Analysis of coagulation factor IX in bioreactor cell culture medium predicts yield and quality of the purified product

Lucia F. Zacchi[1¥], Dinora Roche Recinos[1,2¥], Cassandra L. Pegg[3], Toan K. Phung[3], Mark Napoli[2], Campbell Aitken[2], Vanessa Sandford[2], Stephen M. Mahler[1], Yih Yean Lee[2*], Benjamin L. Schulz[1,3*], Christopher B. Howard[1*]

[1]ARC Training Centre for Biopharmaceutical Innovation, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia.

[2] CSL Limited; Parkville Victoria, 3052, Australia

[3] School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, 4072, Australia

[¥] These authors contributed equally.

[*]Corresponding authors: YihYean.Lee@csl.com.au, b.schulz@uq.edu.au, c.howard2@uq.edu.au

Abstract

Coagulation factor IX (FIX) is a highly complex post-translationally modified human serum glycoprotein and a high-value biopharmaceutical. The quality of recombinant FIX (rFIX), especially complete γ-carboxylation, is critical for rFIX clinical efficacy. Changes in bioreactor operating conditions can impact rFIX production and occupancy and structure of rFIX post-translational modifications (PTMs). We hypothesized that monitoring the bioreactor cell culture supernatant with Data Independent Acquisition Mass Spectrometry (DIA-MS) proteomics would allow us to predict product yield and quality after purification. With the goal of optimizing rFIX production, we developed a suite of MS proteomics analytical methods and used these to investigate changes in rFIX yield, γ-carboxylation, other PTMs, and host cell proteins during bioreactor culture and after purification. Our methods provided a detailed overview of the dynamics of site-specific PTM occupancy and abundance on rFIX during production, which accurately predicted the efficiency of purification and the quality of the purified product from different culture conditions. In addition, we identified new PTMs in rFIX, some of which were near the GLA domain and could impact rFIX GLA-dependent purification efficiency and protein function. The workflows presented here are applicable to other biologics and expression systems, and should aid in the optimization and quality control of upstream and downstream bioprocesses.

Introduction

Maintaining homeostasis of the cardiovascular system requires a fine-tuned blood coagulation machinery that ensures appropriate, rapid, and localized formation of blood clots and their subsequent dissolution [1]. A key step in the coagulation pathway is the formation of the "tenase" complex, which accelerates the production of the fibrin clot. The tenase complex is formed by coagulation factors VIIIa, IXa, and X (where "a" stands for activated), which associate on the phospholipidic membrane of platelets or monocytes through their calcium-binding GLA domains [2-5] (**Fig. 1**). Mutations in the X-linked genes encoding factors VIII and IX lead to bleeding disorders called Hemophilia A and B, respectively [6]. Hemophilia B impacts ~ 1 in 25,000 male births, and is characterized by spontaneous bleeding and an inability to clot [6-8]. Factor IX (FIX) is a key player in the coagulation cascade, and FIX deficiency severely impacts the quality of life of affected individuals.

The standard therapy for Hemophilia B is prophylaxis by intravenous administration of plasma-derived or recombinant FIX (rFIX) [8]. Several rFIX products are available on the market, including BeneFIX (Pfizer, 1997), Rixubis (Baxter, 2013), Alprolix (Biogen Idec, 2014), Ixinity (Emergent Biosolutions, 2015), Idelvion (CSL, 2016), and Refixia/Rebinyn (Novo-Nordisk, 2017) [8]. These rFIX are produced in mammalian expression systems to ensure native post-translational modifications (PTMs), which are required for rFIX activity, stability, and serum half-life [1, 9]. rFIX industrial production is typically 100-1000 fold less for the same cell culture volume than monoclonal antibodies, indicating the presence of substantial biosynthetic bottlenecks [7, 10]. Two well-known biosynthetic bottlenecks in rFIX are proteolysis of the propeptide and the γ-carboxylation of the GLA domain ([11, 12], and reviewed in [10]). The variety and complexity of rFIX's PTMs, together with the low yield, make rFIX a challenging biologic to produce.

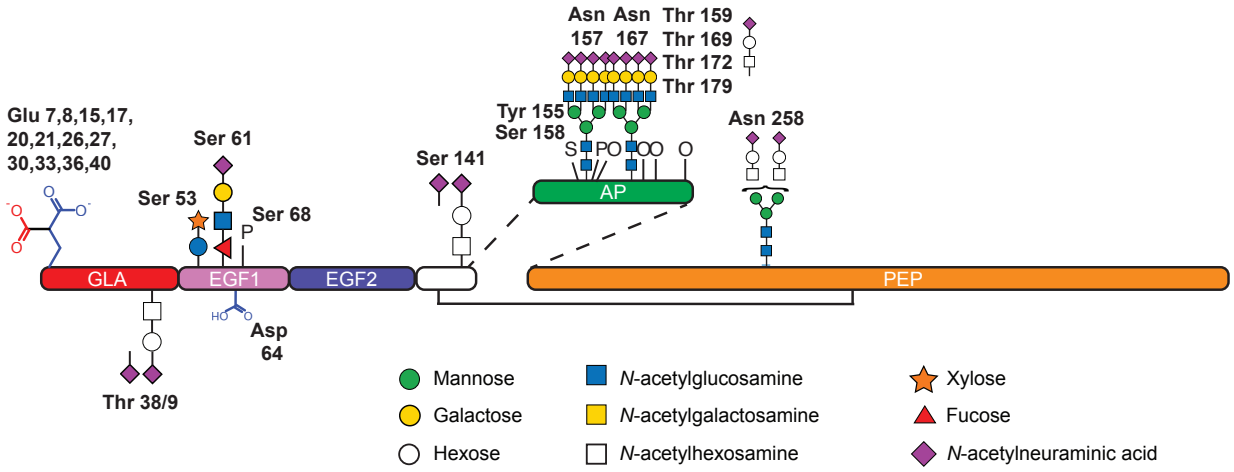


Figure 1: Coagulation Factor IX (FIX) structural domains and post-translational modifications (PTMs). Schematic of mature FIX with previously described PTMs in plasma derived FIX and/or recombinant FIX (figure modified from [13]). The GLA domain (red) contains 12 potential sites of γ-carboxylation on $E^{7\text{-}40}$ and one *O*-glycan at $T^{38/39}$. The EGF-like 1 domain (pink) contains two *O*-glycans on S^{53} and S^{61}, β-hydroxylation of D^{64}, and phosphorylation of S^{68}. The EGF-like 2 domain (violet) has no known PTMs. The short domain (white) linking the EGF-like 2 domain with the activation peptide (AP, green) contains one *O*-glycan at S^{141}. The AP contains two *N*-glycans at N^{157} and N^{167}, four *O*-glycans at T^{159}, T^{169}, T^{172}, and T^{179}, sulfation of Y^{155}, and phosphorylation of S^{158} and T^{159}. The serine protease domain (orange) contains one *N*-linked glycan at N^{258}.

FIX is a highly post-translationally modified glycoprotein [7, 14] (**Fig. 1**). Immature FIX is composed of six structural domains: an N-terminal propeptide, the GLA domain, two consecutive EGF-like domains followed by a short linker, the activation peptide (AP), and the C-terminal serine protease domain (**Fig. 1**). Most described PTMs are on or near the GLA, the EGF-like 1, and the AP domains. The propeptide and the AP are removed through proteolysis events, and both proteolysis are required for function [1, 9, 15]. FIX's propeptide is cleaved by the *trans*-Golgi protease PACE/Furin during transit through the secretory pathway [16, 17]. FIX's AP is removed in the blood by Factor XIa as part of the coagulation cascade (**Fig. 1**), thereby activating the serine protease activity of FIX [18]. A critical PTM on FIX is γ-carboxylation of the GLA domain (**Fig. 1**). The GLA domain is a 46 amino acid long region which contains 12 γ-carboxylated Glu residues in serum derived FIX (pdFIX) [19-21]. γ-carboxylation is the enzymatic addition of a CO_2 moiety to the γ-carbon of Glu residues by the γ-glutamyl carboxylase, a vitamin K dependent enzyme localized in the Endoplasmic Reticulum (ER) [4, 20, 22, 23]. γ-carboxylation of the GLA domain is required for calcium binding and for the formation of the tenase complex and is thus absolutely

essential for FIX function [2, 3, 20]. For this reason, ensuring optimal γ-carboxylation of the GLA domain is a priority for the production of rFIX. The EGF-like 1 domain is *O*-glucosylated at S^{53} [24, 25], *O*-fucosylated at S^{61} [25, 26], β-hydroxylated at D^{64} [27, 28], and phosphorylated at S^{68} [29] (**Fig. 1**). The linker region between the EGF-like 2 and the AP domains is *O*-glycosylated with *O*-GalNAc at S^{141} [13, 30] (**Fig. 1**). The AP is a highly glycosylated 35 amino acids region with 2 *N*-linked glycans at N^{157} [13, 31-34] and N^{167} [13, 31-34], and *O*-GalNAc glycans at T^{159} [25, 35], T^{169} [25, 35], and T^{172} [21] (**Fig. 1**). In addition, the AP is sulfated at Y^{155} [34] and phosphorylated at S^{158} [25, 34] (**Fig. 1**). Finally the serine protease domain is *N*-glycosylated at N^{258} [13] (**Fig. 1**). FIX also contains 11 predicted or verified disulfide bonds including the interchain bond between C^{178} and C^{335} [5] (**Fig. 1**). In addition, rFIX is *O*-glycosylated with *O*-GalNAc at T^{38} or T^{39} and T^{179}, and phosphorylated at T^{159} [25, 36]. The PTMs in FIX are diverse and heterogeneous, and the precise functions of most PTMs are not completely clear.

Measuring PTMs on purified biopharmaceutical proteins is critical, given the key role of PTMs in determining protein stability, half-life, function, and immunogenicity [7, 37-39]. PTMs are routinely measured after product purification as part of the quality control process during manufacture. However, measuring PTMs during bioreactor operation in the bioreactor supernatant has the advantage of providing information on the dynamic biosynthetic capability of the biological system under specific bioprocess conditions. The quality of the product changes during bioreactor operation as the cellular metabolism changes due to cell aging, nutrient depletion, and toxic product accumulation [39-41]. Product purification can also bias the stoichiometry and structure of PTMs in the final product, as is the case for rFIX purification which enriches for γ-carboxylated protein variants [42, 43]. Measuring the product and its PTMs during bioreactor operation could therefore allow for more accurate insights and control of the quantity and quality of the product during biosynthesis.

Mass spectrometry proteomics (MS) is a versatile tool with outstanding applicability to biopharmaceutical PTMs characterization [44]. Data Independent Acquisition (DIA) MS is a powerful analytical tool that allows simultaneous measurement of the relative abundance of proteins and their PTMs in complex protein mixtures [45, 46]. In DIA, all peptides eluting across a liquid chromatography (LC) gradient are fragmented according to pre-determined size

(mass/charge) windows. By summing the abundance of pre-selected fragment ions of interest present in specific windows at specific retention times, it is possible to calculate the stoichiometry of modification and the relative abundance of differently post-translationally modified peptide variants and proteins in each sample [45-51]. DIA is superior to data dependent acquisition (DDA) workflows in that DDA fragmentation is intrinsically biased towards the more abundant peptides, while DIA allows measurement of all detectable analytes, allowing for a full exploration of the precursor landscape in a sample [51, 52].

We hypothesized that using DIA-MS to estimate yield and quality of rFIX in the bioreactor supernatant would allow prediction of yield and quality of the purified product. We developed a suite of LC-MS/MS DIA workflows to measure the relative quantity of rFIX and its PTMs produced in fed batch bioreactor cultures during bioreactor operation and after purification (**Supplementary Fig. S1**). We cultured CHO cells co-expressing wild-type rFIX and PACE/Furin for 13 days in two different commercial feed media (EfficientFeed A or EfficientFeed B nutritional supplements). We analyzed daily samples of clarified supernatants and POROS 50 HQ purified rFIX to characterize the PTMs on rFIX and to measure the relative abundance of rFIX, its PTMs, and host cell proteins (HCPs) during the bioreactor operation and after purification. Our results show that analyzing bioreactor supernatant provided robust data on rFIX yield and quality that correlated with the yield and quality of the purified product. Thus, MS proteomics workflows that monitor the bioprocess protein dynamics have the potential to accelerate upstream product optimization and to aid in quality control process during the production phase.

Methods

Mammalian cell line, seed train, and fed batch conditions

We used a Chinese Hamster Ovary (CHO) K1SV cell line stably expressing a modified version of human Coagulation factor IX (rFIX, accession number P00740 in UniProtKB (www.uniprot.org), with $Q^{2}G$ and $P^{44}V$ amino acid substitutions) integrated using the glutamine synthetase expression system [53], and the protease PACE/Furin (Accession number P09958, UniProtKB) (provided by CSL, Marburg, Germany).

Culture flasks were inoculated at a cell density of 0.3 x 10^6 cells/ mL, and passaged every three days when viable cell density reached 2.5 x 10^6 cells/mL in chemically defined protein-free medium CDCHO (Invitrogen) supplemented with 25 µM methionine sulfoximine and 50 µg/L reduced menadione sodium bisulfite (rMSB, vitamin K, required to support rFIX production) in a 1 L vented cap Erlenmeyer shake flask with a working volume of 300 mL (Corning, Corning, USA). Cells were incubated in an orbital shaking incubator (Kuhner Shaker) at 120 rpm, 37 °C, 5.0% CO_2, and 70% humidity. Two 5 L bioreactors (Sartorius) were seeded at 0.3 x 10^6 cells/mL in 3 L (total working volume) of CD-CHO medium (Invitrogen) supplemented with 50 µg/L rMSB. Bioreactors were fed CHO CD EfficientFeed A (A1023401, Invitrogen, called bioreactor H1) or EfficientFeed B (A1024001, Invitrogen, bioreactor H2) nutrient supplements as a daily bolus starting on day 3 until 10, up to the equivalent of 40% (1.2 L) of the total post-inoculation working volume of the bioreactor, following manufacturer's instructions. Two downflow segmented tri-blade impellers provided mixing and gas bubble dispersing at 200 rpm. The bioreactors were maintained at a pH of 7.1 ± 0.3, 37 °C, and 40% dissolved oxygen (controlled by constant headspace (150 mL/min) and air sparging (9 mL/min) with additional oxygen sparging on demand). Glucose was fed on demand to maintain concentration between 3-6 g/L. Cultures were terminated when cell viability in at least one bioreactor reached below 80%.

