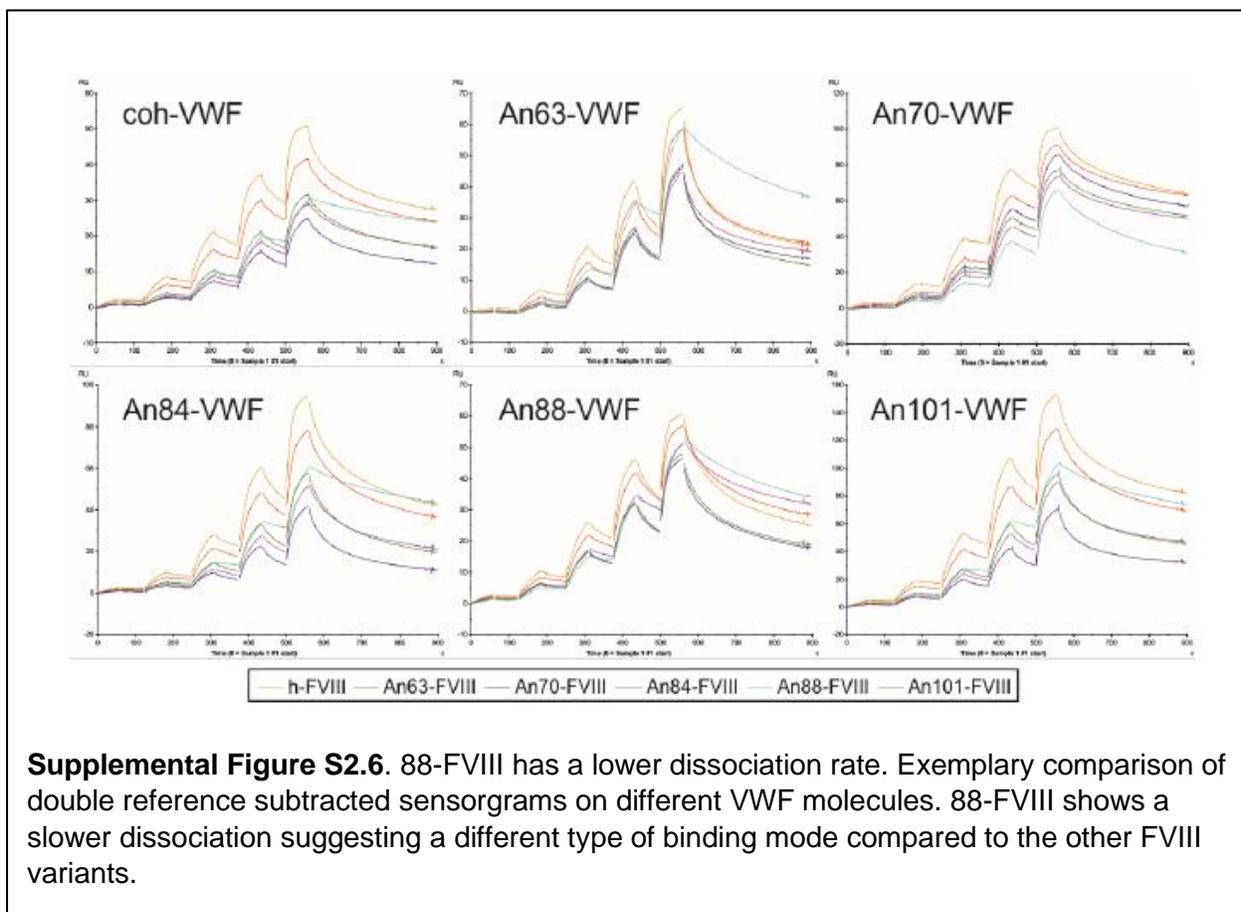


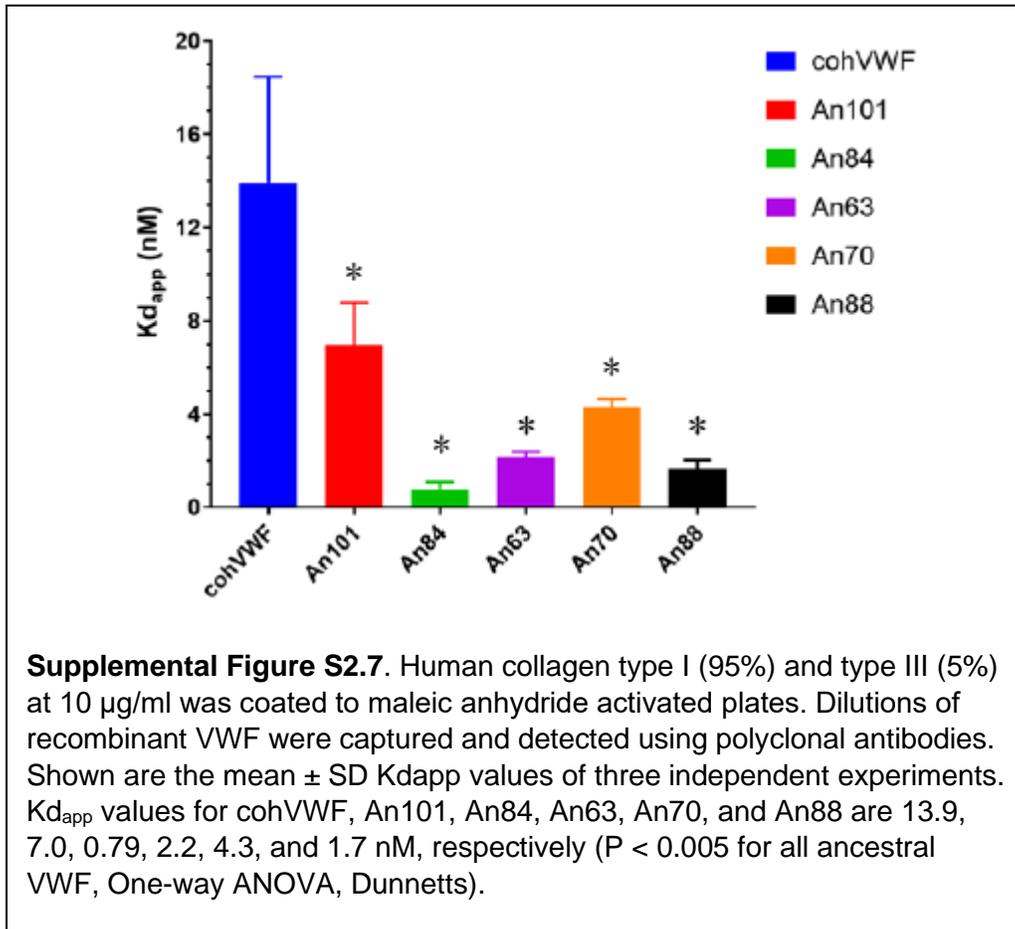
Supplemental Figure S2.3: Total accumulation of mutations within specific subdomains of FVIII A, C or VWF B, D are shown as percent non-identity relative to the size of each domain across ungulate a, b and rodent, c, d, lineages. The common ancestor An-63 VWF and FVIII sequences serve as the reference.





Supplemental Figure S2.5. Classification for inspection of 1:1 Binding fits to SPR sensorgrams. Each single sensorgram was fitted globally and $RI=0$ with the 1:1 Binding model with the BiaEvaluation software. The fits were separately evaluated according to the parameters shown in the figure and a classification for the visual inspection of the fit and the residual plot was assigned.





Chapter 3: Expression of Ancient Coagulation FVIII Molecules and Discovery of Relevant Domains

This research was partially presented as a poster abstract at the American Society of Gene and Cell Therapy Annual Meeting 2020

Christopher Coyle, Jasmine Anickat, Alex Condra, Luke Blackmon, H. Trent Spencer, Eric A. Gaucher, Christopher B. Doering

3.1 Abstract

Clinical hemophilia B gene therapy benefits from incorporating amino acid substitution(s) that increase FIX activity. It is anticipated that FVIII bioengineering would provide comparable advantages. Our group has pioneered efforts to capture the inherent advantageous biological properties of certain animal FVIII orthologs. More recently, we utilized ASR as a platform to study the evolutionary variation in coagulation factors V, VII, VIII, IX, X and VWF and discovered additional functional diversity that can be captured to engineer next-generation gene therapy and protein pharmaceuticals for bleeding disorders. Here, cDNAs corresponding to 9 distinct ancestral BDD FVIII proteins were synthesized de novo. These inferred ancient FVIII variants share 87-98% sequence identity with BDD human FVIII (HSQ). Using transient expression from HEK293T/17 cells, the ancient FVIII variants showed 2 - 27-fold higher FVIII activity than HSQ. The top 3 ancestral FVIII candidates, termed An63, An84 and An70, all had comparable FVIII activity at 27-, 24- and 22-fold greater than HSQ, respectively, as determined by one-stage coagulation assay. Furthermore, codon optimized An63 and An70 in a liver-directed AAV cassette exhibited significantly greater FVIII production than HSQ when transfected into Huh7 cells, suggesting high-expressing ancestral FVIII proteins can be efficiently expressed from liver cells. In an effort to 'humanize' the lead candidate molecules, HSQ domains of were swapped with their cognate ancestral domains. The resulting hybrid expression data suggest that the A1, A2, and A3 domains of An63; the A1, A2, A3, and C2 domains of An84; and the A1, A2, and C2 domains of An70 were required to maintain high activity. Further resolution of the An84 A2 domain identified P747S and E753K as critical substitutions in this region. Collectively, these findings validate the ASR approach as an enabling platform for the development of more potent gene therapy product candidates for hemophilia A, as has been achieved in hemophilia B with factor IX Padua.

3.2 Introduction

FVIII orthologs possess desirable pharmaceutical properties that address many of the shortcomings of human FVIII. For example, porcine FVIII is secreted 10-100 times more efficiently than hFVIII (104). Additionally, previous studies have shown that murine FVIII, upon activation, decays at a much slower rate than human FVIII (114). Even further, ovine, and canine orthologs have increased specific FVIII activity (113; 175). While orthologous FVIII has been used clinically in special circumstances with success (176), exposure to non-human proteins significantly increases the risk of an immunogenic response, rendering the acute treatment no longer effective on a continual basis (177).

ASR is a protein engineering technique that infers sequences of ancient proteins based off extant sequences. The process of ASR begins with gathering extant species sequences for the protein of interest. Using sequence alignment and Bayesian inference computer algorithms, a phylogenetic tree for the predicted evolution the protein is generated, where each node represents the predicted protein sequence of an ancestor (178). We have previously discovered ASR-derived FVIII molecules (129) where top candidates exhibited high specific activity, slow decay, and high biosynthetic efficiency, properties seen in orthologous FVIII. This novel iteration of the phylogenetic tree includes seventeen additional extant sequences and all novel sequences to be studied. To transform candidate An-FVIIIs into viable drug candidates, we utilized a 'humanization' approach, where specific ancestral domains and substitutions were swapped with their cognate human sequence and evaluated.

In line with the neutral theory of evolution (179) we hypothesize that most of the amino acid substitutions in An-FVIII molecules are not required to maintain high activity and that this approach should facilitate the identification of the minimal sequence(s) necessary and sufficient to confer each predictably beneficial pharmaceutical property.

3.3 Materials and Methods

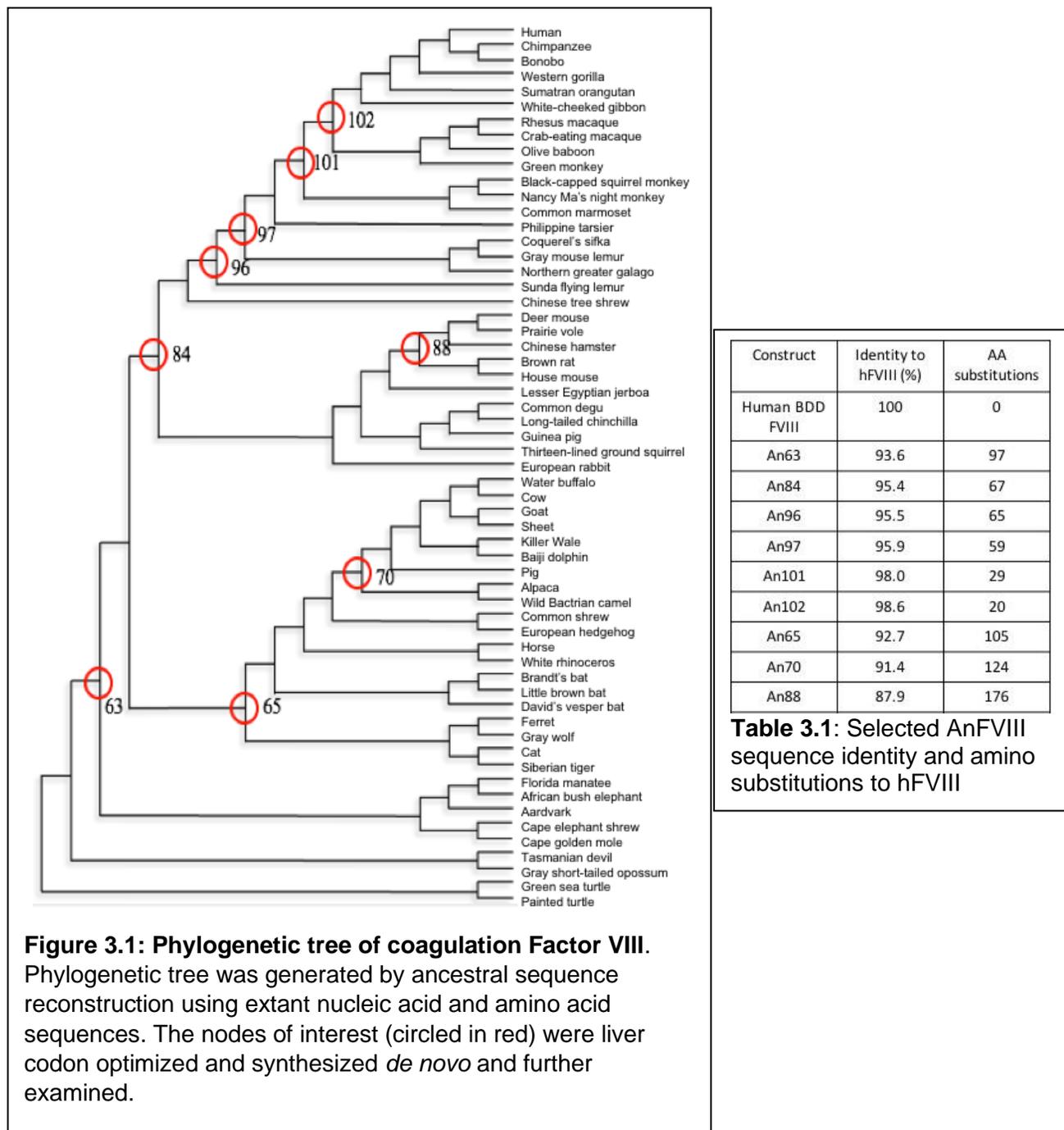
An-FVIII sequence inference and plasmid construction: FVIII ASR was performed as previously described (118; 129). Briefly, fifty-nine extant FIX sequences were aligned using MUSCLE, and an evolutionary tree extending beyond Mammalia was inferred using MrBayes. Ancient mammalian and reptilian sequences were inferred using both DNA and amino acid-based models in PAML (version 4.1). Initial cDNA molecules for AnFVIII 63, 65, 70, 84, 88, 96, 97, 101, and 102 were liver codon optimized (85) and *de novo* synthesized (Genscript, Piscataway, NJ). AnFVIII cDNAs were subcloned into expression vectors under the direction of the elongation-factor 1 α (EF1 α) promoter for initial and hybrid testing, and a previously described liver-specific promoter (85) (HHS4-TTR) for testing in Huh7 cells.

An-FVIII Expression Analysis: Plasmids were prepared from overnight culture of transformed Stbl3 chemically competent cells using a miniprep kit. Low passage HEK293T/17 cells were plated in twenty-four well plates in DMEM/F-12 (Gibco, Waltham, MA) with 10% FBS (Atlanta Biologics, Atlanta, GA). Cells were transfected with plasmids using PEI MAX, according to manufacturer's instructions (Polysciences, Warrington, PA). Twenty-four hours post transfection, wells were washed with Dulbecco's phosphate buffered saline (DPBS; Gibco – Thermo Fisher Scientific, Waltham, MA) and switched to 0.5mL of Freestyle 293 medium (Gibco). After an additional 24 hrs, aliquots of conditioned medium were collected and analyzed via one stage coagulation assay (OSA), as previously described (129).

3.4 Results

Ancestral Sequence Reconstruction and Phylogenetic Tree: The second-generation phylogenetic tree was generated as described in materials and methods can be seen in **figure 3.1**. Fifty-nine extant FVIII sequences were used for the process and the nodes of interest are

circled in red. **Table 3.1** shows the nodes of interest along with their percentage amino acid identity to hFVIII and the total number of amino acid substitutions it has with respect to hFVIII.



Initial testing of second-generation ASR-derived FVIII molecules in vitro: The second generation of ancient FVIII molecules, cloned into plasmids under the direction of the EF1 α promoter were transiently transfected into HEK293T/17 cells, and the conditioned medium was analyzed for FVIII activity by one-stage coagulation assay. The data, presented in **Figure 3.2**, showed that An-FVIII 70, 63, 84, 96, and 97 all had significantly higher activity than HSQ, with top molecules, An70 and An63 exhibiting a ~30-fold improvement in FVIII activity. This suggested that there are amino acid substitutions in these high-expressing ancient molecules that could be added to hFVIII to improve its activity.

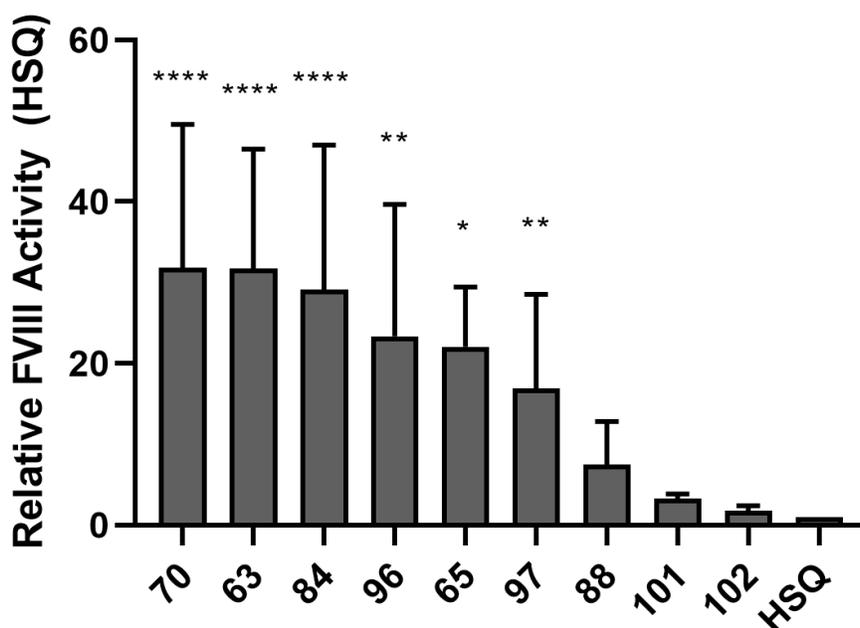
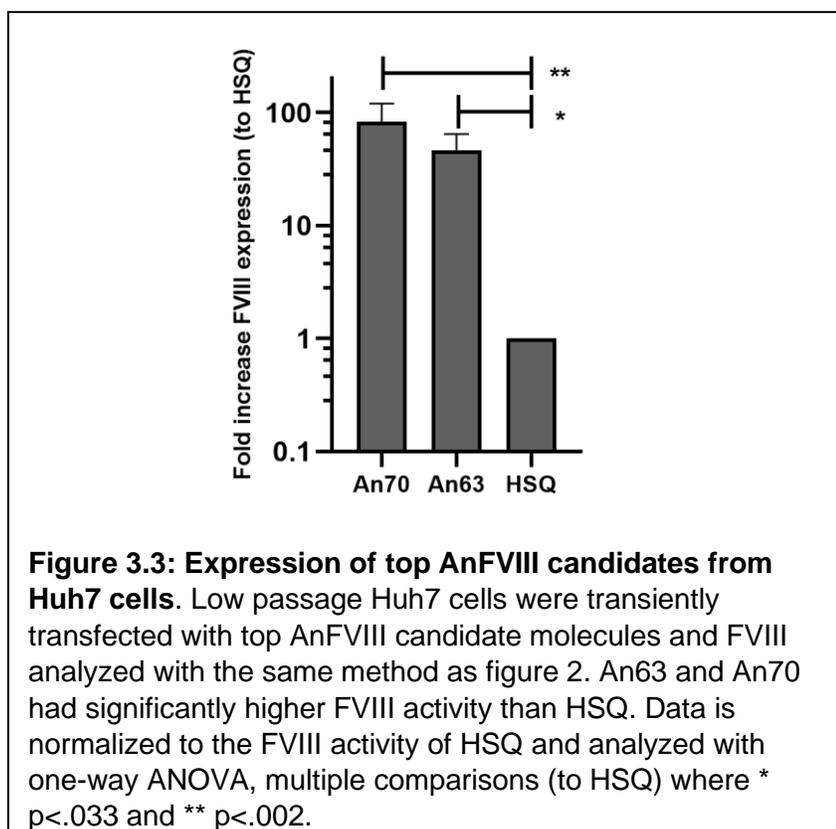


Figure 3.2: Initial Screening of An-FVIII molecules. HEK293T/17 cells were transiently transfected with cDNA corresponding to the selected BDD AnFVIII molecules and examined with one stage aPTT-based coagulation assay. Data is normalized to the FVIII activity of HSQ (which was always within the standard range of detection) and analyzed with one-way ANOVA, multiple comparisons (to HSQ). Here, * $p < .033$, ** $p < .002$, *** $p < .0002$, and **** $p < .0001$.

Expression and activity of top AnFVIII candidates from Huh7 cells

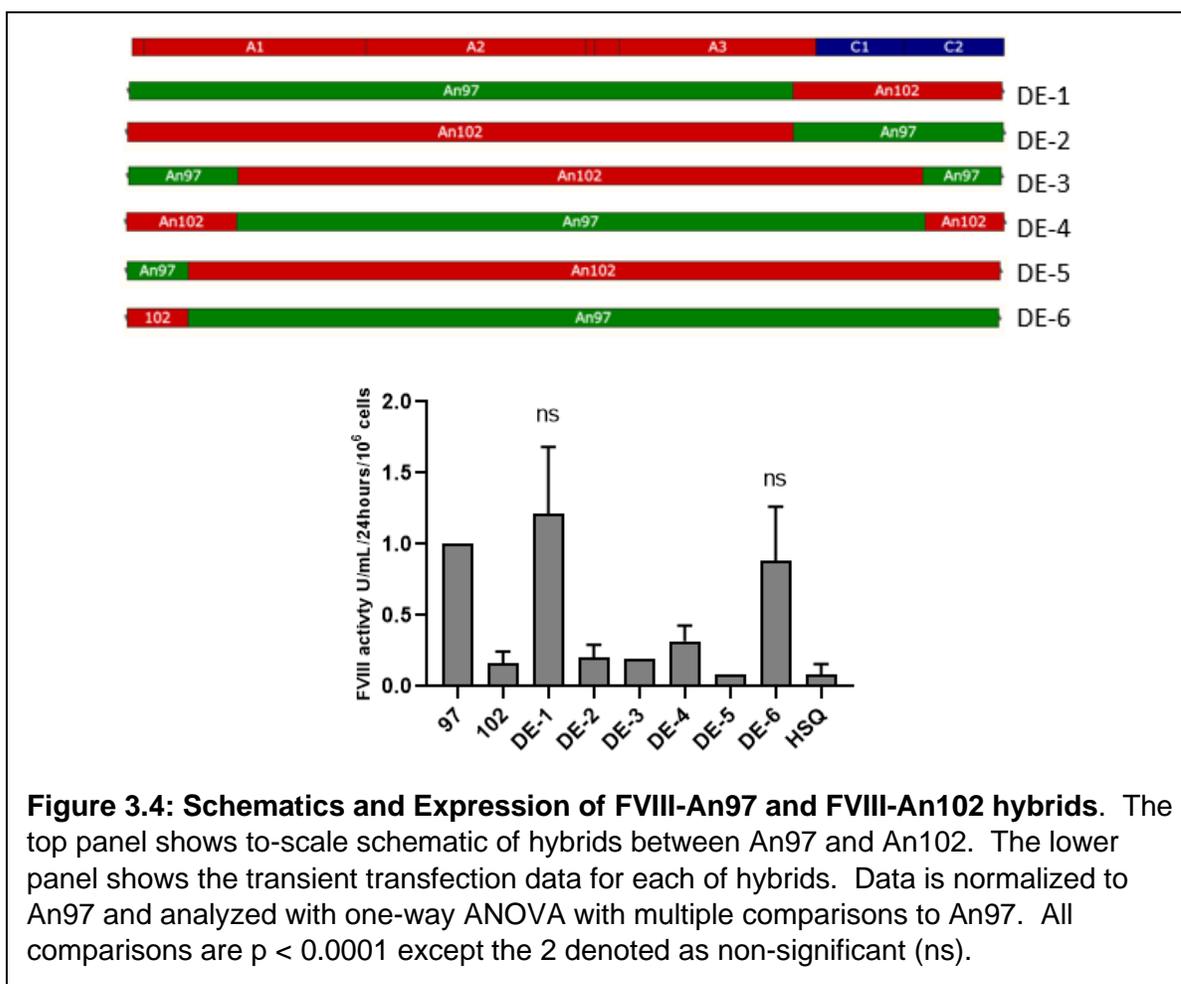
To prepare for the possible use of AnFVIII transgenes in gene therapy, the top two AnFVIII molecules were cloned into an AAV backbone under the control of a liver-specific promoter/enhancer (HHS4-TTR). These plasmids were then transfected into a liver cancer cell

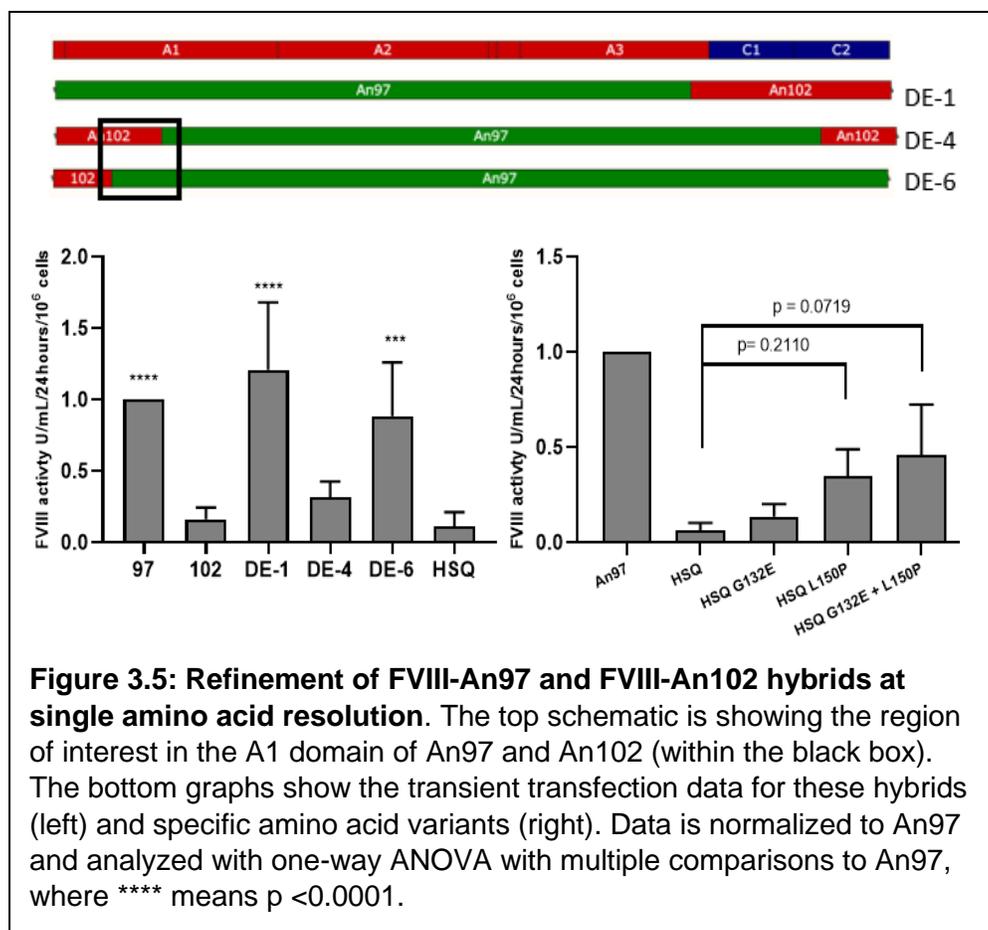


line (Huh7) and the resulting conditioned medium tested for FVIII activity via one-stage coagulation assay (**Figure 3.3**). An70 and An63 had over 50-fold the activity of HSQ when tested in this format, suggesting that they would be compatible with currently available gene therapy technologies.

Humanization studies: After initial testing of the selected AnFVIII nodes, hybrid molecules were made and tested in an effort to determine the relevant domains, regions, and even specific amino acids that were required for top AnFVIII molecules to retain the high expression and/or activity. The first set of hybrids made was between An97 and An102. The hybrids made were

chosen due to available restriction enzyme sites consistent with both transgenes. In **Figure 3.4**, the first set of hybrids are shown in the schematic (to scale). Following testing, hybrids DE-1 and DE-6 were the only molecules that had non-inferior activity compared to the parent molecule, An97. Based on these data, the N-terminal most region of An102 as well as the C-terminal portion of the A3 and C1 and C2 domains did not contain amino acid substitutions that were required to maintain high activity. Upon closer analysis, when comparing DE-4 to DE-6,

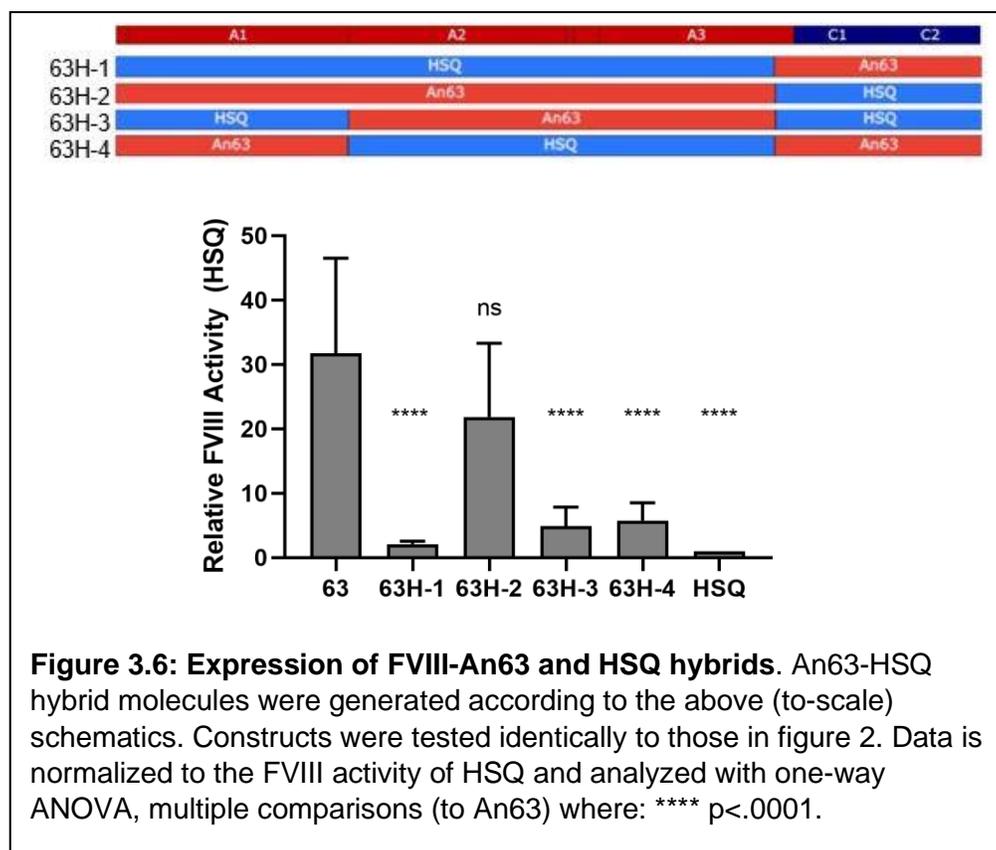




they are identical except for a small portion of An97 sequence within the middle of the A1 domain and the C2 domain. Since the comparison of DE-1 to An97 eliminated C2 residues from contributing to activity, we inferred that the middle-A1 region had amino acid substitutions that improved FVIII activity (**Figure 3.5**). There are just two amino acid substitutions in An97 (with respect to An102) in this region, G132E and L150P. When these two amino acid substitutions were added to HSQ, its activity was modestly improved, although the data are not significant (**Figure 3.5**). The modest improvement, but much lower activity of the HSQ mutant suggested that there are epistatic effects at play, however, it served as a proof of concept that this hybrid method could be used to identify positive amino acid substitutions.