Sample collection and metabolite and physical measurements

The bioreactor was sampled daily for off-line measurements. Total cell density, cell viability, and cell size were measured using a cell counter (Vi-cell Beckman Coulter, Fullerton, CA). Glucose, lactate, ammonium, glutamine, and other metabolites were measured with a NOVA Flex

BioProfile analyzer (NOVA biomedical). pH was measured using Seven Excellence Multiparameter (METTLER TOLEDO). Oxygen (pO_2) and carbon dioxide (pCO_2) partial pressure were measured using a Siemens Blood Gas Analyzer RapidLab 248. Technical replicate samples of 15 mL were collected daily from pre-inoculation day to day 13, centrifuged at 2000 rcf for 10 min, filtered through a 0.2 µm polyethersulfone filter (Pall), and stored at -80 °C in 500 µL aliquots in matrix tubes (Thermo Fisher) for further analysis. Technical replicate bulk sampling of 100 mL aliquots was performed from day 5 to day 13. On termination day (day 13), 2 L of the total working volume was collected and centrifuged at 2000 rcf for 60 min, filtered through a 0.2 µm polyethersulfone filter (Pall), and stored at -80 °C.

Anion exchange chromatography.

rFIX was purified by ÄKTA Pure (GE Healthcare) chromatography controlled with UNICORN 7.0 software. 1.6 L of frozen clarified culture media from both bioreactors was thawed in a 25 °C water bath. The initial conductivity of the media was measured and EDTA levels were adjusted to 35 mM for a final sample conductivity of 15-16 mS/cm, by either adding 0.75 M EDTA and 1.32 M Tris base solution if the conductivity was below 13.5 mS/cm, or by diluting the sample with 20 mM Tris HCl buffer (20 mM Tris HCl buffer pH 7.02, conductivity 1801 µS/cm at 21 °C) if the conductivity was above 13.5 mS/cm. Anion exchange chromatography was performed as previously described [43]. Briefly, two pre-packed columns with POROS 50 HQ resin (0.8 cm ID x 5.0 cm H, V = 2.5 mL, strong anion exchange resin, Repligen GmBH) were first equilibrated with 5 column volumes (CV) of EDTA equilibration buffer (50 mM MES buffer pH 5.02, 100 mM NaCl, 50 mM EDTA, conductivity 15.90 mS/cm at 21.8 °C). The clarified media sample from each bioreactor was applied to each column. Unbound proteins were washed off with 5 CV of NaCl wash buffer (50 mM MES buffer pH 5.04, 195 mM NaCl, 2 mM $CaCl_2 \cdot 2H_2O$, conductivity 20.30 mS/cm at 22.4 °C) and 5 CV of re-equilibration buffer (50 mM Tris base pH 8.52, 100 mM NaCl, conductivity 12.01 mS/cm at 21.8 °C). A gradient elution was performed by mixing re-equilibration buffer with increasing volumes of Elution buffer (50mM Tris base pH 8.52, 100 mM NaCl, 100 mM $CaCl_2 \cdot 2H_2O$, conductivity 29.44 mS/cm at 21.1 °C). The resin was regenerated using Column regeneration buffer (50 mM Tris base, 2 M NaCl, pH 8.5, conductivity 156.3 mS/cm at 21.2 °C). After each experiment, the column was cleaned with 3 CV of 0.5 M NaOH and neutralized using 5 CV of re-equilibration buffer. The elution peak fraction of both purified

samples was collected and concentrated using Amicon Ultra-10 centrifugal filter (10,000 MWCO, Merck) by centrifugation at 4000 rcf for 15 min. Total protein concentration was measured at Absorbance 280 nm using Lunatic (Unchained Labs), concentration was adjusted to 1 mg/mL, and samples were aliquoted in matrix tubes and stored at -80 ˚C for further analysis.

Coomassie blue and western blotting

Purified rFIX samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using a 4-12% Bis-Tris Bolt acrylamide gel (Invitrogen) in 1x MES running buffer (Invitrogen), under non-reducing conditions, following manufacturer's instructions. Protein bands were visualized either with colloidal Coomassie blue (Simply Blue; Invitrogen) or western blotting. For western blotting, proteins were transferred from gels onto a PVDF membrane (Bio-Rad) using a Trans-Blot® Turbo™ Transfer System Transfer Pack (Bio-Rad), following manufacturer's instructions. The membrane was blocked in 2% skim milk (Devondale) dissolved in phosphate buffered saline solution (PBS; Sigma) with 0.05% Tween 20 (Sigma) for 30 min at room temperature. The following antibodies were used: mouse IgG1 anti human Factor IX primary antibody 1C2 (AbCam, 1:1000 dilution) and HRP-conjugated anti mouse IgG (BioRAD, 1:1000 dilution). Membranes were developed with Novex ECL reagent (Invitrogen) and imaged using a Bio-Rad image workstation.

Mass spectrometry sample preparation

Samples from purified rFIX and plasma derived FIX (pdFIX) (F0806, Sigma) were prepared by performing in-solution digestion, as previously described [54]. Briefly, triplicate 5 µg samples of purified rFIX and pdFIX were diluted in denaturation buffer (6 M guanidine hydrochloride, 50 mM Tris HCl buffer pH 8, 10 mM Dithiothreitol (DTT)) in a protein LoBind tube (Eppendorf) and incubated at 30°C for 30 min in a MS100 Thermoshaker at 1500 rpm. Acrylamide was added to a final concentration of 25 mM and samples were incubated at 30 °C for 1 h in a Thermoshaker at 1500 rpm. Additional 5 mM DTT was added to quench acrylamide, and samples were precipitated with 4 volumes of 1:1 methanol:acetone and incubation at -20 °C for 16 h. Solvent was removed by two consecutive centrifugations at 21,000 rcf for 10 min and 1 min, respectively, at room temperature. Samples were air dried at room temperature for ~15 min, resuspended in 50 mM ammonium bicarbonate solution containing one of the following proteases: 0.2 µg trypsin (T6567,

Sigma); 0.17 µg GluC (11420399001, Sigma), 0.17 µg Chymotrypsin (11418467001, Roche), and 0.04 µg AspN (19936721, Roche), and incubated at 37 °C for 16 h in a Thermoshaker at 1500 rpm. When followed by deglycosylation, all proteases were denatured by incubation at 95 °C for 5 min, and trypsin was also inactivated by addition of 1 mM phenylmethylsulfonyl fluoride. Samples were split into two protein LoBind tubes, one set of tubes was frozen at -20 °C, while 300 units of Peptide:N-Glycosidase F (PNGase F, P0704S, New England Biolabs) was added to the other set of tubes and incubated at 37 °C for 16 h in a Thermoshaker at 1500 rpm. Samples were desalted with C18 ZipTips (ZTC18S960, Millipore).

Daily samples collected from bioreactors (day 1 to 13) were prepared with S-Trap columns following manufacturer's instructions (S-Trap C02-mini, Protifi), and as previously described [54]. Briefly, the supernatant samples were thawed on ice, centrifuged at 21000 rcf for 3 min at room temperature to remove aggregates and particulates, and 200 µL from each sample was transferred to protein LoBind tubes containing 200 µL of 2x Lysis buffer (10% sodium dodecyl sulfate (SDS), 100 mM Tris HCl buffer pH 7.55, 20 mM DTT) in triplicate. After incubation at 95 °C for 10 min, samples were cooled to room temperature before adding acrylamide to a final concentration of 25 mM. Samples were incubated at 30 °C for 1 h in a Thermoshaker at 1500 rpm, and acrylamide was quenched by adding additional 5 mM DTT. Samples were acidified by adding phosphoric acid to 1.2% v/v final concentration, diluted 1:7 with S-Trap binding buffer (90% methanol, 100 mM Tris HCl buffer pH 7.1), and loaded onto the S-Trap mini columns. The samples were washed 4 times with 400 µL of S-Trap binding buffer. Samples were then resuspended in 125 µL of 50 mM ammonium bicarbonate with 1 µg of trypsin, and columns were incubated at 37 °C for 15 h in a humidified chamber, without agitation. To recover the peptides, the columns were rehydrated with 80 µL of 50 mM ammonium bicarbonate, incubated at room temperature for 15 min, and centrifuged at 1000 rcf for 1 min at room temperature. This was followed by subsequent elutions with 80 µL of 0.1% formic acid, followed by 80 µL of 50% acetonitrile in 0.1% formic acid. Elutions were pooled and desalted with C18 ZipTips.

High pH reversed-phase peptide fractionation

Peptides were fractionated by high pH reversed-phase fractionation essentially as previously described [55]. Sep-Pak tC18 Vac 1cc (50 mg) cartridges (WAT054960, Waters) were equilibrated

by washing twice with 100% acetonitrile followed by two washes with 0.1% formic acid. 4 µL from each trypsinized bioreactor supernatant sample from each day were pooled, diluted in 0.1% formic acid to a final volume of 500 µL, and applied to the column. The peptides retained on the column were washed with milli-Q water and eluted in 9 x 500 µL fractions of increasing acetonitrile concentration (5-90% (v/v) acetonitrile in 0.1% triethylamine). All solutions applied to the cartridge were eluted by applying positive air pressure. Each fraction was collected in a protein LoBind tube, dried using a Genevac miVac centrifugal vacuum concentrator, and reconstituted in 100 µL of 0.1% formic acid.

Methanolic HCl derivatization

To detect γ-carboxyglutamic acid in peptides by positive ion mode liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS), samples were chemically derivatized essentially as previously described [56, 57], with some modifications. 100 µL of day 13 bioreactor supernatant samples (concentrated in a 0.5 mL 10 kDa Amicon column) and 1 µg of purified rFIX and pdFIX were denatured, reduced, and precipitated overnight as described above for in solution digestion. Samples were completely dried in an acid resistant SpeedVac Savant SPD300DDA (ThermoFisher), were resuspended in 25 µL of 3 M HCl in methanol (90964, Sigma), and incubated at 20 °C for 1 h. The methanol-HCl was removed by completely drying the samples in a SpeedVac Savant at 30 °C for 45 min followed by 40 °C for 20 min. Proteins were resuspended in 50 mM ammonium bicarbonate solution with 0.1 µg trypsin and incubated at 37 °C for 16 h in a Thermoshaker at 1500 rpm. Peptides were desalted with C18 ZipTips.

Mass spectrometry analysis

Desalted peptides were analyzed by LC-ESI-MS/MS using a Prominence nanoLC system (Shimadzu) and a TripleTof 5600 mass spectrometer with a Nanospray III interface (SCIEX) essentially as described [48, 58, 59]. Samples were desalted on an Agilent C18 trap (0.3 x 5 mm, 5 µm) at a flow rate of 30 µL/min for 3 min, followed by separation on a Vydac Everest C18 (300 Å, 5 µm, 150 mm x 150 µm) column at a flow rate of 1 µL/min. A gradient of 10-60% buffer B over 45 min where buffer A = 1 % ACN / 0.1% FA and buffer B = 80% ACN / 0.1% FA was used to separate peptides. Gas and voltage settings were adjusted as required. An MS TOF scan across 350 - 1800 *m/z* was performed for 0.5 s followed by data dependent acquisition (DDA) of up to 20

peptides with intensity greater than 100 counts, across 100 - 1800 *m/z* (0.05 s per spectra) using a collision energy (CE) of 40 +/- 15 V. For data independent acquisition (DIA) analyses (except for derivatized peptides, see below), MS scans across 350 - 1800 *m/z* were performed (0.05 s), followed by high sensitivity DIA mode with MS/MS scans across 50 - 1800 *m/z*, using 26 *m/z* isolation windows for 0.1 sec, across 400 - 1250 *m/z*. For DIA analysis of derivatized peptides, MS scans across 50-1800 *m/z* were performed (0.05 s), followed by high sensitivity DIA mode with MS/MS scans across 50 - 1800 *m/z*, using 6.2 *m/z* isolation windows for 0.04 s, across 400 - 917 *m/z*. CE values for DIA samples were automatically assigned by Analyst software based on *m/z* windows. The following samples were chosen for DDA analysis: 1 replicate per day for days 2, 5, 9, and 13 from each bioreactor, all the high pH fractionation samples, one replicate of each purified rFIX (from each bioreactor) and pdFIX digested with trypsin or trypsin and PNGase F. For DDA analysis of derivatized peptides, 1 replicate from each bioreactor were analyzed.

Mass spectrometric characterization of rFIX

To characterize the PTMs on rFIX, we analyzed the DDA data from purified rFIX and pdFIX digested with trypsin, GluC, AspN, or Chymotrypsin, with or without *N*-glycan removal with PNGase F, using ProteinPilot v5.0.1 (SCIEX) and Byonic v2.13.17 (Protein Metrics). The following parameters were used in ProteinPilot: Sample type, identification; Cysteine Alkylation, acrylamide; Digestion, trypsin or GluC or Chymotrypsin or AspN; Instrument, TripleTOF 5600; ID focus, biological modifications; Search effort, thorough; 1 % global FDR cut-off. The database used contained 5 isoforms of FIX (P00740-1, the UniProtKB canonical sequence with 461 amino acids (**Fig. 1**); the P00740-2 splice isoform, lacking the EGF-like 1 domain, amino acids 93-130; P00740-Δ1-46, lacking the signal sequence and propeptide, amino acids 1-46; P00740-T194A, P00740 containing the $T^{194}A$ common natural variant [60, 61]; and P00740-Q2G_P44V, rFIX $Q^{2}G$, $P^{44}V$ used in this study, and a custom contaminants database from Proteome Discoverer (v2.0.0.802, Thermo Fisher Scientific). For some searches, the database contained only the P00740-1 and P00740-Q2G_P44V isoforms. The parameters used for searches in Byonic varied depending on the enzyme, and their full description can be found in **Supplementary Table S1-Summary**. The list of glycans searched in Byonic included the Byonic database containing 50 common biantennary *N*-glycans at Nglycan (rare1) for trypsin and GluC; or a modified *N*-glycan database combining the *N*-glycan 50 and 57 common biantennary glycoforms from Byonic and

including 14 *N*-glycoforms with LacNAc extensions (LacNAc extensions are typical in CHO cells [62]) (**Supplementary Table S2**) at Nglycan (i.e. at NXS/T) (rare1) for AspN and Chymotrypsin; and a list of *O*-glycan structures previously described for pdFIX (**Supplementary Table S2**) at Oglycan (i.e. at S/T) (rare1 or rare2). The data for these Byonic searches is compiled in **Supplementary Table S1** and the select Byonic peptide-spectrum matches (PSMs) for each PTM of interest can be found in **Supplementary information**.