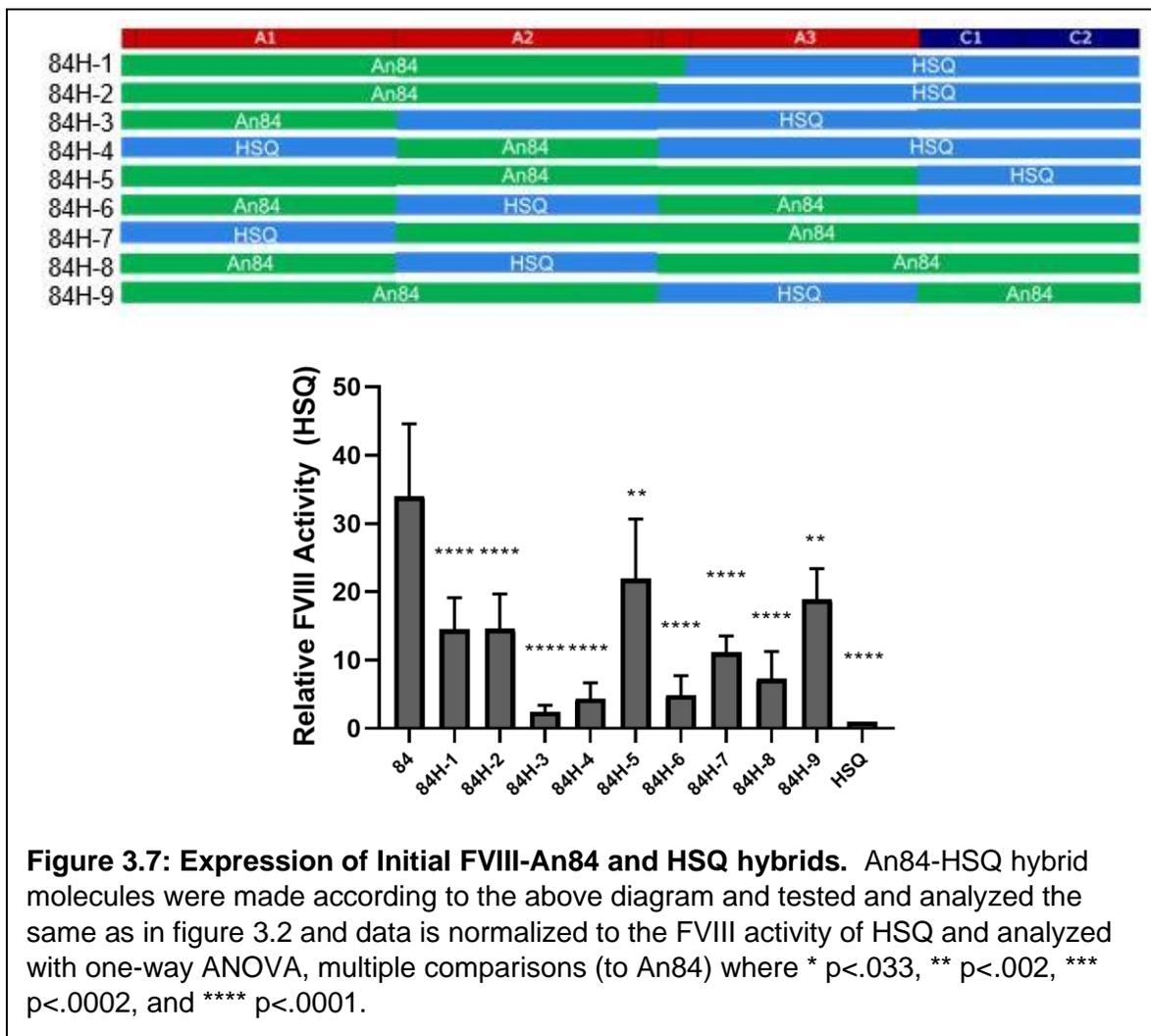
The next set of hybrids made was between An63 and HSQ, since An63 was a top expresser. Since HSQ is the BDD human sequence, it made the most sense to look at any

substitutions in the context of the human sequence, as the minimal number of substitutions with respect to the human sequence is what was desired. The resulting data are seen in **figure 3.6**, where the An63 C1 and C2 domains had no effect on activity, however, the A1, A2 and A3 domains were required to maintain high activity.

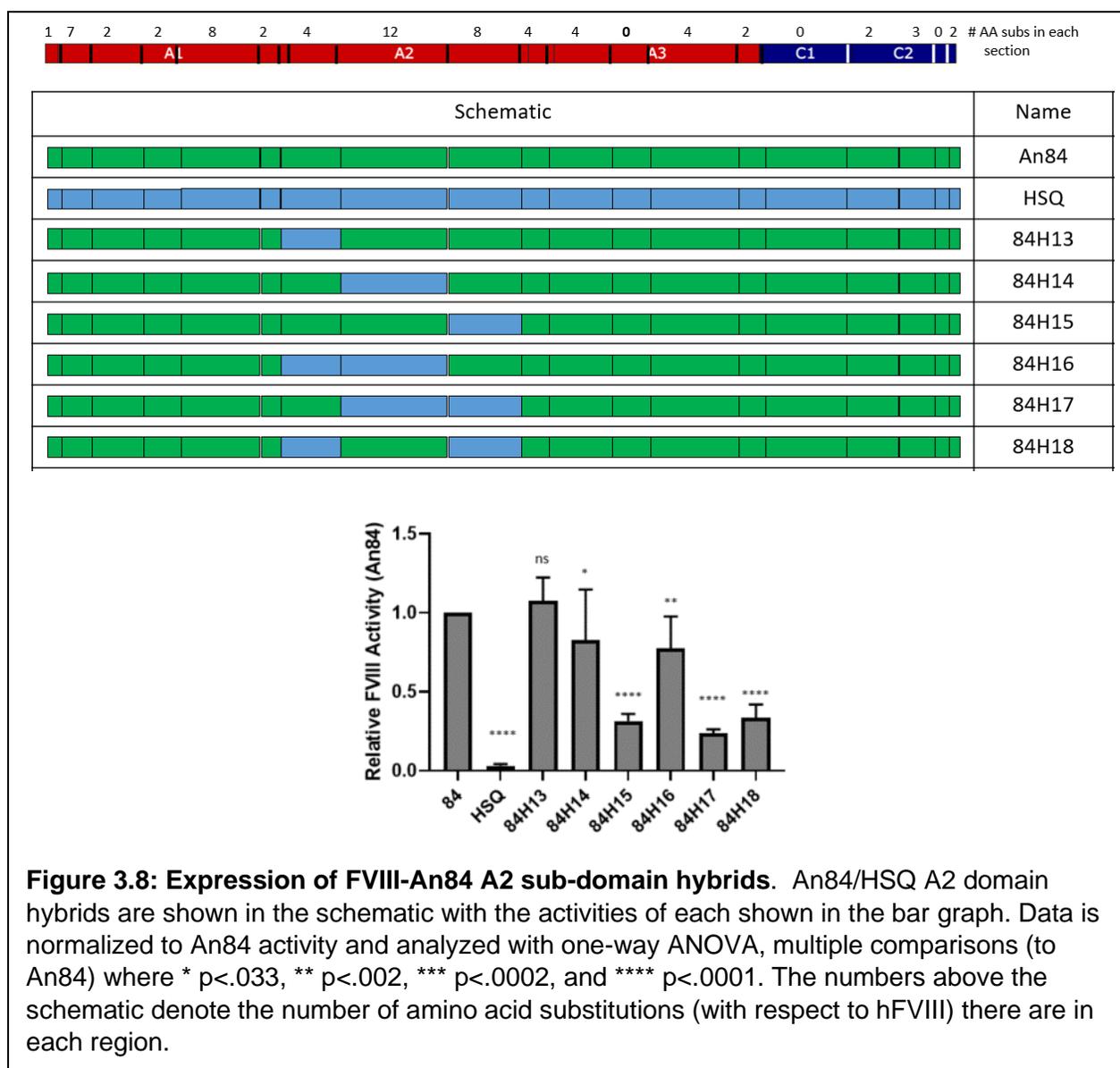


Hybrids between An84 and HSQ were the next set made and tested. **Figure 3.7** shows all nine initial hybrids made in this series along with the appropriately scaled schematics. Here, we can see that each domain is required to maintain high expression (except for the C1 domain, where the sequence is the same as HSQ). Additionally, neither the A1 nor the A2 domain could significantly improve FVIII activity on its own, however the heavy chain of An84 appears to

confer half the high activity. The domains also did not have an equal effect on activity; the A2 domain had the greatest effect, followed by the A1, and the A3 and C2 had an equal effect. These data also suggested, again, that there were epistatic effects, and that negative selection would more likely result in the determination of relevant amino acids.

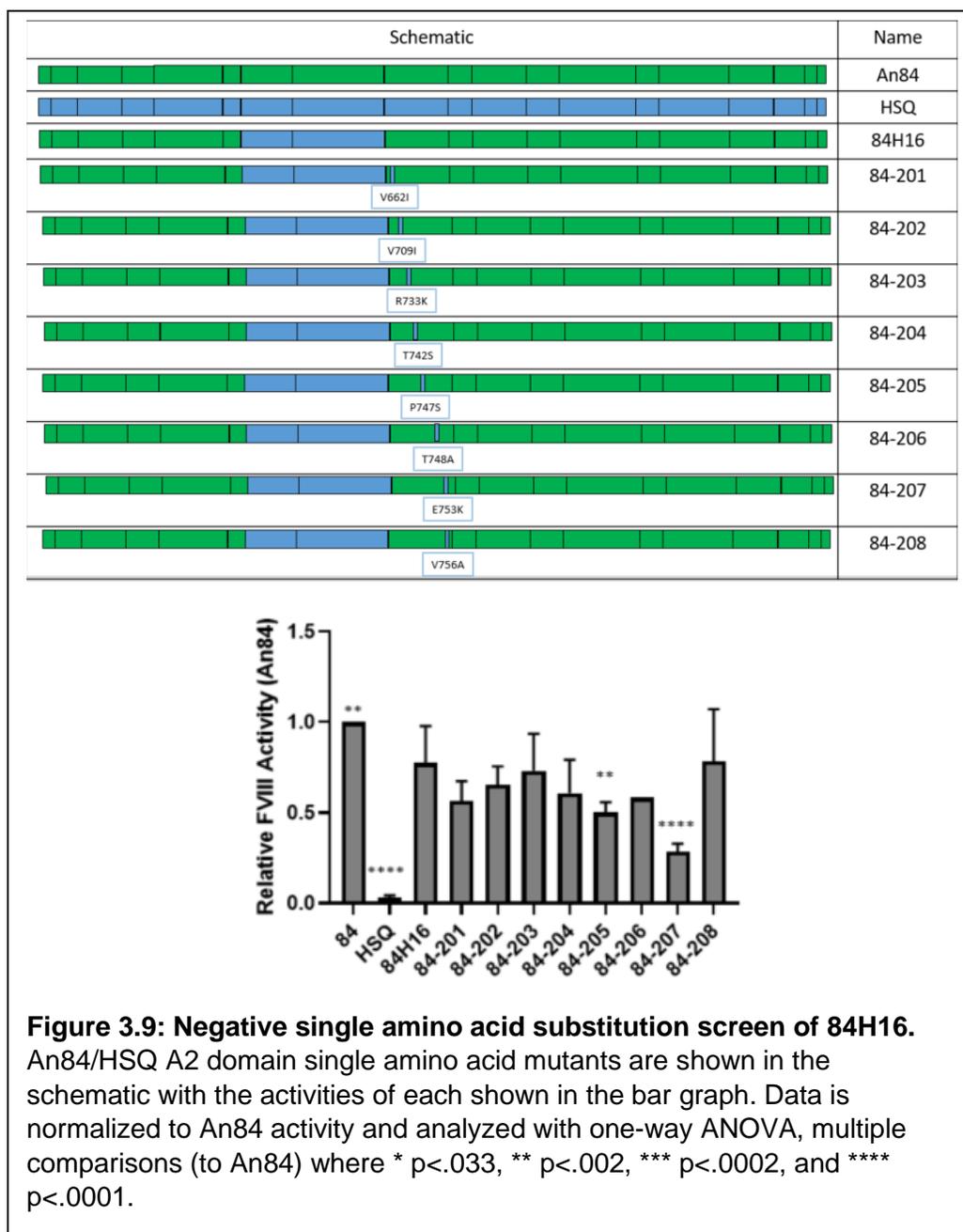


Since the A2 domain of An84 had the largest effect on activity, it was selected for further elucidation. In the next set of six An84/HSQ hybrids, the A2 domain was split into three regions

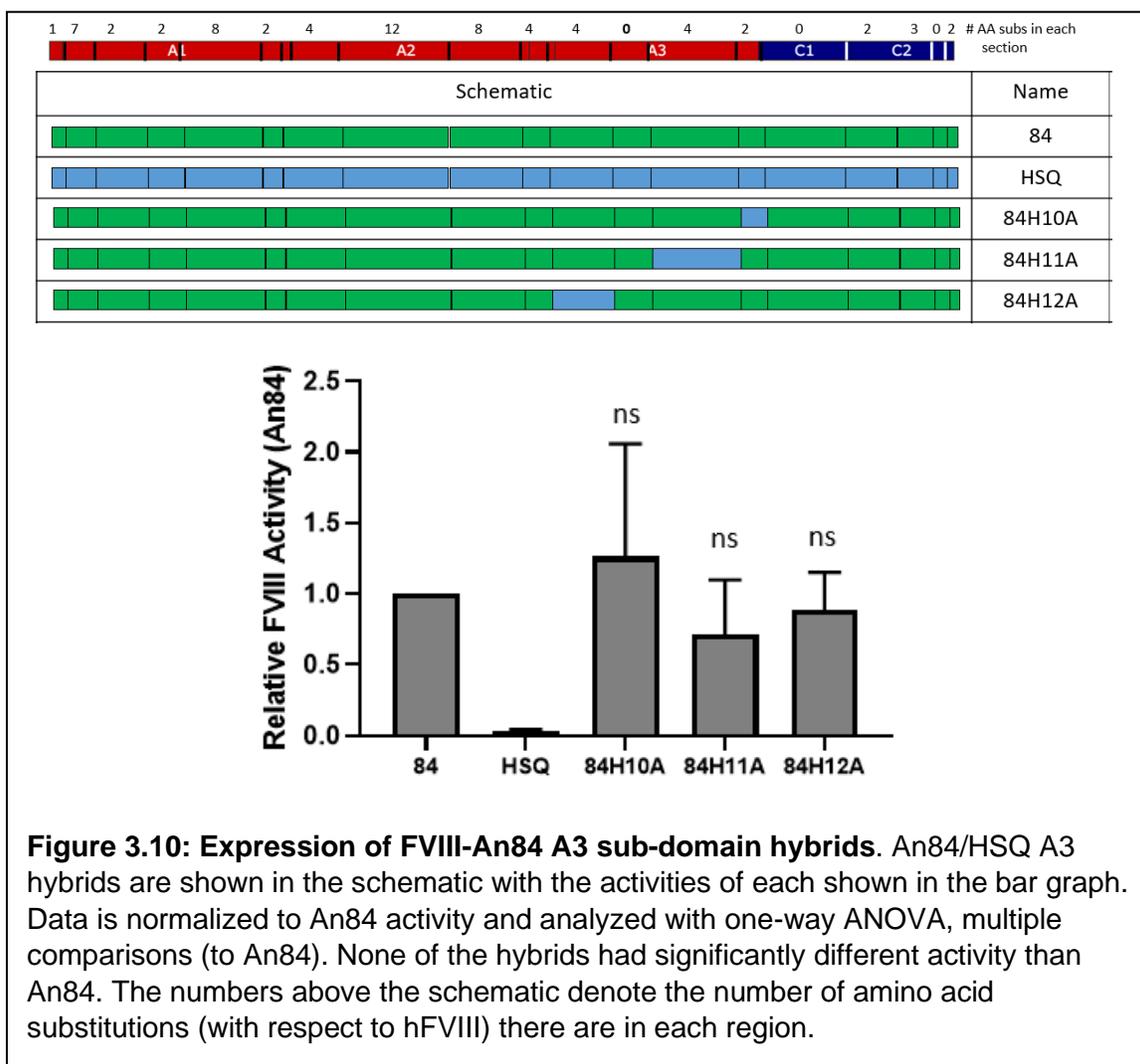


and each combination sub-domain regions alone were evaluated (**Figure 3.8**). It was found that the C-terminal region had a profound effect on activity, as 84H15, 17, and 18 all had exceptionally low activities. Furthermore, middle region has a small and barely significant effect on FVIII activity, while the N-terminal region has no effect on activity.

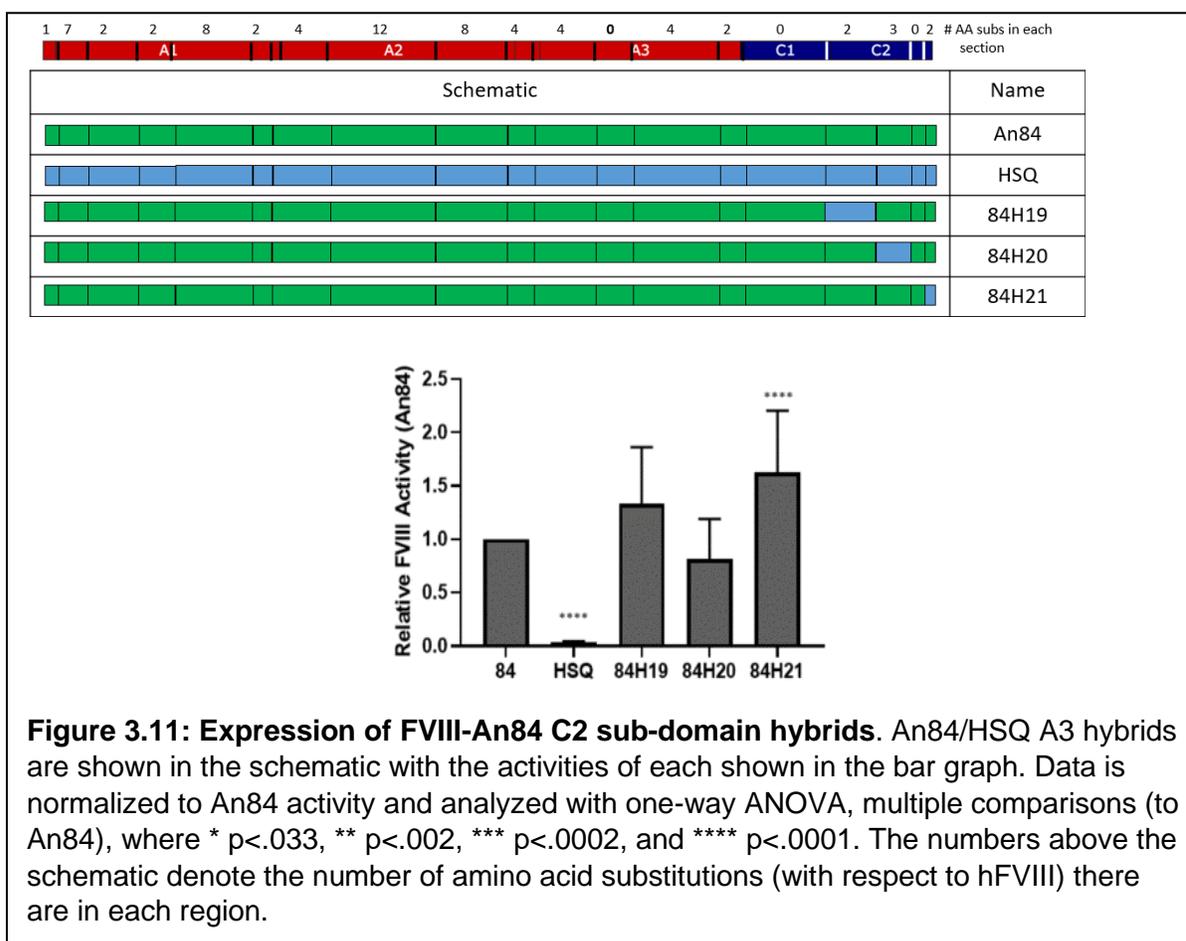
Since the C-terminal region of the An84 A2 domain had the most profound effect on activity, single amino acid mutants were made and tested for each of the amino substitutions in that region (**Figure 3.9**). The statistical analysis showed that there were just two amino acids that contributed significantly to activity in this region, the proline at position 747 and glutamate at position 753.



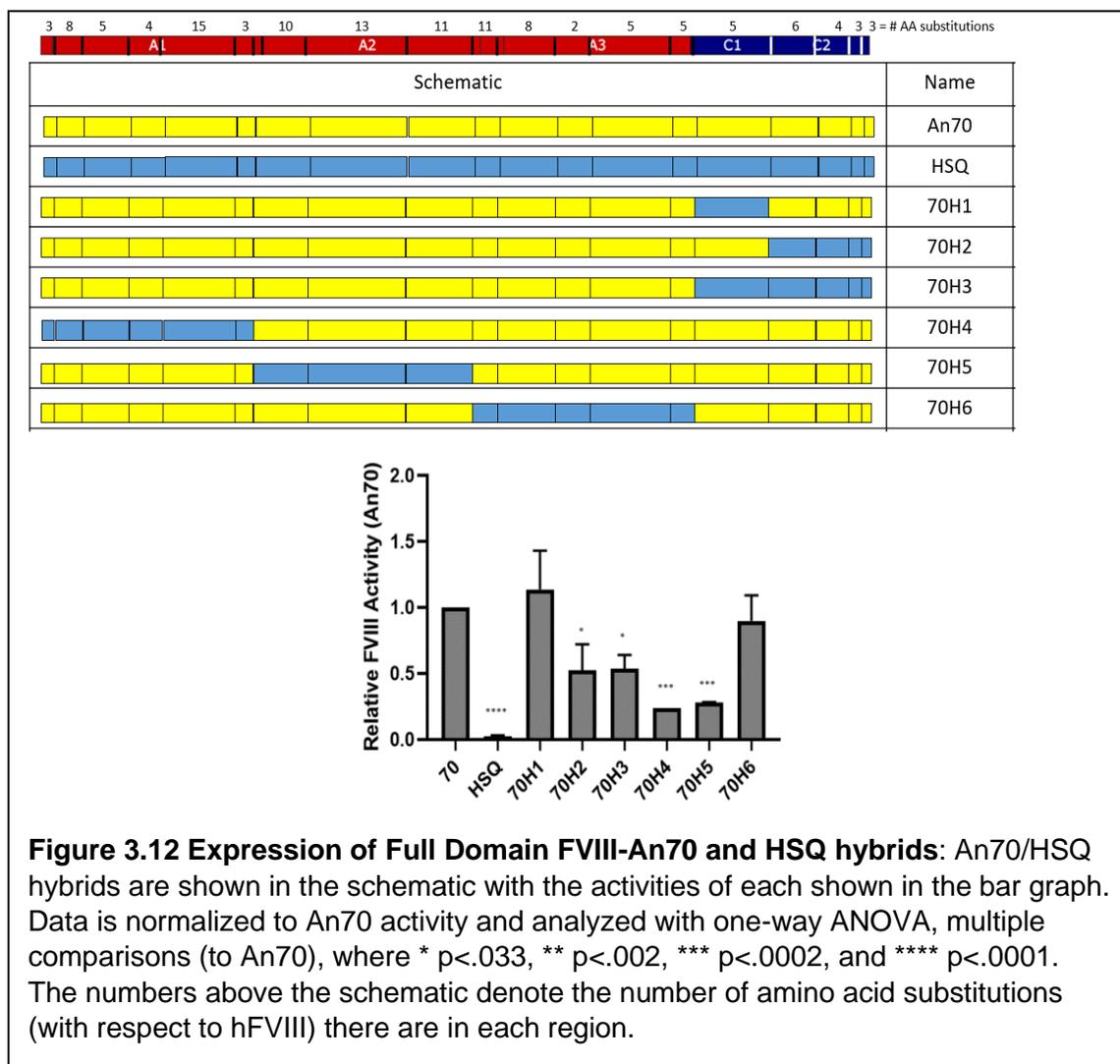
The next domain analyzed of An84 was the A3 domain. Similar to the A2 domain, the An84 A3 domain was split into three regions, and hybrids with each region changed to HSQ sequence were analyzed (**Figure 3.10**). None of the A3 domain hybrid constructs had significantly different activity from An84 suggesting that each of the regions contributes equally.



The final set of An84/HSQ hybrids made and evaluated was with the C2 domain. The data showed in **Figure 3.11** that the two amino acid substitutions in the C-terminal region of the C2 domain of An84 negatively affect activity, as 84H21 has approximately 50% higher activity than An84. The three amino acids in the middle region may have had a positive effect on activity, however it is not statically significant.



Hybrids with An70 and HSQ were also generated (**Figure 3.12**). Here, we can see that only the A1, A2, and C2 domains of An70 are required to maintain high activity. Interestingly, and in contrast to An84, the A1 and A2 domains had an equal effect on activity, while the C2 domain had a lesser, but still significant effect on activity.



3.5 Discussion

ASR has been shown to be a suitable tool for protein engineering biologics and gene therapy products to improve their potency (118; 129). ASR has distinct advantages over other protein engineering techniques, however the most pronounced here are that there is only a small library of proteins generated to test and the nodes are based on functional extant sequences, imparting a high likelihood that the resulting proteins will be functional.

The studies presented here partially proved our hypothesis true: that not all the amino acid substitutions in AnFVIII molecules are required to maintain high activity. This was clearly

observed when the An63 C domains and the An70 A3 and C1 domains were found to have no procoagulant effect compared to their cognate human sequence. When An84 was further examined, the A2 domain was shown to have the largest effect of all the domains, which is in contrast to previous porcine bioengineering efforts, where only the A1 and A3 domains conferred all the procoagulant activity (104). In single amino acid resolution screening of the An84 A2 domain, substitution of a proline at position 747 and glutamate at position 753 conferred the majority of the improved activity, however, further screening would be needed to confirm this.

In the very first set of hybrid molecules, between An97 and An102 (**Figure 3.4**), two amino acid substitutions were found to have a modest, but statistically insignificant effect: G132E and L150P. Interestingly, these two substitutions were determined to be two of the primary five amino acids substitutions from porcine FVIII that confer its activity, although it is G132K in porcine (180). Despite this work remaining incomplete, it shows that these ancestral proteins can be minimized to determine the essential non-human amino acid substitutions.

As a part of Chapter 2, AnFVIII -63, -84, -101, -70, and -88 were purified and had specific activities of: are 20522, 15964, 7809, 19918, and 4233 U/mg, respectively, and HSQ typically has a specific activity of ~5000 U/mg, suggesting that high specific activity partially accounts for the improvement seen in the initial screens, and that hFVIII evolved to be less efficient, possibly due to increased thrombotic risk of stroke as humans became bipedal. It is likely, however, that there is additional activity enhancement due to improved biosynthesis, similarly to porcine FVIII (165). Also of note, An63 was primarily expressed in single chain form, which is likely due to its modification of the PACE-Furin cleavage site in the linker, which has also been previously reported to improve the efficacy of hemophilia A gene therapy (126).

The primary future direction for this work is to gain further resolution of specific amino acid substitutions required to maintain high activity of top candidate AnFVIII molecules.

Additionally, mechanistic studies should be performed to determine if there are multiple mechanisms at play, which is possible (or even likely). Combining both these, it is possible that a novel FVIII molecule that takes unique substitutions from different top AnFVIII nodes could be combined to generate the highest expression FVIII to date.

3.6 Acknowledgements:

This work was supported by grant HL137128 (H.T.S., C.B.D.) from the National Heart, Lung and Blood Institute, National Institutes of Health (NIH) and funding from Hemophilia of Georgia.

Chapter 4

Identification of coagulation factor IX variants with enhanced activity through ancestral sequence reconstruction

This research was published in *Blood Advances*.

Kristopher A. Knight, **Christopher W. Coyle**, Caelan E. Radford, Ernest T. Parker, Andrew Fedanov, Jordan S. Alexander, Fania Szlam, Anatolii Purchel, Michelle Chen, Gabriela Denning, Roman M. Sniecinski, Pete Lollar, H. Trent Spencer, Eric A. Gaucher, Christopher B. Doering

Identification of coagulation factor IX variants with enhanced activity through ancestral sequence reconstruction

Blood Advances. 2021. 5 (17):3333-3343

© 2021 by The American Society of Hematology

Trent Spencer, Eric Gaucher, and Chris Doering conceived the project. Pete Lollar, Trent Spencer, and Chris Doering. designed the experiments and analyzed the data. Kristopher Knight, Chris Coyle, Caelan Radford, Andre Fedanov, Jordan Alexander, Ernest Parker, Gabriela Denning, Fania Szlam, Roman Sniecinski, Anatolii Purchell, and Michelle Chen

performed experiments and analyzed the data. Kristopher Knight and Chris Doering drafted the manuscript. Kristopher Knight, Pete Lollar, Eric Gaucher, and Trent Spencer edited the manuscript.

Note: The experimental contributions of Christopher Coyle are specific to figures: 4.1A, 4.1B, Supplementary Tables S4.2 and S4.3 and Supplementary Figure S3. The data and conclusions from this publication explain the observations and data presented in Chapter 5 and were therefore included in this dissertation. Kristopher Knight, who is responsible for a majority of the data presented in this chapter is first author on this publication and second author on the manuscript presented in chapter 5.

4.1 Abstract

Orthologous proteins contain sequence disparity guided by natural selection. In certain cases, species-specific protein functionality predicts pharmacological enhancement, such as greater specific activity or stability. However, immunological barriers generally preclude use of nonhuman proteins as therapeutics, and difficulty exists in the identification of individual sequence determinants among the overall sequence disparity. Ancestral sequence reconstruction (ASR) represents a platform for the prediction and resurrection of ancient gene and protein sequences. Recently, we demonstrated that ASR can be used as a platform to facilitate the identification of therapeutic protein variants with enhanced properties. Specifically, we identified coagulation factor VIII (FVIII) variants with improved specific activity, biosynthesis, stability, and resistance to anti-human FVIII antibody-based inhibition. In the current study, we resurrected a panel of ancient mammalian coagulation factor IX (FIX) variants with the goal of identifying improved pharmaceutical candidates. One variant (An96) demonstrated 12-fold greater FIX activity production than human FIX. Addition of the R338L Padua substitution further increased An96 activity, suggesting independent but additive mechanisms. After adeno-associated virus 2 (AAV2)/8-FIX gene therapy, 10-fold greater plasma FIX activity was observed in hemophilia B mice administered AAV2/8-An96–Padua as compared with AAV2/8-human FIX–Padua. Furthermore, phenotypic correction conferred by the ancestral variant was confirmed using a saphenous vein bleeding challenge and thromboelastography. Collectively, these findings validate the ASR drug discovery platform as well as identify an ancient FIX candidate for pharmaceutical development.

4.2 Introduction

Collectively, deficiencies in coagulation factor VIII (FVIII) or IX (FIX) represent the most common severe bleeding disorder, hemophilia, and are designated A or B, respectively. Their combined prevalence is estimated at 1 in 3333 newborn males (5). Despite dramatic

improvements in the standard of care through factor replacement therapy, this option remains limited to a minor fraction of the total hemophilia population because of product cost, compliance with lifelong IV therapy, and antidrug antibody responses, clinically referred to as inhibitors.

Gene therapy represents a potentially transformative therapeutic option, and numerous FVIII and FIX gene therapy product candidates are progressing through clinical development. One common aspect of all clinical gene therapy programs is the inclusion of bioengineered elements, including vector capsid, transcriptional regulatory elements, and/or transgenes to maximize product potency and durability while reducing the risks of vector-related toxicities that remain a challenge to adeno-associated virus (AAV) gene therapy achieving 100% normal FIX activity levels. Therefore, further improvement in vector potency seems necessary to unlock the full potential of liver-directed AAV gene therapy for hemophilia B.

Recently, we demonstrated the utility of ancestral sequence reconstruction (ASR) for the discovery of FVIII variants possessing greater expression, specific activity, and active half-life compared with human FVIII while also exhibiting reduced antigenicity to human FVIII (hFVIII) inhibitors (129). Despite extensive research focused on extant FVIII orthologs and their biochemical differences (horizontal comparisons), limited progress has been made toward understanding structure/function relationships at high resolution, and even less has been made toward the translation of preclinical findings into clinical study and practice. Unlike horizontal study of extant protein orthologs, ASR facilitates vertical analyses through the prediction of ancient gene/protein sequences followed by de novo DNA synthesis and laboratory resurrection in the form of recombinant proteins or gene therapy transgene products. ASR also provides a high-resolution mapping solution by enabling empirical comparisons of ancient proteins predicted within sequential branches on a phylogeny (181). Furthermore, ASR generates protein variants that uniformly possess the intended biomolecular function yet can also display

unpredicted or expanded properties. In the current study, ASR was applied to FIX with the goal of harnessing information contained in the ancient and extant vertebrate coagulation systems toward hemophilia B biopharmaceutical development.

4.3 Materials and Methods

Ancestral FIX sequence inference and plasmid construction: FIX ASR was performed as described previously (129). Fifty-nine extant FIX sequences were aligned using MUSCLE, and an evolutionary tree extending beyond Mammalia was inferred using MrBayes. Ancient mammalian and reptilian sequences were inferred using both DNA and amino acid–based models in PAML (version 4.1). On the basis of the amino acid sequences inferred for An102, An97, An96, An88, An84, An70, An65, and An63, complementary DNA (cDNA) sequences were generated using a liver codon optimization (LCO) algorithm described previously and de novo synthesized by GenScript (Piscataway, NJ) to contain flanking 5' *XhoI* and 3' *NotI* restriction sites for subcloning into the mammalian expression vectors ReNeo and pcDNA 3.4. as well as recombinant adeno-associated and lentiviral vector expression plasmids (85). Human FIX-T148 as selected as the wild-type FIX control based on its higher allele frequency and prior functional characterization by our laboratory (182).

AnFIX expression analysis: HEK293T/17 and HepG2 cells were transfected with polyethylenimine or TransIT-X2, respectively, in antibiotic-free media. Transiently transfected cells were washed twice with Dulbecco's phosphate-buffered saline and carefully switched to serum-free media 24 hours after transfection. FIX activity was measured by 1-stage coagulation assay after an additional 24-hour culture period. Stable Expi293F clones were generated, and 18 to 25 FIX-expressing clones per construct were isolated for screening. FIX activity was measured in serum-free media by 1-stage coagulation assay and normalized to cell counts taken at the time of FIX activity determination.

Purification of AnFIX variants: Expi293F clones displaying the greatest rate of FIX production were expanded into Erlenmeyer flasks, and FIX was collected in serum-free media clarified by centrifugation at 1000× g for 25 minutes and stored at −20°C until purification. Recombinant FIX was purified using methods similar to those described previously (183). Elution fractions were analyzed for purity via sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Activity determinations were made using 1-stage coagulation assay, and mass concentrations were measured by A_{280} with A_{320} correction using an estimated extinction coefficient of $64\,590\text{ M}^{-1}\text{ cm}^{-1}$. Purified FIX preparations were aliquoted and stored at −80°C. Comparisons were made to a commercial recombinant human FIX product (BeneFIX; Takeda).

Purification of hFIX-Padua: hFIX-Padua containing serum-free medium was collected and purified over a 50mL IXSelect column according to the manufacturer's instructions. Fractions containing FIX activity were pooled and buffer exchanged into tris-buffered saline (TBS) with 5mM CaCl_2 . Half the pooled material was loaded onto a 5mL Capto MMC Column equilibrated with 20mM sodium citrate, 100mM NaCl, pH 7.0. Material was eluted with 20mM sodium citrate, 1M NaCl, pH 7.0. Capto MMC Eluate was diluted 2-fold with water and loaded onto a 1mL Capto Q column equilibrated with 100mM NaCl 50mM Tris, pH 8.0. Material was eluted on a gradient with 50mM Tris, 100mM NaCl, 100mM CaCl_2 . Elution fractions were analyzed for purity via sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE). Activity determinations were made using one-stage coagulation assay and mass concentrations were measured by A_{280} with A_{320} correction using an estimated extinction coefficient of $63,190\text{ M}^{-1}\text{ cm}^{-1}$. Purified FIX preparations were aliquoted and stored at −80 °C. Purification results are presented in **Supplementary Table S2.2**.

Purification of An96: An96 containing culture serum-free medium was collected, clarified by centrifugation, and initially purified over a XK 16/20 column packed with 30mL of Capto MMC resin equilibrated with 20mM sodium citrate, 100mM NaCl, pH 7.0. Material was eluted with

20mM sodium citrate, 1M NaCl, pH 7.0. Capto MMC eluate was diluted 5x out with 50mM tris buffer (final concentration) and pH adjusted to 8.0 and subsequently loaded onto a 5mL Capto Q column equilibrated with 50mM Tris, 100mM NaCl, pH 8.0. Material was eluted with 50mM tris, pH 8.0 and a CaCl₂ gradient from 0-100mM (top fraction eluted first, around 15mM CaCl₂). Purification results are presented in **Supplementary Table S2.3**.