Label free relative proteomic quantification

To measure rFIX and the host cell proteins (HCPs), we generated a combined ion library by searching all DDA files accompanying the SWATH/DIA experiments (high pH reversed-phase fractions, unfractionated samples from each bioreactor (obtained at days 2, 5, 9, and 13), purified FIX from each bioreactor, and pdFIX) in ProteinPilot v5.0.1 (SCIEX) using the following parameters: Sample type, identification; Cys Alkylation, acrylamide; Digestion, trypsin; Instrument, TripleTOF 5600; ID focus, biological modifications; Search effort, thorough; 1 % global FDR cut-off (**Supplementary Table S3**). The database used included the entire *Cricetulus griseus* proteome UP000001075 (downloaded from Uniprot on 6th February 2019, total 24194 proteins (23887 CHO proteins, the 5 FIX isoforms described above (**Supplementary information**), and a custom contaminants database from Proteome Discoverer). The V^{108}ATVSLPR^{115} trypsin autolysis peptide (421.7584^{2+} *m/z*) was manually added to this library with a confidence of 0.99 (**Supplementary Table S4**). Peptide abundance was measured with PeakView v2.2 (SCIEX) using the following parameters: proteins imported only below 1% global FDR cut-off, shared peptides imported; 6 transitions/peptide; 99% peptide confidence threshold; 1% FDR; 2 min XIC extraction window; and 75 ppm XIC width (**Supplementary Table S5**). The protein intensity output from PeakView was recalculated by eliminating the value of peptides measured with a FDR > 0.01 (**Supplementary Table S5**), using a modified version of our previously described script [63] (**Supplementary information**). To calculate the relative changes in protein abundance through time in the bioreactor, the protein abundance data for each protein in each day was normalized to the abundance of trypsin in that sample, as previously described [47] (**Supplementary Table S5**). The mean (N=3) for each trypsin-normalized protein abundance at each time point was log10 transformed and plotted as a heatmap using GraphPad Prism v7.0a and 8.2.0 for Mac OS X (GraphPad Software, San Diego, California USA). Statistical analysis of

rFIX/trypsin values during bioreactor operation was performed using multiple t-tests, with the two-stage linear set-up procedure of Benjamini, Krieger, and Yekutieli to assign P values, and Q = 1%, in GraphPad Prism. Statistical analyses of the HCPs and rFIX abundance in purified rFIX samples (N = 3) and at days 3, 6, 9, and 13 were performed using MSstats in R [48, 58, 63] (**Supplementary Table S6**). Volcano plots were made in GraphPad Prism.

Quantification of post-translational modifications

To perform relative quantification of PTMs in rFIX from bioreactor supernatant and purified samples, a PeakView ion library specific for FIX PTMs was generated. To do this, we first searched the DDA files corresponding to the SWATH/DIA-MS experiment from tryptic digests of purified rFIX and pdFIX, with and without PNGase F digestion, in Preview (v2.13.17, Protein Metrics), to obtain an overview of the PTMs and mass measurement accuracy. Next, to increase the search efficiency, a merged focused database was generated in Byonic by first performing two separate searches of one sample of purified rFIX from bioreactor H1 and one from H2 digested with trypsin and PNGase F, using the following parameters: fully specific cleavage at the C-terminus of Arg and Lys; 0 missed cleavages allowed; fragmentation type CID low energy; 20.0 ppm precursor tolerance; 40.0 ppm fragment tolerance; 1, 2, and 3 charges applied to charge-unassigned spectra; decoys and contaminants added; 1% protein FDR cut-off; allowing no modifications. The database used was the *Cricetulus griseus* proteome described above, and decoys and contaminants were included. The two focused databased were merged and the amino acid sequences of PNGase F and the rFIX variants P00740-Δ1-46 and P00740-Q2G_P44V were included to generate the merged focused database. Using the merged focused database, the error settings suggested by Preview, and the parameters described in **Supplementary Table S7**, we searched the DDA files in Byonic. To construct the ion library we selected the best PSMs identified in Byonic for each manually validated post-translationally modified tryptic peptide (**Supplementary Table S7 and Supplementary information**). The criteria for selection of the best PSM included: PSM of a fully cleaved tryptic product, PSM within the highest scored PSMs (minimum of 100), highest coverage of *b* and *y* ions, presence of at least one Y ion and corresponding oxonium ions in the case of glycopeptides, and consistent retention time (RT) across the different DDA files (see **Supplementary Table S7-Notes**, for more details). Modified peptides that shared a similar fragmentation pattern and fell in the same SWATH window at a similar RT

were sorted by score, and the highest scoring manually validated PSM was included in the library (**Supplementary Table S7-Selected PSMs**). The MS/MS data, (including fragment ions, intensity, and RT) from the best PSMs was formatted into a PeakView ion library using an in-house written Python script (**Supplementary Table S8**). Peptide abundance was quantified using the following PeakView settings: imported all proteins and shared peptides; measurement of all transitions/peptide and all peptides/protein; 99% peptide confidence threshold; 1% global FDR; 2 min XIC extraction window; 75 ppm XIC width (**Supplementary Table S9**), and the peptide data was filtered out of values with FDR > 0.01 as above (**Supplementary Table S9**). Statistical analysis was performed using one-tailed t-test in Excel (Microsoft).

The ion library for quantification of methylated γ-carboxylated peptides from rFIX was generated essentially as above, with the following changes. The parameters for Byonic searches are described in **Supplementary Table S10-Summary**. The criteria for selection of PSMs was the same as above (with select exceptions), and can be found in **Supplementary Table S10-Notes**. The MS/MS spectra of the PSMs selected to make the ion library can be found in **Supplementary Information**. The ion library generated (**Supplementary Table S11**) was manually modified to include the unfragmented precursors, as these were predominant ions in the DIA MS/MS spectra for the selected PSMs but not in the DDA MS/MS spectra annotated by Byonic, and also to include Y ions with 1 methyl group to account for the potential methylation of the glycan in the TTEFWK methylated glycopeptide precursors identified containing 1 or 2 methyl groups (Y0 = 825.4141 *m/z*, Y1 = 1028.4935 *m/z*, and Y2 = 1190.5463 *m/z*). Peptide abundance was quantified in PeakView using the following settings: measurement of all peptides/protein and all transitions/peptide; 99% peptide confidence threshold; 1% FDR; 6 min XIC extraction window for all peptides except for the CSFEEAR variants, which used 2 min; and 75 ppm XIC width (**Supplementary Tables S11**). The quantified peptide data was filtered out of values with FDR > 0.01 (**Supplementary Table S11**). To calculate the level of carboxylation, the abundance of each variant peptide was first normalized to the abundance of all variations of the same peptide. The intensity of all peptide variants with 0, 1, 2, or 3 carboxyl groups were summed, and the % of each carboxyform was calculated. The data was plotted using GraphPad Prism. Statistical analysis was performed using one-tailed t-test in Excel (**Supplementary Table S11**).

The mass spectrometry raw data and ProteinPilot results are available in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) [64] via the PRIDE partner repository [65] with the dataset identifier PXD018229. The content of each uploaded file is described in **Supplementary Table S17**. A summary diagram of the proteomic workflows used in this work can be found in **Supplementary Fig. S1**.

Bioinformatic analysis

Gene Ontology analysis was performed using ClueGO v2.5.4 [66], a plugin of Cytoscape v.3.7.2 [67], using as universe background the CHO proteome UP000001075 described above.

Results

Performance of CHO cells expressing rFIX in fed-batch bioreactors with different feeds

To test the hypothesis that monitoring bioreactor culture media with DIA-MS proteomics can inform on the yield and quality of the final product, we cultured CHO-K1SV cells that stably co-expressed rFIX and the protease PACE/Furin in two fed-batch bioreactor conditions. The incubation conditions were identical except for the commercial feed used in each bioreactor: EfficientFeed A (bioreactor H1) or B (bioreactor H2). To obtain an overview of the physiological performance of the cultures, we measured cell viability and the concentration of key metabolites (glutamine, ammonium, and lactate) during bioreactor operation (**Fig. 2A**, **Supplementary Fig. S2**, and **Supplementary Table S12**). The viable cell density (VCD) of both bioreactors reached a maximum at days 7 - 8 and then gradually declined (**Fig. 2A**). Similarly, cell viability in both bioreactors remained at ~ 98% until day 7, and then dropped to ~ 93% in H1 and to ~ 77% in H2, leading to the termination of the operations at day 13 (**Fig. 2A**). We also observed lower levels of glutamine and ammonium and a sharp increase in lactate production towards the end of the operation in bioreactor H2 compared to bioreactor H1 (**Supplementary Fig. S2**). Therefore, EfficientFeed A provided better support of cellular viability and metabolism than EfficientFeed B.

Product yield is critical when considering optimization of bioreactor operational parameters. We used DIA-MS to measure the abundance of rFIX in the bioreactor culture supernatant during bioreactor operation. rFIX abundance steadily increased in both bioreactors from days 1 - 8 (**Fig. 2B**). While rFIX abundance in H1 plateaued at day 8, rFIX abundance continued to increase in bioreactor H2 (**Fig. 2B**). Similar results were observed by western blot (**Supplementary Fig. S3**). MSstats comparison of relative rFIX abundance in bioreactors H1 *vs* H2 showed that already at day 6 there was significantly more rFIX in the supernatant of bioreactor H2 compared to H1 and that this difference was largest at day 13 ($P < 10^{-5}$, **Fig. 2C**, **Supplementary Table S6**). Therefore, although feeding with EfficientFeed B (H2) led to lower cell viability, it increased production of rFIX in the bioreactor supernatant.

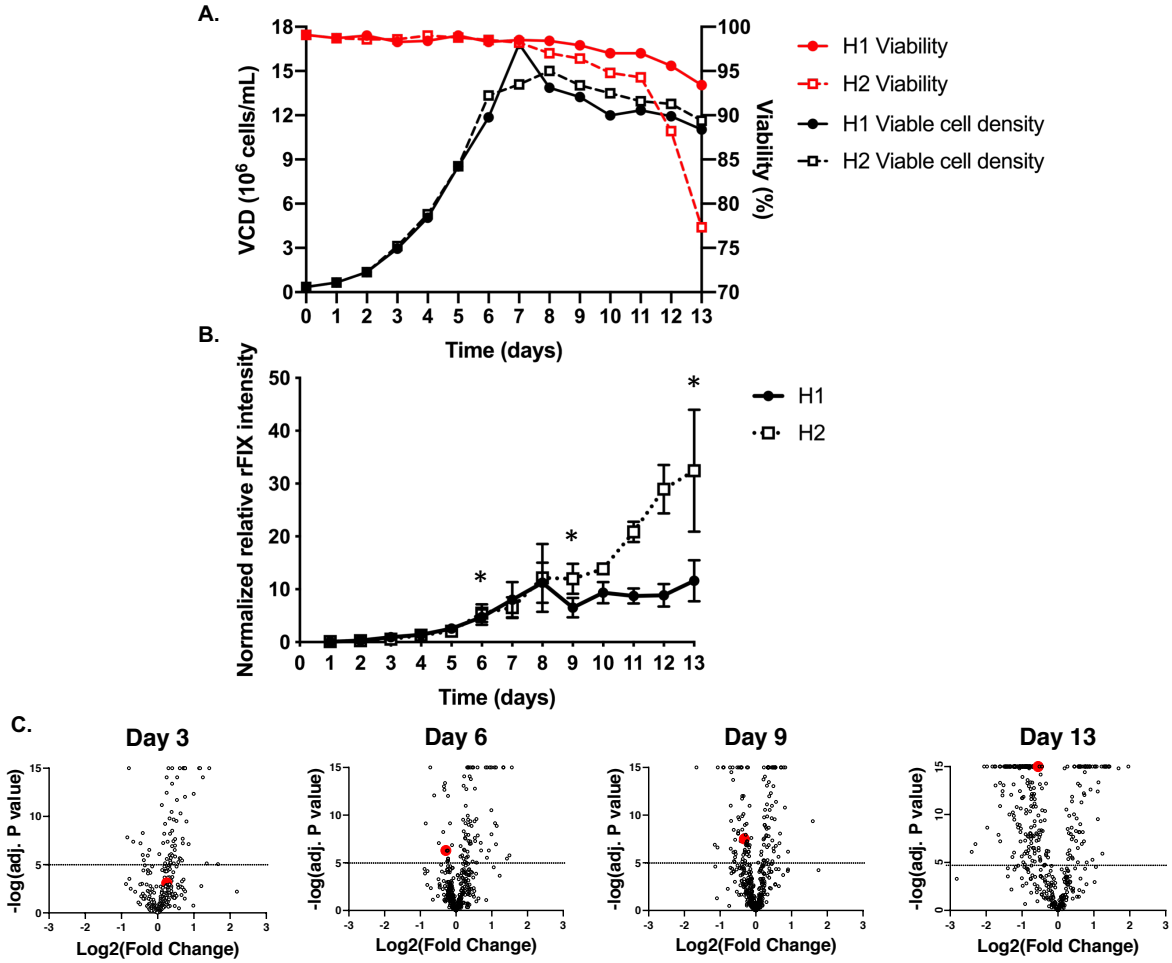


Figure 2. Viability and productivity of CHO cells expressing rFIX in fed batch conditions. CHO cells coexpressing rFIX and PACE/Furin were grown in fed batch bioreactor mode with either EfficientyFeed A (H1) or EfficientFeed B (H2) as feeds. (**A**) Viability (red line) and viable cell density (VCD) (black line) in H1 (solid line, closed circle) and H2 (dotted line, open square). (**B**) Relative rFIX abundance (normalized to trypsin) in the bioreactor supernatant during operation (Mean +/- SEM) (N = 3) (* $P < 10^{-5}$ from MSstats comparisons). **(C)** Volcano plots depicting log2 of the fold change in protein abundance *vs* -log10 of adjusted P value for comparisons of culture media of bioreactor H1 *vs* H2 at days 3, 6, 9, and 13. The dotted horizontal line indicates the value above which the comparisons were significant ($P < 10^{-5}$, N = 3). Each open circle is a unique protein. rFIX is indicated by a red dot.

Predicting γ-carboxylation levels in purified rFIX by analyzing the culture supernatant

One of the key quality attributes of rFIX is γ-carboxylation of the GLA domain [7, 21, 68]. FIX's GLA domain contains 12 Glu residues which are essentially completely modified to γ-carboxyglutamic acid in plasma derived FIX (pdFIX) (**Fig. 3A**) [19-21]. Although similarly functional, CHO produced rFIX is incompletely γ-carboxylated in the last two Glu of the GLA

domain [21, 69, 70]. Due to the critical importance of γ-carboxylation for the physiological function of FIX [68], bioreactor operation and purification procedures seek to optimize rFIX γ-carboxylation.

Fully γ-carboxylated GLA peptides are difficult to detect and identify in positive ion mode LC-ESI-MS/MS. Some of the reasons for this include the negative charge of the carboxyl groups, that γ-carboxylation appears to hinder protease cleavage, and neutral loss of CO_2 upon CID (collision induced dissociation) fragmentation [56, 57, 71, 72]. However, uncarboxylated or partially γ-carboxylated GLA peptides can be detected in positive ion mode LC-ESI-MS/MS and used as a proxy for γ-carboxylation levels. Alternatively, γ-carboxylated peptides can be directly measured by performing methanolic derivatization of the proteins prior to proteolysis [56, 57]. Methylation neutralizes the negative charge of the carboxyl groups, stabilizes the γ-carboxylation preventing neutral loss during fragmentation, and facilitates protease cleavage [56, 57]. Therefore, methylation allows measurement of both incompletely and completely γ-carboxylated peptides. In addition, methylation provides site-specific information on γ-carboxylation, which is valuable for FIX because complete γ-carboxylation of GLA is not required for rFIX function [69]. Therefore, both methodologies provide complementary γ-carboxylation information.

To test how rFIX γ-carboxylation changed throughout bioreactor operation we used DIA-MS to measure rFIX γ-carboxylation in supernatant samples from days 1 to 13 in the underivatized form, and also measured γ-carboxylation of derivatized rFIX at day 13 (**Fig. 3, Supplementary Fig. S4**, and **Supplementary Tables S9** and **S11**). As expected, fully γ-carboxylated GLA peptides were not detectable in the underivatized samples, except for γ-carboxylated TTE^{40}FWK (**Fig. 3A**, **Table 1**, **Supplementary Fig. S6A**, and **Supplementary Table S1**). Measurement of underivatized rFIX allowed quantification of uncarboxylated (+44x0), mono (+44x1), and di γ-carboxylated (+44x2) LE^{7}E^{8}FVQGNLE^{15}R, uncarboxylated CSFE^{26}E^{27}ARE^{30}VFE^{33}NTE^{36}R, and uncarboxylated and γ-carboxylated TTE^{40}FWK (**Fig. 3A,B**, **Table 1**, **Supplementary Figs. S4** and **S6A**, and **Supplementary Table S1**). These results indicated that rFIX in both bioreactors was partially γ-carboxylated. The relative abundance of TTE^{40}FWK carboxyforms in the supernatant of both bioreactors was similar (**Fig. 3** and **Supplementary Fig. S4**). On the other hand, the extent of γ-

carboxylation at E^7, E^8, and E^{15} was lower in bioreactor H2 compared to H1, as we measured more uncarboxylated and partially γ-carboxylated $LE^7E^8FVQGNLE^{15}R$ peptides in the bioreactor H2 supernatant compared to H1, especially towards the end of bioreactor operation (days 9 - 13) (**Fig. 3B**). Methylation greatly increased our ability to detect and reliably quantify γ-carboxylated rFIX, including measurement of fully γ-carboxylated variants of several GLA peptides (**Fig. 3**). In agreement with the measurement of underivatized rFIX, the levels of partially or uncarboxylated methylated $LE^7E^8FVQGNLE^{15}R$, $CSFE^{26}E^{27}AR$, and $E^{30}VFE^{33}NTE^{36}R$ peptides were significantly higher in bioreactor H2 supernatant compared to H1 at day 13 (**Fig. 3A,C-F**, and **Supplementary Table S11**). Indeed, we detected significantly lower levels of fully γ-carboxylated methylated GLA peptides in bioreactor H2 supernatant compared to H1 (**Fig. 3**). Therefore, both analytical approaches unequivocally showed that rFIX in H1 bioreactor supernatant was more efficiently γ-carboxylated than in H2.

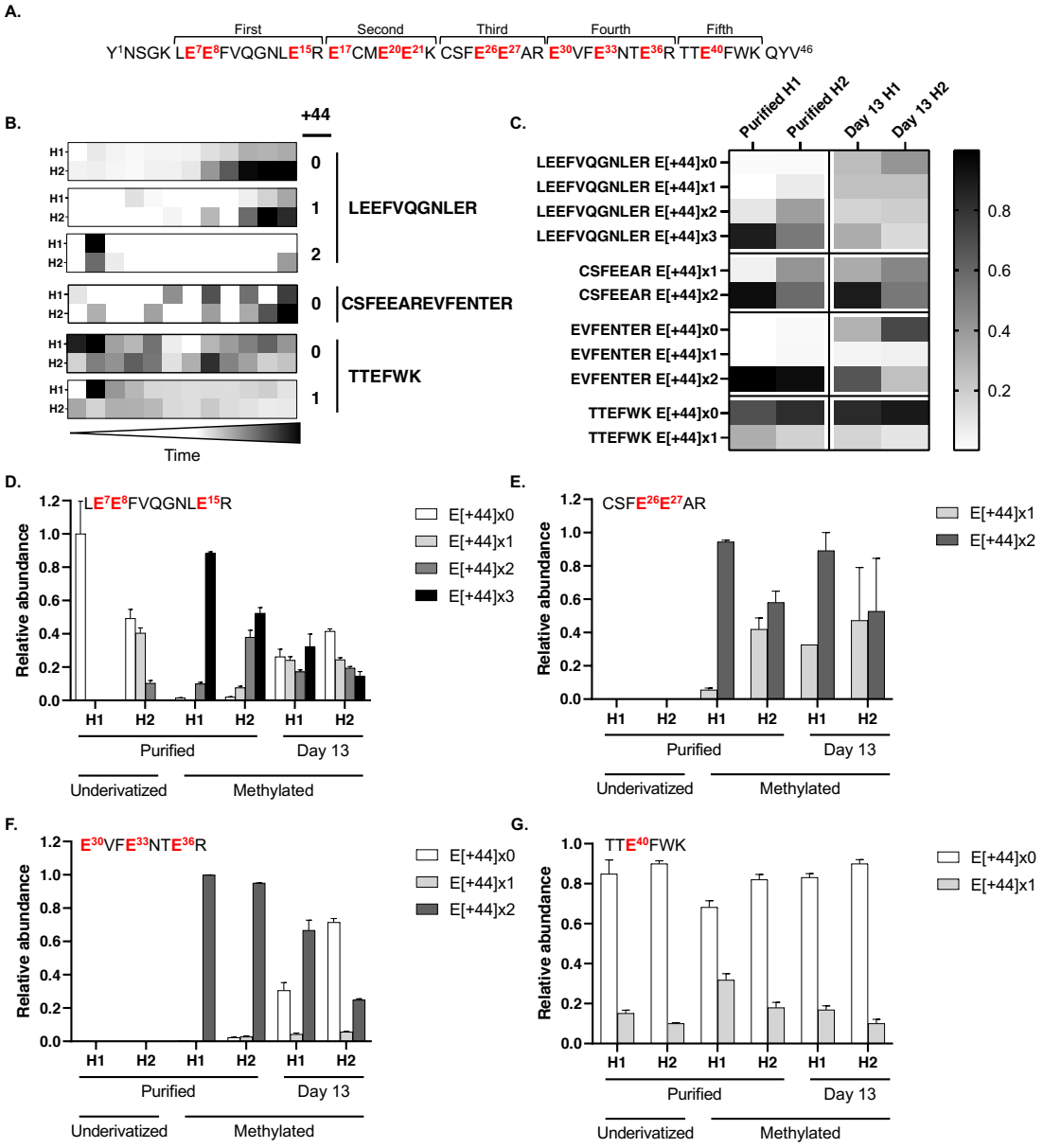


Figure 3: Relative abundance of γ-carboxylated GLA peptides from rFIX during bioreactor operation and after purification measured by DIA-MS. The abundance of rFIX γ-carboxylated tryptic peptides was measured by DIA-MS in samples from bioreactor H1 and H2 supernatant and after rFIX purification. Shown are the results with and without prior derivatization by methylation. **(A)** Amino acid sequence of rFIX GLA domain highlighting the 12 Glu residues (in red) and the five tryptic peptides potentially measurable by LC-ESI-MS/MS. **(B)** Heatmaps depicting the change in relative abundance of unmethylated GLA γ-carboxypeptides containing 0 - 2 γ-carboxylations (+44) over time (13 days) in each bioreactor. **(C)** Heatmaps displaying the mean relative abundance of each of the methylated γ-carboxylated peptide variants in the purified samples and in the bioreactor supernatants at day 13 (N = 2 - 3). **(D-G)** Bar graphs depicting the mean relative intensity of carboxyforms from purified rFIX or rFIX from bioreactor supernatant

at day 13, with or without derivatization by methylation. Multiple variants of each peptide were measured by DIA-MS, containing 0 - 3 γ-carboxylations, 0 - 6 methyl groups, and two different *O*-glycan compositions at $T^{38/39}$ (**Fig. 1**). **(D)** $LE^{7}E^{8}FVQGNLE^{15}R$, **(E)** $CSFE^{26}E^{27}AR$, **(F)** $E^{30}VFE^{33}NTE^{36}R$, and **(G)** $TTE^{40}FWK$. Displayed is the mean +/- SEM (N = 1 - 3). Statistical comparisons can be found in **Supplementary Table S11-G**.

Since rFIX in bioreactor H1 supernatant was more γ-carboxylated than in H2 (**Fig. 3**), we hypothesized that purified H1 rFIX would also be more γ-carboxylated than H2 rFIX. To purify rFIX we used a POROS 50 HQ strong anion-exchange resin, a quaternary polyethyleneimine that binds negatively charged molecules [42, 73]. POROS 50 HQ enriches for γ-carboxylated FIX due to the high negative charge of the γ-carboxyglutamic acids in the GLA domain [7, 43, 57]. We expected that purified rFIX from both bioreactors would be enriched in highly γ-carboxylated forms compared to the bioreactor supernatant. Indeed, the extent of γ-carboxylation in rFIX from bioreactor H1 and H2 was significantly higher after purification than in the culture supernatants (**Fig. 3C-G** and **Supplementary Table S11-G**). However, while purified H1 rFIX was almost completely γ-carboxylated, purified H2 rFIX was not (**Fig. 3C-G** and **Supplementary Fig. S4**). The levels of fully γ-carboxylated GLA peptides were significantly lower in purified H2 rFIX compared to H1 (**Fig. 3** and **Supplementary Table S11-G**) and the abundance of uncarboxylated or partially γ-carboxylated GLA peptides was significantly higher in purified H2 rFIX compared to H1 (**Fig. 3**, **Supplementary Fig. S4**, and **Supplementary Table S11-G**). These results indicated that purification reduced γ-carboxyform heterogeneity and enriched for γ-carboxylated forms in both H1 and H2 rFIX, but did not produce completely γ-carboxylated H2 rFIX (**Fig. 3** and **Supplementary Fig. S4**). Interestingly, while γ-carboxylated peptides $LE^{7}E^{8}FVQGNLE^{15}R$, $CSFE^{26}E^{27}AR$, and $E^{30}VFE^{33}NTE^{36}R$ were enriched after purification (**Fig. 3D-F**), γ-carboxylated $TTE^{40}FWK$ was not strongly enriched after purification (**Fig. 3G**). This suggests that γ-carboxylation of E^{40} is not critical for binding to POROS 50 HQ strong anion-exchange resin. These results are consistent with the low conservation of the 12th GLA residue (E^{40}) in other γ-carboxylated coagulation factors [14], with lack of effect of loss of E^{40} γ-carboxylation on FIX binding to phospholipid membrane and Factor X activation [69], and with incomplete occupancy at this site on pdFIX [13]. In agreement with our hypothesis, purified H1 rFIX was more γ-carboxylated than purified H2 rFIX. It is possible that other PTMs, site-specific contributions, and differences in sample complexity (host cell proteins (HCPs) levels and variety) contributed to the

purification of lower γ-carboxylated rFIX forms in H2. These results indicate that use of a strong anion exchange resin alone does not ensure that the purified rFIX will be completely modified, supporting the need for MS techniques to measure site-specific γ-carboxylation to confirm product quality.

Altogether, we designed a positive ion mode LC-ESI-DIA-MS method to measure site-specific γ-carboxyforms with or without derivatization from the bioreactor supernatant or purified rFIX. The two approaches showed similar results, but derivatization provided a more complete analysis of site-specific rFIX quality. The results showed that the operating conditions of bioreactor H1 favored production of lower amounts of more efficiently γ-carboxylated rFIX compared to bioreactor H2. Importantly, the results demonstrated that measuring yield and γ-carboxylation levels in the bioreactor culture supernatant could indeed predict the bioreactor conditions which led to the highest quality purified product.

Characterization of known and new PTMs on purified rFIX

In addition to γ-carboxylation, FIX is modified by a large number of heterogeneous PTMs, including proteolysis, *N*- and *O*-glycosylation, sulfation, phosphorylation, β-hydroxylation, and disulfide bonds [7, 20] (**Fig. 1**). Changes in expression system or culture conditions can lead to changes in the quantity and quality of recombinant products [74-78], including rFIX [11, 79]. Thus, we hypothesized that in addition to γ-carboxylation, the occupancy and structure of the other PTMs may also be different on rFIX produced in different bioreactor conditions. To test this, we first performed an in-depth DDA-MS proteomic characterization of PTMs on rFIX purified from both bioreactors, and then used the information from the tryptic digests to measure PTM relative abundance both during bioreactor operation and after purification.

Digestion with a range of proteases with orthogonal activity and the addition of PNGase F allowed characterization of the heavily post-translationally modified FIX domains (**Fig. 1**) [7, 10, 14]. These analyses achieved a combined coverage of ~ 85% of the mature rFIX protein sequence (**Supplementary Fig. S5**). We were able to observe most known PTMs on pdFIX and rFIX, and also new PTMs on rFIX (**Figs. 1** and **4)**. Most new PTMs identified were only observed in rFIX (with the exception of the Asp oxidations). Most modifications showed partial occupancy in rFIX

and were generally more heterogeneous than in pdFIX. The compiled data is displayed in **Table 1**, **Fig. 4A**, **Supplementary Figs. S5** and **S6**, **Supplementary Table S1**, and **Supplementary information**. Below we summarize the results and highlight some of the novel discoveries.

FIX contains a propeptide region (T^{-18}VFLDHENANKILNRPKR^{-1}) that is cleaved during secretion by the endoprotease PACE/Furin [16, 17]. The propeptide holds a binding site for the γ-glutamyl carboxylase and is required for γ-carboxylation of the GLA domain [80-82], but failure to cleave the propeptide leads to inactive FIX [9, 15]. We did not identify peptides covering the propeptide region in pdFIX, which is consistent with complete proteolytic processing of pdFIX. However, we identified peptides covering the T^{-18}VFLDHENANKIL^{-6} propeptide region in rFIX from both bioreactors (**Table 1**, **Supplementary Fig. S5**, and **Supplementary Table S1**). These results indicated that PACE/Furin activity in both bioreactor cultures was insufficient to fully process rFIX's propeptide.