AnFIX glycan and activation studies: Recombinant hFIX and An96-Padua were treated with the endoglycosidase, PNGase F, according to the manufacturer instructions and resolved under denaturing reactions conditions prior to SDS-PAGE and visualization by Coomassie blue staining. FIX cleavage by activated factor XI (FXIa) was performed by incubating 20 µg FIX with 40 ng plasma derived FXIa in the presence of 5mM CaCl₂ and a final volume of 168 µL at 37° C. At specified timepoints, 21 µL of the reaction mixture was removed and immediately quenched by addition of 6x SDS-PAGE loading buffer with dithiothreitol (DTT) and analyzed using 12% SDS-PAGE and Coomassie blue visualization.

Thrombin generation assay: Calibrated automated thrombography was performed using a Thrombinoscope (Diagnostica Stago, Inc., Parsippany, NJ). Briefly, 80 µL of warmed FIX-deficient plasma was placed in a 2HB transparent round-bottom 96-well plate. Twenty microliters of thrombin calibrator or PPP-Reagent Low was added to either calibration wells or test wells, respectively. FIX was added at the specified concentration in triplicate wells and the program initiated per manufacturer instructions.

FIX activity measurements: FIX activity in conditioned supernatant, purified recombinant FIX preparations, and mouse citrated plasma was performed by activated partial thromboplastin time (aPTT)-based one-stage coagulation assay, as previously described (184). Briefly, 5 µL of the test article was diluted in 50 µL FIX-deficient plasma. aPTT (50 µL) was then added, and samples were incubated at 37 °C for 3 min. Following incubation, 50 µL of 20 mM CaCl was added. Time to clot formation was determined by automated viscometric measurement. FIX

activity was determined by linear regression analysis of the clotting time versus the logarithm of the reciprocal plasma dilution. Chromogenic activity assays (Rox Factor IX, Diapharma, West Chester, OH) were performed according to the manufacturer's protocols.

FIX ELISA: FIX levels in conditioned supernatant and purified recombinant FIX preparations were measured by human FIX-specific ELISA using the Abcam – Human Factor IX SimpleStep ELISA Kit according to the manufacturer's instructions. Briefly, citrated plasma samples were collected by retroorbital bleeding method and quantified by colorimetric sandwich ELISA against a standard curve prepared from lyophilized native human FIX protein using a SpectraMax 96 well plate reader (Molecular Devices, San Jose, CA). FIX levels in test samples were quantified by 4-parameter curve fit to the standard curve using the Softmax Pro analysis software (Molecular Devices, San Jose, CA).

AAV-FIX production and characterization: Recombinant AAV2(ITR)/8(capsid) vectors were manufactured and titrated via quantitative polymerase chain reaction (qPCR) by ViGene (Rockville, MD). Additionally, the AAV2/8-HHS4 vector preps were subjected to size-exclusion chromatography/multiangle light scattering (SEC-MALS) analysis using an Infinity II high-performance liquid chromatography system (Agilent, Santa Clara, CA), 18-angle DAWN (Wyatt Technologies, Santa Barbara, CA), Optilab (Wyatt Technologies), AAV column (Wyatt Technologies), and ASTRA 7.3 software (Wyatt Technologies) (185).

SEC-MALS AAV Analysis: All AAV samples were screened with a DynaPro Plate Reader for the presence of large aggregates before injecting onto the HPLC system. An Agilent 1260 Infinity II HPLC system was employed with a Wyatt WTC-050S5 column (7.8 x 300 mm) and the corresponding guard column. Phosphate-buffered saline was used as the mobile phase at a flow rate of 0.5 mL/min. The volume of each injection was 10 μ L for AAV-hFIX-Padua and AAV-An96 and 30 μ L AAV-An96-Padua due to lower concentration of the latter. The detection system consisted of the Agilent HPLC's UV-Vis detector measuring at wavelengths of 260 nm

and 280 nm, an 18-angle DAWN® Multi Angle Light Scattering (MALS) detector with a WyattQELS® embedded online dynamic light scattering (DLS) module, and an Optilab® dRI detector. Data from the MALS, DLS, UV (both wavelengths), and dRI detectors were collected and processed using ASTRA® 7.3 software with HPLC manager. All three samples provided excellent signal/noise. Quantitation with UV280 and dRI provided capsid content and particle concentration. DNA and capsid molar mass obtained from MALS analysis confirmed identity of the AAV-DNA (predicted complete vector genome) conjugate and serves as an internal control. Protein capsid molar masses obtained were in excellent agreement with theoretical values for AAV8. The data presented are the mean and standard deviation of two independent infections.

In vivo FIX gene transfer: The AAV2/8 vector design consisted of synthetic liver-directed promoters (either hepatic combinatory bundle [HCB] or HHS4-TTR), a minute virus of mouse intron, a minimal synthetic β -globin polyadenylation sequence, and either the hFIX-148T-LCO, An-96-LCO, or An96-Padua-LCO transgene. Mice were weighed and randomized. AAV vector was diluted to the desired vector genomes (vg) per kilogram or vector particle (vp) per kilogram dose in 100 μ L of sterile DPBS with 0.001% Pluronic F68. Blinded AAV doses were administered to hemophilia B mice age 8 to 12 weeks. Blood plasma was collected biweekly after AAV administration. FIX activity was measured by 1-stage coagulation assay using a standard curve generated from pooled citrated human plasma (FACT). Experimental and control mice were subjected to saphenous vein bleed challenge as previously described or thromboelastography as described in this report.

Thromboelastography (TEG): Whole citrated blood was assayed utilizing the TEG 5000 Thrombelastograph Hemostasis Analyzer system (Haemonetics, Switzerland) according to the manufacturer instructions for native activation.

ED₅₀ determination: Hemostatic challenge was performed via saphenous vein challenge modified from ones previously described (186; 187). ED₅₀ estimation was made using the up-

and-down method as described by Dixon et al (188; 189). FIX doses were prepared immediately preceding intravenous injection. Eight to twelve-week old hemophilia B mice (B6.129P2-F9tm1Dws/J, The Jackson Laboratory) were weighed and anesthetized using isoflurane, in a supine position on a heating blanket. Ventral hind limb hair was removed from the right leg, and a 23-G needle was used to make an entry hole to the vein 8–10 mm distal to the femoral vein, midway between the femoral vein and the distal saphenous branch. Cellulose swabs were used to absorb pooled blood by gently touching the forming blood drop away from the clot site every 30 s until hemostasis was achieved (1 – 3 min for all mice). Immediately following hemostasis, an approximate 1 mm longitudinal distal cut was made using student vannas spring scissors by inserting one blade into the vessel as far as possible, using the punch hole as entry. The ability of the mouse to achieve hemostasis was assessed by determining the number of bleeding episodes and the time to clot after disrupting with a blunted 30-G needle in the direction of the blood flow over a 30 min period. Mice displaying an average time for hemostasis > 3 standard deviations (s.d.) of the mean value obtained for wild-type C57Bl/6 mice in an identical challenge, without exogenous FIX administration, were categorized as ‘bleeders’.

Animal Studies: All animal studies were approved by the Emory University Institutional Animal Care and Use Committee and performed within the Emory University Department of Animal Research facilities. The experimenter was blinded to AAV vector identification throughout the studies. Mice were randomized using www.random.org. Male and female mice, 8 – 12 weeks of age were utilized for studies with recombinant FIX. For all AAV-FIX studies, only male mice were utilized due to sex-specific differences in AAV8 murine liver transduction efficiency.

Statistical analysis: All data analysis, graphing, and statistical calculations were performed using GraphPad Prism 7.04 software (GraphPad Software, San Diego, CA). Comparisons made between AnFIX constructs and hFIX with or without Padua were performed

by 1-way analysis of variance. Individual comparisons among the groups were made by post hoc multiple comparison testing (Tukey or Holm-Sidak as specified).

4.4 Results

FIX ASR screen: ASR of FIX was performed as previously described to infer a phylogenetic tree beyond the class *Mammalia* (**Figure 4.1A**) (129). AnFIX sequences representing 8 nodes sharing 83% to 98% amino acid identity with hFIX were reconstructed using the LCO algorithm described previously and de novo cDNA synthesis (85). **Figure 4.1B** and **supplemental Table S4.1** illustrate the amino acid sequence variation in the reconstructed AnFIX sequences compared with hFIX.

As an initial screen for AnFIX candidates with enhanced properties, the 1-stage coagulation assay was used. This assay allows for the detection of potential variants with more efficient FIX biosynthesis (i.e., protein translation and secretion) and/or enhanced specific activity, but it does not distinguish between the 2 possibilities. Inferred ancient cDNA sequences were cloned into a mammalian expression plasmid and transfected into cell lines used for recombinant coagulation factor manufacturing. FIX expression levels, defined in the current study as the FIX activity (IU) accumulation per 24 hours per million cells, were measured after 24 hours in serum-free media by 1-stage coagulation assay. Two of the 8 AnFIX constructs, designated An96 and An97, demonstrated 12- and 7.5-fold greater FIX activity, respectively, than an identically codon-optimized hFIX (**Figure 4.2A**).

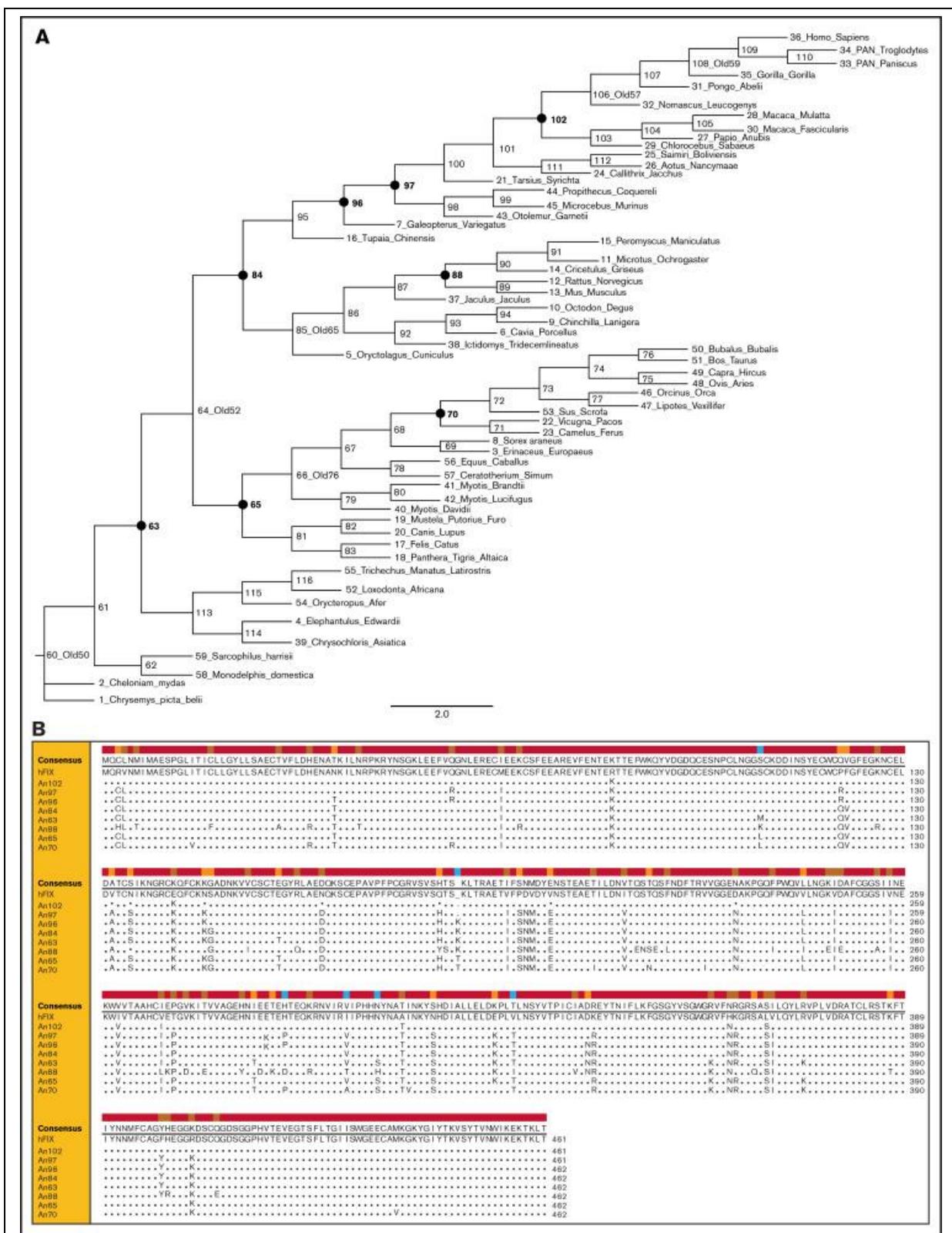


Figure 4.1: AnFIX phylogeny and sequences. (A) Phylogeny and ancestral FIX sequences were constructed using extant FIX sequence information. A phylogram of the tree is shown for display purposes. Closed circles and bold numbers denote node designations for reconstructed AnFIX molecules. (B) Amino acid sequence alignment of AnFIX sequences was generated using Clone Manager 9 software. Sequence variations from hFIX are shown in the lower rows.

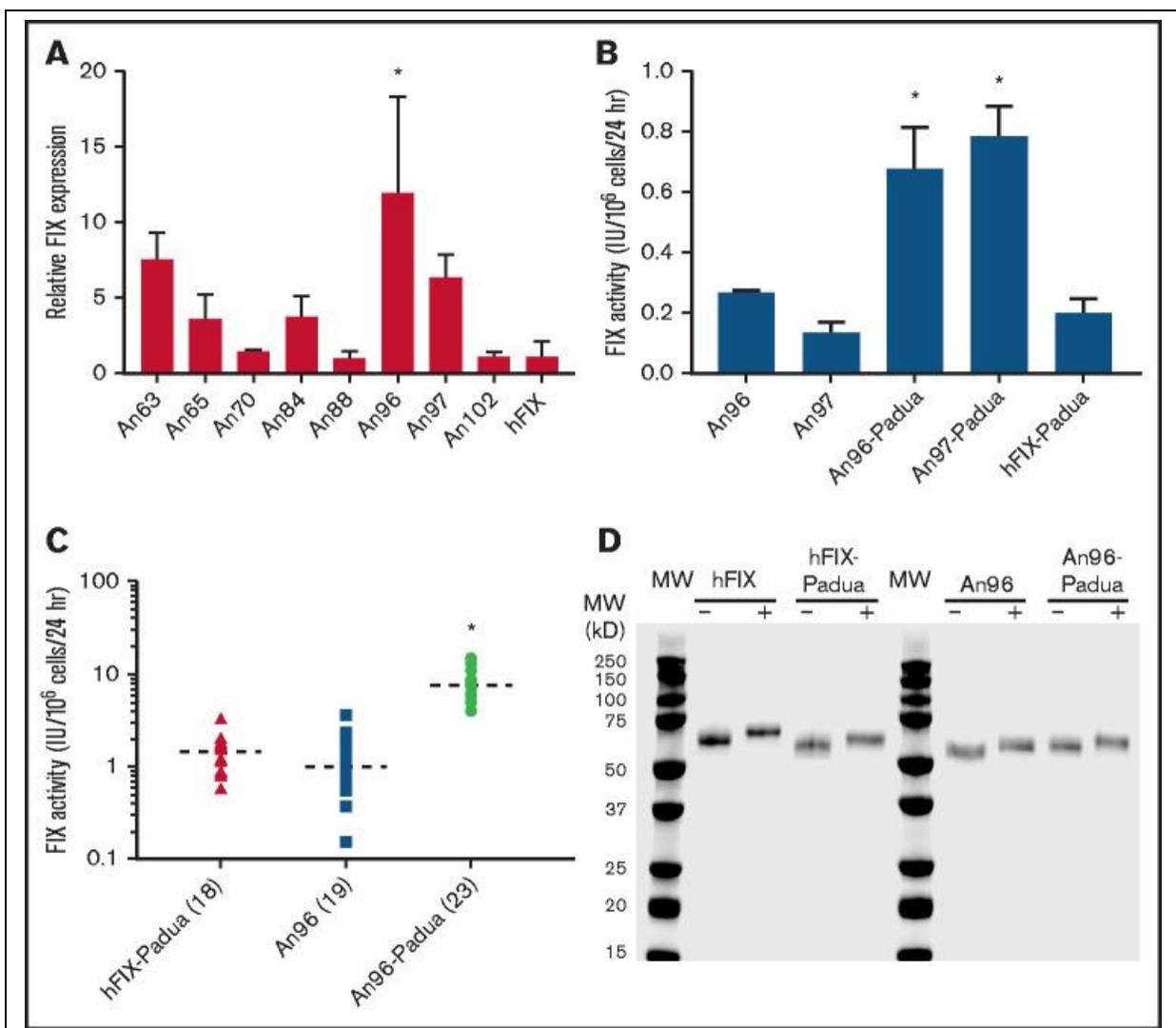


Figure 4.2: Recombinant AnFIX expression and characterization. Recombinant AnFIX expression was determined in transient (A-B) and stable (C) mammalian cell expression systems with and without R383L (Padua) mutation (B-C) by measuring FIX activity in the conditioned medium by 1-stage coagulation assay and manually counting the cells present in each well by hemocytometer to obtain a FIX activity (IU) accumulated per million cells per 24 hours (expression value). Error bars indicate sample standard deviation. Closed shapes represent individual clones, with sample size in parentheses and mean as a dashed line. Statistical analysis was performed using a 1-way analysis of variance with post hoc Tukey multiple comparisons test. (D) Highly purified recombinant hFIX, hFIX-Padua, An96, and An96-Padua were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and visualized by Coomassie blue staining under reducing (+) and nonreducing (–) conditions after incubation at 95°C for 5 minutes with or without dithiothreitol, respectively. *AnFIX variants with significant expression above the other variants ($P < .05$).

Samelson-Jones et al (190) recently reported detailed biochemical studies on recombinant hFIX-Padua showing that interaction with activated FVIII (FVIIIa) is critical to specific activity enhancement, which is not observed when FVIIIa is replaced with the bispecific mimetic Hemlibra. On the basis of the earlier described benefits observed in clinical gene therapy trials incorporating R338L and the finding that inferred AnFIX sequences do not possess a substitution at R338, R338L (Padua) was incorporated into An96 and An97 to determine if the relevant mechanisms conferring increased activity to An96 and An97 overlapped, inhibited, or synergized with the seven- to eightfold increase in the FVIIIa cofactor-dependent specific activity of hFIX-Padua (190; 191). An96-Padua and An97-R338L displayed FIX expression levels up to fourfold higher than hFIX R338L, indicative of the presence of functional residues independent of and additive to the known specific activity benefit conferred by R338L (**Figure 4.2B**). Given that there are only 2 amino acid differences between the mature An96 and An97 variants and An96 trended toward higher expression in the initial screen (**Figure 4.2A**), we conservatively selected An96 and An96-Padua as the lead candidates for further study.

Generation and characterization of clonal cell lines stably expressing recombinant proteins represents a more rigorous and reliable method than transient transfection (103-105; 113; 129; 190-193). Stable HEK293 cell lines expressing hFIX-Padua, An96, and An96-Padua were generated, and a collection of 18 to 23 clones each were assessed for FIX production. An96-Padua demonstrated fivefold greater mean productivity compared with hFIX-Padua and An96 without the Padua mutation, supporting the independent and additive enhancement of An96 substitutions with R338L (**Figure 4.2C**). Commercial enzyme-linked immunosorbent assay detection of An96 was limited and did not generate a parallel concentration-signal curve to that obtained with hFIX (**supplemental Figure S4.1**). Therefore, it was not possible to estimate the specific activity of the AnFIX variants in transfection

experiments to establish whether the increased activity resulted from improved biosynthesis/secretion or specific activity enhancement.

Biochemical and pharmacological studies of An96 FIX: Conditioned media from the top expressing hFIX-Padua, An96, and An96-Padua HEK293 stable clones were harvested for recombinant FIX purification. In total, 17 543 IU of An96-Padua FIX activity was harvested in 2.86 L of conditioned media (6.13 IU/mL). Mixed-mode and anion exchange chromatography were employed sequentially to purify recombinant An96 and An96-Padua to near homogeneity as evidenced by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (**Table 4.1; Figure 4.2D; Supplementary Tables S4.2 and S4.3**). The specific activity of recombinant An96-Padua (6450 IU/mg) was ~32 times greater than that of commercial recombinant hFIX (200 IU/mg), 3.2-fold greater than that of hFIX-Padua (2013 IU/mg), and 3.1-fold greater than that of An96 (2090 IU/mg) by 1-stage coagulation assay. The observation that the specific activity of An96 and that of hFIX-Padua are similar, while that of An96-Padua is an additional 3.1- to 3.2-fold higher, suggests that the primary mechanism of increased An96 productivity is enhanced specific activity and not increased biosynthesis/secretion. Enhanced specific activity of An96-Padua was further demonstrated using calibrated automated thrombography (**Figure 4.3A-E; Supplementary Figure S4.3**). Respective An96-Padua, An96, hFIX-Padua, and recombinant hFIX thrombograms also support the procoagulant activity relationship of An96-Padua > An96 ≈ hFIX-Padua > hFIX. No differences were observed in the activation kinetics of An96-Padua compared with hFIX-Padua upon treatment with FXIa (**supplemental Figure S4.2**).

Table 4.1: Purification of recombinant An96-Padua

Sample	Volume (mL)	A_{280}^*	Total A_{280}	Activity IU/mL	IU	IU/ A_{280}	Yield, %	Fold purification
Harvest	2860	4.307	12318	6.13	17543	1.42	100	1
Multimodal chromatography peak	100	0.240	24	134.7	13471	561	76	395
Anion exchange chromatography peak	10	0.768	7.68	748.5	7484	975	43	687

* A_{280} measurements are A_{320} corrected.

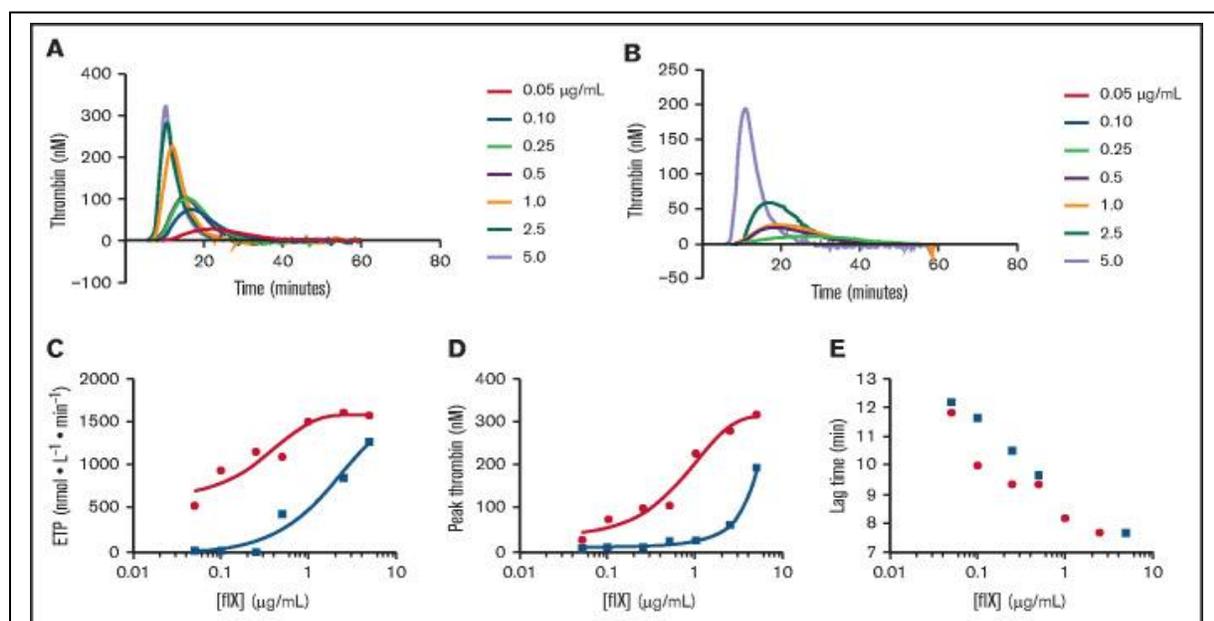


Figure 4.3: Thrombin generation analysis. Calibrated automated thrombography was performed using varying concentrations of recombinant hFIX (blue squares and lines) or An96-Padua (red circles and lines) spiked into hemophilia B plasma spanning the severe hemophilia B boundary ($\sim 1\%$) to 100% normal pooled plasma hFIX range (0.05-5 $\mu\text{g/mL}$). Concentration-response curves of endogenous thrombin potential (ETP) (C), peak thrombin level (D), and lag time (E) for recombinant An96-Padua and hFIX are depicted.

Next, we sought to determine if the enhanced in vitro properties of An96 and An96-Padua also translated to in vivo activity enhancement in the setting of recombinant FIX infusion

into hemophilia B mice. The up-and-down staircase method combined with a saphenous vein bleed challenge was used to estimate 50% effective dose (186-189). Consistent with and even exceeding the in vitro findings, An96-Padua demonstrated an 50% effective dose of 4.8 U/kg, or 740 ng/kg, compared with 8.7 U/kg, or 44 µg/kg, for both hFIX and hFIX-Padua (**supplemental Figure S4.4**). Given that hemophilia B mice were dosed using equivalent FIX activity and not protein concentration, these results convey a 60- and sixfold potency enhancement (on a mass basis) over hFIX and hFIX-Padua, respectively, for An96-Padua in an in vivo model of hemophilia B.

AAV-AnFIX gene therapy: On the basis of the improved activity of recombinant An96-Padua in coagulation assays and murine infusion studies, An96-Padua was investigated in the setting of liver-directed AAV gene therapy. Two sets of AAV2/8 vectors were constructed (**Figure 4.4A**). The first contained a minimal liver-specific synthetic 146bp promoter, termed HCB, while the second contained a larger more potent promoter termed HHS4 (85; 182). Vectors encoding the LCO-An96, LCO-An96-Padua, or LCO-hFIX-Padua transgene driven by either promoter were produced in HEK293 cells after transient triple-plasmid transfection and purified using iodixanol gradient ultracentrifugation-based separation and concentration. Vector titers were obtained by qPCR of DNA obtained from the final vector preparations (**Table 4.2**). Additionally, physical characterization of the HHS4 promoter-containing vectors was performed using SEC-MALS analysis. As shown in **Table 4.2**, the total AAV vector concentrations varied from $(3.37 \pm 0.28) \times 10^{12}$ to $(2.11 \pm 0.01) \times 10^{13}$ vp/mL. The ratio of full/total capsids ranged from 0.86 ± 0.05 to 0.99 ± 0.34 . The molar mass of each vector, capsid, and nucleic acid genome is reported in **Table 4.3**. On the basis of these findings, the qPCR titer data provided by the manufacturer seemed to overestimate the full capsid titer by 3.4- to 3.7-fold.

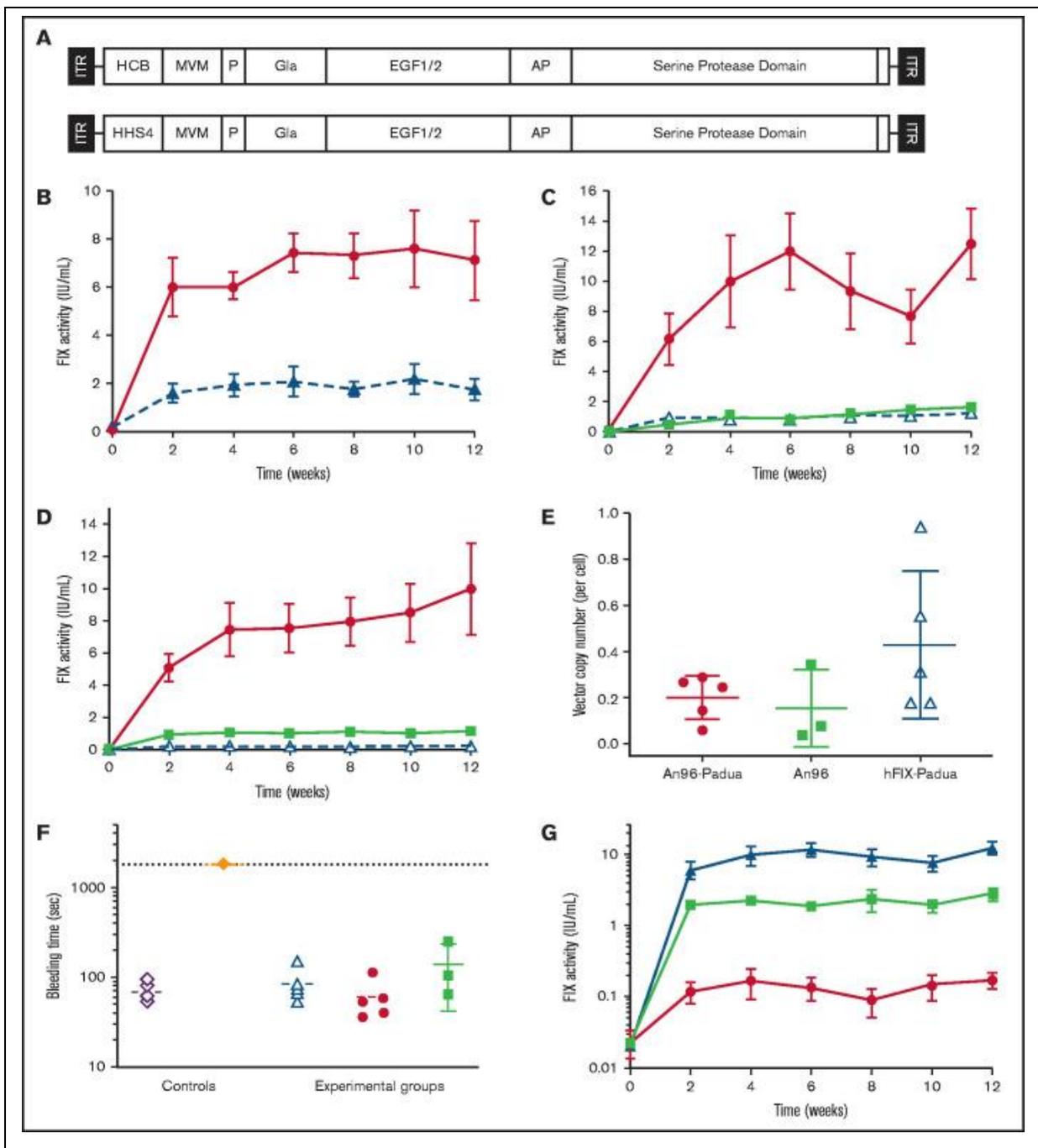


Figure 4.4: AAV-AnFIX gene therapy. (A) AAV vectors incorporating AAV2 inverted terminal repeats, either the HCB (146 nt) or HHS4 (294 nt) promoter, a minute virus of mice (MVM) intron (92 nt), a FIX transgene, and a synthetic β -globin polyadenylation signal are depicted. The predicted HCB- and HHS4-containing ssDNA AAV genome sizes are 2003 and 2151 nt, respectively. (B) Hemophilia B mice were injected IV with 5×10^{12} vg/kg AAV2/8 encoding An96 Padua-LCO (closed circles) or hFIX-Padua (open triangles) driven by the highly compact, liver-directed HCB promoter ($n = 6$ per group). Plasma was collected biweekly for 12 weeks, and FIX activity was determined by 1-stage clotting assay. (C) Hemophilia B mice were injected IV with a 37-fold lower dose (1.4×10^{11} vp/kg) of AAV2/8-HHS4-An96-Padua (closed circles), AAV2/8-HHS4-An96 (closed squares), or AAV2/8-HHS4-hFIX-Padua (open triangles; $n = 4-6$ per group). Plasma was collected biweekly for 12 weeks, and FIX activity was determined by 1-stage clotting assay (C) or chromogenic assay (Rox Factor IX; Diapharma, West Chester, OH) (D). (E) AAV vector copy number was determined from liver tissue genomic DNA using qPCR. No significant differences were observed among the groups (1-way analysis of variance [ANOVA]; $P = .2$). (F) Saphenous vein bleeding challenge was performed on mice from each group (from left to right, open diamonds represent wild-type C57Bl/6; closed diamonds, untreated hemophilia B; open triangles, hFIX-Padua transgene; closed circles, An96-Padua transgene; closed squares, An96 transgene). The average time to clot for each mouse was measured over 30 minutes. Comparisons among the groups were made by 1-way ANOVA and post hoc Holm-Sidak testing. All groups were significantly different to the untreated hemophilia B (negative control) group ($P < .05$). (G) Dose-response curves were generated by administration of \log_{10} doses of AAV2/8-HHS4-An96-Padua at 1.4×10^9 (circles), 1.4×10^{10} (squares), or 1.4×10^{11} vp/kg (triangles) to hemophilia B mice ($n = 5-7$ per group). Plasma FIX activity was measured by 1-stage clotting assay at biweekly intervals.

Table 4.2: AAV Vector Quality Attributes

AAV sample	qPCR titer, vg/mL	Total AAV concentration, Cp/mL	Capsid occupancy SEC-MALS, vg/Cp	Capsid occupancy qPCR, vg/Cp
hFIX-Padua	7.74×10^{13}	$2.1 \pm 0.1 \times 10^{13}$	0.86 ± 0.05	3.67
An96	2.52×10^{13}	$7.5 \pm 0.5 \times 10^{12}$	0.86 ± 0.03	3.36
An96-Padua	1.15×10^{13}	$3.4 \pm 0.3 \times 10^{12}$	0.99 ± 0.34	3.41

Values are mean and standard deviation of 2 injections except An96-Padua ($n = 1$ because of lower titer and limited sample availability). Cp, capsid particle.

Table 4.3: AAV Vector Molar Mass and Composition

AAV sample	Vector M_w , MDa	Capsid M_w , MDa	Nucleic acid M_w , kDa
hFIX-Padua	4.53 ± 0.04	3.86 ± 0.01	666 ± 36
An96	4.42 ± 0.08	3.81 ± 0.05	609 ± 32
An96-Padua	4.62 ± 0.08	3.74 ± 0.05	914 ± 390

Values are mean and standard deviation of 2 injections except An96-Padua (n = 1 because of lower titer and limited sample availability).