Table 1. Post-translational modifications observed on pdFIX and rFIX from bioreactors H1 and H2.

FIX Domain	Residue(s)	Post-translational modification	pdFIX	H1 rFIX	H2 rFIX	Ref
Propeptide	T^{-18}VFLDHENANKIL^{-41}	Proteolysis		*	*	[16, 17]
GLA	E^{7}, and/or E^{8}, and/or E^{15}	Carboxylation		*p	*p	[19]
	E^{17}, and/or E^{20}, and/or E^{21}	Carboxylation			*p	[19]
	E^{26} and/or E^{27}	Carboxylation		*p	*p	[19]
	E^{30}, and/or E^{33}, and/or E^{36}	Carboxylation		*p	*p	[19]
	E^{40}	Carboxylation	*	*p	*p	[19]
	$T^{38/39}$	*O*-glycan $HexNAc_1Hex_1NeuAc_1$ @		*p	*p	This study
		O-glycan $HexNAc_1Hex_1NeuAc_2$ @		*p	*p	[25]
		O-glycan $HexNAc_1Hex_1NeuAc_1NeuGc_1$ @		*p	*p	This study
	Y^{45}	Sulfation/phosphorylation		*p	*p	This study
EGF-like 1	D^{47}	Oxidation		*p	*p	This study
	D^{49}	Oxidation	*p	*p	*p	This study
	S^{53}	*O*-glycan Hex_1Xyl_1	*			[24, 25]
		O-glycan Hex_1Xyl_2	*	*p	*p	[21, 24, 25]
	S^{61}	*O*-glycan $Fuc_1HexNAc_1Hex_1NeuAc_1$	*	*p	*p	[21, 24, 25]
		O-glycan $Fuc_1HexNAc_1Hex_1NeuGc_1$		*p	*p	This study
	S^{53}, S^{61}, S^{68}	*O*-glycan Hex_1Xyl_2, $Fuc_1HexNAc_1Hex_1NeuAc_1$, Fuc_1 @		*p	*p	This study
	D^{64}	β-hydroxylation	*p	*p	*p	[21, 27, 28]
	S^{68}	Phosphorylation	*p			[29]
EGF-like 2	D^{85}	Oxidation	*p	*p	*p	This study
	D^{104}	Oxidation	*p	*p	*p	This study
	$S^{102/110}$, T^{112}	*O*-glycans Hex_1Xyl_2, $Fuc_1HexNAc_1Hex_1NeuAc_1$ @		*p	*p	This study
Linker	S^{141}	*O*-glycan $HexNAc_1Hex_1NeuAc_1$	*p	*p	*p	[13]
		O-glycan $HexNAc_1Hex_1NeuAc_2$	*p	*p	*p	[13, 30]

AP[$]	Y^{155}/S^{158}	Sulfation/phosphorylation @	*p	*p	*p	[25, 34]
	N^{157}	*N*-glycan	*	*	*	[13, 31-34]
	T^{159}	*O*-glycan $HexNAc_1Hex_1NeuAc_1$ @		*p		[25, 35]
		O-glycan $HexNAc_1Hex_1NeuAc_2$ @	*p	*p	*p	[25, 36]
	N^{167}	*N*-glycan	*	*	*	[13, 31-34]
	$T^{169/172}$	*O*-glycan $HexNAc_1Hex_1NeuAc_1$ @	*p			[21, 25, 35]
		O-glycan $HexNAc_1Hex_1NeuAc_2$ @		*p	*p	[21, 25, 36]
	T^{179}	*O*-glycan $HexNAc_1Hex_1NeuAc_1$		*p	*p	[21, 25, 36]
		O-glycan $HexNAc_1Hex_1NeuAc_2$		*p	*p	[21, 25, 36]
		O-glycan $HexNAc_1Hex_1NeuAc_1NeuGc_1$		*p		This study
Protease	D^{186}	Oxidation	*p	*p		This study
	D^{203}	Oxidation	*p	*p	*p	This study
	N^{258}	*N*-glycan $HexNAc_4Hex_5NeuAc_2$	*p			[13]
	D^{276}	Oxidation	*p	*p	*p	This study
	D^{292}	Oxidation	*p	*p	*p	This study
	D^{359}	Oxidation	*p	*p		This study
	D^{364}	Oxidation	*p	*p		This study

* = only occupied peptides identified

*p = at least one partially occupied peptide identified.

@ It was not possible to determine the precise location of the PTM.

[$] Occupancy inferred after enzymatic removal of *N*-glycans.

We were able to obtain 87% coverage of the GLA domain in H1 and H2 rFIX and 11% in pdFIX (**Supplementary Fig. S5**). We identified a variety of peptides covering the entire GLA region in both rFIX, with 0 to 3 γ-carboxyglutamic acids per peptide (3 being generally the maximum possible occupancy). Consistent with the DIA-MS data (**Fig. 3**) H2 rFIX showed higher heterogeneity of γ-carboxyglutamic acid occupancy compared to H1 rFIX (**Supplementary Table S1** and **Supplementary Fig. S6A**). In contrast to the heterogeneity observed with rFIX, we only observed peptides containing γ-carboxylated E^{40} in pdFIX (**Supplementary Fig. S6A**, **Table 1**, **Supplementary Tables S1**). These results are consistent with higher levels of γ-carboxylation in pdFIX compared to rFIX [21, 69, 70], and in H1 rFIX compared to H2 rFIX (**Fig. 3** and **Supplementary Figs. S4** and **S6A**). We observed the $T^{38/39}$ *O*-glycan in both H1 and H2 rFIX with the reported monosaccharide composition $HexNAc_1Hex_1NeuAc_2$ [25] (**Fig. 1**) and also with the new compositions $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_1NeuGc_1$ (**Fig. 4A, Supplementary Fig. S6A**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). The $T^{38/39}$ *O*-glycan was not detected in pdFIX, and was only observed when E^{40} was not γ-carboxylated (**Supplementary Table S1**). Because γ-carboxylation occurs in the ER and *O*-glycosylation in the Golgi Complex, this result suggests that γ-carboxylation of E^{40} prevented *O*-glycosylation at $T^{38/39}$. This would also explain why this *O*-glycan has not been previously observed in pdFIX, since pdFIX is essentially completely γ-carboxylated [13, 19]. Additionally, we identified a previously undescribed sulfation/phosphorylation event on rFIX at Y^{45} (**Fig. 4B**, **Supplementary Fig. S6B**, **Table 1**, **Supplementary Table S1**, and **Supplementary Information**). The CID fragmentation pattern of Y^{45} modified peptides showed predominantly fragment ions with neutral loss of SO_3, supporting the presence of a sulfation at Y^{45} [83] (**Fig. 4B**). However, some MS/MS spectra also contained a phosphotyrosine immonium ion (216.043 *m/z*) [84] (**Fig. 4B**, inset). Other analytical workflows, including higher resolution positive ion mode or negative ion mode ESI-MS/MS analyses would be required to distinguish between sulfation or phosphorylation at Y^{45} [83-85] Analysis of the MS1 spectra of $F^{41}WKQY^{45}VDGDQCE^{54}$ peptides showed that Y^{45} is largely unmodified in rFIX from both H1 and H2 bioreactors (**Supplementary Fig. S6B**). The Y^{45} sulfation/phosphorylation was not observed in pdFIX. Overall, the GLA domain of both H1 and H2 rFIX showed higher heterogeneity than pdFIX in terms of types of PTMs, PTM composition, and occupancy. These PTMs included lower levels of γ-carboxylation

and higher levels of *O*-glycosylation in rFIX compared to pdFIX, and the presence of a novel C-terminal sulfation/phosphorylation event unique to rFIX.

We were able to obtain 100% coverage of the region containing the EGF-like 1, EGF-like 2, and linker domains in all FIX variants (**Supplementary Fig. S5**). The peptides covering these domains were considerably more heterogeneous in rFIX than in pdFIX, showing a larger variety of PTM types and compositions than pdFIX, and showing partial occupancy for all PTMs in rFIX (**Table 1** and **Supplementary Table S14**). We detected all the previously known PTMs in these domains and identified several new PTMs (**Fig. 1** and **4**, **Table 1**, **Supplementary Fig. S6**, **Supplementary Table S1**, **Supplementary information**). In the EGF-like 1 domain we identified the previously undescribed, low abundant, partial oxidation of one or both D^{47} and D^{49} in the peptide $F^{41}WKQYVD^{47}GD^{49}QCE^{54}$ in both H1 and H2 rFIX and in pdFIX (**Fig. 4A, Supplementary Fig. S6B**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). In addition to the previously described $Fuc_1HexNAc_1Hex_1NeuAc_1$ *O*-glycan at S^{61} [25, 26], we identified a minor fraction of $Fuc_1HexNAc_1Hex_1NeuGc_1$ in H1 and H2 rFIX (**Fig. 4A**, **Table 1**, **Supplementary Table S1**, **Supplementary information**). We also observed a peptide carrying the typical Hex_1Pent_2 and $Fuc_1HexNAc_1Hex_1NeuAc_1$ *O*-glycans but with an additional Fuc_1 from rFIX from both bioreactors (**Fig. 4A**). The extra fucose could be attached to one of the known *O*-glycans or it could be a mono *O*-fucosylation of S^{68} (inspection of peptide Y ions did not allow for an unequivocal distinction). Mono *O*-fucosylation of recombinant proteins expressed in CHO, including within EGF-like domains, has been previously reported [86-89]. We observed the β-hydroxylation of D^{64} in pdFIX [27, 28] and in both H1 and H2 rFIX (**Fig. 1**, **Supplementary Fig. S6C**, **Table 1**, and **Supplementary Table S1**). Both H1 and H2 rFIX were more β-hydroxylated at D^{64} than pdFIX, as previously reported for rFIX (**Supplementary Fig. S6C**) [21, 90, 91]. We observed the S^{68} phosphorylation in pdFIX [29], but not in H1 and H2 rFIX (**Fig. 4A**, **Supplementary Fig. S6C**, **Table 1**, **Supplementary Table S1**, **Supplementary information**). No PTM has been reported so far for the EGF-like 2 domain of FIX. However, we identified low occupancy oxidations at D^{85} and D^{104} and two possible new sites of *O*-glycosylation at $S^{102/110}$ and T^{112} in both H1 and H2 rFIX (**Fig. 4A**,**C**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). The D^{85} and D^{104} oxidations were also observed in pdFIX. The $S^{102/110}$ and T^{112} sites contained *O*-glycans with masses consistent with a Hex_1Pent_2 *O*-Glu and a

$Fuc_1HexNAc_1Hex_1NeuAc_1$ *O*-Fuc glycans in the $Q^{97}FCKNSADNKVVCSCTE^{113}$ peptide (**Fig. 4C**, **Supplementary Table S1**, and **Supplementary information**). The presence of multiple Y ions and diagnostic oxonium ions in the MS/MS spectra were also consistent with these compositions (**Fig. 4C**). These data suggest that the EGF-like 2 domain in rFIX contains similar *O*-glycans as the EGF-like 1 domain, but in lower abundance. *O*-glucosylation and *O*-fucosylation aid in EGF domain stabilization and function [88, 92], which supports the presence of these glycans at this location. Overall, we were able to identify all previously reported PTMs for the EGF-like and linker domains, as well as new PTMs. With the exception of oxidation at D^{49}, D^{85}, and D^{104}, which were identified in both pdFIX and rFIX, all other new PTMs appeared exclusively in rFIX from H1 and H2 bioreactors: oxidation at D^{47}, $Fuc_1HexNAc_1Hex_1NeuGc_1$ *O*-Fuc glycan likely at S^{61}, a difucosylated $S^{53}S^{61}S^{68}$ glycopeptide, and *O*-glucosylation and *O*-fucosylation of $S^{102/110}/T^{112}$ in the EGF-like 2 domain.

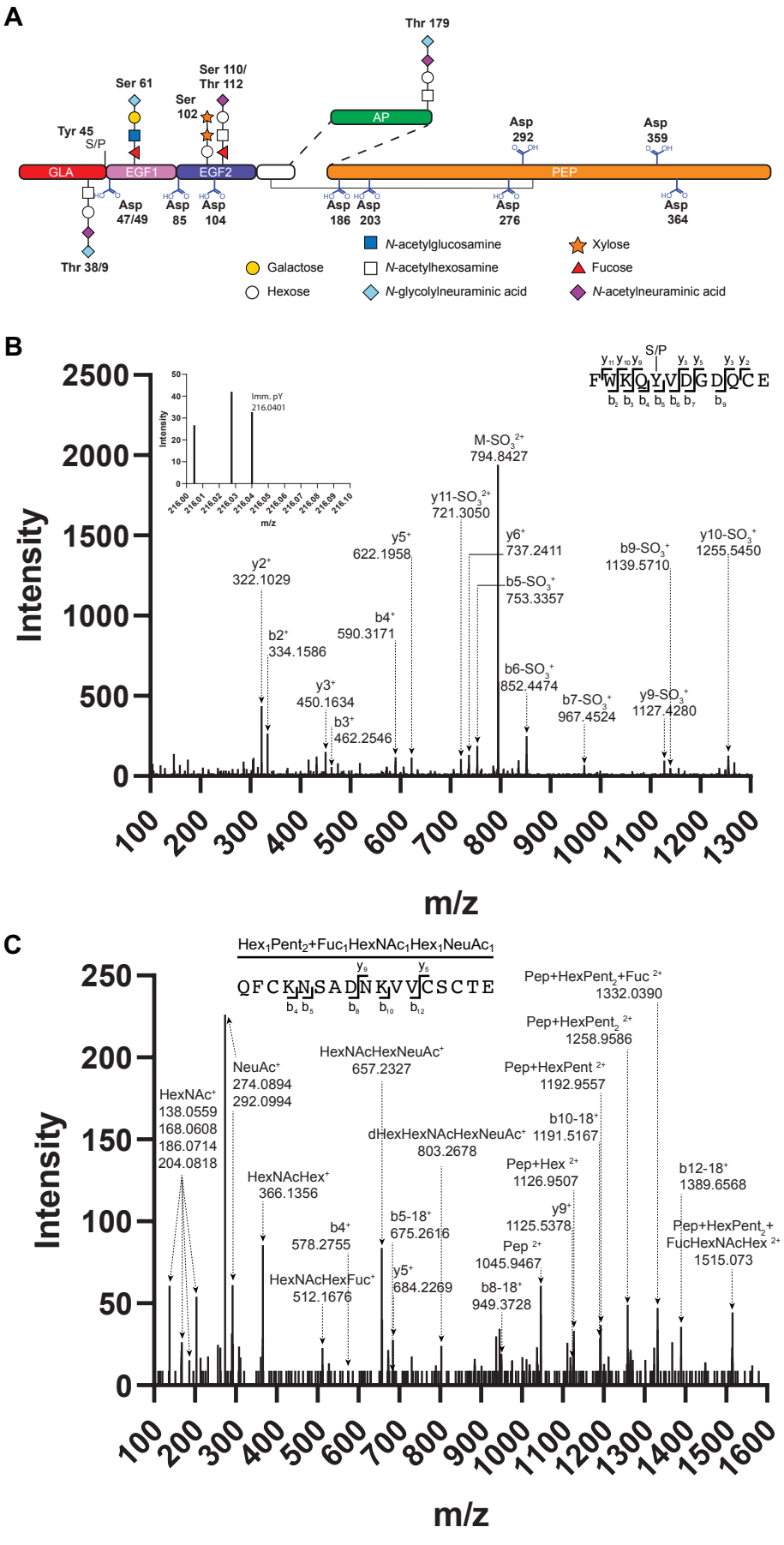


Figure 4: New post-translational modifications identified on rFIX from bioreactors H1 and H2. (A) Schematics of FIX containing the new PTMs identified on rFIX in this study. (B,C) CID fragmentation of select GluC rFIX peptides. **(B)** $F^{41}WKQYVDGDQCE^{54}$ peptide with sulfation/phosphorylation (S/P) at Y^{45} (observed precursor *m/z* value 834.8189^{2+}, Δ 2.4 ppm). The inset shows the phosphotyrosine immonium ion (pY, 216.0401 *m/z*, Δ 2.9 ppm). **(C)** $Q^{97}FCKN(+1)SADN(+1)KVVCSCTE^{113}$ glycopeptide with Hex_1Xyl_2 and $Fuc_1HexNAc_1Hex_1NeuAc_1$ *O*-glycans attached to $S^{102/110}$ and T^{112} (observed precursor *m/z* value 1107.1067^{3+}, Δ 7.56 ppm). Pep = peptide.

We obtained a coverage of 100% of the AP in all FIX variants, and identified several known PTMs and one *O*-glycan composition (**Fig. 1** and **Supplementary Fig. S5**). FIX AP is modified with 4 *O*-glycans, 2 *N*-glycans, one sulfation, and one phosphorylation (**Fig. 1**) [7, 21, 25, 31, 32, 34, 35]. The *N*-glycans in the AP are fucosylated or afucosylated tri- and tetra-antennary sialylated structures [31], while the *O*-glycans are mono and disialylated $GalNAc_1Gal_1$ structures [93]. The large size and number of glycans make characterization of the AP by LC-ESI-MS/MS difficult. To determine if the AP was glycosylated and to be able to observe other PTMs we treated samples with PNGase F. De-*N*-glycosylation allowed for the detection of several peptides deamidated at N^{157} or N^{167} in rFIX H1 and H2, indicating that these peptides were previously *N*-glycosylated (**Supplementary Table S1** and **Supplementary information**). These deamidated peptides were observed with and without other previously described modifications [21, 25, 35]. Partial occupancy of sulfation/phosphorylation and *O*-glycans on the AP has been previously described [21]. We observed the previously reported sulfation/phosphorylation of Y^{155}/S^{158} [21, 25, 34] in pdFIX and H1 and H2 rFIX (**Fig. 1**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). We also observed the previously reported $HexNAc_1Hex_1NeuAc_2$ *O*-glycan at S^{158}/T^{159} [25, 35] on pdFIX and H1 and H2 rFIX, and $HexNAc_1Hex_1NeuAc_1$ on H1 rFIX. We observed peptides containing one $HexNAc_1Hex_1NeuAc_2$ *O*-glycan at one of the $T^{169/172}$ sites in H2 rFIX and one $HexNAc_1Hex_1NeuAc_1$ on pdFIX (**Fig. 1**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). Finally, we observed *O*-glycans with the known compositions $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$, and the new composition $HexNAc_1Hex_1NeuAc_1NeuGc_1$ at T^{179} in H1 and H2 rFIX (**Fig. 4A**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). Overall, several known AP PTMs were identified on both pdFIX and rFIX, and one new *O*-glycan composition was identified in T^{179} on rFIX. All FIX variants showed partial PTM occupancy except for the inferred completely occupied *N*-glycosites.

Finally, we obtained a coverage of 99.5%, 87.6%, and 97% of the protease domain in rFIX from H1, H2, and pdFIX, respectively (**Supplementary Fig. S5**). We observed the N^{258} *N*-glycan with the composition $HexNAc_4Hex_5NeuAc_2$ on pdFIX [13], but not rFIX (**Fig. 1**, **Supplementary Fig. S6E**, **Table 1**, **Supplementary Table S1**, and **Supplementary information**). Deglycosylation with PNGase F did not help in identifying the presence of the *N*-glycan at N^{258} in rFIX because deamidation was observed in N^{258} containing peptides also without enzymatic deglycosylation (**Supplementary Table S1** and **Supplementary information**). Indeed, we note that Asn deamidation was identified throughout pdFIX and both rFIX, in Asn outside sequons and without PNGase F treatment (**Supplementary Table S1**). In addition, we observed several Asp oxidations with partial occupancy in pdFIX and H1 rFIX, and also in some cases in H2 rFIX, including at D^{186}, D^{203}, D^{276}, D^{292}, D^{359}, and D^{364} (**Fig. 4A**, **Table 1**, **Supplementary Table S1**, **Supplementary information**). Together, our data extended the repertoire of PTMs identified in the protease domain.

In summary, we observed most of the known PTMs of pdFIX in rFIX purified from both H1 and H2 bioreactors, including γ-carboxylation, glycans, sulfation/phosphorylation, and oxidation, as well as several new PTMs (**Fig. 4A**). Many new PTMs were localized at or near the GLA domain, and could impact rFIX folding, function, and/or purification efficiency. Other new PTMs were localized in the EGF-like 2 domain and may impact the stability and function of this domain. We also observed higher heterogeneity in PTMs on rFIX compared to pdFIX, both in terms of occupancy and composition. This higher PTM heterogeneity of rFIX may be the consequence of the artificial high expression CHO system combined with the changing metabolic status of cells during a fed batch incubation, compared to a more stable native, physiological expression of pdFIX.

Predicting the abundance of PTMs in purified rFIX by analyzing the culture supernatant

We showed above that measuring rFIX γ-carboxylation levels in the bioreactor supernatant could predict the levels of γ-carboxylation of the purified protein (**Fig. 3**). While γ-carboxylation is a key PTM on FIX, other PTMs also have important functional roles and their composition and abundance can also be dependent on changes in the expression system and bioprocess parameters

[3, 5, 9, 10, 14, 21, 27, 68, 92, 94-99]. Therefore, it would be advantageous to be able to predict the relative abundance and composition of different types of PTMs on purified rFIX by analyzing the culture supernatant. To do this, we first identified high confidence tryptic peptides containing sites of PTM on purified rFIX and used them to build an ion library for DIA-MS (**Supplementary Table S8**). The ion library contained information to measure *O*-glycosylation at $T^{38/39}$, S^{53}, S^{61}, and S^{141}; *N*-glycosylation at N^{258}; and oxidation at D^{64}, D^{85}, D^{104}, and D^{203}. We used the ion library to interrogate the DIA-MS data from the purified rFIX samples and the culture supernatants from both bioreactors. To calculate the relative abundance of each PTM variant, we normalized the data for each variant to the abundance of rFIX in each sample (**Supplementary Table S9**). Because the DIA data was acquired with the standard 26 Da windows, the measurements for D^{104} and D^{203} oxidations could not be used because the unoxidized and oxidized versions of these peptides fell in the same SWATH window and the precursors eluted at a similar RT (**Supplementary Table S9**). This was not an issue for the D^{64} variants, which eluted at different RT. We could also not quantify oxidized D^{85} because of the presence of a second, unidentified peak at similar RT and in the same SWATH window sharing a large number of the transitions selected to measure the oxidized D^{85} peptide. In total, 8 peptide variants were reliably quantified (FDR < 0.01), accounting for 4 post-translationally modified sites (**Fig. 5** and **Supplementary Table S9**).

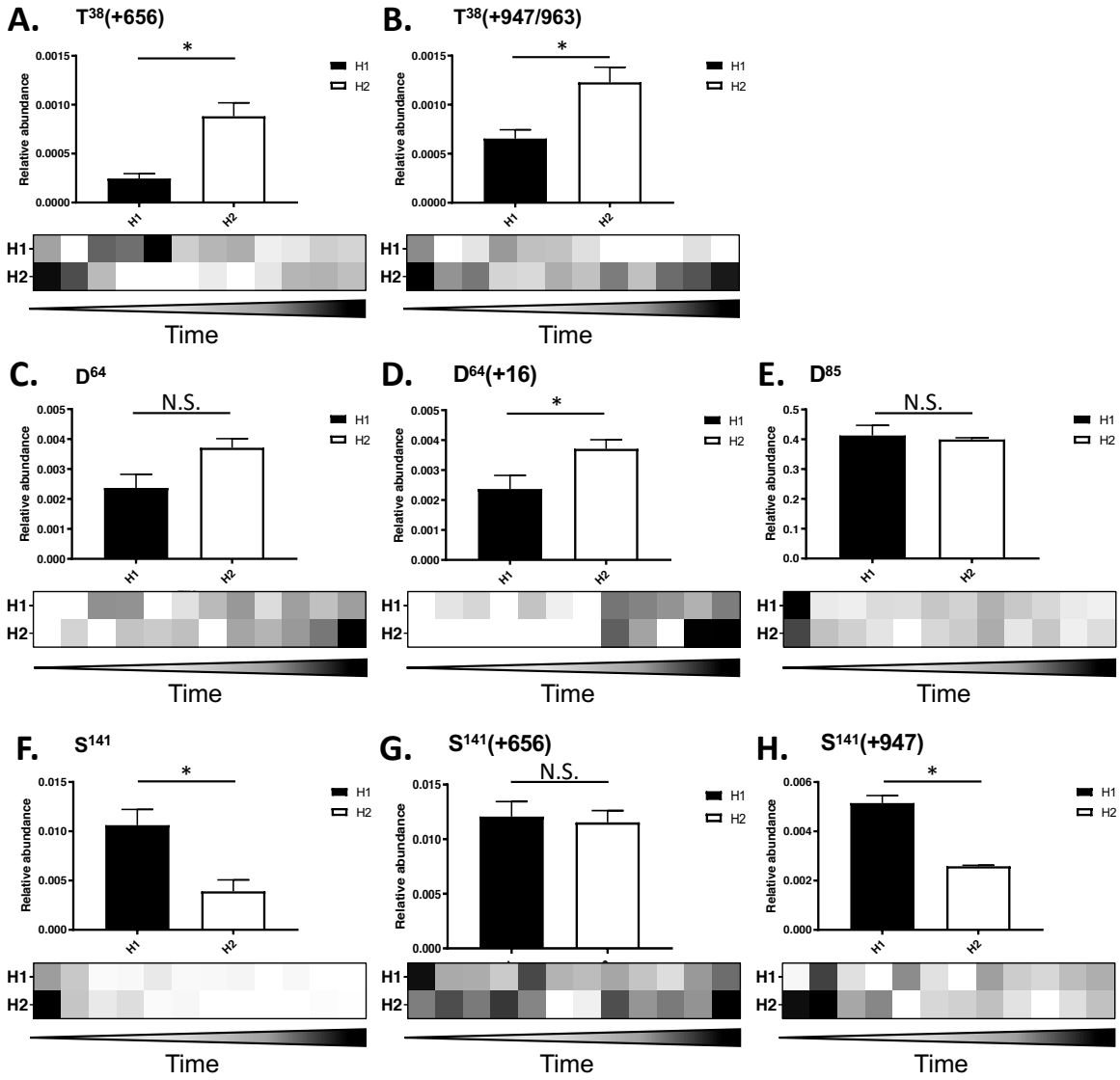


Figure 5: Relative abundance of post-translational modifications (PTMs) on rFIX during bioreactor operation and after purification measured by DIA-MS. Bar graphs depict peptide relative abundance in purified samples from H1 (black) or H2 (white) bioreactor. The heatmaps show peptide relative abundance in each bioreactor supernatant over time (days 1-13). Values correspond to the mean +/- SEM (bar graphs) or the mean (heatmaps) (N = 2 - 3) of the normalized abundance of each post-translationally modified rFIX peptide relative to the abundance of rFIX in each sample. Shown are: **(A)** $T^{38/39}$ *O*-glycan $HexNAc_1Hex_1NeuAc_1$ (+656) **(B)** $T^{38/39}$ *O*-glycan $HexNAc_1Hex_1NeuAc_2$ (+947) and $HexNAc_1Hex_1NeuAc_1NeuGc_1$ (+963) **(C)** D^{64} with no oxidation. **(D)** D^{64} β-hydroxylation (+16). **(E)** D^{85} with no oxidation. **(F)** S^{141} unglycosylated. **(G)** S^{141} *O*-glycan $HexNAc_1Hex_1NeuAc_1$ (+656) and low abundant S^{141} unglycosylated peptide. **(H)** S^{141} *O*-glycan $HexNAc_1Hex_1NeuAc_2$ (+947). One-tailed t-test: * $p < 0.05$, N.S. = not significant.