Identical AAV2/8 vectors encoding the HCB promoter driving either an LCO-An96-Padua or LCO-hFIX-Padua transgene were administered to hemophilia B mice at a dose of 5×10^{12} vg/kg based on qPCR titer, because SEC-MALS data were not available (n = 6 mice per group; **Figure 4.4A**). Analysis of plasma FIX activity over the 12 weeks post-administration demonstrated fourfold greater mean FIX activity in mice administered AAV-An96-Padua compared with mice administered AAV-hFIX-Padua (**Figure 4.4B**). The minimal HCB promoter was designed for an AAV-FVIII platform, where vector genome size was a major constraint. Because AAV-FIX vectors do not have the same size limitation, a larger, stronger HHS4 promoter was incorporated in place of HCB. Using a significantly lower dose of 1.4×10^{11} vp/kg based on SEC-MALS (or $\sim 5 \times 10^{11}$ vg/kg via qPCR), AAV2/8-HHS4-An96-Padua-LCO-treated mice displayed FIX activity levels ~ 10 -fold higher than AAV-An96-LCO- or AAV-hFIX-Padua-LCO-treated hemophilia B mice (n = 4-6 mice per group) over the course of the 12-week study (**Figure 4.4C**). The latter 2 vectors produced similar plasma FIX activity levels by 1-stage clotting assay, again demonstrating that An96 sequences confer an enhancement similar to, but independent of, R338L. Previous studies documented a 1.8-fold discrepancy between the 1-stage clotting assay and chromogenic assay results for hFIX-Padua, with the 1-stage clotting assay producing the higher result (194). Consistent with this observation, plasma samples from animals treated with AAV-hFIX-Padua displayed >50% lower FIX activity across all timepoints when measured using a chromogenic assay (**Figure 4.4D**). In contrast, AAV-An96 and AAV-An96-Padua samples produced more consistent results, although slightly lower activity was

observed for An96-Padua using the chromogenic assay. These data again support the conclusion of a differential enhancement mechanism for An96 as compared with R338L. However, because of the lack of parallelism between dose-response curves for An96 in a parallel fashion to using a commercial enzyme-linked immunosorbent assay, it is not possible to discriminate between greater An96/An96-Padua production from the genetically modified cells or enhanced specific activity or the possibility of a combination of both.

Consistent with previous results using similar doses of AAV2/8 vectors in mice, liver vector copy number per diploid genome equivalent was <1 , and no significant differences were observed between the groups (**Figure 4.4E**). Before euthanasia, a subset of mice from each group were subjected to the saphenous vein bleeding challenge to assess the level of phenotypic correction. Each of the vectors conferred significant phenotypic correction that was indistinguishable from wild-type C57Bl/6 mice (**Figure 4.4F**). Given that the 1.4×10^{11} vp/kg dose resulted in supraphysiologic FIX activity levels, a decreasing dose-response study was performed to further characterize vector potency and estimate the minimal effective dose. AAV2/8-HHS4-An96-Padua-LCO vector was administered at doses ranging down to 1.4×10^9 vp/kg ($n = 5-7$ mice per group). A clear dose response was observed with mice displaying durable FIX activity levels in the ranges of 10, 2, and 0.2 IU/mL for the 1.4×10^{11} , 1.4×10^{10} , and 1.4×10^9 vp/kg groups, respectively, over the 12-week study period (**Figure 4.4G**). The extended durability of plasma FIX activity in all current AAV-FIX hemophilia B mouse groups suggest the absence of immune responses to the transduced cells or transgene products. TEG analysis of AAV2/8-An96-Padua-treated mice was performed to provide additional evidence of hemostatic correction (**supplemental Figure S4.5**). Other than the TEG angle in the 1.4×10^9 vp/kg AAV-HHS4-An96-Padua group, no other TEG parameters measured for any of the treatment groups were significantly different from those obtained for positive control C57Bl/6

mice. These data demonstrate that low levels of An96-Padua promote stable clot formation comparable to that in wild-type control mice.

4.5 Discussion

An emerging concept in the development of protein drugs and gene therapies is the requirement for bioengineering of the native protein and nucleic acid sequences to improve therapeutic performance. For example, virtually all gene therapies are codon optimized to improve production of the respective transgene products. This engineering is performed without a priori knowledge of the respective codons or other RNA sequences that are limiting expression or a weak understanding of the mechanism of action (195). However, the striking improvements observed upon empirical testing serve as evidence that many human cDNA and messenger RNA sequences have not evolved to maximize protein output (85; 196; 197). This can be rationalized by the common requirement for endogenous proteins to maintain a concentration and specific activity window of functionality where there exists sufficient activity to avoid deficiency and not enough to cause pathogenic effects. Coagulation FVIII and FIX are examples of this principle. Deficiencies <50% of the normal levels (~1 and 90 nM, respectively) result in bleeding severity that inversely correlates with their concentration. On the other end of the spectrum, levels >150% of normal increase the lifetime risk of thrombotic events, which may serve as a selective disadvantage, as has been suggested previously by us and others (129; 190; 198).

When proteins and genes are taken out of their endogenous biosynthetic context and used as pharmaceuticals, the parameters established by nature may not be optimal. In gene therapy, vector potency is the key optimization parameter. Although significant progress has been made toward making more efficient gene transfer vectors, clinical AAV gene therapy often requires vector doses equal to or vastly exceeding the total number of cells in the human body to achieve efficacy. Some dose levels seem near or even exceed the maximum tolerated dose

based on reported adverse events (199). Therefore, optimization of the internal vector components is advantageous. Recently, we reported nucleic acid design strategies aimed at optimizing transgene expression specifically for the target cell type of interest. For example, LCO generated significant improvements in the expression of hFIX as well as several FVIII variants, including an AnFVIII (85; 129). The second arm of internal vector bioengineering, optimization of the transgene product sequence, has been the slowest to advance into clinical testing. Two main challenges exist to this approach. The first relates to immunogenicity risks associated with incorporation of nonhuman protein sequences. The development of anti-protein drug antibodies, referred to as inhibitors in the case of FVIII and FIX, has been observed clinically in many settings, including monoclonal antibody therapy, coagulation factor deficiencies, and inborn errors of metabolism. In some cases, this is not completely unexpected because of the use of xenoproteins (eg, murine monoclonal antibodies) or the naïve immune status of the patient in the case of null mutations (ie, cross-reactive material negative). However, in the setting of gene therapy, where the therapeutic protein arises from within the body, the immunogenicity risks are less clear, and no predictive preclinical models have been established. To date, no inhibitors have been observed in >100 hemophilia A and B patients who have received liver-directed AAV gene therapy in at least a dozen independent clinical trials. However, all patients had extensive prior exposure to FVIII and FIX replacement products, without a history of inhibitor development. It is important to recognize that in a majority of these cases, the gene therapy product was both codon optimized as well as protein sequence optimized to include B domain deletion in the case of FVIII or the Padua mutation in the case of FIX. In the current study, no AAV-An96-Padua or AAV-human FIX-Padua vector-treated animals were observed to have undergone a rapid decrease in plasma FIX activity indicative of an inhibitory humoral immune response to the bioengineered transgene product. However, the clinical immunogenicity risks of ASR variants such as An96-Padua remain to be established and

likely will require clinical testing in the relevant delivery settings to understand product-specific risks (eg, protein infusion or gene therapy).

A second challenge to protein sequence engineering is the limited, high-resolution structure/function information available and the lack of effective methods for *in silico* prediction of functional protein enhancements. Instead, protein engineering methods have relied heavily on high-throughput screening after random mutagenesis or directed evolution. Although these techniques have led to major improvements in commercial biotechnology products, they have yet to penetrate protein therapeutics and gene therapies, presumably because of the structural and mechanism-of-action complexities mentioned previously. As in the case of the Padua mutation, an effective strategy has been to use information from nature. Similarly, we and others have developed bioengineered therapeutic candidates using information obtained from interspecies protein diversity as well as inferred ancient protein sequences (107; 126; 129; 159). Employing ASR, FVIII variants with improved biosynthesis, specific activity, and stability were identified without any information other than extant FVIII sequences found in publicly available databases. The power of this approach relies on the ability to rapidly span an evolutionary space where protein functionality is uniformly retained and often enhanced, presumably to thrive under physiological, sociological, and environmental pressures.

ASR was applied to FIX with the goal of identifying FIX variants that can improve the potency of AAV-FIX gene therapy vectors. In the initial screen of 8 AnFIX variants inferred from nodes corresponding to early placental mammals, supraprimates, primates, rodents, carnivores, and ungulates, 2 AnFIX variants (An96 and An97), inferred as the last common ancestors of humans and lemurs, were discovered to possess 10-fold enhanced FIX activity beyond that achieved with hFIX-Padua while retaining >90% sequence identity. These nodes are not too distant but do not overlap with the AnFVIII node that possessed the greatest activity enhancement, An84 (previously designated An53 by Zakas *et al* (129)). Comparison of the

AnFIX sequences with other reported bioengineered hFIX variants reveals that novel functional residues are likely to be identified. For example, none of the residues addressed by Quade-Lyssy *et al* (200) are altered in An96, and only 1 matching substitution is shared with the TripleL variant described by Kao *et al* (201) (V86A). Of note, An96 does possess an alternative substitution, E277K, compared with the E277A in the TripleL variant described by Kao *et al*. On the basis of the specific activity estimates provided for the Triple and TripleL variants (2258 and 4309 IU/mg), An96 and An96-Padua display higher specific activity, suggesting a role for additional substituted residues present in the ancestral variants.

In the previous AnFVIII studies as well as the current AnFIX study, these inferred ancient variants were studied in the context of human plasma, purified human or animal coagulation factors, or in vivo murine hemophilia models. Therefore, interspecies compatibility of the interacting clotting factors was essential and consistently observed in both studies. Although, it has not been investigated yet, it will be of interest to study the activity of the FIX An96 variant in the presence of the FVIII An84 variant, which may help define the mechanisms of enhancement. The complexity of the FVIII and FIX proteins as well as the substrate, FX, and the requirement for their tripartite, macromolecular interaction on a phospholipid membrane interaction has contributed to the difficulty in determining high-resolution structure/function relationships. It is possible that ASR may enable significant advancements in this longstanding area of basic research investigation. However, as a direct approach to protein drug and gene therapy optimization, the FIX ASR data reported in the current study validate this powerful approach.

One aspect of AAV gene therapy that makes intervector comparisons uncertain is the lack of universal AAV standards and methods for assessing vector titer and quality (202). Even within individual clinical programs using identical vector designs but different manufacturing lots, striking outcome variability has been observed. In both preclinical and clinical studies, AAV

vectors are administered on a vg/kg dose basis using vector titers (vg/mL) determined using qPCR or digital droplet PCR. Other vector characteristics such as particle size, molecular weight, and empty/full AAV genome-containing capsids typically are not reported despite the potential impact on performance. As an attempt to address this critical issue in the current study, we performed SEC-MALS analysis on the AAV2/8-HHS4-FIX vectors. SEC-MALS provides a rapid, automated, label-free, nondestructive method to measure the 3 primary quality attributes of AAV vectors (ie, molar mass, composition, and concentration). All 3 AAV-HHS4-FIX vector preparations displayed similar physical characteristics. SEC-MALS also revealed that vector titers seemed to be overestimated several-fold by qPCR (**Table 4.3**). Because of the limited vector samples available in this study, it will be important to verify these results using multiple vector designs and independent manufacturing runs in future studies.

The primary goal of the current study was to identify FIX variants that can improve the potency of clinical AAV-FIX gene therapy vectors. Clinical data suggest that additional improvements beyond hFIX-Padua will be required to achieve complete correction (ie, 100% normal FIX activity levels) at vector doses that can avoid liver inflammation and demonstrate long-term safety. Furthermore, the ability to lower the viral dose through potency optimization will translate to lower manufacturing cost per dose and greater product supply. As the AAV gene therapy field moves toward multiple administrations of vector products, the cumulative vector dose may be an important safety aspect as well. Collectively, our data demonstrate ASR to be a successful strategy to improve recombinant FIX and AAV-FIX performance, which should be directly translatable to other clinical protein drug and gene therapy development endeavors.

4.6: Acknowledgments

This work was supported by grants HL137128 (H.T.S., C.B.D.) and U54 HL141981 (P.L.) from the National Heart, Lung and Blood Institute, National Institutes of Health (NIH);

grant R01AR069137 (E.A.G.) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH; and grant RGP0041 (E.A.G.) from the Human Frontier Science Program.

Conflict-of-interest disclosure: C.E.R., H.T.S., E.A.G., and C.B.D. are inventors on a patent application describing the ancestral FIX technology filed by Emory University, Children's Healthcare of Atlanta, and Georgia Institute of Technology. H.T.S. and C.B.D. are inventors on a patent for liver-directed codon-optimization and promoter technology filed by Emory University and Children's Healthcare of Atlanta. H.T.S. and C.B.D. are cofounders of Expression Therapeutics and own equity in the company. Expression Therapeutics has obtained licenses for the ancestral FIX, liver codon optimized FIX, and synthetic liver-directed promoter intellectual property. The terms of this arrangement have been reviewed and approved by Emory University in accordance with its conflict-of-interest policies. The remaining authors declare no competing financial interests.

4.7: Supplemental Information

Supplementary Table S4.1: AnFIX sequence variation

FIX	Substitutions in complete sequence*	Substitutions in mature protein	Mature protein identity (%)
Human	0	0	100
An102	8	7	98
An97	40	37	91
An96	43	39	90
An84	42	38	90
An65	46	42	89
An63	48	44	89
An70	54	48	88
An88	74	65	84

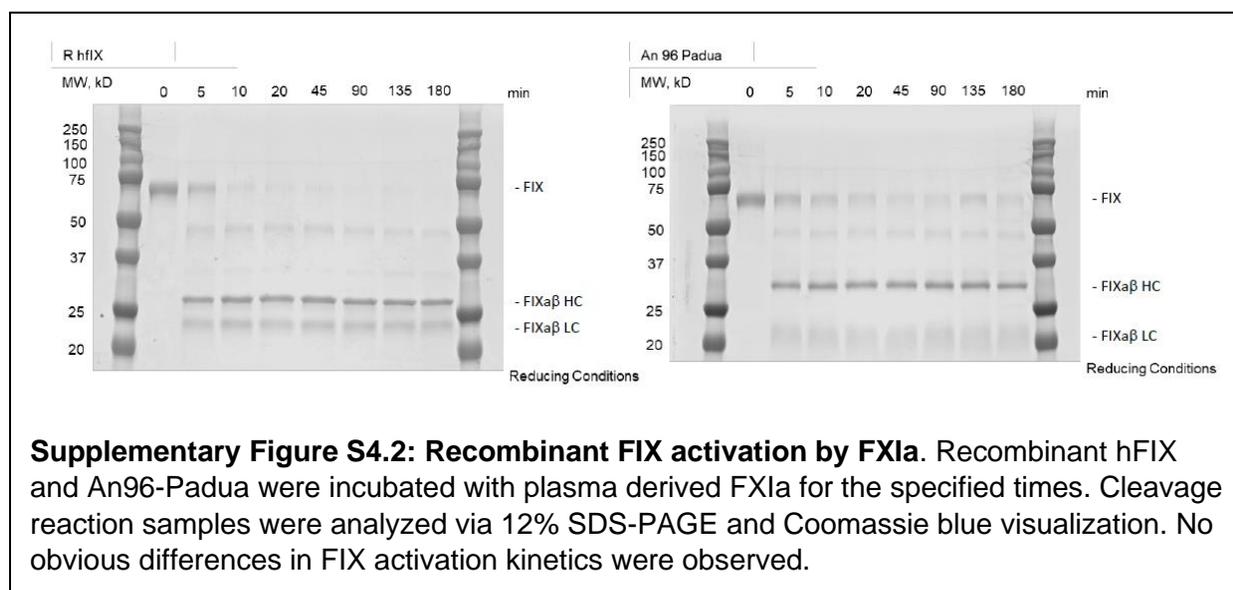
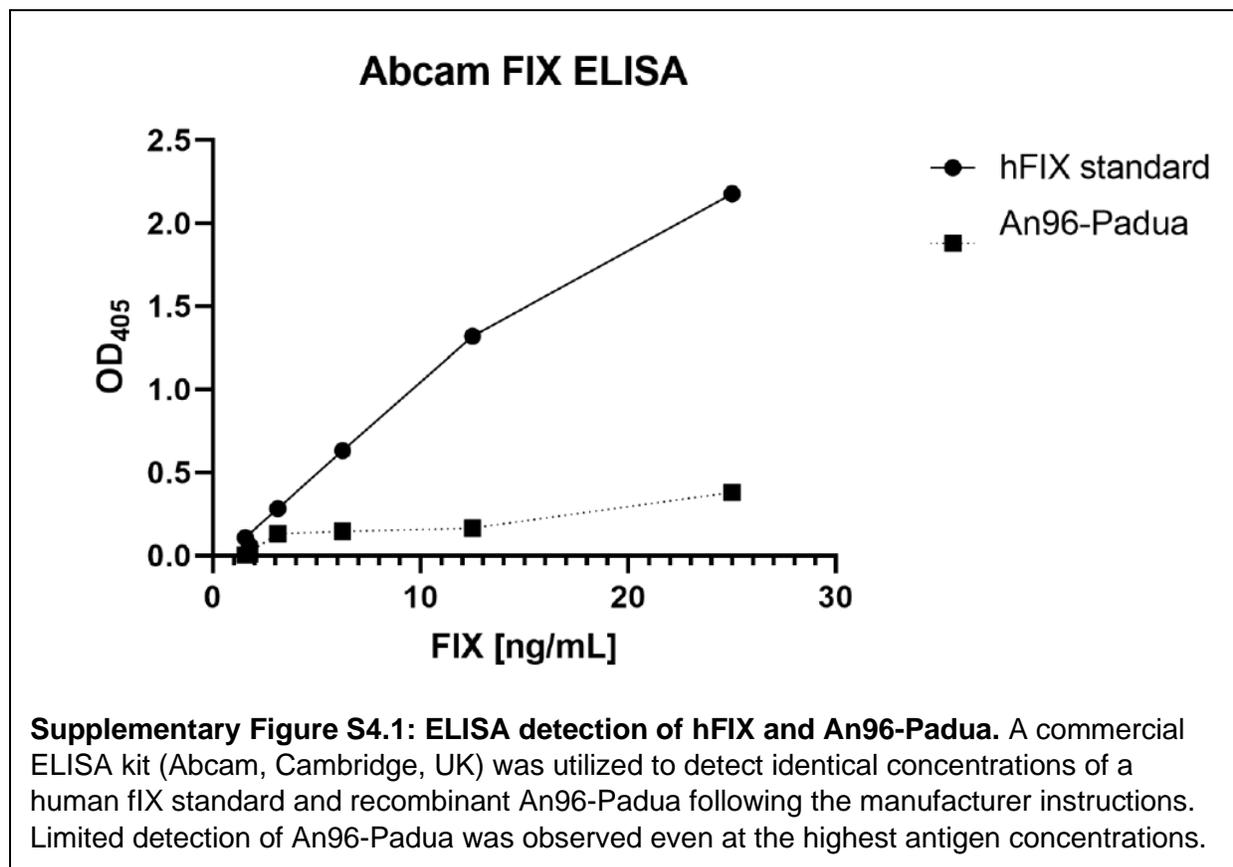
* Includes signal prepeptide and propeptide

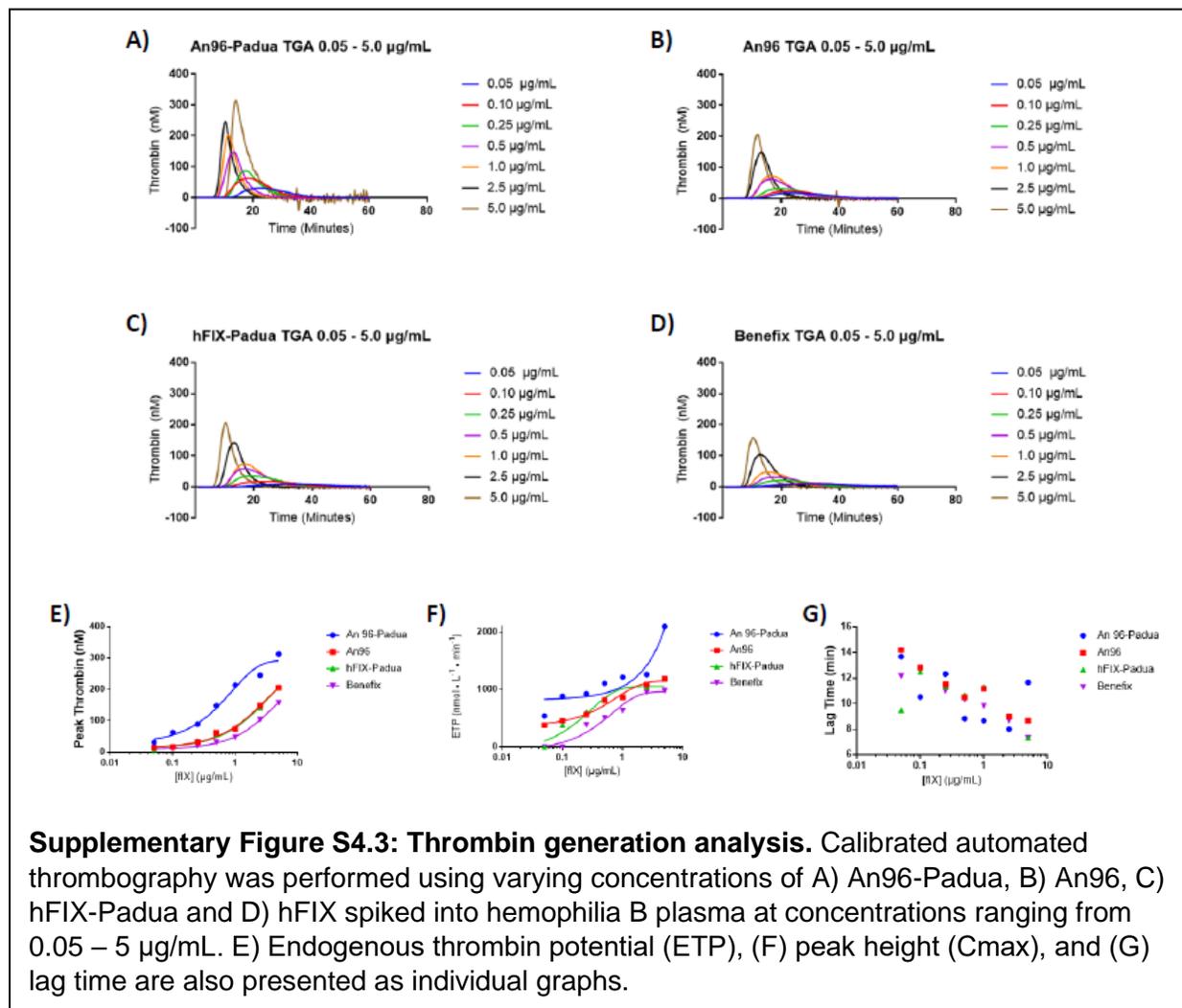
Supplementary Table S4.2: hFIX-Padua purification table

Sample	Volume (mL)	A280	Total A280	Activity (IU/mL)	IU total	IU/A280	% yield (IU total over IU total)	Fold Purification
Harvest (IX-select load)	1850	4.301	7960	5.13	9490	1.19	100	1
IXSelect Elution and final product	3.00	10.4	31.2	2850	8540	274	90	230
MMC Peak	50	0.100	5.02	33	1650	328	35	275
Q Peak (Best Fraction)	0.325	1.02	0.332	1550	504	1520	11	1280

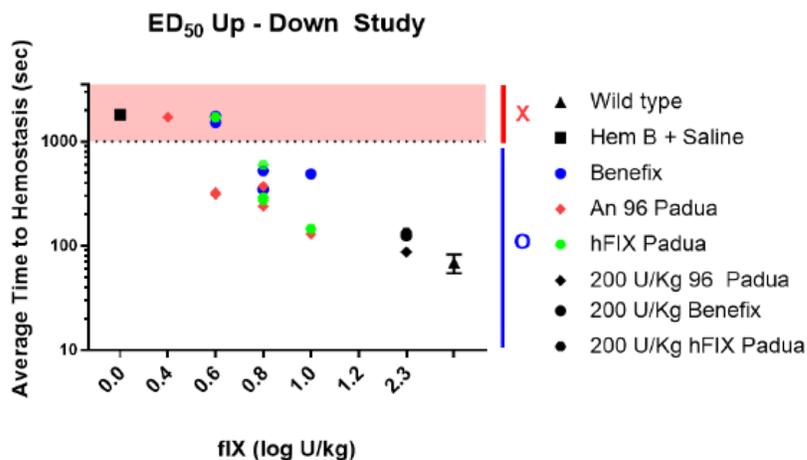
Supplementary Table S4.3: An96 purification table

Sample	Volume (mL)	A280/mL	Total A280	Activity (IU/mL)	IU total	IU/A280	% yield	Fold Purification
Harvest	1900	4.301	8170	6.29	11900	1.46	100	1
MMC Peak	240	0.516	124	46.6	11200	90.3	94	61.9
Q Peak	5.00	0.228	1.14	350	1750	1540	15	1050





A.



B.

FIX Efficacy: Staircase Assay - Saphenous Vein bleed challenge									
Dixon Up-and-Down Method for Small Samples									
FIX	Dose (u/kg)	Log(Dose)	Response: X = bleed, O = no bleed						
An 96 Padua	25.1	1.4							
	15.8	1.2							
	10	1.0	O						
	6.3	0.8		O				O	
	3.98	0.6			O		X		O
	2.51	0.4				X			
	1.58	0.2							
Hem B + Saline	0	0	X						
Hem B + An96	200	2.3	O						
$ED_{50} = Xf + k*d \quad d = 0.2$ $N' = 7 \quad N = 5 \quad OXOXO$ $Xf = 0.6, \quad k = .305$ $\log ED_{50} = 0.6 + (0.305*0.2) = 0.661$ $ED_{50} = 10^{0.661} = 4.58 \text{ U/kg}$									

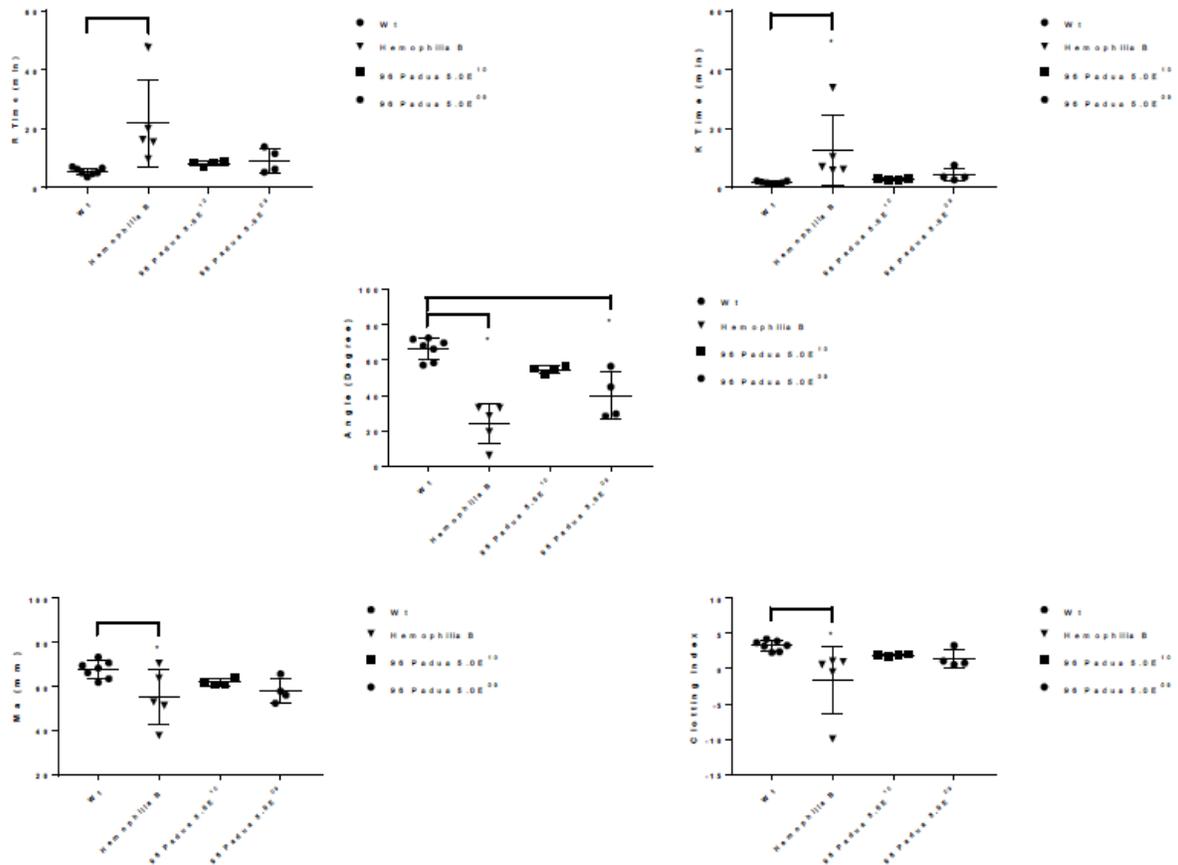
C.

FIX Efficacy: Staircase Assay - Saphenous Vein bleed challenge									
Dixon Up-and-Down Method for Small Samples									
FIX	Dose (u/kg)	Log(Dose)	Response: X = bleed, O = no bleed						
Benefix	25.1	1.4							
	15.8	1.2							
	10	1.0	O						
	6.3	0.8		O		O		O	
	3.98	0.6			X		X		
	2.51	0.4							
	1.58	0.2							
Hem B + Saline	0	0	X						
Hem B + Benefix	200	2.3	O						
$ED_{50} = Xf + k*d \quad d = 0.2$ $N' = 6 \quad N = 5 \quad OXOXO$ $Xf = 0.8, \quad k = .701$ $\log ED_{50} = 0.8 + (0.701*0.2) = 0.9402$ $ED_{50} = 10^{0.9402} = 8.71 \text{ U/kg}$									

D.

hFIX Padua Specific Activity = 1904.461159									
FIX Efficacy: Staircase Assay - Saphenous Vein bleed challenge									
Dixon Up-and-Down Method for Small Samples									
FIX	Dose (u/kg)	Log(Dose)	Response: X = bleed, O = no bleed						
hFIX Padua	25.1	1.4							
	15.8	1.2							
	10	1.0	O						
	6.3	0.8		O		O		O	
	3.98	0.6			X		X		
	2.51	0.4							
	1.58	0.2							
Hem B + Saline	0	0	X						
Hem B + hFIX Padua	200	2.3	O						
$ED_{50} = Xf + k*d \quad d = 0.2$ $N' = 6 \quad N = 5 \quad OXOXO$ $Xf = 0.8, \quad k = .701$ $\log ED_{50} = 0.8 + (0.701*0.2) = 0.9402$ $ED_{50} = 10^{0.9402} = 8.71 \text{ U/kg}$									

Supplementary Figure S4.4: ED₅₀ determinations for hFIX (A and B), An96-Padua (A and C) and hFIX-Padua (A and D). ED₅₀ estimation was made using the up-and-down method as described by Dixon *et al.* Hemostatic challenge was performed via saphenous vein challenge modified from ones previously described. Mice displaying an average time for hemostasis > 3 standard deviations (s.d.) of the mean value obtained for wild-type C57Bl/6 mice in an identical challenge, without exogenous FIX administration, were categorized as 'bleeders'.



Supplementary Figure S4.5: TEG of whole blood obtained from control and AAV2/8-An96-Padua treated mice. TEG was performed using whole citrated blood and the TEG 5000 Thrombelastograph Hemostasis Analyzer system according to the manufacturer instructions for native activation. All output parameters are displayed in the various figure panels. Statistical comparisons were made using one-way ANOVA and post hoc Holm-Sidak multiple comparisons testing. Connecting bars and asterisks (*) denote cases where $p < 0.05$. Other than Angle in the 1.4×10^9 vp/kg AAV-HHS4-AN96-Padua group, all other parameters were not significantly different than positive control C57Bl/6 mice for the each of the treatment groups.

Chapter 5

Humanization and functional characterization of enhanced coagulation factor IX variants identified through ancestral sequence reconstruction

This research was accepted for publication in *Journal of Thrombosis and Haemostasis*.

Christopher W. Coyle, Kristopher A. Knight, Harrison C. Brown, Stephan N. George, Gabriela Denning, Gianna M. Branella, Kenneth C. Childers, P. Clint Spiegel, H. Trent Spencer, Christopher B. Doering

Humanization and functional characterization of enhanced coagulation factor IX variants identified through ancestral sequence reconstruction

© 2023 by International Society on Thrombosis and Haemostasis

5.1 Abstract

Laboratory resurrection of ancient coagulation factor IX (FIX) variants generated through ancestral sequence reconstruction (ASR) led to the discovery of an ancient FIX variant, designated An96, that possesses enhanced specific activity independent of, and additive to that provided by human p.Arg384Lys, referred to as FIX-Padua. The goal of the current study was to identify the amino acid substitution(s) responsible for the enhanced activity of An96 and create a humanized An96 FIX transgene for gene therapy application. Reductionist screening approaches, including domain swapping, and scanning residue substitution, were employed and guided by one-stage FIX activity (OSA) assays. In vitro characterization of top candidates included recombinant protein production and high purity preparation, specific activity determination and enzyme kinetic analysis. Final candidate sequences were packaged into adeno-associated viral (AAV) vectors and delivered to hemophilia B mice. Five of 42 total amino acid substitutions in An96 appear sufficient to retain the enhanced activity of An96 in an otherwise human FIX variant. Additional substitution of the Padua variant further increases the specific activity 5-fold. This candidate, designated ET9, demonstrates 51-fold greater specific activity than hFIX. AAV2/8-ET9 treated hemophilia B mice produced plasma FIX activities equivalent to those observed previously for AAV2/8-An96-Padua, which were 10-fold higher than AAV2/8-hFIX-Padua. Starting from computationally inferred ancient FIX sequences, novel amino acid substitutions conferring activity enhancement were identified and translated into an AAV-FIX gene therapy cassette demonstrating high potency. This ASR discovery and sequence mapping refinement approach represents a promising platform for broader protein drug and gene therapy candidate optimization.

5.2 Introduction

Factor IX (FIX) is a zymogen of a vitamin K-dependent serine protease critical to the coagulation network. Deficiencies in FIX activity result in hemophilia B, an X-linked recessive

genetic disorder with a worldwide annual incidence of 1 in 25,000 male births (5). Historically, the treatment paradigm for hemophilia B has involved intravenous infusions of plasma-derived or recombinant FIX to prevent and/or treat bleeding episodes. However, in 2023 the first gene therapy for hemophilia B was approved (203). This product consists of an adeno-associated viral (AAV) vector encoding a codon-optimized variant of human FIX (hFIX), termed hFIX-Padua, driven by a liver-specific promoter (99). The Padua variant was not discovered by rational drug design, but instead was identified serendipitously in a family in Padua, Italy presenting with inherited thrombophilia linked to a single nucleotide change in the *F9* gene (c. 31134G>T) that results in the FIX variant NP_000124.1:p.Arg384Leu, herein referred to as “Padua.” (191). It was determined that this single amino acid substitution increases the specific activity of activated hFIX (hFIXa) 7 – 8-fold. Currently, all clinical stage AAV gene therapies for hemophilia B utilize this variant as it was shown to elevate FIX activity levels from the moderate to low mild hemophilia B ranges (achieved with wild-type hFIX in 1st generation AAV gene therapies) to the upper mild to supraphysiologic ranges in 2nd generation gene therapies depending on the dose, AAV serotype used, and transgene cassette configuration (204; 205). However, global utilization may be limited as the first approved AAV-FIX gene therapy comes at a price tag of \$3.5 million dollars in the United States (206).