We found significant differences in the abundance of different rFIX PTMs between H1 and H2 samples. There was significantly higher mono (+656) and disialylated (+947/+963) $T^{38/39}$ *O*-glycan in purified H2 rFIX compared to H1 rFIX (**Fig. 5A,B**). We note that the $HexNAc_1Hex_1NeuAc_2$ (+947) and $HexNAc_1Hex_1NeuAc_1NeuGc_1$ (+963) glycoforms were quantified together, since their precursors had a similar *m/z* (879.8644^{2+} *vs* 887.8618^{2+}) that fell on the same SWATH window, and eluted at similar RT (**Supplementary Fig. S6A**). D^{64} was significantly more β-hydroxylated in purified H2 rFIX compared to H1 (**Fig. 5C,D** and **Supplementary Fig. 6C**). There was no significant difference in the abundance of unoxidized D^{85} peptide in the purified rFIX from H1 and H2 bioreactors (**Fig. 5E**). Finally, the S^{141} peptide was significantly less glycosylated in H1 rFIX compared to H2 rFIX (**Fig. 5F**), but the S^{141} *O*-glycan was significantly more sialylated (+947) in H1 rFIX compared to H2 rFIX (**Fig. 5H**). We note that the $HexNAc_1Hex_1NeuAc_1$ (+656) glycosylated and unglycosylated S^{141} peptides were quantified together, since their precursors had a similar *m/z* that fell in the same SWATH window (621.309^{3+} *m/z vs* 603.3461^{2+} *m/z*) and eluted at similar RT (**Supplementary Fig. S6D**). However, the unoccupied peptide (603.3461^{2+} *m/z*) had lower relative intensity than the glycosylated variant (621.309^{3+} *m/z*), and the most intense ions measured for the glycopeptide were the Y ions, indicating that the signal measured for these precursors corresponds mostly to the glycopeptide. Therefore, the different feeds led to significant relative quantitative differences in glycan occupancy, glycan sialylation, and β-hydroxylation between purified H1 and H2 rFIX, and these differences were site-specific.

To test if the differences in PTM abundance observed above could have been predicted by measuring those PTMs in the bioreactor, we quantified these same PTMs in the bioreactor supernatants throughout the 13 days of operation. Consistent with the purified rFIX results, we found that the abundance of the mono (+656) and disialylated (+947/+963) $T^{38/39}$ *O*-glycans was higher in the bioreactor H2 compared to H1 in the later days of the culture (**Fig. 5A,B,** heatmaps). There was also more β-hydroxylated D^{64} in bioreactor H2 compared to H1 in the later days of the process (**Fig. 5D**, heatmap). Also in agreement with analysis of purified rFIX, we observed generally similar amounts of unoxidized D^{85} in the H1 bioreactor and the H2 bioreactor throughout most of the 13 day process (**Fig. 5E**, heatmap). We were able to detect the unoccupied S^{141} peptide in the earlier days during bioreactor process in both H1 and H2 bioreactors, but only weakly in the later days, suggesting that the S^{141} site became occupied as the process progressed (**Fig. 5F,**

heatmap). Unmodified S^{141} peptide was still detectable in the H1 bioreactor in later days compared to the H2 bioreactor, consistent with the lower levels of occupancy of S^{141} measured in purified H1 rFIX compared to H2 rFIX (**Fig. 5F**). The relative abundance of the two S^{141} *O*-glycoforms varied throughout the process (**Fig. 5G,H** heatmaps). Consistent with the purified rFIX results, there was more disialylated S^{141} *O*-glycopeptide in the later days of the process in bioreactor H1 compared to H2 (**Fig. 5H**, heatmap, **Supplementary Fig. S6D**). Therefore, we were able to directly measure the abundance of several rFIX PTMs in bioreactor supernatant samples, and these site-specific PTM profiles were generally consistent with purified rFIX.

We observed that some peptide variants were measurable even in the earlier days of the bioreactor process, when rFIX concentration was low, while other peptide variants were difficult to measure even in the later days of the bioprocess. This was especially the case in the H1 bioreactor, in which rFIX relative abundance was lower than in the H2 bioreactor (although the number and variety of HCPs was also lower in H1 bioreactor compared to H2). Even though GluC provided a better coverage of rFIX PTMs (**Supplementary Table S1**), we performed the quantification studies using trypsin, a standard choice in the field due to its robustness and frequency of cleavage sites. However, future studies may benefit from the use of GluC for quantification of rFIX PTMs [100].

In summary, using DIA-MS to measure rFIX PTMs in the bioreactor supernatant and after purification, and consistent with the results for the γ-carboxylation measurements (**Fig. 3**), we observed a strong correlation between the relative abundance of the rFIX's PTMs measured in the bioreactor supernatant and in the purified material (**Fig. 5** and **Supplementary Fig. S4**). Together, these data show that measurements in the bioreactor can predict measurements in the purified forms.

Predicting purity of purified rFIX by measuring changes in host cell proteins during bioreactor operation

Host cell proteins (HCPs) present in the bioreactor supernatant are secretory and intracellular proteins released into the culture medium by the cells through secretion and cell lysis. A decrease in cellular viability is usually accompanied by an increase in the diversity and abundance of HCPs in the culture supernatant [101, 102]. Bioreactor H2 showed greater loss of viability compared to

H1 during the last days of the bioprocess (**Fig. 2A**). Thus, we predicted that the HCP content in H2 culture supernatant would be higher than in H1. To test this idea, we used DIA-MS to measure changes in HCP abundance in the culture supernatant. Using a combined ion library (**Supplementary Table S3**) containing 768 proteins identified in the bioreactor culture supernatants (global FDR < 0.01), we confidently measured the abundance of 694 of these proteins throughout the process (**Supplementary Tables S3-5**). We observed that proteome complexity and relative abundance of most proteins increased through time in both bioreactors, as expected since proteins accumulate in the culture supernatant in a fed batch bioreactor (**Supplementary Fig. S7**, and **Supplementary Table S5**). To compare the proteomic differences between both bioreactors we performed MSstats analysis on the DIA-MS data at different time points throughout the process (days 3, 6, 9, and 13) (**Fig. 2C** and **Supplementary Table S6**). We confidently measured an increasing number of proteins in both bioreactors as the processes progressed, from 293 total proteins at day 3 to 458 total proteins at day 13 (FDR < 0.01) (**Figs. 2C** and **6, Supplementary Fig. S7**, and **Supplementary Table S6**). Most proteins showed no significant difference in abundance between bioreactors H1 and H2 up to day 9 (**Fig. 6**, 225-322 proteins or ~ 70-76%, $P > 10^{-5}$). However, at day 13 most of the proteins measured were significantly more abundant in the H2 bioreactor compared to H1 (203 proteins, 44.3%, $P < 10^{-5}$, **Fig. 6**, **Supplementary Table S12**). Gene Ontology analysis of the most abundant proteins in bioreactor H2 culture supernatant at day 13 showed that these proteins were associated with the cytoskeleton, the secretory pathway, the lysosome and proteasome (proteases, hydrolases, isomerases, among others), metabolism (glycolysis, amino acid biosynthesis), and apoptosis (**Supplementary Tables S6** and **S13**). Therefore, there were large proteomic changes throughout bioreactor operation associated with using the different EfficientFeeds, which exacerbated as the cultures progressed. The more dramatic proteomic and metabolic changes observed in bioreactor H2 began at approximately day 11 (**Figs. 2, 6,** and **Supplementary Figs. S2** and **S7**), when the batches were no longer being fed (last feed was at day 10). Together, these results suggest that unmet metabolic demands in bioreactor H2 led to higher loss of viability, which in turn led to higher HCP release into the supernatant in bioreactor H2 compared to H1.

Day	More abundant in H1	More abundant in H2	Not significantly different
3	59 (20.1%)	9 (3.1%)	225 (76.8%)
6	68 (18.4%)	20 (5.4%)	282 (76.2%)
9	65 (14.3%)	67 (14.8%)	322 (70.9%)
13	79 (17.2%)	203 (44.3%)	176 (38.4%)

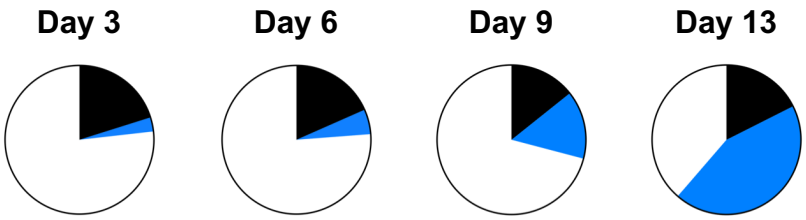


Figure 6: DIA-MS relative quantification of host cell proteins in the supernatant during bioreactor operation. The table and pie charts show the number (and %) of proteins that were either significantly more abundant in H1 (black), significantly more abundant in H2 (blue), or not significantly different (white) at days 3, 6, 9, and 13 of the bioreactor process ($P < 10^{-5}$, $N = 3$).

HCPs in the bioreactor supernatant can impact the purification of recombinant products and the quality of the final product [103, 104]. Because of the more efficient metabolism, higher cell viability, and lower complexity of HCPs in the supernatant of bioreactor H1 **(Figs. 2** and **6)** we predicted that the purified H1 sample would have lower co-purifying HCP complexity and abundance relative to rFIX compared to H2. To test if the H1 purified sample had lower HCP complexity than the H2 purified sample we performed independent DDA searches of the purified H1 and H2 rFIX samples (**Supplementary Tables S14** and **S15**). As expected, we confidently identified 18 co-purifying HCPs in the H1 sample and 51 HCPs in the H2 sample (1% global FDR) (**Supplementary Tables S14-16**). Most of the proteins co-purifying with H1 rFIX also co-purified with H2 rFIX (13 proteins) (**Supplementary Table S16**). Except for histone H2B, all the other 12 common protein contaminants were secretory proteins or intracellular proteins that can be potentially secreted such as carboxypeptidase, lipoprotein lipase (LPL), clusterin, BiP, thrombospondin-1, vitamin K-dependent protein (S), procollagen C-endopeptidase enhancer 1, serine protease HTRA1, HSP90α [105], peroxiredoxin [106], and thioredoxin reductase 1 [107] (**Supplementary Table S16**). Several of these proteins bind calcium, including carboxypeptidase, LPL, peroxiredoxin, thrombospondin-1, and vitamin K-dependent protein S, which contributes to explaining their co-purification with rFIX in our system (**Supplementary Table S16**). Other

proteins like BiP, serine protease HTRA1, and histones are common CHO HCP contaminants observed during downstream recombinant product purification [108]. Thirty-five additional proteins co-purified only with H2 rFIX, most of which were intracellular proteins (23 out of 35 proteins, ~ 66%, **Supplementary Tables S15** and **S16**). Therefore, purification using POROS 50 HQ resin enriched for rFIX (**Supplementary Fig. S8** and **Supplementary Table S6**), but additional proteins co-purified with rFIX in both samples, with the H2 sample being more complex. The increased complexity of the purified H2 sample, and the predominant presence of intracellular proteins and chaperones were consistent with increased cellular stress and cell lysis in bioreactor H2 compared to H1 (**Figs. 2** and **6**, **Supplementary Fig. S7**, and **Supplementary Table S16**). Thus, in agreement with our hypothesis, the lower complexity of HCPs in the bioreactor supernatant correlated with lower complexity of HCPs after purification.

To test the idea that the lower abundance of HCPs in bioreactor supernatants would lead to higher purity of rFIX relative to HCPs after purification we performed DIA-MS of the purified samples (**Supplementary Table S5**). As expected, rFIX was significantly more abundant after purification in the H1 sample compared to the H2 sample ($P < 10^{-5}$, **Fig. 7**, box, red bar, **Supplementary Table S6**). This result was validated by non-reducing SDS-PAGE of purified rFIX samples (**Supplementary Fig. S8**). This result is interesting because rFIX was less abundant relative to HCPs in the H1 bioreactor supernatant compared to the H2 bioreactor, indicating a differential enrichment during purification (**Fig. 2B** and **7**, blue bar). This differential enrichment was likely driven by a combination of the higher γ-carboxylation levels in H1 rFIX compared to H2 and the lower HCP contaminant complexity in the H1 supernatant compared to H2. There were 12 other proteins with significantly different abundance between both bioreactors ($P < 10^{-5}$) (**Fig. 7** and **Supplementary Table S6**). We observed significantly more thrombospondin-1 and carboxypeptidase in the purified H1 rFIX sample compared to the purified H2 rFIX sample (**Fig. 7**, red bars, **Supplementary Tables S6** and **S16**, in bold). Thrombospondin-1 and carboxypeptidase are calcium binders and were also more abundant in the supernatant from the H1 bioreactor compared to the H2 bioreactor at day 13 (**Fig. 7**, blue bars). Conversely, we observed significantly more histone H2B, histone H4, BiP, peroxiredoxin-1, clusterin, glutamine synthetase, translation initiation factor eIF3c, ADE2, and nucleolar phosphoprotein p130 in the purified H2 rFIX sample compared to the purified H1 rFIX sample (**Fig. 7**, red bars, **Supplementary Tables**

S6 and **S16**, in italics). Most of these proteins do not bind calcium, and it is possible that the increased load of intracellular proteins in bioreactor H2 interfered with rFIX purification, leading to enrichment of non-calcium binding intracellular proteins. Thus, as predicted, H1 bioreactor process conditions led to a higher relative yield of purified rFIX and lower relative abundance and complexity of HCPs than H2 conditions.

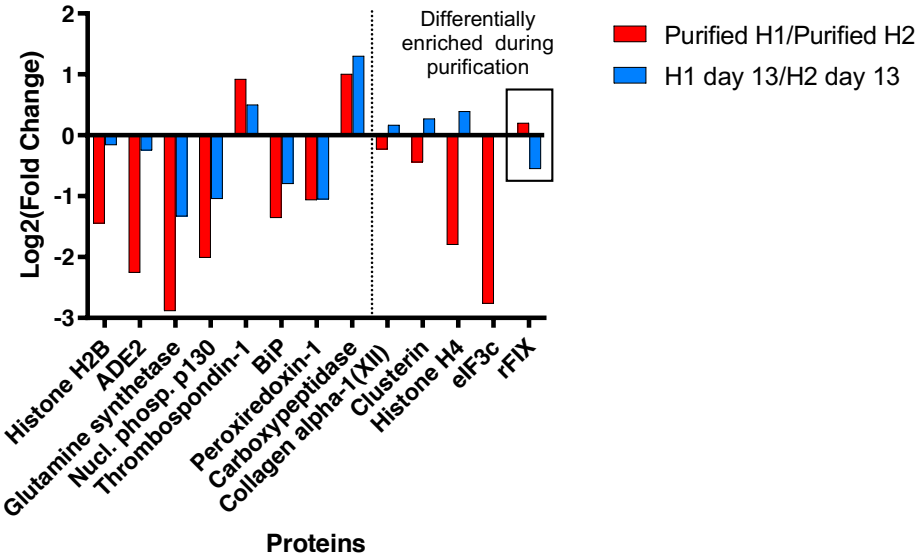


Figure 7: Comparison of changes in relative abundance of rFIX and co-purifying HCPs before and after purification. Comparison of the changes in relative abundance (Log2(Fold Change)) of the 13 proteins that were significantly different in purified H1 *vs* H2 samples before and after purification analyzed by MSstats (H1 *vs* H2, and H1 at day 13 *vs* H2 at day 13) ($P < 10^{-5}$, N = 3). rFIX comparison is highlighted with a box.