Other than the inclusion of FIX-Padua in hemophilia B gene therapies, all other approved bioengineered FIX products involve addition of protein sequences or chemical conjugates that increase the circulating half-life of hFIX (e.g., immunoglobulin Fc-domain, albumin, or glycopegylation) (115). None of these enhancements are being pursued in clinical gene therapy trials to increase vector potency and thereby reduce vector requirements. However, progress towards further enhancing FIX activity has been demonstrated in preclinical studies where specific substitutions or combinations of substitutions including p.Val132Ala, p.Glu323Ala, p.Arg364Tyr, p.Arg384Ala/Glu/Leu, and p.Tyr389Arg have been shown to improve

the specific activity of recombinant and/or gene therapy produced hFIX (207-209). Building off our hypothesis that most protein drug and gene therapy transgene product performances can be enhanced above that of native human sequences, our group has taken an evolutionary approach utilizing the ancestral sequence reconstruction (ASR) platform to discover sequence enhancements existing in nature (172). This approach allows for rapid scanning of the evolutionary landscape in terms of both ortholog diversity and natural selection trajectories. Recently, we reported our initial ASR findings for coagulation factor VIII (FVIII), FIX, and von Willebrand factor (VWF) (118; 129; 210). Subsequently, others have applied the ASR approach to pharmaceutically relevant entities beyond coagulation factors such as Cas9 and base editors (211-214). In each of our studies, we identified sets of amino acid substitutions that confer enhancement to the biosynthetic efficiency, specific activity and/or half-lives of these critical coagulation factors. Within each inferred ASR variant exists positive, neutral, and potentially negatively acting substitutions. Thus, additional mapping studies are required to identify the specific substitutions that are necessary and sufficient to confer the functional enhancements observed in initial ASR screens. Ideally, next-generation protein drugs and gene therapies will contain the fewest alterations possible from the human protein thereby minimizing the risk of immunogenicity and neutralizing anti-drug antibodies.

FIX-An96 is a representative early mammalian FIX estimated to have been in existence prior to the Cretaceous-Paleogene boundary (118). Our studies of laboratory resurrected FIX-An96 demonstrate the variant to have a specific activity enhancement comparable to FIX-Padua without a substitution at p.Arg384Leu. Instead, FIX-An96 possesses 42 non-human substitutions spread over each FIX domain [signal peptide(sp) – propeptide – γ -carboxyglutamic acid-rich domain (Gla) – epidermal growth factor-like (EGF) domain 1 (EGF1) – EGF domain 2 (EGF2) – linker – activation peptide (ap) – protease (pro)]. The mechanism of FIX-An96 enhanced specific activity was shown to be independent of, and additive to, p.Arg384Leu as

recombinant An96-Padua displayed a specific activity of 6,450 IU/mg, which is 3.2-fold greater than that of hFIX-Padua (2013 IU/mg), and 3.1-fold greater than that of An96 (2090 IU/mg). This enhancement also translated to higher *in vivo* plasma FIX activities following AAV2/8-FIX-An96 ± Padua delivery (118). In the current study, we sought to map the critical substitutions necessary and sufficient to confer FIX-An96 comparable activity in a humanized FIX variant predicted to be safer for clinical use.

5.3 Materials and Methods

Engineering of human/ancestral hybrid FIX expression plasmids: All cDNA constructs were purchased from Integrated DNA Technologies (Coralville, IA) as gblocks gene fragments. They were subcloned into a liver-directed expression AAV cassette (85; 182) and/or pcDNA3.4 plasmid using *XhoI* and *NotI* restriction sites. Single amino acid variants were created utilizing overlap extension PCR (215) with primers according to Supplementary Table S4. Plasmid identity was confirmed with Sanger DNA sequencing (Azenta Life Sciences, Piscataway, New Jersey).

In vitro expression and testing of FIX constructs: FIX expression plasmids were transfected into low passage Huh7 cells (Creative Biolabs, Shirley, NY) with TransIT-X2 (Mirus Bio, Madison, Wisconsin) in antibiotic-free medium supplemented with 15µg/mL vitamin K₁ (Sigma Aldrich, St. Louis, MO), according to the Mirus Bio instructions. Twenty-four hours post transfection, wells were washed with Dulbecco's phosphate buffered saline (DPBS; Gibco – Thermo Fisher Scientific, Waltham, MA) and switched to 0.5mL of Freestyle 293 medium (Gibco) supplemented with 15µg/mL vitamin K₁. After an additional 24 hr, aliquots of conditioned medium were collected and analyzed via one stage coagulation assay (OSA) as described below. Data were normalized to activity present in the conditioned media from hFIX expressing cells and referred to as 'relative FIX expression rate.'

Purification of FIX variants: Expi293F Cells (Gibco) were transiently transfected with linearized FIX transgene cassettes in the pcDNA3.4 backbone. Two days following transfection, cells were moved to a T75 flask and geneticin G417 sulfate was added at a concentration of 400 μ g/mL until antibiotic selection was complete (~2 weeks). Polyclonal selected cell lines were then expanded into Erlenmeyer Flasks in Freestyle F-17 medium (Gibco) with 100 μ g/mL G418 and 15 μ g/mL vitamin K. Conditioned medium was collected every 48-72 hrs, subjected to centrifugation at 1,250 x g for 15 min, filtered through a 0.22 μ m polyethersulfone (PES) membrane and stored at -20°C with 0.02% NaN₃. FIX proteins were purified as previously described (118), but briefly involved an initial capture step using a Capto MMC resin (Cytiva, Marlborough, Massachusetts) and final polishing step using a RESOURCE Q column with a CaCl₂ elution gradient (183; 216). FIX concentrations were determined using absorbance at 280nm and an estimated molar extinction coefficient. For biochemical studies of hFIX, the commercial product, BeneFIX (Pfizer, New York, NY), was utilized. FIX activity measurement and thrombin generation assays of purified proteins were performed as previously described (118).

FIX activation: FIX zymogen was mixed with 1:100 to 1:1000 ratios of hFXIa (Prolyix, Essex Junction, Vermont) for 2-3 hrs at 37°C. FXIa was removed via immunoprecipitation with an anti-FXIa antibody (GMA-070, Green Mountain Antibodies, Burlington, Vermont) and protein G-agarose beads (Pierce Protein Biology, ThermoFisher Scientific, Waltham, Massachusetts).

FIXa enzyme kinetic studies with peptide substrate: Spectrozyme FIXa substrate was purchased from BioMedica Diagnostics (Windsor, NS, Canada). For experiments where the substrate concentration was kept constant, 1 mM of substrate was mixed with 0-800nM in the recommended buffer (50 mM TRIS, 100 mM NaCl, 5 mM CaCl₂, pH 7.4, with 33% (v/v) ethylene glycol) and immediately put on a plate reader spectrometer to measure the change in

absorbance at 405 nm ($\Delta 405$ nm). Initial velocities were obtained using Softmax Pro 4.3.1 Software (Graphpad Software, San Diego, CA).

Tenase enzyme kinetic studies: Recombinant proteins were purchased from Prolytix (Essex Junction, Vermont). Protein purity was assessed by SDS-PAGE and Coomassie-blue visualization. For FVIII/FVIIIa, Advate (Takeda, Lexington, MA), a recombinant full-length FVIII was utilized. PCPS lipids were made as previously described (217). FVIII (5 nM) was activated with 50 nM of thrombin for 30 sec, followed by thrombin inhibition with recombinant hirudin (75 nM) for 15 sec. FVIIIa (1 nM final concentration) was added to a mixture of FIXa (0.1 nM final concentration) and PCPS lipids (20 μ M) in HBS buffer with 10mM CaCl_2 . After allowing the FXase complex to form for 15 sec, FX was added (various concentrations) and aliquots of the reaction were quenched. The slope of each FXa generation curve was plotted against the FX concentration used and Michaelis-Menten kinetic parameters were fit using GraphPad Prism. Experiments where the FIXa-FVIIIa K_d was measured were performed as described above, except 400nM of FX was used, FVIIIa concentration was varied from 0-40nM, and the hirudin concentration was a 1.5x molar excess of the FVIII concentration.

FIXa modeling: Molecular modeling was initially performed with AlphaFold (218). The model was subsequently aligned with a previously described model of the intrinsic FXase complex based on the homologous prothrombinase complex (PDB code: 7TPP) (219; 220). Structural figures of the lead candidate, designated ET9, were generated with PyMOL Molecular Graphics System, Version 2.0 (Schrödinger).

Adeno-associated viral vector production and testing: FIX transgenes were liver codon optimized (LCO) as previously described (85) and subcloned into an AAV transfer plasmid utilizing AAV2 ITRs. LCO FIX transgenes were under the direction of an HHS4-TTR enhancer-promoter and included a minute virus of mouse (MVM) intron, and a minimal synthetic β -globin polyadenylation sequence. Plasmids were transformed into Stbl3 competent cells and

following monoclonal colony selection, grown to a volume of 500mL in terrific broth with 200 µg/mL ampicillin. Plasmids were prepared using a Qiagen plasmid Mega kit (Qiagen, Germantown, MD) and sequence was confirmed with Sanger sequencing. Additionally, intact ITR sequences were confirmed by *SmaI/AhdI* restriction enzyme (New England Biolabs, Ipswich, MA) digestion and ITR Sanger sequencing (Azenta Life Sciences, Piscataway, New Jersey). Plasmids were then packaged into single-stranded AAV8-capsid viral vectors by Charles River Laboratories (Rockville, MD). Viral vectors were titered using qPCR with QuantStudio3 Real-Time PCR system and PowerUP SYBER Green (Applied Biosystems, Waltham, MA) and interpolated from a plasmid standard.

Male hemophilia B mice (exon 1-3 disrupted, Jackson Laboratory strain #004303), aged 7 to 14 weeks, were injected intravenously into the lateral tail vein with varying doses of AAV (previously diluted into a volume of 100 µL of PBS with 0.001% pluronic F-68). Blood was collected from the retro-orbital plexus into 0.38% (m/v) sodium citrate (final concentration). Plasma was isolated and FIX activity determined by OSA as described above.

Statistical analysis: All data analysis, statistical testing and graph creation were performed using GraphPad Prism 9.5.1 software (GraphPad Software, San Diego, CA).

5.4 Results

Mapping functional residues necessary for FIX-An96 enhanced activity: FIX-An96 possesses 42 amino acid substitutions with respect to hFIX (118). We hypothesized that not all 42 substitutions are required to maintain the enhanced specific activity of FIX-An96. As an initial screen for the domain locations of the functional substitutions, a panel of hybrid FIX-An96/hFIX molecules were generated and cloned into a mammalian expression plasmid under the control of a synthetic liver-specific promoter (**Figure 5.1A**). The resulting molecules were tested *in vitro* for expression and activity by transient transfection into low passage Huh-7 cells and assay of

the conditioned media by FIX OSA (**Figure 5.1B**). This assay allows for the detection of hybrid proteins with more efficient FIX biosynthesis (*i.e.*, protein translation and/or secretion) or enhanced specific activity, but it does not distinguish between the two possibilities. Of the initial 14 hybrids generated, only 3 (FIX-108, FIX-109, and FIX-112) displayed FIX activity in the conditioned medium at levels non-inferior to FIX-An96. The common FIX-An96 domains specific to these 3 variants are the EGF2 and protease domains, suggesting that each domain contains amino acid substitutions necessary for the activity enhancement observed for FIX-An96.

Within the EGF2 domain, there are 5 amino acid substitutions, one of which was previously described (p.Val132Ala) to confer a specific activity enhancement (208). Of note, many extant mammalian species outside of the human/primate lineage and including rodents and ungulates (hoofed mammals) naturally share alanine at this position. Constructs containing only the FIX-An96 N-terminal (FIX-115) or C-terminal (FIX-114) halves of the protease domain were included to interrogate this largest domain at higher resolution (**Figure 5.1A**). While neither FIX-114 nor FIX-115 matched FIX-An96 performance in the transient transfection assay, FIX-114 displayed partial enhancement, possibly limited only by the lack of p.Val132Ala. Therefore, three additional constructs were generated to reintroduce p.Val132Ala and interrogate protease subdomain substitutions within the FIX-An96 C-terminal half: FIX-114 + p.Val132Ala (FIX-201), N-terminal FIX-An96 subdomain of FIX-114 + p.Val132Ala (FIX-202), and C-terminal An96 subdomain of FIX-114 + p.Val132Ala (FIX-203) (**Figure 5.1A**). FIX-201 displayed enhancement indistinguishable from the parent FIX-An96 variant. However, both FIX-

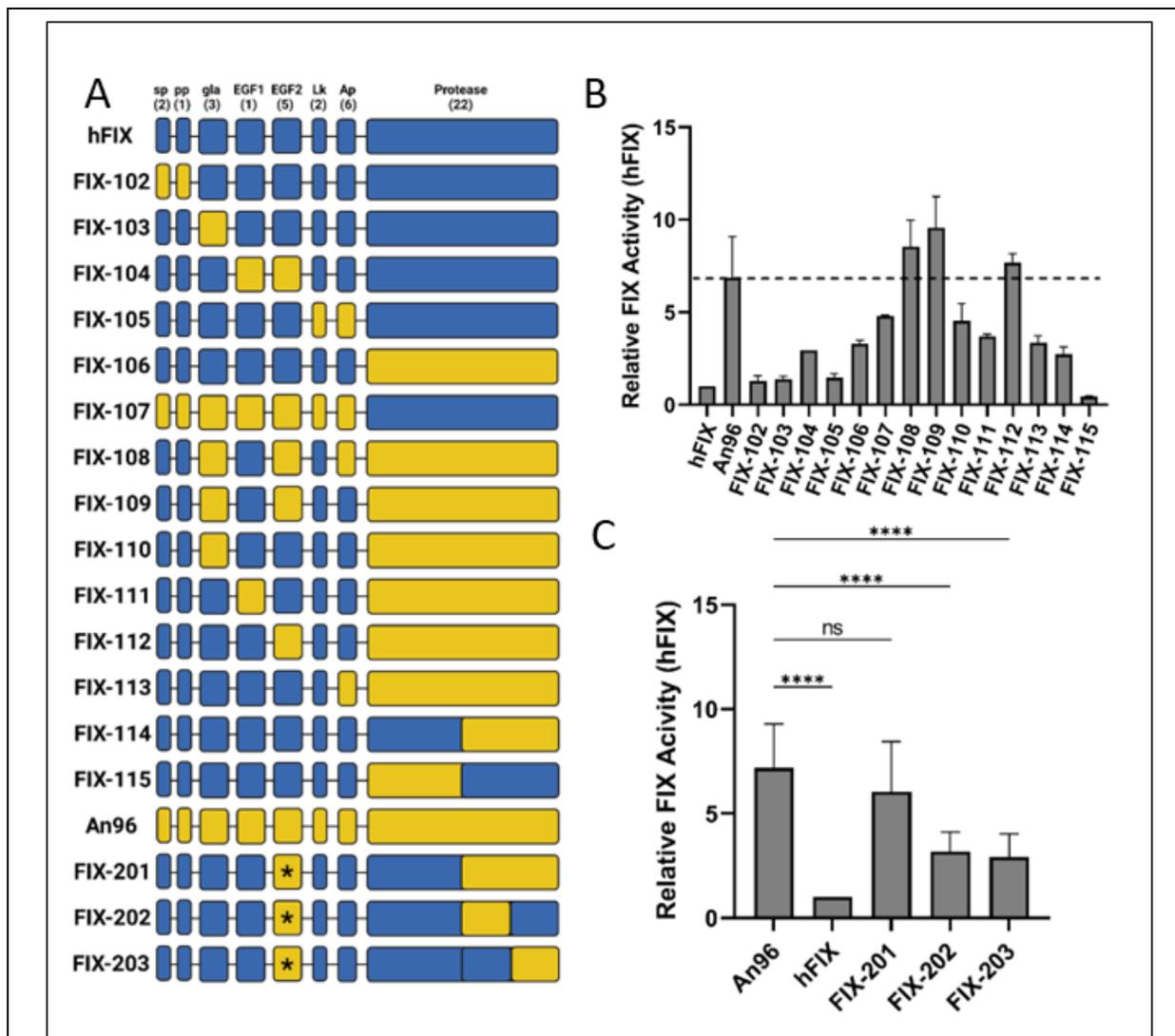


Figure 5.1: Ancestral / human FIX hybrid schematics and expression. A) Schematics of the 17 initial An96-hFIX hybrids generated are shown. Blue regions indicate hFIX sequence and gold regions represent FIX-An96 sequence. The * represents addition of the p.Val132Ala variant. The numbers in parentheses are the numbers of amino acid substitutions with respect to hFIX sequence. B) Relative FIX activities were obtained by first measuring FIX activity production rates (IU/10⁶ cells/24 hr) using timed media collection, measurement of FIX activity by OSA, and cell counting within each well. The rates were then normalized to the values obtained for hFIX and presented as 'Relative FIX Activity'. The dashed horizontal line denotes the relative expression rate for FIX-An96 (positive control). C) FIX expression rates were similarly assessed for the C-terminal protease domain hybrids again relative to hFIX. Error bars indicate sample standard deviation. Statistical analysis was performed using a one-way ANOVA with post-hoc Tukey multiple comparisons test. Asterisks denote hybrids with significant expression below FIX-An96 where * p<.033, ** p<.002, *** p<.0002, and **** p<.0001. Three or more independent experiments with two technical replicates each were performed in B) and C).

202 and FIX-203 demonstrated < 50% of the parent FIX-An96 enhancement suggesting that

gain-of-function substitutions are present within each subdomain of the C-terminal half of the FIX-An96 protease domain (**Figure 5.1C**).

Collectively, there are 11 amino acid substitutions in these two subdomains relative to hFIX. Of note, none of these substitutions occur at Arg384. To identify the functional FIX-An96 protease domain substitutions, a negative screen was performed whereby single amino acids in FIX-201 were substituted for their hFIX counterpart anticipating that substitution away from gain of function residues would result in a decrease in FIX enhancement compared to the parent molecule (**Supplementary Table S5.1** and **Figure 5.2**). Six of the 11 FIX-An96 C-terminal protease domain substitutions reduced the FIX enhancement: p.Lys323Glu, p.Trp326Val, p.Asn338Asp, p.Arg339Lys, p.Arg362Lys, and p.Ser367Leu. The other five substitutions (p.Ser313Asn, p.Asn361His, p.Ile368Val, p.Tyr399Phe, and p.Lys404Arg) conferred no significant effect.

Having observed the substitutions within FIX-201 to be necessary for the FIX-An96 enhancement, a final panel of constructs was generated. Each construct contained p.Val132Ala plus at least one An96 C-terminal protease domain substitution (**Supplementary Table S5.2**). Although each of these humanized FIX-An96 constructs possessed enhancement significantly above hFIX, the two variants that contained only 2 or 3 substitutions in the C-terminal protease domain (designated FIX-401 and FIX-402) displayed reduced enhancement compared to FIX-An96 (**Figure 5.3A**). However, the two constructs that contained 4 and 5 substitutions within the C-terminal protease domain (designated FIX-403 and FIX-404, respectively) demonstrated enhancement indistinguishable from FIX-An96 revealing that p.Val132Ala, p.Glu323Lys,

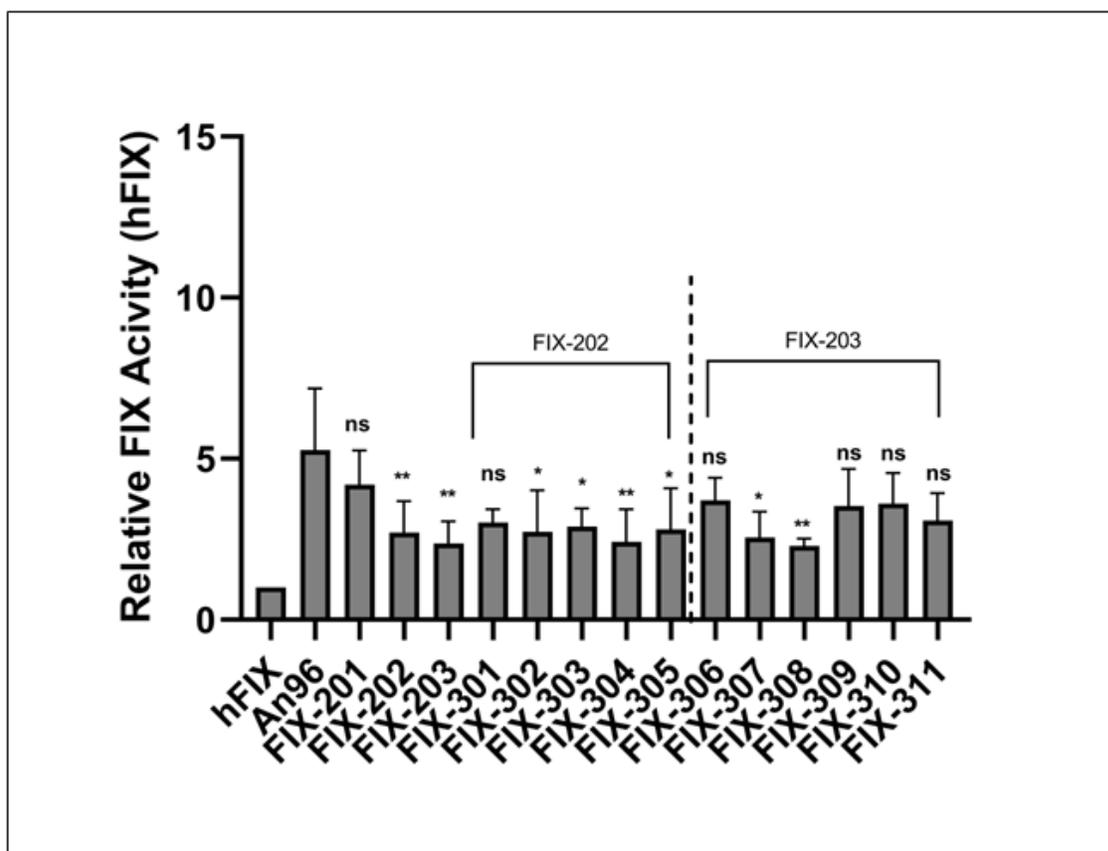
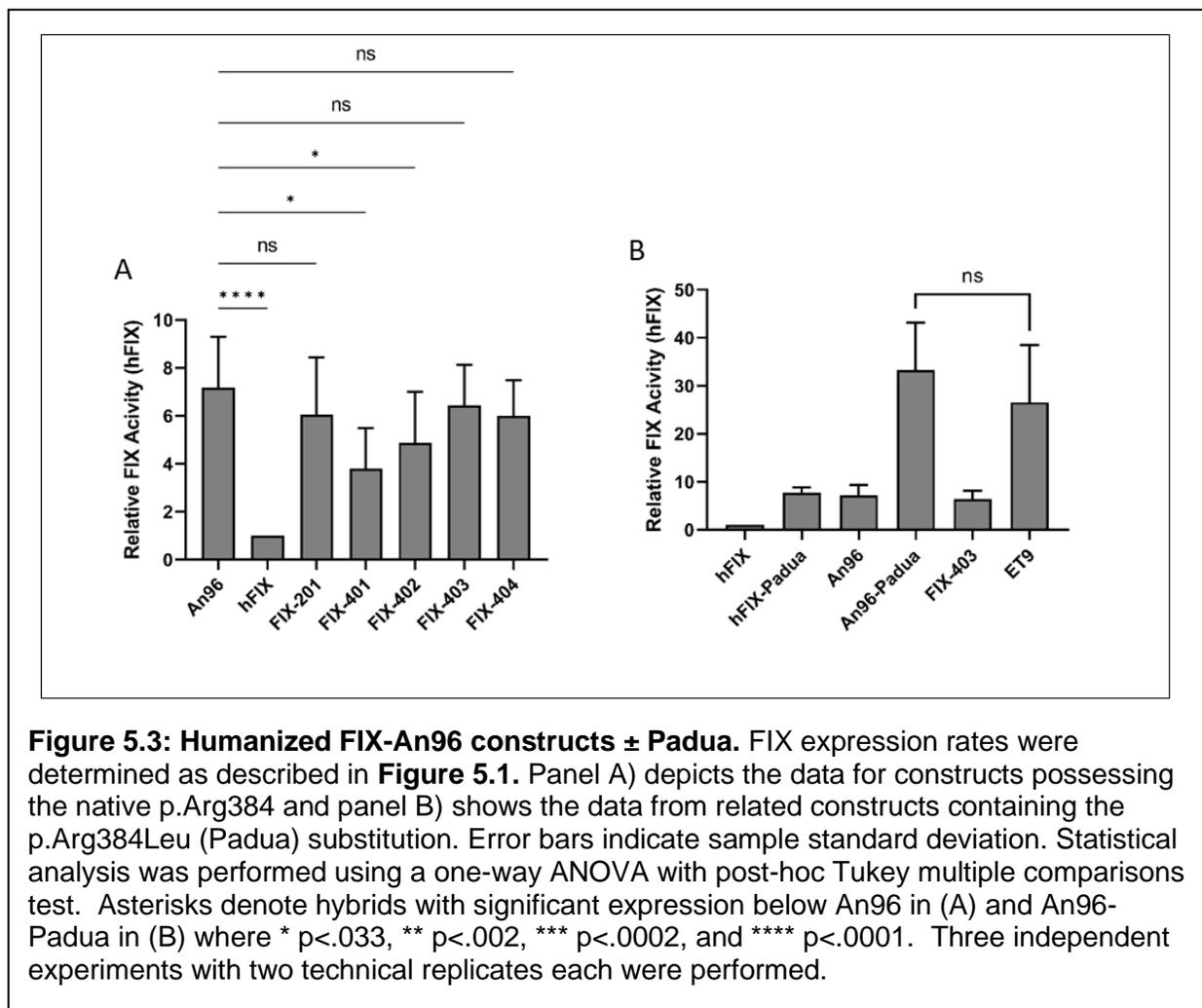


Figure 5.2: Negative single amino acid substitution screen of FIX-201. FIX expression rates of single amino acid variants listed in **Supplementary Table S5.1** were determined as described in **Figure 5.1**. Brackets indicate the parental hybrid from which each subvariant was derived. Error bars indicate sample standard deviation. Statistical analysis was performed using a one-way ANOVA with post-hoc Tukey multiple comparisons test. Asterisks denote hybrids with significant expression below An96 where * $p < .033$, ** $p < .002$, *** $p < .0002$, and **** $p < .0001$. Three independent experiments with two technical replicates each were performed.

p.Asp328Asn, p.Lys362Arg, and p.Leu367Ser (FIX-403) are necessary and sufficient to recreate the enhancement of FIX-An96 in a molecule with 99% identity to hFIX (**Figure 5.3A**).

Addition of Padua Variant (Arg384Leu) to FIX-403: Previously, we demonstrated the specific activity enhancement conferred by the Padua variant is additive to the enhancement present in FIX-An96. To confirm that this observation remained valid for FIX-403, p.Arg384Leu was introduced into FIX-403 to create a construct designated ET9. Similar to our previous findings for FIX-An96 ± Padua, ET9 demonstrated 4 – 5-fold greater enhancement over hFIX-

Padua or FIX-An96 and equivalent enhancement to that described previously for FIX-An96-Padua (**Figure 5.3B**) (118).



In order to address the possibility of increased FIX biosynthesis (transcription, translation, and secretion) as the mechanism driving enhanced FIX activity measurements, an ELISA was performed on conditioned media samples from transiently transfected Huh7 (**Supplementary Figure S5.1A**) and HEK293T/17 (**Supplementary Figure S5.1B**) cells. While FIX-403 showed moderately (< 2-fold) higher FIX antigen concentrations compared to hFIX, hFIX-Padua, and ET9 when expressed from Huh7 cells, the antigen levels of all four constructs were equivalent when expressed from HEK293T/17 cells. These data do not support the

hypothesis that the 5- and 25-fold enhanced relative activities of FIX-403 and ET9 observed in the transient transfection assay (**Figure 5.3B**) result from more efficient biosynthesis.

Procoagulant activities of recombinant FIX variants: To further interrogate the mechanism(s) driving FIX-An96, FIX-403, and ET9 activity enhancement, stable polyclonal cell populations expressing each FIX variant were generated, conditioned media was collected, and high purity recombinant FIX preparations were generated as described previously (118). Subsequently, the specific activities of each preparation were determined from OSA measurements of serial dilutions of each recombinant FIX preparation (**Supplementary Table S5.3** and **Figure 5.4A**). Nearly parallel activity curves were observed for each variant supporting the validity of the activity comparisons. As expected, overlapping curves were generated for the parental:humanized progeny pairs, An96:FIX-403 and An96-Padua:ET9, again supporting the conclusion that the full activity enhancement was conferred by identified residues. Both FIX-An96 and FIX-403 display specific activities (2,182 and 1,681 IU/mL, respectively) equivalent to hFIX-Padua (2,108 IU/mg). Substitution of p.Arg384Leu into FIX-403 (*i.e.*, ET9) further increased the specific activity to 10,644 IU/mg, which is indistinguishable to that of An96-Padua (10,553 U/mg). Notably, ET9 specific activity is 5-fold greater than that of the hFIX-Padua, and 51-fold higher than commercial recombinant hFIX.

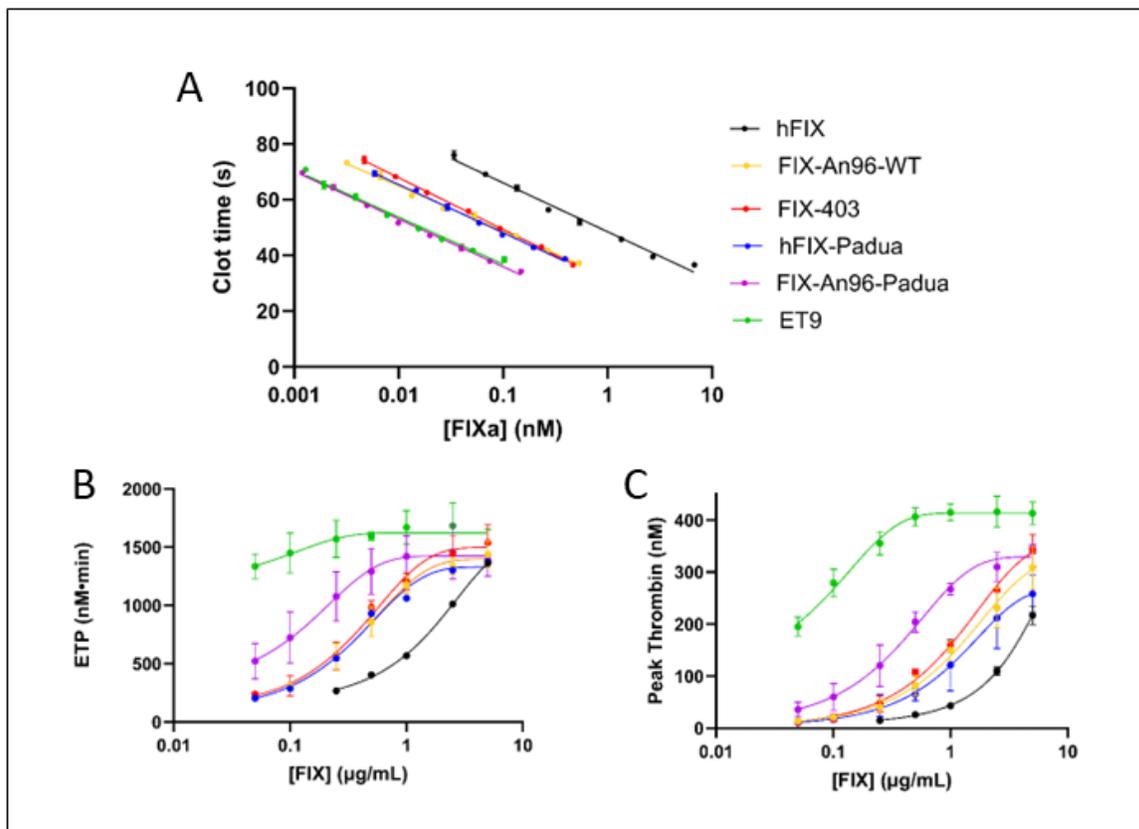


Figure 5.4: OSA activity curves and thrombin generation assays for recombinant FIX variants ± Padua. A) Dilutions (0.001 – 6.8 nM) of FIX variants were added to FIX-deficient plasma and time to fibrin clot formation was measured by OSA. Slopes were calculated by linear regression of data obtained from three independent experiments ($R^2 > 0.98$). Error bars indicate sample standard deviation. Slopes (95% CI) for hFIX, hFIX-Padua, FIX-403, An96, An96-Padua and ET9 were: -17.54 (-18.52 to -16.57), -17.34 (-17.99 to -16.69), -18.53 (-19.04 to -18.03), 16.15 (-16.80 to -15.49), -17.12 (-17.78 to -16.46) and -16.93 (-17.94 to -15.93), respectively. Endogenous thrombin potential (B) and peak thrombin (C) as determined by thrombin generation assay. Data points indicate the average of 3 independent experiments. Error bars indicate standard deviation.

The thrombin generation assay (TGA) represents a complementary and more comprehensive assessment of procoagulant activity as it takes place over a significantly longer timescale than the OSA and produces a multiparameter readout including a time-dependent thrombin generation curve (thrombogram), peak thrombin concentration, and endogenous thrombin potential (ETP). TGA parameter comparisons of recombinant FIX variants added to hemophilia B plasma revealed similar relative procoagulant activity trends as those observed by OSA with one striking difference, which was consistently higher procoagulant activity of ET9

compared to An96Padua which was even more pronounced at the lowest FIX concentrations. While in the OSA, ET9 and An96Padua were identical, in the TGA, the following FIX activity ranking was observed: ET9 > An96Padua > FIX-403 \approx An96 \approx hFIX-Padua > hFIX (**Figure 5.4B-C** and **Supplementary Figure S5.2**).

FIX-403 and ET9 enzyme kinetics: Others have shown that the specific activity enhancement conferred by hFIX-Padua (p.Arg384Leu) as well as p.Arg384Ala, is not apparent in FIXa chromogenic peptide substrate assays (190; 209). Similarly, activated hFIX, hFIX-Padua, FIX-403, and ET9 all displayed similar reaction velocities towards a synthetic FIXa substrate (**Supplementary Figure S5.3**). Next, the activities of activated FIX-403 and ET9 were compared to hFIXa and hFIXa-Padua using a reconstituted FXase assay with purified components (**Figure 5.5A**). In this assay, proteolytic cleavage of a chromogenic FXa substrate

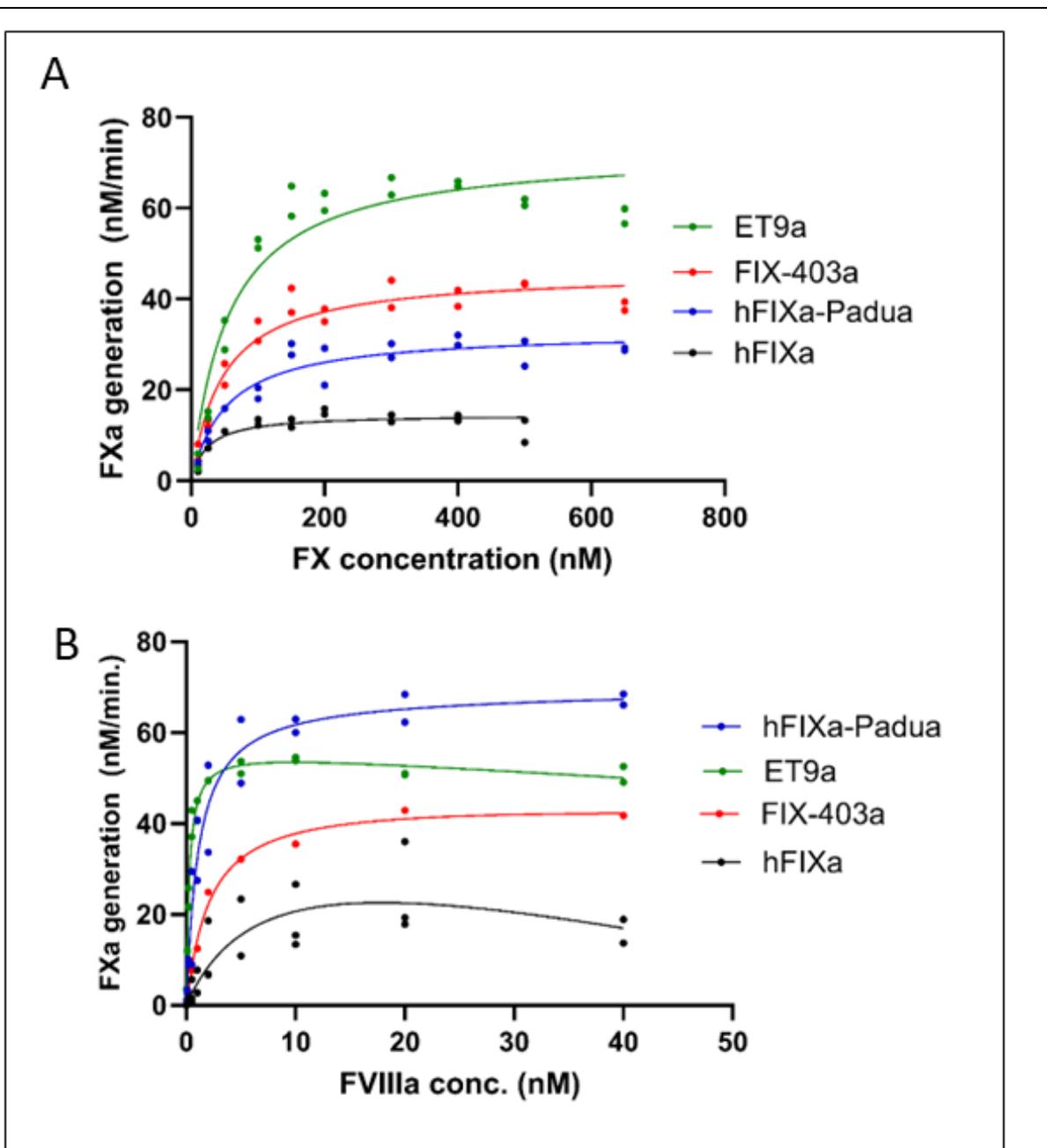


Figure 5.5: Tenase activity and FVIIIa binding among FIXa variants. Using purified components, FXase complexes were assembled and analyzed for FIXa kinetic parameters (A) as well as FVIIIa binding (B). Data points represent values obtained from 2 independent experiments. (A) Lines represent the Michaelis-Menten fit of the data. R^2 values of 0.79, 0.91, 0.94, and 0.92 were obtained for hFIX, hFIX-Padua, FIX-403, and ET9, respectively. (B) Lines represent one-site total binding fit of the data. R^2 values of 0.79, 0.96, 0.98, and 0.98, were obtained for hFIXa, hFIXa-Padua, FIXa-403, and ET9a, respectively.

is utilized as an indirect measurement of FXase product (FXa) accumulation, which can be

interpolated from a standard curve generated using known concentrations of highly purified FXa and saturating concentrations of Xa substrate. Applying the Michaelis-Menten kinetic model, specific parameters, K_m and k_{cat} are estimated and presented in **Table 5.1**. This analysis reveals that the ASR identified amino acid substitutions appear to operate through enhancement of substrate turnover with FIX-403a displaying a k_{cat} 3.1 times higher than hFIXa (7.67 s^{-1} versus 2.44 s^{-1} , respectively) and 1.4 times higher than hFIXa-Padua (5.49 s^{-1}). ET9a displays the greatest k_{cat} enhancement at 12.17 s^{-1} , which is 5-fold greater than hFIXa and 2.2 and 1.6-fold greater than hFIXa-Padua and FIX-403a, respectively. Similar to previous observations, substitution of the Padua mutation variant results in K_m elevations (190) of 2.2 and 1.2-fold, respectively, for hFIXa-Padua and ET9a. Catalytic efficiency estimates (k_{cat}/K_m) of hFIXa and hFIXa-Padua were similar with FIX-403a and ET9a being 1.5 and 2.1-fold higher, respectively.

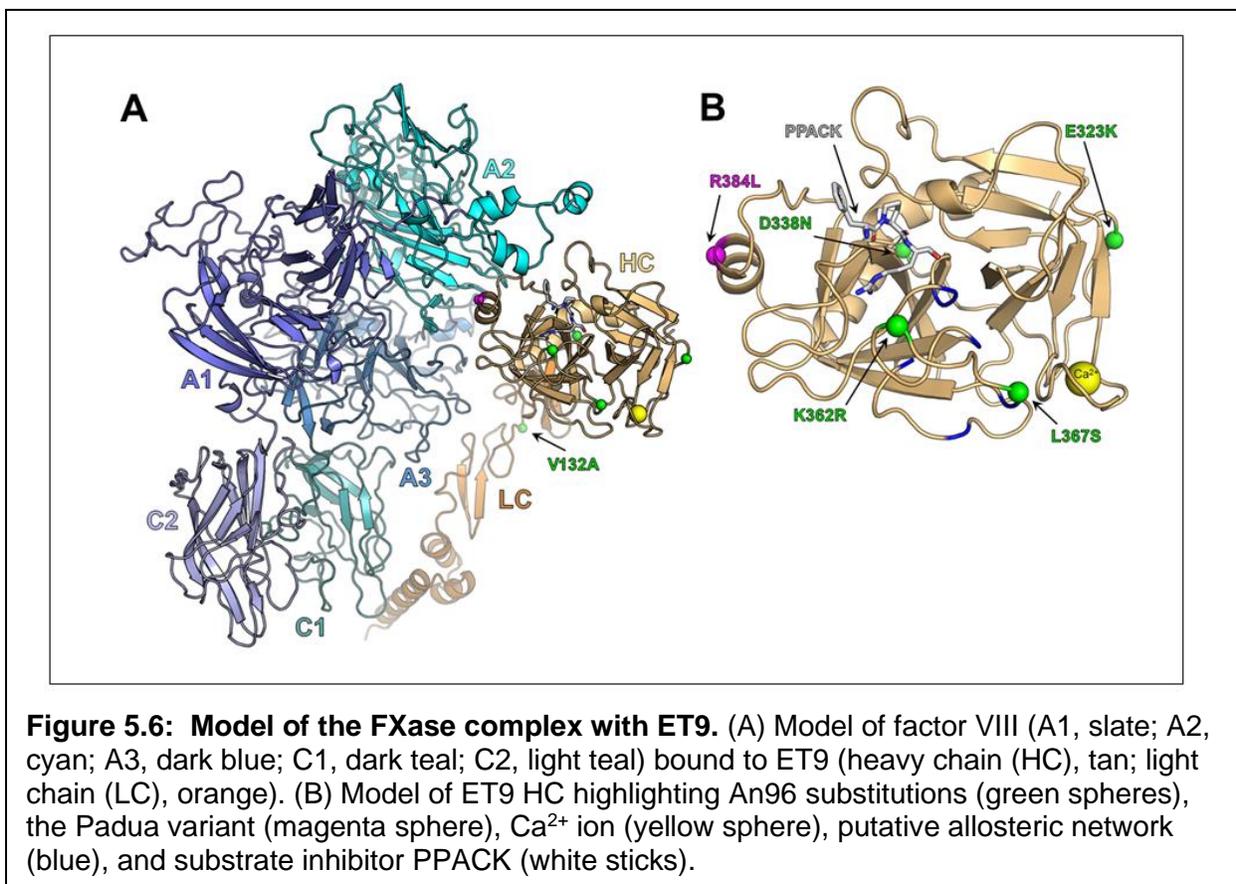
Table 5.1: Michaelis-Menten kinetic parameters of hyperactive variants in the purified tenase assay

FIXa variant	K_m (nM)	95% CI	k_{cat} (s^{-1})	95% CI	V_{max} (nM/min)	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_D^{app} (nM)	95% CI
hFIX	23.7	13.1 to 40.6	2.44	2.17 to 2.73	14.61	1.03×10^8	7.80	1.76 to 125.3
hFIX-Padua	52.9	35.6 to 77.7	5.49	5.02 to 6.04	32.94	1.04×10^8	1.08	0.59 to 2.00
FIX-403	47.6	34.6 to 65.0	7.67	7.14 to 8.27	46.03	1.61×10^8	2.08	1.37 to 3.21
ET9	56.1	38.4 to 80.9	12.17	11.1 to 13.4	73.01	2.18×10^8	0.26	0.20 to 0.34

Similar to the p.Arg384Leu and p.Arg384Ala variants previously described, the activity enhancements observed for FIX-403 and ET9 also appear to require FVIIIa (190; 209). The initial step in FVIIIa cofactor activity involves FVIIIa-FIXa complex formation driven by direct affinity of the two macromolecules. Since FXase activity is directly proportional to the concentration of FVIIIa-FIXa complex and addition of FVIIIa results in a >1000-fold increase in FXase reaction rate, the binding affinity of FIXa for FVIIIa can be estimated by monitoring FXa generation over a range of FVIIIa (or FIXa) concentrations as previously described (190; 209).

To determine the binding affinity of ASR-based variants enhancement for FVIIIa, a similar modified FXase assay design was utilized (**Figure 5.5B**). Using this approach, apparent K_d (K_d^{app}) estimates obtained for hFIXa, hFIXa-Padua, FIXa-403, and ET9a were: 7.80, 1.08, 2.08, and 0.26 nM, respectively. Consistent with the relative procoagulant activities of the FIX variants, the relative order of FVIIIa binding affinities as well as the magnitude of K_d^{app} differences follow a similar trend suggesting a role for improved FVIIIa binding contributing to the activity enhancement.

Molecular modeling of FIX-403 non-human substitutions: To structurally rationalize the FIX-403 non-human substitutions, we queried the five amino acid residue changes, along with the Padua variant using AlphaFold (218). The resultant model was subsequently superimposed with a previously modeled FXase complex (**Figure 5.6A**), which is a model that was initially generated through sequence homology to the recently determined prothombinase complex and further refined for energetic minimization and conformational docking, along with the addition of the PPACK FIXa inhibitor (**Figure 5.6B**) (219). The sole light chain substitution is p.Val132Ala, which removes hydrophobicity from the EGF1 and EGF2 interface. Within the catalytic domain, the remaining substitutions are spread throughout the FIXa structure and do not make direct

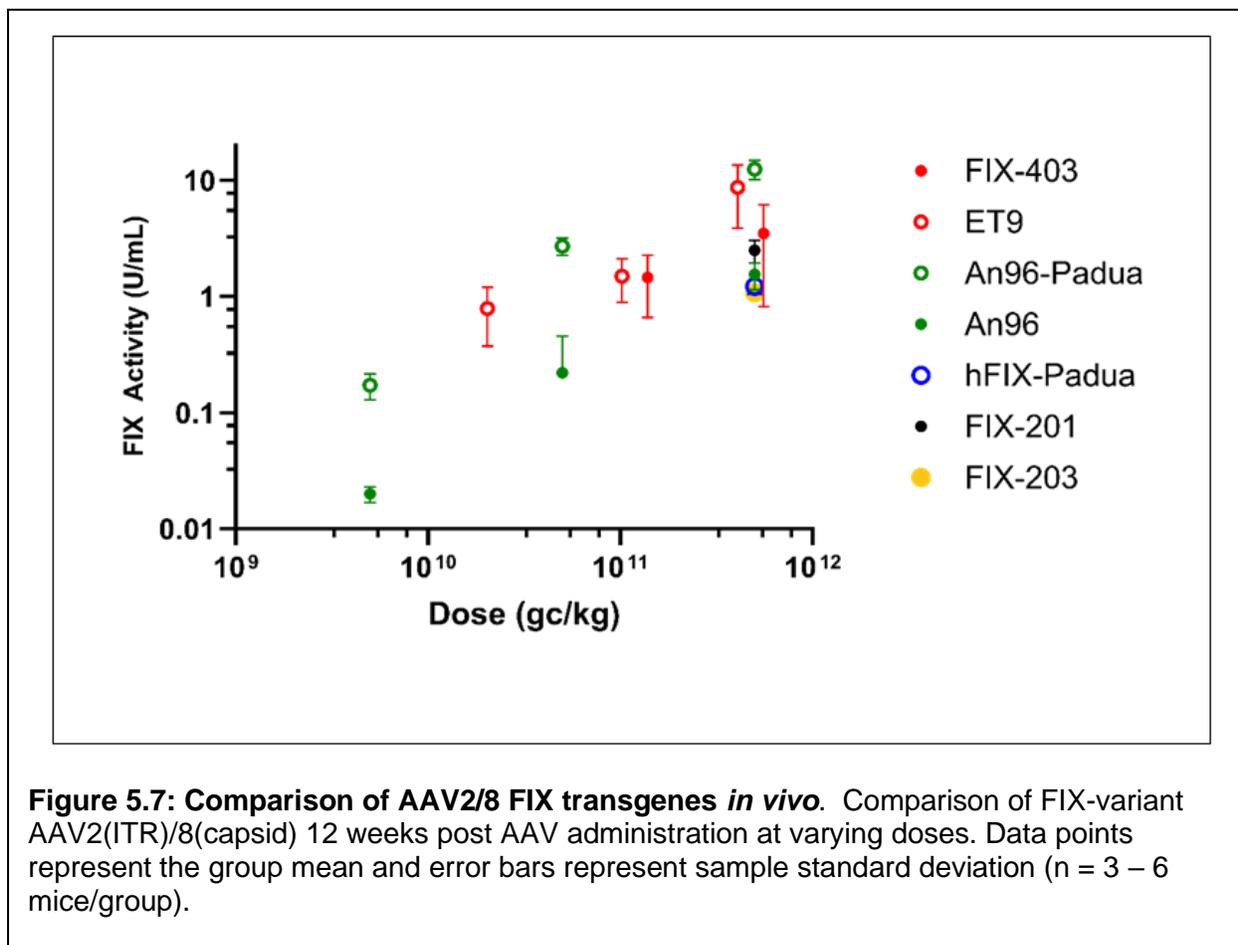


contact with the protease active site. The substitutions, p.Glu323Lys, p.Lys362Arg, and p.Leu367Ser are all solvent-exposed; both p.Lys362Arg and p,Leu367Ser are adjacent to a putative allosteric network of amino residues connecting the Ca²⁺ loop to the substrate binding pocket (221). In contrast to the solvent-exposed substitutions, p.Asp338Asn is partially buried in the structure of the catalytic domain and resides on the N-terminus of an alpha helix that packs against the EGF2 domain.

In vivo testing of ET9: Previously, we showed that the enhanced activity of An96 measured in vitro using OSA and TGA translated to 60- and 6-fold in vivo potency enhancement over hFIX and hFIX-Padua, respectively (118). In the current study, we utilized identical methodology, up-and-down staircase method combined with a saphenous vein bleed challenge, to determine the 50% effective dose (ED₅₀) of ET9. Consistent with our previous findings for An95-Padua and hFIX-Padua, in the current study, ET9 displayed an ED₅₀ 10-times lower than

hFIX-Padua (0.256 µg/kg and 2.34 µg/kg, respectively) (**Supplementary Figure S5.4**). These results further validate the translational potential of ET9 in clinical protein replacement and gene therapy applications.

To investigate the *in vivo* performance of FIX-403 and ET9 in a gene therapy setting, liver-directed and codon-optimized transgene cassettes were packaged into AAV8 capsid vector particles and injected into hemophilia B mice at varying doses. Bi-weekly post-AAV-FIX injection, blood was collected, and FIX activity determined by OSA and compared to our previously published data for AAV-hFIX-Padua , AAV-An96, and AAV-An96-Padua (**Figure 5.7**). Consistent with our *in vitro* observations, *in vivo* performance of the AAV2/8-ET9 vector across dose range between 10^9 – 10^{12} gc/kg was overlapping with that observed previously for AAV2/8-An96-Padua. At the highest dose tested (4.1×10^{11} gc/kg), mean AAV2/8-ET9 activity was 8.7 IU/mL, which is ~7-higher than observed previously for AAV2/8-hFIX-Padua at a dose of 5×10^{11} gc/kg (1.2 IU/mL). Furthermore, AAV2/8-ET9 achieved plasma FIX activity levels similar to the latter vector at a dose level 25-times lower (0.79 IU/mL at 2×10^{10} gc/kg).



5.5 Discussion

In gene therapy candidates, the importance of potency as a critical attribute has grown in appreciation as clinical data continue to accumulate. Observations of vector related toxicities and the relatively slow progress made towards overcoming manufacturing challenges such as scale limitations and cost of goods remain major commercialization challenges (222-224). Currently, there are three primary approaches to improving vector potency, 1) increasing gene transfer efficiency, 2) optimizing transgene mRNA transcription and/or stability, and 3) enhancing the performance of the transgene product through protein bioengineering. AAV-FIX gene therapies represent the state of the art in terms of optimizations at all three levels, *i.e.*, capsid serotype transition from AAV2 to AAV3, 5, 8 and variants thereof, promoter/transgene

(synthetic liver-directed promoters and codon optimized FIX transgenes with reduced CpG content), and transgene product engineering (inclusion of the Padua variant) (85; 118; 200; 208; 225-227). Together, these modifications appear to be critical to the clinical successes observed.

The first approved AAV-FIX gene therapy for hemophilia B is the most expensive single drug developed to date (228). One strategy for reducing cost is the development of more potent transgene cassettes that allow for lower dosing and therefore more doses produced per manufacturing campaign. Our team has focused on the development of bespoke optimization strategies for gene therapy transgene cassettes at the nucleic acid and amino acid levels (85; 118; 129; 135; 182). For the latter, we have explored extant and ancient (inferred) interspecies diversity as a discovery method followed by identification of functional sequences and engineering them into the relevant human transgene products. Although this approach comes with a theoretical risk of increased immunogenicity, we and others have not observed this outcome in preclinical and clinical studies to date. In contrast, existing data frequently support the notion that leading gene therapy approaches, such as liver-directed AAV gene therapy and hematopoietic stem cell lentiviral gene therapy, promote immune tolerance to the transgene products (for review see Patel *et al.*) (229; 230). Therefore, *a priori*, and consistent with the outcomes observed using the Padua variant, protein engineering need not be considered off-limits during the development of future gene therapy candidates.

Previously, we utilized the ASR approach to identify FIX variants with enhanced properties and discovered a top variant, FIX-An96 that displayed 11-fold higher specific activity than hFIX. Addition of the Padua variant to FIX-An96 (An96-Padua) further boosted the specific activity to a level nearly 60-fold higher than hFIX, likely representing the most active FIX variant described to date (115). In the current study, we systematically identified the amino acids necessary and sufficient for the enhancement observed in FIX-An96 by removing them from FIX-An96 and subsequently engineering them into hFIX. The resulting construct, FIX-403,

possessed 5 non-human amino acid substitution located in two domains; EGF2 (p.Val132Ala) and protease (p.Glu323Lys, p.Asp338Asn, p.Lys362Arg, and p.Leu367Ser). Variant p.Val132Ala has been described previously by Chang *et al* (231) in a comprehensive EGF1/2 domain alanine scanning experiment and shown to increase k_{cat} in the presence of FVIIIa. Interestingly, p.Val132Ala is naturally present in most non-human mammals and the majority of ancestral FIX variants inferred during our previous study (118). Despite being categorized as non-human substitutions, p.Val132Ala and p.Glu323Lys have a minor allele frequencies of 1.09×10^{-5} and 8.53×10^{-4} , respectively (232; 233). Additionally, Lin *et al* described the combinations of p.Val132Ala with p.Glu323Ala and p.Arg384Ala/Leu having the additive benefits of each individual residue. It is intriguing that using orthologous approaches (*i.e.*, alanine scanning and ASR) two common residues, p.Val132 and p.Glu323, were identified as positions where substitutions result in gain of function. These data combined with the identification and characterization of the Padua gain of function substitution by Stafford, Simioni, Arruda and Samelson-Jones further support the hypothesis that FIX activity has been suppressed by negative selection during human evolution (191; 209; 234). Furthermore, ET9 builds upon the TripleL (p.Val132Ala/Glu323Ala/Arg384Leu) variant described by Kao *et al* to have a 22-fold specific activity difference (4308.8 ± 182.0 versus 197.9 ± 8.7 IU/mg) by OSA and antigen ELISA (201). Through modification of p.Glu323Lys, as opposed to alanine in TripleL, and addition of the p.Asp338Asn, p.Lys362Arg, and p.Leu367Ser substitutions, ET9 specific activity, as determined by clotting assay and spectrophotometric protein determination, increased another 2.5-fold. Kao *et al* also showed that p.Val132Ala/Glu323Ala/Arg384Ala (or Leu) enhances FIXa performance by both increasing the apparent affinity for FVIIIa (10-fold decrease in K_d^{app}) and increasing k_{cat} (<3-fold) (115; 201). Kinetic studies of hFIXa-403 and ET9a uncovered a further increase in k_{cat} (~5-fold increase) with little alteration of K_m other than the consistent increase with addition of p.Arg384Leu as has been observed previously (190; 201; 209). Molecular modeling was performed to provide speculative mechanistic information as

follows. p.Val132Ala possibly disrupts the Van der Waals packing at the hydrophobic interface between the EGF1 and EGF2 domains. Valine is considerably more hydrophobic than alanine (235) and has substantially more surface area, which likely accounts for the relatively large effect observed by us and others. p.Asp338Asn may destabilize the alpha helix that packs against the EGF2, as the negative charge of the carboxyl group of aspartic acid serves as a helix capping residue that stabilizes the N-terminal positive charge of a helix macrodipole. Both of these substitutions may function by lowering the energetic barrier to putative conformational changes required for catalytic activity upon FVIIIa binding. p.Glu323Lys is completely exposed on the opposing face of the FVIIIa/FIXa interface and thus may be involved in electrostatically steering FIXa towards FVIIIa binding. Tighter binding between FIXa and FVIIIa results in higher FVIIIa-FIXa complex concentrations, thus potentially improving FX turnover and is the proposed mechanism by which p.Arg384Leu exerts its effects (190). As stated above, p.Lys362Arg and p.Leu367Ser substitutions reside in the region of the FIXa catalytic domain proposed to possess a signal transduction network of interactions connecting the Ca²⁺ loop to the substrate binding pocket (221). p.Lys362Arg may form a salt bridge with p.Glu434 stabilizing the sodium binding loops and the active site, possibly rectifying a known shortcoming of FIXa: a poorly formed active site (236-238). p.Lys362Arg is also located on the autolysis loop predicted to regulate binding to FX binding and activation. Previous studies have shown p.Lys362Ala to reduce FX activation (239). Lastly, p.Leu367Ser may stabilize the Ca²⁺ active conformation and signal relay to the active site. Here, the p.Leu367Ser substitution forms a putative hydrogen bond with p.Glu375 located on the Ca²⁺ loop.

These ASR substitutions, along with our previously described enhanced ancestral FVIII (termed An53) (129) are predicted to have existed along a common ancestral lineage ~90 million years ago. Therefore, it is reasonable to speculate that the ancient mammalian hemostatic system may have required more active secondary hemostasis components. This

corresponds to a pre-platelet, thrombocyte era where primary hemostasis was presumably less efficient and more efficient secondary hemostasis perhaps advantageous. The development of invasive placentation, unique to mammals, and evolution towards live birth has been proposed as the driving force for platelet evolution (240). By providing a highly effective primary hemostatic mechanism, platelets combined with higher activity coagulation factors may have tipped the hemostatic balance towards thrombosis overtaking as the dominant selective force. If accurate, one would predict to observe a gradual dampening of secondary hemostatic efficiency through the stepwise introduction of amino acid substitutions that confer reduced activity in critical coagulation factors, such as FVIII and FIX. Although this type of evolutionary mechanistic speculation is intriguing, no such understanding is required to utilize ASR to explore the diversity and capture functionality evolution has achieved for the purpose of designing improved protein drugs and gene therapies. The current study provides additional proof of concept for this promising drug development approach.

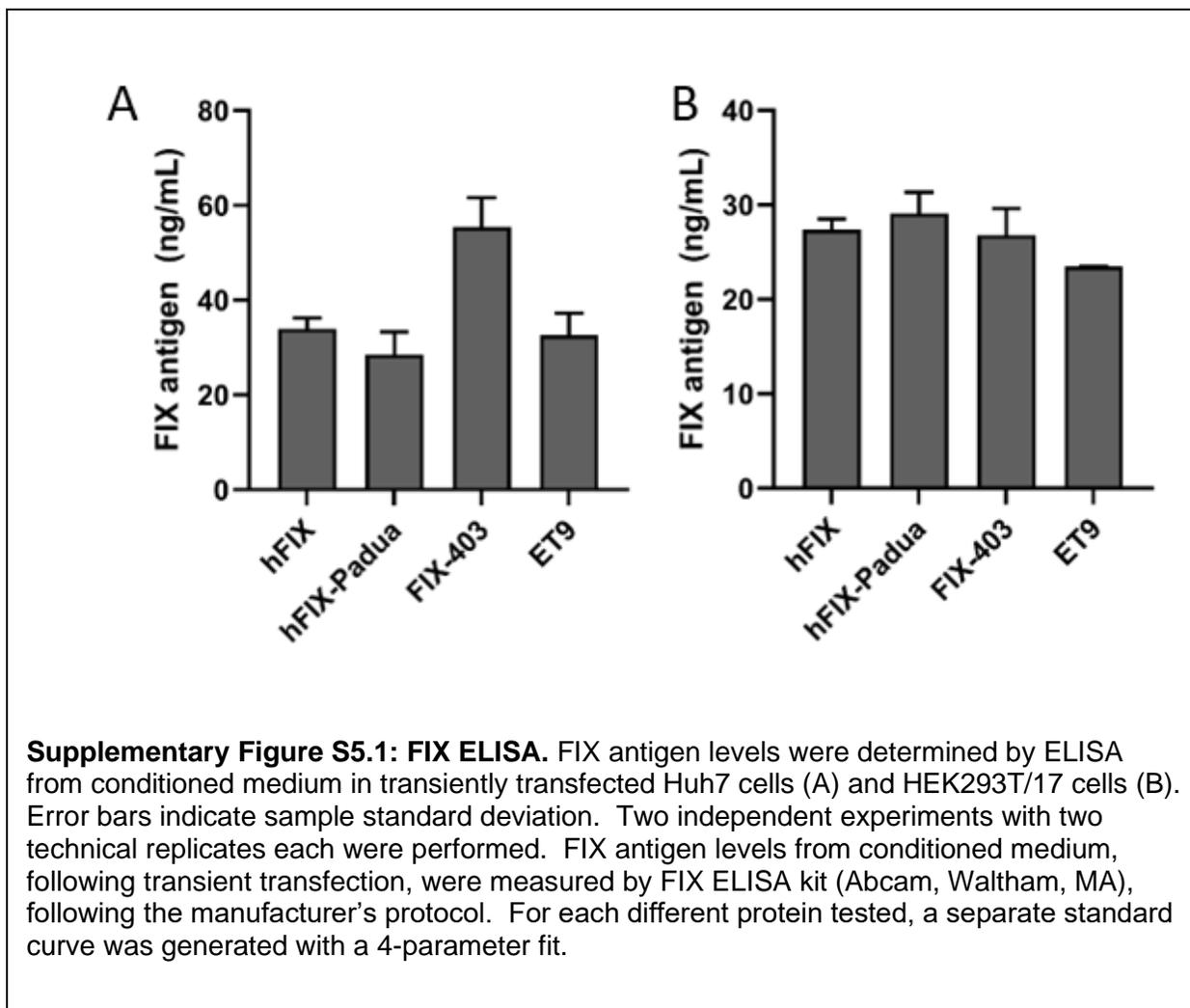
5.6 Acknowledgements

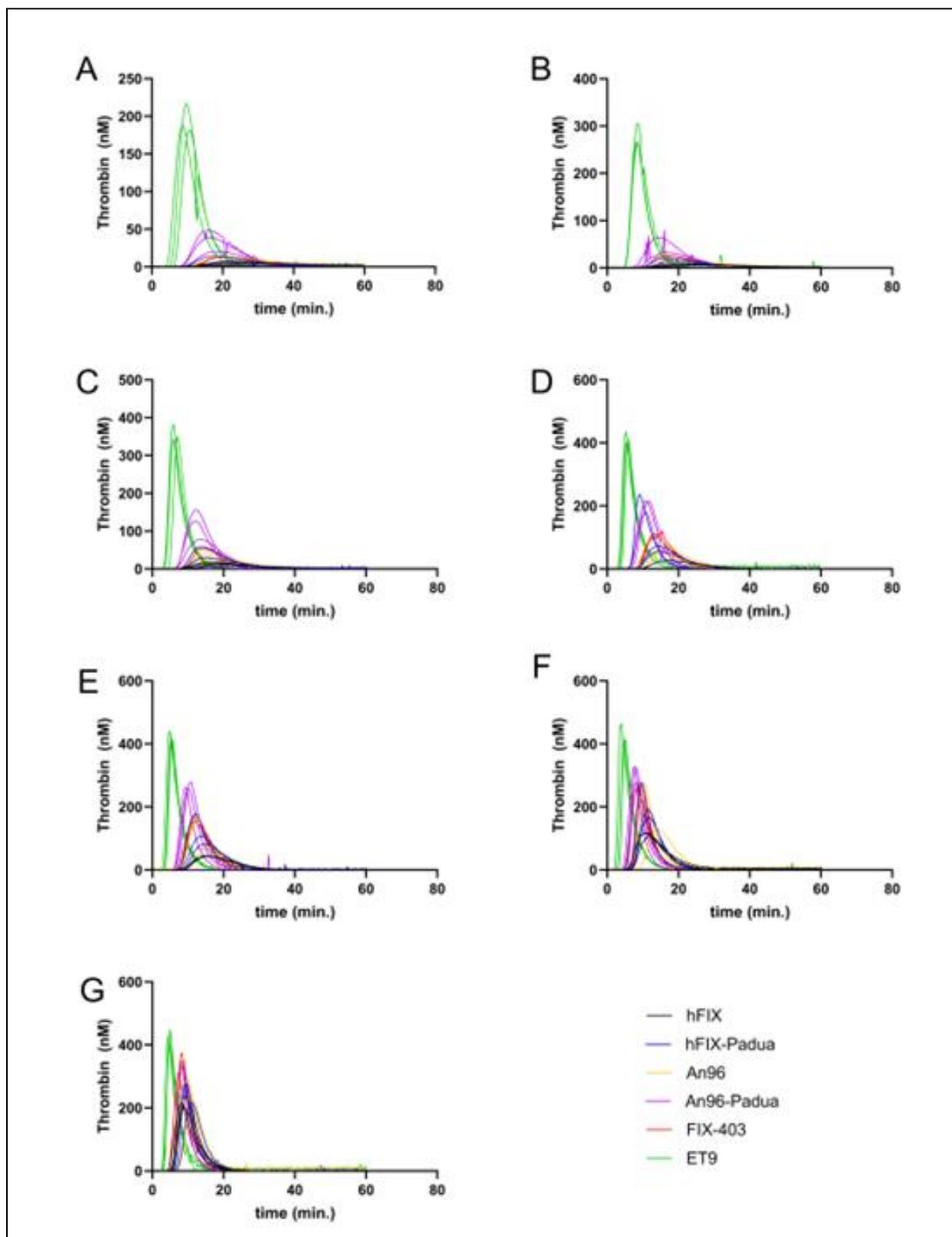
This work was supported by funding from the National Institutes of Health: National Heart, Lung and Blood Institute (HL137128 and U54 HL141981 to H.T.S. and C.B.D., R15HL135658 and U54HL141981 to P.C.S.), the National Hemophilia Foundation Judith Graham Pool Postdoctoral Research Fellowship to K.C.C., and Hemophilia of Georgia, Gene Therapy Program grant to H.T.S. and C.B.D. Graphic in Figure 1A was created with Biorender.com.

Conflict of Interest Disclosure: H.C.B., G.D., S.N.G., C.B.D., H.T.S., K.A.K. and C.W.C. are inventors on patents and patent applications describing the ancestral FIX technology filed by Expression Therapeutics, Emory University, Children's Healthcare of Atlanta, and Georgia Institute of Technology. C.B.D., H.T.S. and H.C.B. are inventors of liver-directed codon-optimization and promoter technology filed by Emory University and Children's Healthcare of

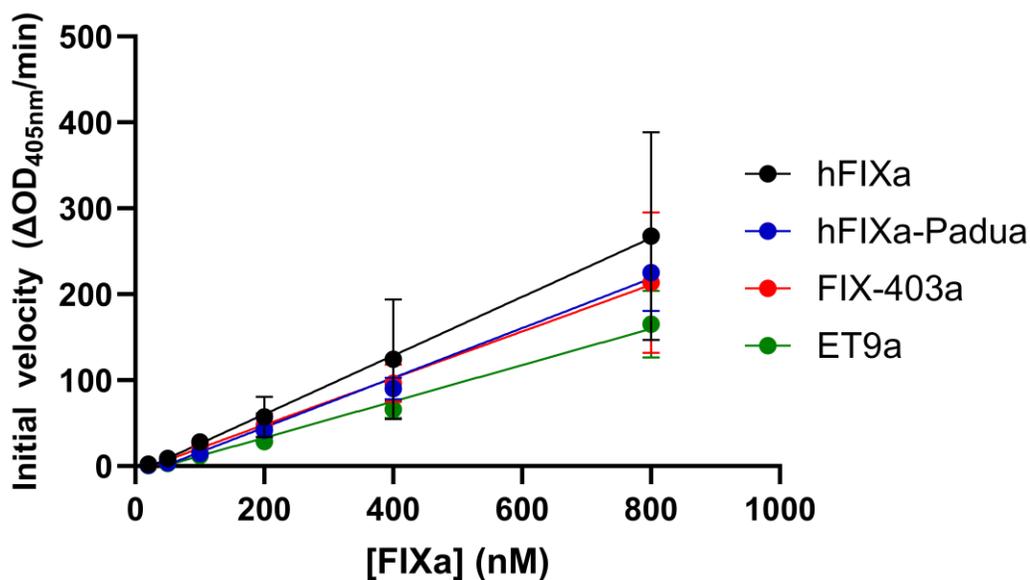
Atlanta. H.T.S. and C.B.D. are cofounders of Expression Therapeutics, Inc., and own equity in the company. H.C.B., G.D. and S.N.G. are employees of Expression Therapeutics, Inc., and own equity in the company. Expression Therapeutics, Inc. has obtained licenses for FIX-An96, liver codon optimized FIX, and synthetic liver-directed promoter intellectual property. K.C.C., P.C.S., and G.M.B., declare no conflicts of interest. The terms of these arrangements have been reviewed and approved by Emory University in accordance with its conflict-of-interest policies.

5.7 Supplementary Information

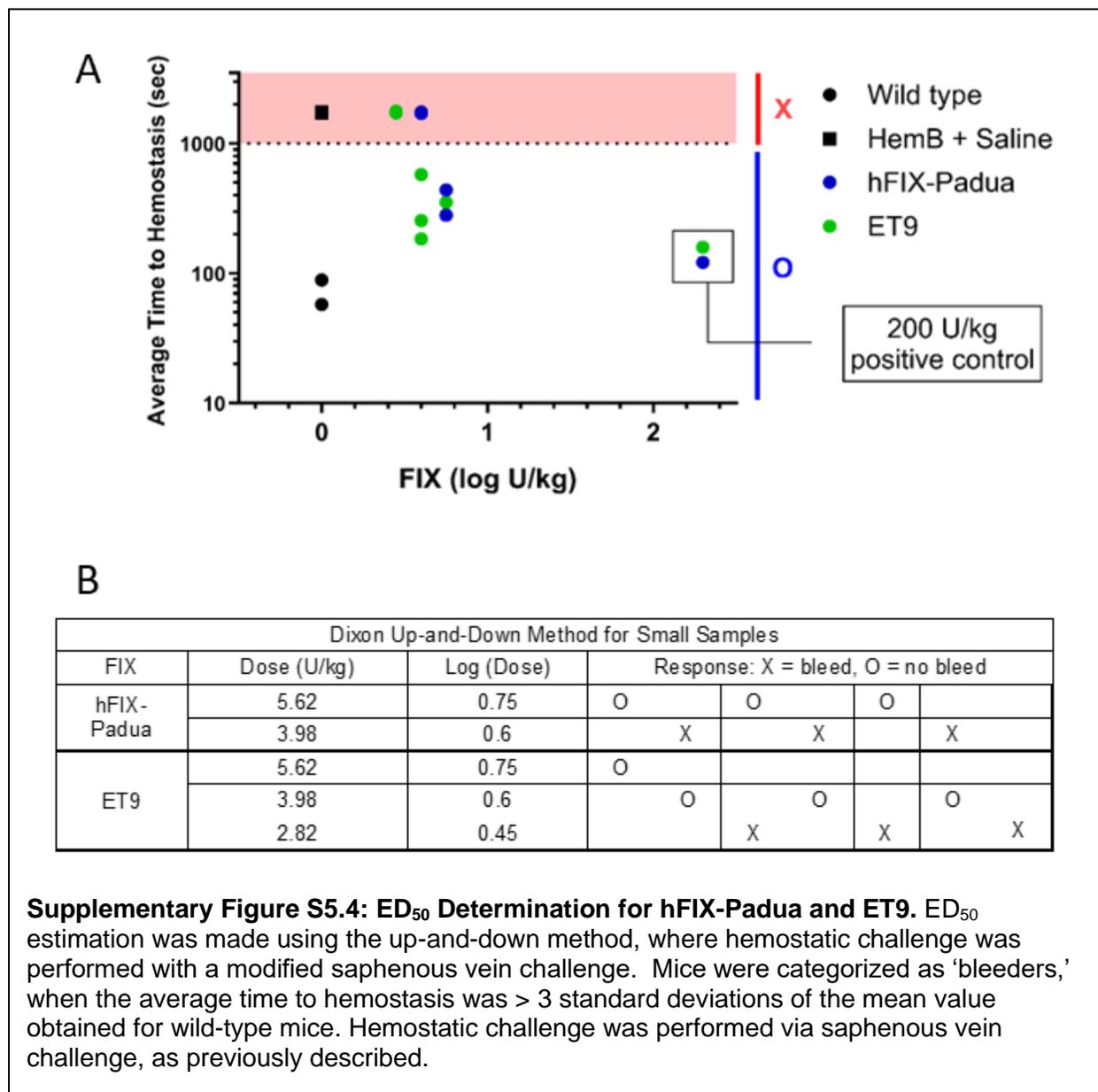




Supplementary Figure S5.2: Individual Thrombograms of FIX Variants. Calibrated automated thrombography was performed using (A) 0.05 $\mu\text{g/mL}$, (B) 0.1 $\mu\text{g/mL}$, (C) 0.25 $\mu\text{g/mL}$, (D) 0.5 $\mu\text{g/mL}$, (E) 1 $\mu\text{g/mL}$, (F) 2.5 $\mu\text{g/mL}$, and (G) 5 $\mu\text{g/mL}$ of each FIX variant. Three independent experiments with two technical replicates each were performed. Each curve represents the mean of the two technical replicates and all three curves generated from the three independent experiments are displayed.



Supplementary Figure S5.3: Amidolytic activity of recombinant FIX variants. Zero – 800nM of FIXa was mixed with a FIXa-specific substrate and initial velocity measured by change in absorbance at 405nm. Slope values obtained by linear regression were 0.34, 0.29, 0.27 and 0.21 $\Delta 405\text{nm}/\text{min}/\text{nM}$ FIXa for hFIX, hFIX-Padua, FIX-403 and ET9, respectively. Data points represent the mean and error bars indicate sample standard deviation. $N \geq 2$ for each FIX variant.



Supplementary Table S5.1: FIX-201 single amino acid substitutions

Molecule name	Amino acid Substitution
FIX-301	p.Ser313Asn
FIX-302	p.Lys323Glu
FIX-303	p.Thr326Val
FIX-304	p.Asn338Asp
FIX-305	p.Arg339Lys
FIX-306	p.Asn361His
FIX-307	p.Arg362Lys
FIX-308	p.Ser367Leu
FIX-309	p.Ile368Val
FIX-310	p.Tyr399Phe
FIX-311	p.Lys404Arg

Amino acid substitutions in Supplementary Table S1 are denoted with FIX-201 as the parent molecule and the substituted amino acid being the human residue at that position.

Supplementary Table S5.2: Minimally humanized FIX-An96 variants

Construct	FIX Construct	Substitutions
FIX-401	304 + 308	p.V132Ala, p.Asp338Asn (304), p.Leu367Ser (308)
FIX-402	304, 307, 308	p.V132Ala , p.Asp338Asn (304), p.Lys362Arg (307), p.Leu367Ser (308)
FIX-403	302, 304, 307, 308	p.V132Ala, p.Glu323Lys (302), p.Asp338Asn (304), p.Lys362Arg (307), p.Leu367Ser (308)
FIX-404	302, 303, 304, 305, 307, 308	p.V132Ala, p.Glu323Lys (302), Val326Thr (303), p.Asp338Asn (304), p.Lys339Arg (305), p.Lys362Arg (307), p.Leu367Ser (308)

Amino acid substitutions in Supplementary Table S1 are denoted with hFIX as the parent molecule and the substituted amino acid being the An96 residue at that position.

Supplementary Table S5.3: Specific activity of FIX variants

FIX Construct	Specific activity (IU/mg)
hFIX	207
hFIX-Padua	2,108
An96	2,182
An96-Padua	10,553
hFIX-403	1,681
hFIX-403-Padua ("ET9")	10,644

Supplementary Table S4: Mutagenic Primers used to develop each single amino acid variant from Supplementary Table S5.1

Primer name	Primer Sequence (5' → 3')
FIX Master Forward	GACAACATCACCCAGTCTAC
FIX Master Reverse	GGCCACAACCTCCTCATAAAG
one F	CAACAAGTACAACCACGACATC
two F	TGGAGCTGGACGAGCCCCTGACC
three F	AGCCCCTGGTCCTGAACAGCTAC
four F	ATCTGCATCGCCGACAGGGA
five F	TCTGCATCGCCAACAAGGAGTA
six F	CGTGTTCCACCGCGGCCGGTCC
seven F	ACGCGTGTTCAACAAGGGCCGGTCC
eight F	CGGTCCGCCCTCATCCTGCAGTA
nine F	GTCCGCCAGCGTCCTGCAGTA
ten F	GTTCTGCGCTGGATTCCACGA
eleven F	ACCACGAGGGAGGGAGGGACTC
one R	GATGTCGTGGTTGTACTIONTGTG
two R	GGTCAGGGGCTCGTCCAGCTCCA

three R	GTAGCTGTTTCAGGACCAGGGGCT
four R	TCCCTGTCTGGCGATGCAGAT
five R	TACTCCTTGTTGGCGATGCAGA
six R	GGACCGGCCGCGGTGGAACACG
seven R	GGACCGGCCCTTGTTGAACACGCGT
eight R	TACTGCAGGATGAGGGCGGACCG
nine R	TACTGCAGGACGCTGGCGGAC
ten R	TCGTGGAATCCAGCGCAGAAC
eleven R	GAGTCCCTCCCTCCCTCGTGGT

Chapter 6

General Discussion

6.1 Summary of Results

The field of gene therapy, and especially the use of AAV vectors has rapidly expanded within the past several decades, with several FDA approvals occurring during the period covered by the experiments presented in this dissertation. Despite numerous approvals and positive outcomes for hundreds of patients, significant limitations remain. For example, commercial manufacturing of viral vector-based gene therapies remains one of the major bottlenecks in this field. While significant efforts are underway to improve production yield and product quality beyond what was initially translated essentially 'as-is' from research laboratories, improving the potency of each viral particle through transgene product engineering has remained relatively ignored. However, just like improving manufacturing yield, a more potent transgene facilitates lower doses required and thus more patients can be treated from a single manufacturing run. Numerous protein engineering methods exist; however, ASR is the primary method utilized in this dissertation. Other protein engineering methods such as rational design, random mutagenesis, and directed evolution have not been as successful for engineering coagulation factors and each have drawbacks which include the need for high-resolution structure/function information and inability to effectively screen hundreds of thousands or millions of mutants. ASR generates a small set (typically < 100) of ancient, computationally-predicted sequences that can then be easily tested. Since the ancestral proteins are based on functional extant sequences, there is a high likelihood that they will be functional as well.

In Chapter 2, ASR was utilized to examine the co-evolution of FVIII and VWF and identify ancestral VWF variants with enhanced properties. FVIII and VWF both play critical roles in mammalian hemostasis and the studies presented provide molecular evidence that FVIII and VWF have evolved together with activities changing in a reciprocal manner; where FVIII evolved towards lower specific activity and biosynthesis VWF evolved towards higher specific activity and biosynthesis. This is consistent with the hypothesis presented in Chapter 5 for the observed

decline in specific activity of FIX over time along the mammalian lineage resulting in *homo sapiens*. The hypothesis is an extension of the hypothesis of platelet evolution which states that platelets are a requirement for invasive placentation observed in placental mammals. As platelets and VWF work together in the primary phase of hemostasis, it is logical that they evolve together towards more efficient activity to prevent death due to hemorrhage during childbirth. However, this drastic improvement in primary hemostasis could require a dampening of the secondary, coagulation factor driven, phase of hemostasis to avoid placing the hemostatic balance too far in the procoagulant direction where thrombosis related mortality becomes driving factor. Specifically in the studies presented in Chapter 2, several novel observations were made. For example, AnFVIII-VWF binding was not greatest between cognate ancestral pairs, except in one instance (An88), but low nanomolar to high picomolar binding affinities were observed for each pair. An70-VWF, which is derived from the undulate lineage had the tightest average affinity and An63-VWF and An84-VWF exhibited the lowest. AnVWF variants were also effective *in vivo*, similarly to previously described AnFVIII variants and suggesting a fundamental conservation of the activities of each coagulation factor among existing and ancient mammals (129).

Chapter 3 investigates a novel iteration of ASR-derived FVIII variants. In the initial iteration of this project, 42 extant sequences were utilized to build the phylogenetic tree and resultant variants had improved specific activity, biosynthesis, and stability of activated co-factor. The second-generation phylogenetic tree, which was utilized in Chapters 2, 3, and 4, utilized an additional 17 extant species and expanded beyond class *Mammalia*. An initial screen of the AnFVIII variants showed that they were all active and trended towards lower activity as they moved down the lineage towards hFVIII. The goal of this work was to identify specific amino acids in top candidate AnFVIII variants that conferred the improvements seen in specific activity. The top 3 AnFVIII variants were An63-FVIII, An84-FVIII, and An70-FVIII. Numerous

hybrid variants with hFVIII were cloned and tested and regions within each candidate AnFVIII were determined to be necessary. In An63-FVIII, the A1, A2 and A3 domains were required to maintain high activity, but the C1 and C2 domains were not. In An70-FVIII, the A1, A2, and C2 domains were required, but the A3 and C1 were not. In An84-FVIII, all 4 domains were required to maintain high activity, as the C1 domain did not have any substitutions with respect to the hFVIII sequence. The A2 domain of An84 appeared to have the largest effect on activity compared to the others, so further resolution was obtained in determining the relevant sequences by first examining sub-domains, and then single amino acid substitutions. Through a negative selection screening approach, P747S and E753K were determined to be significant contributors to activity, although further testing is needed to determine if they are the only contributors.

Chapter 4 utilized ASR similarly to Chapter 2 and Chapter 3 to identify FIX variants with improved pharmaceutical properties. Upon initial screening, several variants, An63-FIX, An96-FIX, and An97-FIX had improved expression and the highest activity variant, An96-FIX was selected as the top candidate molecule moving forward. An96-FIX had about 10-fold greater specific activity than commercial hFIX. An96-FIX did not contain the widely used and previously discussed Padua variant (R384L) and its addition to An96-FIX had an additive effect. This molecule, termed An96-Padua exhibited ~32-fold greater specific activity than hFIX and significantly improved pro-hemostatic properties as identified by thrombin generation assay. Liver-directed AAV-based gene therapy led to significantly higher FIX activity levels *in vivo* than hFIX-Padua, which is the highest activity FIX variant currently used in clinical trials.

Chapter 5 builds off the work done in Chapter 4 by identifying the necessary and sufficient amino acid substitutions within FIX-An96 to confer high activity. FIX-An96 has a 91% sequence identity with respect to the human sequence, and thus contains 42 non-human amino acid substitutions. Using a reductionist, negative selection screening approach, it was

determined that only 5 of the total 42 amino acid substitutions in FIX-An96 were required to maintain high activity. This variant, termed FIX-403, had was fully human except for the amino acid substitutions: V132A, E332K, D328N, K362R, and L367S. Identically to FIX-An96, when the Padua variant was introduced to FIX-403 (termed "ET9"), it had the same additive effect. When exploring the mechanism of increased specific activity through kinetic experiments, it was determined that the ancestral amino acid substitutions primarily operate through enhancement of substrate turnover demonstrated by marked increases in k_{cat} . Additionally, when FIXa-FVIIIa binding studies were performed, binding affinity for FIXa variants followed the trend of: hFIXa > hFIXa-Padua = FIXa-403 > ET9a, where the magnitude of increase of affinity followed the magnitude of differences observed in the specific activities. Lastly, in a liver-directed AAV-based gene therapy setting, ET9 showed equal potency to An96-Padua.

6.2 Implications of Findings

The clinical utility of protein drugs and gene therapies is based on their performance, which can often be improved through bioengineering. Taken in whole, this dissertation utilized ASR to successfully examine the evolution of critical components of the coagulation network: VWF, FVIII, and FIX. The vertebrate coagulation network is thought to have evolved from an initial simple system in jawless hagfish and though numerous genome duplication events and the unique selective became increasing complex and unique for each extant and extinct species. This work provides positive evidence towards the theory that evolution of the coagulation network in humans evolved to become less efficient. Hemostasis is a highly complex and fine-tuned process, and exists on a balance, where one end is hemorrhage and the other is thrombosis. When platelets appeared, the potency of primary hemostasis was increases, therefore, to lower thrombotic risk, secondary hemostasis needed to compensate by becoming less efficient. Interestingly, our lab had performed identical ASR studies to FVII, which primarily operates through the extrinsic pathway, which initiates coagulation. FVII-ASR variants

were not notably better or worse than hFVII, suggesting that intrinsic pathway was more tunable through evolution than the extrinsic pathway, as the intrinsic pathway primarily amplifies coagulation, and the extrinsic pathway initiates it.

Additionally, this work, along with our lab's (129) and other's (212) previously published studies demonstrated the utility of using ASR as a platform for drug development. As stated above, the selective pressure towards a less efficient coagulation network left humans with some of the worst performing FVIII and FIX orthologs, and using ASR, we were able to resurrect ancient and better forming FVIII and FIX variants. The neutral theory of evolution states contends that: "at the molecular level, most evolutionary changes and polymorphisms within species are not caused by natural selection, but by random genetic drift" (241). Recognizing this theory, we were able to determine that just 12% of the amino acid substitutions present in FIX-An96 were required to maintain its high activity, suggesting that the other substitutions existed just due to random genetic drift. Determining the relevant causative amino acid substitutions is important for immunogenicity considerations and improvement of structure-activity-relationship knowledge. Previous coagulation factor bioengineering efforts have resulted in patients developing anti-drug antibodies (inhibitors) to the modified FVII (173) and FIX (242) variants, highlighting how important it is to minimize non-human sequences. In liver-directed gene therapy for hemophilia A and B, there have been no reports of patients developing inhibitors, even when FIX-Padua was used for hemophilia B patients. It is unclear if this is due to utilizing unmodified FVIII or single substitution FIX variants, or due to patient selection bias (no patients with a history of inhibitors have been allowed to enroll in trials) and the immune privileged-nature of the liver (243).

Clinical gene therapies for hemophilia A and B are currently dosed in the $10^{11} - 10^{13}$ vg/kg range, which is equivalent to or greatly exceeding the total number of cells in the human body. One of the major issues facing the field of viral vector-based gene therapy is

manufacturing inefficiency, which are also a major factor in the extreme cost associated with gene therapies. Various techniques such as viral capsid optimization and codon optimization are currently utilized to maximize the potency of viral vectors, however, bioengineering the transgene provides another optimization parameter. This work has shown that utilizing FIX-An96-Padua or ET9 as the transgene for liver-directed AAV gene therapy markedly improves plasma FIX expression compared to hFIX and hFIX-Padua at equivalent vector doses. And all else being equal, this could be translated to an approximate 5-fold increase in the doses able to be made from a single manufacturing run, which can lower costs and improve accessibility. A more potent treatment also reduces the number of viral particles administered to the patient, which can lower side effects.

6.3 Limitations and Future Directions

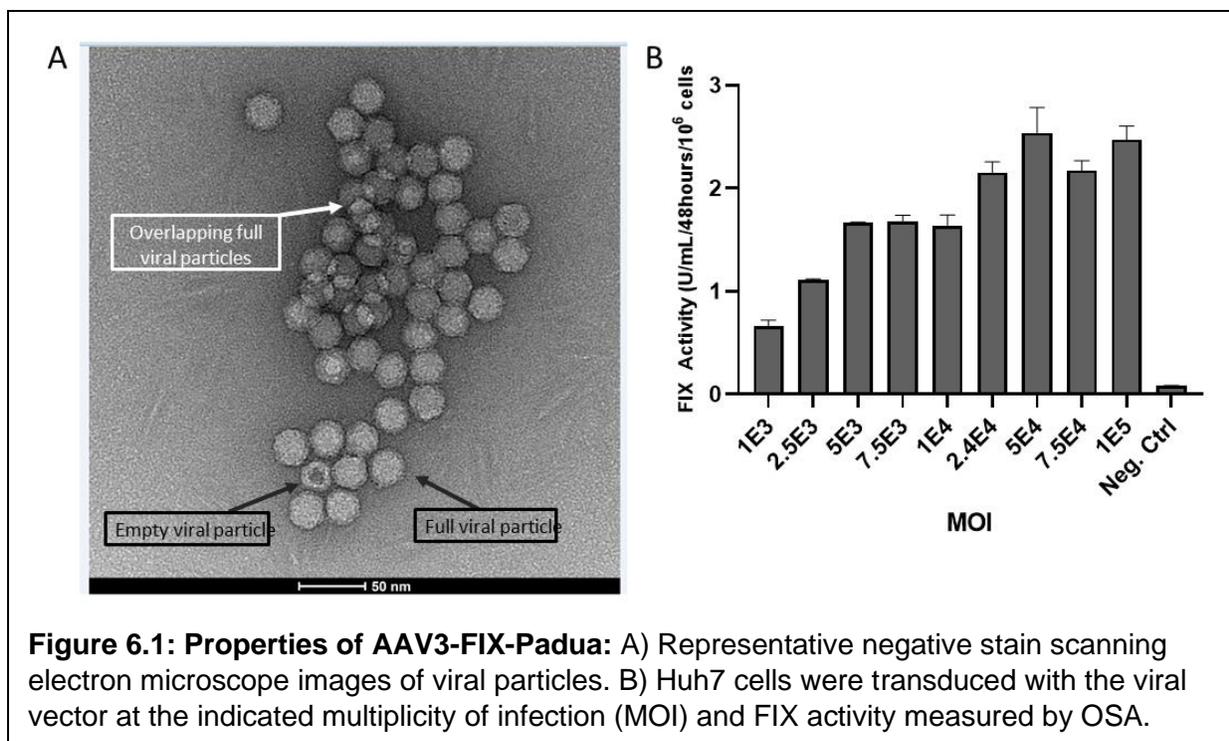
One general limitation for all of Chapter's 2, 3 and 4 is that only a few selected nodes were examined for additional analysis. These nodes were chosen to include several along the human lineage plus a few others in undulate and rodent lineages as well as common ancestors for all three. By failing to examine every generated ASR node, it is possible that additional enhancements could have been missed.

For the FVIII ASR derived variants, total resolution of all critical amino acid substitutions was not obtained. This was due to several factors including: attempting to minimize multiple different AnFVIII variants simultaneously and the large size of FVIII (and therefore many more substitutions to evaluate). Based on our AnFIX amino acid elucidation, we believe that it is possible to reach single amino acid resolution, it will just require additional time and resources. Another limitation for ASR-derived FVIII and FIX is the lack of data examining their interactions with natural coagulation inhibitors. FVIIIa is broken down primarily by activated protein C, with protein S as a cofactor and thrombin to a lesser extent. FIXa is inhibited by antithrombin III. It is

possible that AnFVIII and AnFIX variants could have resistance to these respective coagulation regulators, which could lead to an increase in thrombotic risk.

One of the future directions for this work would be to examine the co-evolution of FVIII and FIX, similarly to how VWF and FVIII were examined in Chapter 2. We know very generally that both FVIII and FIX evolved to be less efficient in human evolution, but the specific mechanisms underlying this observation remain elusive. A future direction for the AnFVIII work (besides elucidation of critical amino acid residues) would be to combine domains from top AnFVIII variants An63, An84, and An70 to see if they have separate and additive mechanisms for improving expression/activity.

The data presented in Chapters 4 and 5 tells a complete story as it culminates in the development of ET9 and the underlying mechanism for improved activity. The next step here would be to move AAV-ET9 into preclinical development. The AAV3 serotype has the highest transduction efficiency for human hepatocytes of all natural AAV serotypes, although it is hindered by poor yields compared to other serotypes (244). Our lab and its associated company, Expression Therapeutics have developed compact and highly active liver-directed promoters along with a proprietary codon optimization algorithm that, all together, improve the potency of gene therapies for bleeding disorders. These elements, along with improvements to AAV3 production have resulted in clinical grade AAV3-hFIX-Padua being produced (182). This viral vector exhibited 96.6% packaging efficiency (calculated by ratio of empty:full capsids) (**Figure 6.1A**) and a dose-dependent response with a peak, *in vitro* (**Figure 6.1B**). The high potency properties of this AAV could be combined with ET9 to generate a superior product.



6.4 Conclusion

This dissertation characterized ASR-derived coagulation factors to isolate their improved properties and the relevant sequence determinants. Here, it was shown that ASR is a viable method for bioengineering coagulation factors and examining their co-evolution. In line with the neutral theory of molecular evolution, we determined that many amino acid substitutions in ancestral coagulation factors had no effect on activity. This led to the development of ET9, a high activity, humanized FIX variant with vastly improved specific activity, which is now a candidate to begin IND enabling studies for the treatment of hemophilia B.

References:

1. Smith SA, Travers RJ, Morrissey JH. 2015. How it all starts: Initiation of the clotting cascade. *Crit Rev Biochem Mol Biol* 50:326-36
2. Berntorp E, Fischer K, Hart DP, Mancuso ME, Stephensen D, et al. 2021. Haemophilia. *Nat Rev Dis Primers* 7:45
3. McRae S. 2011. In *Mechanisms of Vascular Disease: A Reference Book for Vascular Specialists*. University of Adelaide Press: Adelaide, Australia. Number of.
4. Osterud B, Rapaport SI. 1977. 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* 74:5260-4
5. Iorio A, Stonebraker JS, Chambost H, Makris M, Coffin D, et al. 2019. Establishing the Prevalence and Prevalence at Birth of Hemophilia in Males: A Meta-analytic Approach Using National Registries. *Ann Intern Med* 171:540-6
6. White GC, 2nd, Rosendaal F, Aledort LM, Lusher JM, Rothschild C, Ingerslev J. 2001. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost* 85:560
7. Fischer K, Ljung R, Platokouki H, Liesner R, Claeysens S, et al. 2014. Prospective observational cohort studies for studying rare diseases: the European PedNet Haemophilia Registry. *Haemophilia* 20:e280-6
8. Otto JC. 1996. An account of an hemorrhagic disposition existing in certain families. *Clin Orthop Relat Res*:4-6
9. Biggs R, Douglas AS, Macfarlane RG, Dacie JV, Pitney WR, Merskey. 1952. Christmas disease: a condition previously mistaken for haemophilia. *Br Med J* 2:1378-82
10. Hoyer LW. 1994. Hemophilia A. *N Engl J Med* 330:38-47
11. Mannucci PM. 2002. Hemophilia and related bleeding disorders: a story of dismay and success. *Hematology Am Soc Hematol Educ Program*:1-9
12. Nilsson IM. 1993. Experience with prophylaxis in Sweden. *Semin Hematol* 30:16-9
13. Oldenburg J. 2015. Optimal treatment strategies for hemophilia: achievements and limitations of current prophylactic regimens. *Blood* 125:2038-44
14. Oldenburg J, Mahlangu JN, Kim B, Schmitt C, Callaghan MU, et al. 2017. Efficacy of Prophylaxis in Hemophilia A with Inhibitors. *N Engl J Med* 377:809-18
15. Mahlangu J, Oldenburg J, Paz-Priel I, Negrier C, Niggli M, et al. 2018. Efficacy of Prophylaxis in Patients Who Have Hemophilia A without Inhibitors. *N Engl J Med* 379:811-22
16. Ar MC, Balkan C, Kavaklı K. 2019. Extended Half-Life Coagulation Factors: A New Era in the Management of Hemophilia Patients. *Turk J Haematol* 36:141-54
17. Rodriguez-Merchan EC. 2020. The cost of hemophilia treatment: the importance of minimizing it without detriment to its quality. *Expert Rev Hematol* 13:269-74
18. Dimichele D. 2002. Inhibitors: resolving diagnostic and therapeutic dilemmas. *Haemophilia* 8:280-7
19. Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, et al. 1984. Characterization of the human factor VIII gene. *Nature* 312:326-30
20. Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, et al. 1984. Structure of human factor VIII. *Nature* 312:337-42
21. Turner NA, Moake JL. 2015. Factor VIII Is Synthesized in Human Endothelial Cells, Packaged in Weibel-Palade Bodies and Secreted Bound to ULVWF Strings. *PLoS One* 10:e0140740

22. Orlova NA, Kovnir SV, Vorobiev, II, Gabibov AG, Vorobiev AI. 2013. Blood Clotting Factor VIII: From Evolution to Therapy. *Acta Naturae* 5:19-39
23. Zhang B, Kaufman RJ, Ginsburg D. 2005. LMAN1 and MCFD2 form a cargo receptor complex and interact with coagulation factor VIII in the early secretory pathway. *J Biol Chem* 280:25881-6
24. Michnick DA, Pittman DD, Wise RJ, Kaufman RJ. 1994. Identification of individual tyrosine sulfation sites within factor VIII required for optimal activity and efficient thrombin cleavage. *J Biol Chem* 269:20095-102
25. Canis K, Anzengruber J, Garenaux E, Feichtinger M, Benamara K, et al. 2018. In-depth comparison of N-glycosylation of human plasma-derived factor VIII and different recombinant products: from structure to clinical implications. *J Thromb Haemost*
26. Lenting PJ, van Mourik JA, Mertens K. 1998. The life cycle of coagulation factor VIII in view of its structure and function. *Blood* 92:3983-96
27. Anzengruber J, Feichtinger M, Bärnthaler P, Haider N, Ilas J, et al. 2019. How Full-Length FVIII Benefits from Its Heterogeneity – Insights into the Role of the B-Domain. *Pharmaceutical Research* 36:77
28. Adamson R. 1994. Design and operation of a recombinant mammalian cell manufacturing process for rFVIII. *Ann Hematol* 68 Suppl 3:S9-14
29. Turecek PL, Johnsen JM, Pipe SW, O'Donnell JS. 2020. Biological mechanisms underlying inter-individual variation in factor VIII clearance in haemophilia. *Haemophilia* 26:575-83
30. Lenting PJ, CJ VANS, Denis CV. 2007. Clearance mechanisms of von Willebrand factor and factor VIII. *J Thromb Haemost* 5:1353-60
31. Fay PJ. 2004. Activation of factor VIII and mechanisms of cofactor action. *Blood Rev* 18:1-15
32. Fay PJ, Haidaris PJ, Smudzin TM. 1991. Human factor VIIIa subunit structure. Reconstruction of factor VIIIa from the isolated A1/A3-C1-C2 dimer and A2 subunit. *J Biol Chem* 266:8957-62
33. Gale AJ, Cramer TJ, Rozenshteyn D, Cruz JR. 2008. Detailed mechanisms of the inactivation of factor VIIIa by activated protein C in the presence of its cofactors, protein S and factor V. *J Biol Chem* 283:16355-62
34. Andersson HM, Arantes MJ, Crawley JT, Luken BM, Tran S, et al. 2010. Activated protein C cofactor function of protein S: a critical role for Asp95 in the EGF1-like domain. *Blood* 115:4878-85
35. Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K. 1985. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochemistry* 24:3736-50
36. Girolami A, Ferrari S, Cosi E, Santarossa C, Randi ML. 2018. Vitamin K-Dependent Coagulation Factors That May be Responsible for Both Bleeding and Thrombosis (FII, FVII, and FIX). *Clin Appl Thromb Hemost* 24:42s-7s
37. Gao W, Xu Y, Liu H, Gao M, Cao Q, et al. 2020. Characterization of missense mutations in the signal peptide and propeptide of FIX in hemophilia B by a cell-based assay. *Blood Adv* 4:3659-67
38. Orlova NA, Kovnir SV, Vorobiev, II, Gabibov AG. 2012. Coagulation Factor IX for Hemophilia B Therapy. *Acta Naturae* 4:62-73
39. Mann DM, Stafford KA, Poon MC, Matino D, Stafford DW. 2021. The Function of extravascular coagulation factor IX in haemostasis. *Haemophilia* 27:332-9
40. Gailani D, Geng Y, Verhamme I, Sun MF, Bajaj SP, et al. 2014. The mechanism underlying activation of factor IX by factor XIa. *Thromb Res* 133 Suppl 1:S48-51
41. Smith SB, Gailani D. 2008. Update on the physiology and pathology of factor IX activation by factor XIa. *Expert Rev Hematol* 1:87-98
42. Zögg T, Brandstetter H. 2011. Chapter 2 - Complex Assemblies of Factors IX and X Regulate the Initiation, Maintenance, and Shutdown of Blood Coagulation. In *Progress in Molecular Biology and Translational Science*, ed. E Di Cera, 99:51-103: Academic Press. Number of 51-103 pp.

43. Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, et al. 1995. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science* 270:475-80
44. Muul LM, Tuschong LM, Soenen SL, Jagadeesh GJ, Ramsey WJ, et al. 2003. Persistence and expression of the adenosine deaminase gene for 12 years and immune reaction to gene transfer components: long-term results of the first clinical gene therapy trial. *Blood* 101:2563-9
45. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, et al. 2000. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288:669-72
46. Cavazzana M, Six E, Lagresle-Peyrou C, André-Schmutz I, Hacein-Bey-Abina S. 2016. Gene Therapy for X-Linked Severe Combined Immunodeficiency: Where Do We Stand? *Hum Gene Ther* 27:108-16
47. Somia N, Verma IM. 2000. Gene therapy: trials and tribulations. *Nat Rev Genet* 1:91-9
48. Wagner JA, Reynolds T, Moran ML, Moss RB, Wine JJ, et al. 1998. Efficient and persistent gene transfer of AAV-CFTR in maxillary sinus. *Lancet* 351:1702-3
49. Kuzmin DA, Shutova MV, Johnston NR, Smith OP, Fedorin VV, et al. 2021. The clinical landscape for AAV gene therapies. *Nat Rev Drug Discov* 20:173-4
50. Ylä-Herttuala S. 2012. Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. *Mol Ther* 20:1831-2
51. Momin AA, Bankar MP, Bhoite GM. 2016. Study of Common Genetic Variant S447X in Lipoprotein Lipase and Its Association with Lipids and Lipoproteins in Type 2 Diabetic Patients. *Indian J Clin Biochem* 31:286-93
52. Senior M. 2017. After Glybera's withdrawal, what's next for gene therapy? *Nature Biotechnology* 35:491-2
53. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, et al. 2017. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* 390:849-60
54. Lowes LP, Alfano LN, Arnold WD, Shell R, Prior TW, et al. 2019. Impact of Age and Motor Function in a Phase 1/2A Study of Infants With SMA Type 1 Receiving Single-Dose Gene Replacement Therapy. *Pediatr Neurol* 98:39-45
55. Zaidman CM, Proud CM, McDonald CM, Lehman KJ, Goedeker NL, et al. 2023. Delandistrogene Moxeparvovec Gene Therapy in Ambulatory Patients (Aged ≥ 4 to < 8 Years) with Duchenne Muscular Dystrophy: 1-Year Interim Results from Study SRP-9001-103 (ENDEAVOR). *Ann Neurol* 94:955-68
56. Keam SJ. 2022. Eladocogene Exuparvovec: First Approval. *Drugs* 82:1427-32
57. Atchison RW, Casto BC, Hammon WM. 1965. ADENOVIRUS-ASSOCIATED DEFECTIVE VIRUS PARTICLES. *Science* 149:754-6
58. Xiao X, Li J, Samulski RJ. 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J Virol* 72:2224-32
59. Ghosh S, Brown AM, Jenkins C, Campbell K. 2020. Viral Vector Systems for Gene Therapy: A Comprehensive Literature Review of Progress and Biosafety Challenges. *Applied Biosafety* 25:7-18
60. Issa SS, Shaimardanova AA, Solovyeva VV, Rizvanov AA. 2023. Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. *Cells* 12
61. Srivastava A, Lusby EW, Berns KI. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol* 45:555-64
62. Pereira DJ, McCarty DM, Muzyczka N. 1997. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J Virol* 71:1079-88

63. Elmore ZC, Patrick Havlik L, Oh DK, Anderson L, Daaboul G, et al. 2021. The membrane associated accessory protein is an adeno-associated viral egress factor. *Nature Communications* 12:6239
64. Naumer M, Sonntag F, Schmidt K, Nieto K, Panke C, et al. 2012. Properties of the adeno-associated virus assembly-activating protein. *J Virol* 86:13038-48
65. Yang J, Zhou W, Zhang Y, Zidon T, Ritchie T, Engelhardt JF. 1999. Concatamerization of adeno-associated virus circular genomes occurs through intermolecular recombination. *J Virol* 73:9468-77
66. Naso MF, Tomkowicz B, Perry WL, Strohl WR. 2017. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* 31:317-34
67. Clément N, Grieger JC. 2016. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol Ther Methods Clin Dev* 3:16002
68. Wang D, Tai PWL, Gao G. 2019. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* 18:358-78
69. Ferrari FK, Samulski T, Shenk T, Samulski RJ. 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* 70:3227-34
70. McCarty DM, Fu H, Monahan PE, Toulson CE, Naik P, Samulski RJ. 2003. Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. *Gene Ther* 10:2112-8
71. Hordeaux J, Hinderer C, Goode T, Katz N, Buza EL, et al. 2018. Toxicology Study of Intra-Cisterna Magna Adeno-Associated Virus 9 Expressing Human Alpha-L-Iduronidase in Rhesus Macaques. *Mol Ther Methods Clin Dev* 10:79-88
72. MacLachlan TK, Milton MN, Turner O, Tukov F, Choi VW, et al. 2018. Nonclinical Safety Evaluation of scAAV8-RLBP1 for Treatment of RLBP1 Retinitis Pigmentosa. *Mol Ther Methods Clin Dev* 8:105-20
73. Thomsen G, Burghes AHM, Hsieh C, Do J, Chu BTT, et al. 2021. Biodistribution of onasemnogene abeparvovec DNA, mRNA and SMN protein in human tissue. *Nat Med* 27:1701-11
74. Chen SJ, Sanmiguel J, Lock M, McMenemy D, Draper C, et al. 2013. Biodistribution of AAV8 vectors expressing human low-density lipoprotein receptor in a mouse model of homozygous familial hypercholesterolemia. *Hum Gene Ther Clin Dev* 24:154-60
75. Sun K, Liao MZ. 2022. Clinical Pharmacology Considerations on Recombinant Adeno-Associated Virus-Based Gene Therapy. *The Journal of Clinical Pharmacology* 62:S79-S94
76. Fong S, Yates B, Sihn CR, Mattis AN, Mitchell N, et al. 2022. Interindividual variability in transgene mRNA and protein production following adeno-associated virus gene therapy for hemophilia A. *Nat Med* 28:789-97
77. Davidson MD, Khetani SR. 2020. Intermittent Starvation Extends the Functional Lifetime of Primary Human Hepatocyte Cultures. *Toxicol Sci* 174:266-77
78. Fong S, Handyside B, Sihn CR, Liu S, Zhang L, et al. 2020. Induction of ER Stress by an AAV5 BDD FVIII Construct Is Dependent on the Strength of the Hepatic-Specific Promoter. *Mol Ther Methods Clin Dev* 18:620-30
79. Nathwani AC. 2022. Gene therapy for hemophilia. *Hematology Am Soc Hematol Educ Program* 2022:569-78
80. Nault JC, Datta S, Imbeaud S, Franconi A, Mallet M, et al. 2015. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* 47:1187-93
81. Nowrouzi A, Penaud-Budloo M, Kaepffel C, Appelt U, Le Guiner C, et al. 2012. Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. *Mol Ther* 20:1177-86

82. Chandler RJ, LaFave MC, Varshney GK, Trivedi NS, Carrillo-Carrasco N, et al. 2015. Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J Clin Invest* 125:870-80
83. Nguyen GN, Everett JK, Kafle S, Roche AM, Raymond HE, et al. 2021. A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. *Nat Biotechnol* 39:47-55
84. Lutz S, Iamurri SM. 2018. Protein Engineering: Past, Present, and Future. *Methods Mol Biol* 1685:1-12
85. Brown HC, Zakas PM, George SN, Parker ET, Spencer HT, Doering CB. 2018. Target-Cell-Directed Bioengineering Approaches for Gene Therapy of Hemophilia A. *Mol Ther Methods Clin Dev* 9:57-69
86. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, et al. 2015. Codon optimality is a major determinant of mRNA stability. *Cell* 160:1111-24
87. Faust SM, Bell P, Cutler BJ, Ashley SN, Zhu Y, et al. 2013. CpG-depleted adeno-associated virus vectors evade immune detection. *J Clin Invest* 123:2994-3001
88. Martino AT, Suzuki M, Markusic DM, Zolotukhin I, Ryals RC, et al. 2011. The genome of self-complementary adeno-associated viral vectors increases Toll-like receptor 9-dependent innate immune responses in the liver. *Blood* 117:6459-68
89. Li C, Samulski RJ. 2020. Engineering adeno-associated virus vectors for gene therapy. *Nat Rev Genet* 21:255-72
90. Martino AT, Basner-Tschakarjan E, Markusic DM, Finn JD, Hinderer C, et al. 2013. Engineered AAV vector minimizes in vivo targeting of transduced hepatocytes by capsid-specific CD8+ T cells. *Blood* 121:2224-33
91. Louis Jeune V, Joergensen JA, Hajjar RJ, Weber T. 2013. Pre-existing anti-adeno-associated virus antibodies as a challenge in AAV gene therapy. *Hum Gene Ther Methods* 24:59-67
92. Li X, Wei X, Lin J, Ou L. 2022. A versatile toolkit for overcoming AAV immunity. *Front Immunol* 13:991832
93. Herzog RW, Hagstrom JN, Kung SH, Tai SJ, Wilson JM, et al. 1997. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. *Proc Natl Acad Sci U S A* 94:5804-9
94. Herzog RW, Yang EY, Couto LB, Hagstrom JN, Elwell D, et al. 1999. Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector. *Nat Med* 5:56-63
95. Snyder RO, Miao C, Meuse L, Tubb J, Donahue BA, et al. 1999. Correction of hemophilia B in canine and murine models using recombinant adeno-associated viral vectors. *Nat Med* 5:64-70
96. Manno CS, Chew AJ, Hutchison S, Larson PJ, Herzog RW, et al. 2003. AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. *Blood* 101:2963-72
97. Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, et al. 2006. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* 12:342-7
98. Mingozi F, High KA. 2013. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* 122:23-36
99. Pipe SW, Leebeek FWG, Recht M, Key NS, Castaman G, et al. 2023. Gene Therapy with Etranacogene Dezaparvovec for Hemophilia B. *N Engl J Med* 388:706-18
100. Nathwani AC, Reiss U, Tuddenham E, Chowdhary P, McIntosh J, et al. 2018. Adeno-Associated Mediated Gene Transfer for Hemophilia B: 8 Year Follow up and Impact of Removing "Empty Viral Particles" on Safety and Efficacy of Gene Transfer. *Blood* 132:491-

101. Shah J, Kim H, Sivamurthy K, Monahan PE, Fries M. 2023. Comprehensive analysis and prediction of long-term durability of factor IX activity following etranacogene dezaparvovec gene therapy in the treatment of hemophilia B. *Curr Med Res Opin* 39:227-37
102. Roth DA, Tawa NE, Jr., O'Brien JM, Treco DA, Selden RF. 2001. Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A. *N Engl J Med* 344:1735-42
103. Doering CB, Healey JF, Parker ET, Barrow RT, Lollar P. 2002. High level expression of recombinant porcine coagulation factor VIII. *J Biol Chem* 277:38345-9
104. Doering CB, Healey JF, Parker ET, Barrow RT, Lollar P. 2004. Identification of porcine coagulation factor VIII domains responsible for high level expression via enhanced secretion. *J Biol Chem* 279:6546-52
105. Spencer HT, Denning G, Gautney RE, Dropulic B, Roy AJ, et al. 2011. Lentiviral vector platform for production of bioengineered recombinant coagulation factor VIII. *Mol Ther* 19:302-9
106. Johnston JM, Denning G, Doering CB, Spencer HT. 2013. Generation of an optimized lentiviral vector encoding a high-expression factor VIII transgene for gene therapy of hemophilia A. *Gene Ther* 20:607-15
107. Doering CB, Denning G, Shields JE, Fine EJ, Parker ET, et al. 2018. Preclinical Development of a Hematopoietic Stem and Progenitor Cell Bioengineered Factor VIII Lentiviral Vector Gene Therapy for Hemophilia A. *Hum Gene Ther* 29:1183-201
108. Rangarajan S, Walsh L, Lester W, Perry D, Madan B, et al. 2017. AAV5-Factor VIII Gene Transfer in Severe Hemophilia A. *N Engl J Med* 377:2519-30
109. Ozelo MC, Mahlangu J, Pasi KJ, Giermasz A, Leavitt AD, et al. 2022. Valoctocogene Roxaparvovec Gene Therapy for Hemophilia A. *N Engl J Med* 386:1013-25
110. Blair HA. 2022. Valoctocogene Roxaparvovec: First Approval. *Drugs* 82:1505-10
111. Pittman DD, Marquette KA, Kaufman RJ. 1994. Role of the B domain for factor VIII and factor V expression and function. *Blood* 84:4214-25
112. Sabatino DE, Freguia CF, Toso R, Santos A, Merricks EP, et al. 2009. Recombinant canine B-domain-deleted FVIII exhibits high specific activity and is safe in the canine hemophilia A model. *Blood* 114:4562-5
113. Zakas PM, Gangadharan B, Almeida-Porada G, Porada CD, Spencer HT, Doering CB. 2012. Development and characterization of recombinant ovine coagulation factor VIII. *PLoS One* 7:e49481
114. Doering C, Parker ET, Healey JF, Craddock HN, Barrow RT, Lollar P. 2002. Expression and characterization of recombinant murine factor VIII. *Thrombosis and haemostasis* 88:450-8
115. Samelson-Jones BJ, Arruda VR. 2019. Protein-Engineered Coagulation Factors for Hemophilia Gene Therapy. *Mol Ther Methods Clin Dev* 12:184-201
116. Simioni P, Tormene D, Tognin G, Gavasso S, Bulato C, et al. 2009. X-Linked Thrombophilia with a Mutant Factor IX (Factor IX Padua). *New England Journal of Medicine* 361:1671-5
117. Sugahara Y, Catalfamo J, Brooks M, Hitomi E, Bajaj SP, Kurachi K. 1996. Isolation and characterization of canine factor IX. *Thromb Haemost* 75:450-5
118. Knight KA, Coyle CW, Radford CE, Parker ET, Fedanov A, et al. 2021. Identification of coagulation factor IX variants with enhanced activity through ancestral sequence reconstruction. *Blood Adv* 5:3333-43
119. Swofford DL, Maddison WP. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87:199-229
120. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368-76

121. Grant MA, Beeler DL, Spokes KC, Chen J, Dharaneeswaran H, et al. 2017. Identification of extant vertebrate Myxine glutinosa VWF: evolutionary conservation of primary hemostasis. *Blood* 130:2548-58
122. Doolittle RF, Jiang Y, Nand J. 2008. Genomic evidence for a simpler clotting scheme in jawless vertebrates. *J Mol Evol* 66:185-96
123. Jiang Y, Doolittle RF. 2003. The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes. *Proc Natl Acad Sci U S A* 100:7527-32
124. Doolittle RF. 2009. Step-by-step evolution of vertebrate blood coagulation. *Cold Spring Harb Symp Quant Biol* 74:35-40
125. Healey JF, Parker ET, Barrow RT, Langley TJ, Church WR, Lollar P. 2009. The comparative immunogenicity of human and porcine factor VIII in haemophilia A mice. *Thromb Haemost* 102:35-41
126. Nguyen GN, George LA, Siner JI, Davidson RJ, Zander CB, et al. 2017. Novel factor VIII variants with a modified furin cleavage site improve the efficacy of gene therapy for hemophilia A. *J Thromb Haemost* 15:110-21
127. Siner JI, Iacobelli NP, Sabatino DE, Ivanciu L, Zhou S, et al. 2013. Minimal modification in the factor VIII B-domain sequence ameliorates the murine hemophilia A phenotype. *Blood* 121:4396-403
128. Zakas PM, Vanijcharoenkarn K, Markovitz RC, Meeks SL, Doering CB. 2015. Expanding the ortholog approach for hemophilia treatment complicated by factor VIII inhibitors. *J Thromb Haemost* 13:72-81
129. Zakas PM, Brown HC, Knight K, Meeks SL, Spencer HT, et al. 2017. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat Biotechnol* 35:35-7
130. Zurbano MJ, Escolar G, Heras M, Ordinas A, Castillo R. 2000. Differential aspects of the glycoprotein Ib-von Willebrand factor axis in human and pig species. *Haematologica* 85:514-9
131. Muchitsch EM, Dietrich B, Rottensteiner H, Auer W, Nehrbass D, et al. 2010. Preclinical testing of human recombinant von Willebrand factor: ADAMTS13 cleavage capacity in animals as criterion for species suitability. *Semin Thromb Hemost* 36:522-8
132. Brinkhous KM, Thomas BD, Ibrahim SA, Read MS. 1977. Plasma levels of platelet aggregating factor/von Willebrand factor in various species. *Thromb Res* 11:345-55
133. Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451:704-7
134. Gaucher EA, Thomson JM, Burgan MF, Benner SA. 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425:285-8
135. Doering CB, Denning G, Dooriss K, Gangadharan B, Johnston JM, et al. 2009. Directed engineering of a high-expression chimeric transgene as a strategy for gene therapy of hemophilia A. *Mol Ther* 17:1145-54
136. Berber E, Ozbil M, Brown C, Baslar Z, Caglayan SH, Lillicrap D. 2017. Functional characterisation of the type 1 von Willebrand disease candidate VWF gene variants: p.M771I, p.L881R and p.P1413L. *Blood Transfus* 15:548-56
137. Pruss CM, Golder M, Bryant A, Hegadorn CA, Burnett E, et al. 2011. Pathologic mechanisms of type 1 VWD mutations R1205H and Y1584C through in vitro and in vivo mouse models. *Blood* 117:4358-66
138. Golder M, Pruss CM, Hegadorn C, Mewburn J, Laverty K, et al. 2010. Mutation-specific hemostatic variability in mice expressing common type 2B von Willebrand disease substitutions. *Blood* 115:4862-9

139. Pruss CM, Notley CR, Hegadorn CA, O'Brien LA, Lillicrap D. 2008. ADAMTS13 cleavage efficiency is altered by mutagenic and, to a lesser extent, polymorphic sequence changes in the A1 and A2 domains of von Willebrand factor. *Br J Haematol* 143:552-8
140. Pruss CM, Golder M, Bryant A, Hegadorn C, Haberichter S, Lillicrap D. 2012. Use of a mouse model to elucidate the phenotypic effects of the von Willebrand factor cleavage mutants, Y1605A/M1606A and R1597W. *J Thromb Haemost* 10:940-50
141. Stakiw J, Bowman M, Hegadorn C, Pruss C, Notley C, et al. 2008. The effect of exercise on von Willebrand factor and ADAMTS-13 in individuals with type 1 and type 2B von Willebrand disease. *J Thromb Haemost* 6:90-6
142. Karlsson R, Katsamba PS, Nordin H, Pol E, Myszka DG. 2006. Analyzing a kinetic titration series using affinity biosensors. *Anal Biochem* 349:136-47
143. Myszka DG. 1997. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Curr Opin Biotechnol* 8:50-7
144. Favalaro EJ, Mohammed S. 2014. Towards improved diagnosis of von Willebrand disease: comparative evaluations of several automated von Willebrand factor antigen and activity assays. *Thromb Res* 134:1292-300
145. Ni Y, Nesrallah J, Agnew M, Geske FJ, Favalaro EJ. 2013. Establishment and characterization of a new and stable collagen-binding assay for the assessment of von Willebrand factor activity. *Int J Lab Hematol* 35:170-6
146. Pazos F, Valencia A. 2001. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* 14:609-14
147. Swystun LL, Georgescu I, Mewburn J, Deforest M, Nesbitt K, et al. 2017. Abnormal von Willebrand factor secretion, factor VIII stabilization and thrombus dynamics in type 2N von Willebrand disease mice. *J Thromb Haemost* 15:1607-19
148. Ghosh A, Vo A, Twiss BK, Kretz CA, Jozwiak MA, et al. 2012. Characterization of zebrafish von Willebrand factor reveals conservation of domain structure, multimerization, and intracellular storage. *Adv Hematol* 2012:214209
149. Muia J, Zhu J, Greco SC, Vanhoorelbeke K, Gupta G, et al. 2019. Phylogenetic and functional analysis of ADAMTS13 identifies highly conserved domains essential for allosteric regulation. *Blood* 133:1899-908
150. Deng W, Wang Y, Druzak SA, Healey JF, Syed AK, et al. 2017. A discontinuous autoinhibitory module masks the A1 domain of von Willebrand factor. *J Thromb Haemost* 15:1867-77
151. Nichols TC, Bellinger DA, Merricks EP, Raymer RA, Kloos MT, et al. 2010. Porcine and canine von Willebrand factor and von Willebrand disease: hemostasis, thrombosis, and atherosclerosis studies. *Thrombosis* 2010:461238
152. Rietveld IM, Lijfering WM, le Cessie S, Bos MHA, Rosendaal FR, et al. 2019. High levels of coagulation factors and venous thrombosis risk: strongest association for factor VIII and von Willebrand factor. *J Thromb Haemost* 17:99-109
153. Yee A, Gildersleeve RD, Gu S, Kretz CA, McGee BM, et al. 2014. A von Willebrand factor fragment containing the D'D3 domains is sufficient to stabilize coagulation factor VIII in mice. *Blood* 124:445-52
154. Yee A, Oleskie AN, Dosey AM, Kretz CA, Gildersleeve RD, et al. 2015. Visualization of an N-terminal fragment of von Willebrand factor in complex with factor VIII. *Blood* 126:939-42
155. Chiu PL, Bou-Assaf GM, Chhabra ES, Chambers MG, Peters RT, et al. 2015. Mapping the interaction between factor VIII and von Willebrand factor by electron microscopy and mass spectrometry. *Blood* 126:935-8
156. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW. 2012. Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet* 8:e1003072

157. Harms MJ, Eick GN, Goswami D, Colucci JK, Griffin PR, et al. 2013. Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc Natl Acad Sci U S A* 110:11475-80
158. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. *Nature* 481:360-4
159. Kratzer JT, Lanaspá MA, Murphy MN, Cicerchi C, Graves CL, et al. 2014. Evolutionary history and metabolic insights of ancient mammalian uricases. *Proc Natl Acad Sci U S A* 111:3763-8
160. Hendrikse NM, Holmberg Larsson A, Svensson Gelius S, Kuprin S, Nordling E, Syrén PO. 2020. Exploring the therapeutic potential of modern and ancestral phenylalanine/tyrosine ammonia-lyases as supplementary treatment of hereditary tyrosinemia. *Sci Rep* 10:1315
161. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β -lactamases. *J Am Chem Soc* 135:2899-902
162. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* 100:1056-61
163. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, et al. 2007. The delayed rise of present-day mammals. *Nature* 446:507-12
164. Vlot AJ, Koppelman SJ, van den Berg MH, Bouma BN, Sixma JJ. 1995. The affinity and stoichiometry of binding of human factor VIII to von Willebrand factor. *Blood* 85:3150-7
165. Brown HC, Gangadharan B, Doering CB. 2011. Enhanced biosynthesis of coagulation factor VIII through diminished engagement of the unfolded protein response. *J Biol Chem* 286:24451-7
166. Kaufman RJ, Pipe SW, Tagliavacca L, Swaroop M, Moussalli M. 1997. Biosynthesis, assembly and secretion of coagulation factor VIII. *Blood Coagul Fibrinolysis* 8 Suppl 2:S3-14
167. Marquette KA, Pittman DD, Kaufman RJ. 1995. A 110-amino acid region within the A1-domain of coagulation factor VIII inhibits secretion from mammalian cells. *J Biol Chem* 270:10297-303
168. Burton M, Nakai H, Colosi P, Cunningham J, Mitchell R, Couto L. 1999. Coexpression of factor VIII heavy and light chain adeno-associated viral vectors produces biologically active protein. *Proc Natl Acad Sci U S A* 96:12725-30
169. Meyer D, Fressinaud E, Gaucher C, Lavergne JM, Hilbert L, et al. 1997. Gene defects in 150 unrelated French cases with type 2 von Willebrand disease: from the patient to the gene. INSERM Network on Molecular Abnormalities in von Willebrand Disease. *Thromb Haemost* 78:451-6
170. Ahmad F, Jan R, Kannan M, Obser T, Hassan MI, et al. 2013. Characterisation of mutations and molecular studies of type 2 von Willebrand disease. *Thromb Haemost* 109:39-46
171. Lohmeier HK, Slobodianuk TL, Kanaji S, Haberichter SL, Montgomery RR, Flood VH. 2020. von Willebrand factor variant D1472H has no effect in mice with humanized VWF-platelet interactions. *Blood Adv* 4:4065-8
172. Lazarus RA, Scheiflinger F. 2017. Mining ancient proteins for next-generation drugs. *Nat Biotechnol* 35:28-9
173. Mahlangu JN, Weldingh KN, Lentz SR, Kaicker S, Karim FA, et al. 2015. Changes in the amino acid sequence of the recombinant human factor VIIa analog, vatreptacog alfa, are associated with clinical immunogenicity. *J Thromb Haemost* 13:1989-98
174. Lamberth K, Reedtz-Runge SL, Simon J, Klementyeva K, Pandey GS, et al. 2017. Post hoc assessment of the immunogenicity of bioengineered factor VIIa demonstrates the use of preclinical tools. *Sci Transl Med* 9
175. Sabatino DE, Freguia CF, Toso R, Santos A, Merricks EP, et al. 2009. Recombinant canine B-domain-deleted FVIII exhibits high specific activity and is safe in the canine hemophilia A model. *Blood* 114:4562-5

176. Lillcrap D, Schiviz A, Apostol C, Wojciechowski P, Horling F, et al. 2015. Porcine recombinant factor VIII (Obizur; OBI-1; BAX801): product characteristics and preclinical profile. *Haemophilia*
177. Pratt KP. 2016. Engineering less immunogenic and antigenic FVIII proteins. *Cellular immunology* 301:12-7
178. Merkl R, Sterner R. 2016. Ancestral protein reconstruction: techniques and applications. *Biol Chem* 397:1-21
179. Kimura M. 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc Natl Acad Sci U S A* 76:3440-4
180. Cao W, Dong B, Horling F, Firman JA, Lengler J, et al. 2020. Minimal Essential Human Factor VIII Alterations Enhance Secretion and Gene Therapy Efficiency. *Mol Ther Methods Clin Dev* 19:486-95
181. Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat Commun* 7:12847
182. Brown HC, Doering CB, Herzog RW, Ling C, Markusic DM, et al. 2020. Development of a Clinical Candidate AAV3 Vector for Gene Therapy of Hemophilia B. *Hum Gene Ther* 31:1114-23
183. Kovnir SV, Orlova NA, Shakhparonov MI, Skryabin KG, Gabibov AG, Vorobiev, II. 2018. A Highly Productive CHO Cell Line Secreting Human Blood Clotting Factor IX. *Acta Naturae* 10:51-65
184. Markusic DM, Nichols TC, Merricks EP, Palaschak B, Zolotukhin I, et al. 2017. Evaluation of engineered AAV capsids for hepatic factor IX gene transfer in murine and canine models. *J Transl Med* 15:94
185. McIntosh NL, Berguig GY, Karim OA, Cortesio CL, De Angelis R, et al. 2021. Comprehensive characterization and quantification of adeno associated vectors by size exclusion chromatography and multi angle light scattering. *Sci Rep* 11:3012
186. Buyue Y, Whinna HC, Sheehan JP. 2008. The heparin-binding exosite of factor IXa is a critical regulator of plasma thrombin generation and venous thrombosis. *Blood* 112:3234-41
187. Pastoft AE, Lykkesfeldt J, Ezban M, Tranholm M, Whinna HC, Lauritzen B. 2012. A sensitive venous bleeding model in haemophilia A mice: effects of two recombinant FVIII products (N8 and Advate®). *Haemophilia* 18:782-8
188. Dixon WJ. 1965. The Up-and-Down Method for Small Samples. *Journal of the American Statistical Association* 60:967-78
189. Dixon WJ. 1991. Staircase bioassay: the up-and-down method. *Neurosci Biobehav Rev* 15:47-50
190. Samelson-Jones BJ, Finn JD, George LA, Camire RM, Arruda VR. 2019. Hyperactivity of factor IX Padua (R338L) depends on factor VIIIa cofactor activity. *JCI Insight* 5
191. Simioni P, Tormene D, Tognin G, Gavasso S, Bulato C, et al. 2009. X-linked thrombophilia with a mutant factor IX (factor IX Padua). *N Engl J Med* 361:1671-5
192. Baranyi L, Doering CB, Denning G, Gautney RE, Harris KT, et al. 2013. Rapid generation of stable cell lines expressing high levels of erythropoietin, factor VIII, and an antihuman CD20 antibody using lentiviral vectors. *Hum Gene Ther Methods* 24:214-27
193. Dooriss KL, Denning G, Gangadharan B, Javazon EH, McCarty DA, et al. 2009. Comparison of factor VIII transgenes bioengineered for improved expression in gene therapy of hemophilia A. *Hum Gene Ther* 20:465-78
194. George LA, Sullivan SK, Giermasz A, Rasko JEJ, Samelson-Jones BJ, et al. 2017. Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med* 377:2215-27
195. Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* 20:604-13
196. McIntosh J, Lenting PJ, Rosales C, Lee D, Rabbanian S, et al. 2013. Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. *Blood* 121:3335-44

197. Nathwani AC, Gray JT, Ng CY, Zhou J, Spence Y, et al. 2006. Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. *Blood* 107:2653-61
198. Bertina RM. 2003. Elevated clotting factor levels and venous thrombosis. *Pathophysiol Haemost Thromb* 33:395-400
199. 2020. High-dose AAV gene therapy deaths. *Nat Biotechnol* 38:910
200. Quade-Lyssy P, Abriss D, Milanov P, Ungerer C, Königs C, et al. 2014. Next generation FIX muteins with FVIII-independent activity for alternative treatment of hemophilia A. *J Thromb Haemost* 12:1861-73
201. Kao CY, Yang SJ, Tao MH, Jeng YM, Yu IS, Lin SW. 2013. Incorporation of the factor IX Padua mutation into FIX-Triple improves clotting activity in vitro and in vivo. *Thromb Haemost* 110:244-56
202. Kotin RM, Wright JF. 2019. Recombinant Adeno-Associated Virus Quality Control for Non-Clinical and Clinical Vectors: How an Unregulated Commercial Sector Can Compromise Development of New Gene Therapies. *Hum Gene Ther* 30:1447-8
203. Heo YA. 2023. Etranacogene Dezaparvovec: First Approval. *Drugs* 83:347-52
204. George LA. 2022. Factor IX Padua for haemophilia B gene addition: universal adaptation and repeated success. *Lancet Haematol* 9:e465-e6
205. Samelson-Jones BJ. 2022. Worldwide use of factor IX Padua for hemophilia B gene therapy. *Mol Ther* 30:2394-6
206. Naddaf M. 2022. Researchers welcome \$3.5-million haemophilia gene therapy - but questions remain. *Nature* 612:388-9
207. Nichols TC, Levy H, Merricks EP, Raymer RA, Lee ML. 2020. Preclinical evaluation of a next-generation, subcutaneously administered, coagulation factor IX variant, dalcinonacog alfa. *PLoS One* 15:e0240896
208. Kao CY, Lin CN, Yu IS, Tao MH, Wu HL, et al. 2010. FIX-Triple, a gain-of-function factor IX variant, improves haemostasis in mouse models without increased risk of thrombosis. *Thromb Haemost* 104:355-65
209. Chang J, Jin J, Lollar P, Bode W, Brandstetter H, et al. 1998. Changing residue 338 in human factor IX from arginine to alanine causes an increase in catalytic activity. *J Biol Chem* 273:12089-94
210. Zakas PM, Coyle CW, Brehm A, Bayer M, Solecka-Witulska B, et al. 2021. Molecular coevolution of coagulation factor VIII and von Willebrand factor. *Blood Adv* 5:812-22
211. Blazeck J, Karamitros CS, Ford K, Somody C, Qerqez A, et al. 2022. Bypassing evolutionary dead ends and switching the rate-limiting step of a human immunotherapeutic enzyme. *Nat Catal* 5:952-67
212. Hendrikse NM, Sandegren A, Andersson T, Blomqvist J, Makower Å, et al. 2021. Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome. *iScience* 24:102154
213. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, et al. 2018. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* 36:843-6
214. Alonso-Lerma B, Jabalera Y, Samperio S, Morin M, Fernandez A, et al. 2023. Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins. *Nat Microbiol* 8:77-90
215. Heckman KL, Pease LR. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* 2:924-32
216. Ribeiro DA, Passos DF, Ferraz HC, Castilho LR. 2013. Anion-exchange purification of recombinant factor IX from cell culture supernatant using different chromatography supports. *J Chromatogr B Analyt Technol Biomed Life Sci* 938:111-8

217. Parker ET, Doering CB, Lollar P. 2006. A1 subunit-mediated regulation of thrombin-activated factor VIII A2 subunit dissociation. *J Biol Chem* 281:13922-30
218. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583-9
219. Childers KC, Peters SC, Spiegel PC, Jr. 2022. Structural insights into blood coagulation factor VIII: Procoagulant complexes, membrane binding, and antibody inhibition. *J Thromb Haemost* 20:1957-70
220. Ruben EA, Summers B, Rau MJ, Fitzpatrick JAJ, Di Cera E. 2022. Cryo-EM structure of the prothrombin-prothrombinase complex. *Blood* 139:3463-73
221. Zögg T, Brandstetter H. 2009. Structural basis of the cofactor- and substrate-assisted activation of human coagulation factor IXa. *Structure* 17:1669-78
222. Ertl HCJ. 2022. Immunogenicity and toxicity of AAV gene therapy. *Front Immunol* 13:975803
223. Shupe J, Zhang A, Odenwelder DC, Dobrowsky T. 2022. Gene therapy: challenges in cell culture scale-up. *Curr Opin Biotechnol* 75:102721
224. Michael Meagher P, Mani Krishnan, Claire Davies, PhD 2021. *Uncharted Territory: Top Challenges Facing Gene Therapy Development*.
<https://www.genengnews.com/topics/uncharted-territory-top-challenges-facing-gene-therapy-development-2/>
225. Nair N, De Wolf D, Nguyen PA, Pham QH, Samara-Kuko E, et al. 2021. Gene therapy for hemophilia B using CB 2679d-GT: a novel factor IX variant with higher potency than factor IX Padua. *Blood* 137:2902-6
226. Ran G, Chen X, Xie Y, Zheng Q, Xie J, et al. 2020. Site-Directed Mutagenesis Improves the Transduction Efficiency of Capsid Library-Derived Recombinant AAV Vectors. *Mol Ther Methods Clin Dev* 17:545-55
227. Kumar SRP, Xie J, Hu S, Ko J, Huang Q, et al. 2021. Coagulation factor IX gene transfer to non-human primates using engineered AAV3 capsid and hepatic optimized expression cassette. *Mol Ther Methods Clin Dev* 23:98-107
228. Naddaf M. 2022. *\$3.5-Million Hemophilia Gene Therapy Is World's Most Expensive Drug*.
<https://www.scientificamerican.com/article/3-5-million-hemophilia-gene-therapy-is-worlds-most-expensive-drug/>
229. Patel SR, Lundgren TS, Spencer HT, Doering CB. 2020. The Immune Response to the fVIII Gene Therapy in Preclinical Models. *Front Immunol* 11:494
230. Samelson-Jones BJ, Arruda VR. 2020. Translational Potential of Immune Tolerance Induction by AAV Liver-Directed Factor VIII Gene Therapy for Hemophilia A. *Front Immunol* 11:618
231. Chang YJ, Wu HL, Hamaguchi N, Hsu YC, Lin SW. 2002. Identification of functionally important residues of the epidermal growth factor-2 domain of factor IX by alanine-scanning mutagenesis. Residues Asn(89)-Gly(93) are critical for binding factor VIIIa. *J Biol Chem* 277:25393-9
232. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434-43
233. McVey JH, Rallapalli PM, Kemball-Cook G, Hampshire DJ, Giansily-Blaizot M, et al. 2020. The European Association for Haemophilia and Allied Disorders (EAHAD) Coagulation Factor Variant Databases: Important resources for haemostasis clinicians and researchers. *Haemophilia* 26:306-13
234. Samelson-Jones BJ, Finn JD, Raffini LJ, Merricks EP, Camire RM, et al. 2021. Evolutionary insights into coagulation factor IX Padua and other high-specific-activity variants. *Blood Adv* 5:1324-32
235. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105-32

236. Neuenschwander PF, Deadmond KJ, Zepeda K, Rutland J. 2012. Correlation of factor IXa subsite modulations with effects on substrate discrimination. *J Thromb Haemost* 10:382-9
237. Hopfner KP, Brandstetter H, Karcher A, Kopetzki E, Huber R, et al. 1997. Converting blood coagulation factor IXa into factor Xa: dramatic increase in amidolytic activity identifies important active site determinants. *Embo j* 16:6626-35
238. Hopfner KP, Lang A, Karcher A, Sichler K, Kopetzki E, et al. 1999. Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding. *Structure* 7:989-96
239. Yang L, Manithody C, Olson ST, Rezaie AR. 2003. Contribution of basic residues of the autolysis loop to the substrate and inhibitor specificity of factor IXa. *J Biol Chem* 278:25032-8
240. Martin JF, Wagner GP. 2019. The origin of platelets enabled the evolution of eutherian placentation. *Biol Lett* 15:20190374
241. Ohta T. 2013. Neutral Theory. In *Brenner's Encyclopedia of Genetics (Second Edition)*, ed. S Maloy, K Hughes:67-8. San Diego: Academic Press. Number of 67-8 pp.
242. Mahlangu J, Levy H, Lee M, Del Greco F. 2021. Efficacy and safety of subcutaneous prophylaxis with dalcinonacog alfa in adults with haemophilia B. *Haemophilia* 27:574-80
243. Keeler GD, Markusic DM, Hoffman BE. 2019. Liver induced transgene tolerance with AAV vectors. *Cellular immunology* 342:103728
244. Ling C, Yin Z, Li J, Zhang D, Aslanidi G, Srivastava A. 2016. Strategies to generate high-titer, high-potency recombinant AAV3 serotype vectors. *Mol Ther Methods Clin Dev* 3:16029