Altogether, our results indicated that bioreactor H1 outperformed H2 in terms of metabolic and cell viability profile, and in terms of rFIX quality (γ-carboxylation and protein co-contaminants) and purification efficiency. We were not only able to analyze the abundance and complexity of the proteome as it shifted during the bioprocess and after purification, but we were also able to monitor the abundance of diverse PTMs with site specificity, including the critically important γ-carboxylation of the GLA domain, and showed that their relative abundance in the supernatant correlated with their relative abundance post-purification. Collectively, the results indicate that a detailed DIA-MS analysis of the culture supernatant during bioreactor operation can predict the quality and yield of the purified product.

Discussion

rFIX is commercialized as a replacement therapy for Haemophilia B [8, 10]. Considerable effort has been devoted to optimize rFIX bioprocess and expression systems [7, 10-12, 109-113]. FIX is highly post-translationally modified with a variety of heterogeneous PTMs, and GLA domain γ-carboxylation is the most functionally relevant PTM [14, 19-21]. Changes in bioprocess operation conditions can lead to changes in occupancy and composition of PTMs, including the level of GLA γ-carboxylation [8, 10-12, 14, 42, 70, 79, 95-98, 109, 110, 113-115]. As we show here, assessment of rFIX yield and quality post-purification cannot provide a complete picture of the bioreactor performance, because rFIX purification enriches for highly γ-carboxylated product. A more effective strategy to analyze bioreactor performance is to measure yield and PTM abundance and heterogeneity in the bioreactor supernatant. To do this, we developed and implemented a suite of DIA-MS proteomics workflows to characterize and accurately monitor rFIX yield and quality in the bioreactor supernatant. We applied these workflows to compare two 13-days fed batch bioprocesses which only differed in the chemical composition of the feeds. We observed large differences in rFIX yield and in the type and relative quantity of rFIX PTMs between both fed-batches, including γ-carboxylation. Our results indicate that measurements of rFIX quality and quantity during bioreactor operation provide a real-time assessment of culture performance and allow an accurate estimation of post-purification results.

The most critical PTM on FIX, and also one of the most challenging to observe by positive ion mode LC-MS/MS, is γ-carboxylation [56, 57, 71, 72]. Here, we show for the first time that DIA-MS is an excellent tool to measure rFIX γ-carboxylation both in the bioreactor supernatant and after purification. To measure rFIX γ-carboxylation we used two complementary techniques which showed similar results: direct measurement of uncarboxylated or partially γ-carboxylated GLA peptides and measurement of all GLA peptide γ-carboxyforms after methanolic derivatization (**Fig. 3** and **Supplementary Fig. S4**). Direct GLA peptide measurement in positive ion mode can provide an accurate estimation of rFIX uncarboxylation levels [72]. However, derivatization provided site-specific γ-carboxylation information and allowed for monitoring of the fully γ-carboxylated peptides, giving a more comprehensive overview of rFIX γ-carboxylation. The

results of underivatized and derivatized carboxypeptide measurements were consistent and emphasized critical differences in the quality of rFIX produced in both fed batches (**Fig. 3** and **Supplementary Fig. S4**). The γ-carboxylation measurements also demonstrated that the purification strategy was unable to distinguish between fully and substantially uncarboxylated GLA rFIX, highlighting a weakness of the enrichment procedure. It is possible that other PTMs on the GLA and EGF-like domains (such as the *O*-glycans at $T^{38/39}$, the sulfation/phosphorylation at Y^{45}, and the oxidation at $D^{47/49}$, **Fig. 4** and **Table 1**) could impact the conformation of the GLA domain and EGF-like domains or interact with the POROS 50 HQ resin, facilitating purification of partially γ-carboxylated species. For example, H2 rFIX showed significantly higher sialylation levels at $T^{38/39}$ compared to H1 rFIX (**Fig. 5**), and these additional negative charges at the end of the GLA domain could be partially responsible for the purification of incompletely γ-carboxylated H2 rFIX. Together, we demonstrate that DIA-positive ion mode MS provided an accurate depiction of rFIX γ-carboxylation levels both in the supernatant and after purification, and that these workflows can greatly aid in optimization and quality control procedures for rFIX production.

Similar to other coagulation factors, FIX is modified by a large number of PTMs (**Fig. 1**) which impact its biosynthetic efficiency, function, and/or pharmacokinetic properties [3, 5, 9, 10, 14, 20, 21, 27, 68, 92, 94-99]. For example, rFIX *N*-linked glycans (native and engineered) influence rFIX *in vivo* half-life [94, 116-119] and may modulate rFIX activation [31], while rFIX EGF-like 1 *O*-glycans stabilize the domain and likely participate in protein-protein interactions [92, 99]. Changes in the expression system and incubation conditions alter the number and variety of PTMs in recombinant proteins [7, 10, 11, 21, 70, 109, 114], and new, unexpected PTMs may appear. Indeed, through a detailed MS proteomic analysis we identified new PTMs on purified rFIX, including a sulfation/phosphorylation on Y^{45}, several Asp oxidations spread throughout the molecule, and new *O*-glycan sites and compositions (**Fig. 4** and **Table 1**). The new Y^{45} modification was located within the $F^{41}WXXY^{45}$ aromatic amino acid stack at the end of the GLA domain, which is conserved in Factors VII, IX, and X, Protein C, and Prothrombin [20]. Studies on Prothrombin showed that the hydrophobic stack contributes to the Ca^{2+} induced folding of the GLA domain and protects the C^{18}-C^{23} disulfide bond from solvent exposure (in Prothrombin the residues are C^{17}-C^{22} and Y^{44}) [120]. Tyr *O*-sulfation is performed by the *trans*-Golgi network tyrosylprotein

sulfotransferase [121], while Tyr phosphorylation is performed by tyrosine kinases located throughout the secretory pathway [122]. Therefore, the impact of the Y^{45} modification on GLA folding would likely depend on whether the Y^{45} modification occurs before or after the folding of the GLA domain. In addition, the hydrophobic stack may interact with phospholipidic membranes [20], helping the formation of the tenase complex. Therefore, the presence of a sulfation/phosphorylation event on Y^{45} may affect important roles of the GLA aromatic stack and FIX function. Mature FIX contains 18 Asp residues, and D^{64} is β-hydroxylated [27, 28]. We identified 10 additional oxidized Asp in rFIX (11 and 9 total oxidized Asp in rFIX from bioreactor H1 and H2, respectively), and almost all of these were also identified in pdFIX (**Fig. 4** and **Table 1**). None of these additional Asp oxidations had been previously reported for pdFIX. All oxidations showed partial occupancy and, with the exception of D^{64} β-hydroxylation, they were in lower abundance compared to the non-oxidized peptide (**Supplementary Fig. S6B,C** and data not shown). It is possible that these modifications are an artefact of sample processing [123, 124]. Oxidative artefacts have been described on Met and Cys (both sulfur containing amino acids), and Trp and His, but not on Asp ([123-126] and www.unimod.org). Alternatively, the modifications may have been incorporated during the bioprocess, either in the cells or in the bioreactor [127], and they may be physiologically relevant to FIX's activity. Finally, our glycan analyses of purified rFIX revealed a portion of *O*-glycans decorated with NeuGc (**Fig. 4**, **Supplementary Fig. S6A**, and **Table 1**), a non-human type of sialic acid that is undesirable for biotherapeutic production due to its negative impact on efficacy and safety [37]. We observed a higher level of structural heterogeneity in rFIX *O*-glycans compared to pdFIX (**Table 1**), which is expected for product from a fed batch process. We also identified potential new *O*-glycan sites, including the intriguing presence of an *O*-Glu and *O*-Fuc in the EGF-like 2 domain (**Fig. 4C**, **Table 1**, and **Supplementary information**). These EGF-like 2 glycans would not be attached to Ser/Thr within the known *O*-glucosylation $C^1XSX(P/A)C^2$ [24, 128] or *O*-fucosylation $C^2XXXX(S/T)C^3$ [88, 129, 130] consensus sequences (the superscript indicates the Cys residue within the EGF-like domain). However, recent studies uncovered new *O*-glucosyltransferases that can *O*-glucosylate Ser residues between the 3rd and 4th Cys of EGF-like domains [131], precisely where we map the new EGF-like 2 Hex_1Pent_2 *O*-glycan in rFIX. In fact, human Coagulation Factor X is *O*-glucosylated at S^{106} in the EGF-like 2 domain, also between the 3rd and 4th Cys residues and in a position equivalent to the S^{102} residue in rFIX [132]. These *O*-glycans could be attached by yet

uncharacterized *O*-glycosyltransferases or by known enzymes with different substrate and/or sugar donor specificity able to glycosylate with low efficiency a wider range of substrates, as shown for the POGLUT1 *O*-glucosyltransferase, responsible for *O*-glucosylation at $C^1XSX(P/A)C^2$ [24, 128, 133]. Alternatively, *O*-glycans at EGF-like 2 domain may negatively impact rFIX biosynthesis, leading to low occupancy of these sites in the secreted molecule. The *O*-glucosylation and *O*-fucosylation of EGF-like domains are important for EGF domain stability and the interaction with binding partners [87, 92, 99], which supports their presence also in the EGF-like 2 domain of rFIX. Altogether, our results demonstrate that a thorough analysis of the composition and abundance of all measurable PTMs in rFIX is required to confidently assess the impact of optimization procedures and for quality control purposes.

One of the highlights of this study was the development of DIA-MS workflows to measure rFIX yield and PTMs in bioreactor supernatant (**Supplementary Figure S1**). Measuring product quantity and quality in the bioreactor is important because purification of rFIX enriches for γ-carboxylated rFIX, thereby altering the relative abundance of γ-carboxylated rFIX, and potentially also of other PTMs, in the samples. Post-purification analysis may not always be representative of the metabolic and biosynthetic capabilities of the expression system and bioprocess parameters utilized. We show in **Fig. 7** a clear example of how post-purification analysis did not correlate with bioreactor performance: the relative abundance of rFIX in bioreactor H2 was significantly higher than in bioreactor H1 at the end of the fed batch (**Fig. 1B** and **Fig. 7**), but there was more purified H1 rFIX relative to HCPs than purified H2 rFIX (**Fig. 1C** and **Fig. 7**). This is likely because H2 rFIX was also significantly less γ-carboxylated than H1 rFIX (**Fig. 3** and **Supplementary Fig. S4**), and because there was a larger variety of contaminating HCPs in the H2 supernatant compared to H1 (**Supplementary Fig. S7** and **Supplementary Table S6**). If purification efficiency was the only measure of rFIX yield and quality available, we could only conclude that H1 bioreactor conditions allowed for the production of more γ-carboxylated rFIX, but we would have no understanding of how the feeds impacted rFIX biosynthesis as a whole. Even more, this conclusion would be based on the false assumption that only highly γ-carboxylated species were purified with POROS 50 HQ (**Fig. 3**). However, we showed here that a significant proportion of purified H2 rFIX was not highly γ-carboxylated (**Fig. 3**). Without data from the bioreactor supernatant describing rFIX yield, HCP complexity, and the critically detailed site-

specific analysis of γ-carboxylation and PTMs in rFIX, our ability to optimize bioreactor operation conditions would be limited and biased.

As the world population increases and ages, there is a greater need to lower the cost of biotherapeutic production. Medical and biotechnological advances have also led to an increase in the number of biotherapeutics with numerous diverse PTMs in the production pipeline. Many PTMs require complex biosynthetic pathways, leading to lower yields. The biopharmaceutical industry faces the challenge of overcoming biosynthetic bottlenecks while optimizing expression systems and bioprocessing parameters to make more of a better product, under time pressure. The DIA-MS workflows developed here are specifically designed to assist in the assessment of bioreactor performance to accelerate identification of factors impacting biosynthetic capability and capacity. These workflows are versatile and can be modified to suit the monitoring of any biologic of choice for optimization and quality control purposes.

Acknowledgements

We thank Dr Amanda Nouwens and Peter Josh from the Mass Spectrometry Facility at the School of Chemistry and Molecular Biosciences of The University of Queensland for their help and advice. This project utilized the scientific and technical assistance of the National Biologics Facility (NBF) at the University of Queensland. NBF is supported by Therapeutic Innovation Australia (TIA). TIA is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program. LFZ was funded by a Promoting Women Fellowship from the University of Queensland, Australia. BLS was funded by an Australian National Health and Medical Research Council RD Wright Biomedical (CDF Level 2) Fellowship APP1087975. This work was funded by an Australian Research Council Discovery Project DP160102766 to BLS and an Australian Research Council Industrial Transformation Training Centre IC160100027 to SMM, BLS, CBH, VS, and YYL.

Contributions

LFZ, DRR, CLP, MN, CA, YYL, BLS, and CBH designed the study. LFZ, DRC, and TKP performed the experiments. LFZ, DRC, and BLS wrote the manuscript. All authors edited and approved the final manuscript. LFZ, VS, SMM YYL, BLS, and CBH obtained the funding.

References

1. Macfarlane, R.G., AN ENZYME CASCADE IN THE BLOOD CLOTTING MECHANISM, AND ITS FUNCTION AS A BIOCHEMICAL AMPLIFIER, *Nature*, 1964. 202: 498-9.
2. Muntean, W. and Leschnik, B., Factor VIII influences binding of factor IX and factor X to intact human platelets, *Thromb Res*, 1989. 55(5): 537-48.
3. Huang, M., Rigby, A.C., Morelli, X., Grant, M.A., Huang, G., Furie, B., Seaton, B., and Furie, B.C., Structural basis of membrane binding by Gla domains of vitamin K-dependent proteins, *Nat Struct Biol*, 2003. 10(9): 751-6.
4. Nelsestuen, G.L., Role of gamma-carboxyglutamic acid. An unusual protein transition required for the calcium-dependent binding of prothrombin to phospholipid, *J Biol Chem*, 1976. 251(18): 5648-56.
5. Zogg, T. and Brandstetter, H., Structural basis of the cofactor- and substrate-assisted activation of human coagulation factor IXa, *Structure*, 2009. 17(12): 1669-1678.
6. Goodeve, A.C., Hemophilia B: molecular pathogenesis and mutation analysis, *J Thromb Haemost*, 2015. 13(7): 1184-95.
7. Orlova, N.A., Kovnir, S.V., Vorobiev, II, and Gabibov, A.G., Coagulation Factor IX for Hemophilia B Therapy, *Acta Naturae*, 2012. 4(2): 62-73.
8. Peters, R. and Harris, T., Advances and innovations in haemophilia treatment, *Nat Rev Drug Discov*, 2018. 17(7): 493-508.
9. Bristol, J.A., Freedman, S.J., Furie, B.C., and Furie, B., Profactor IX: the propeptide inhibits binding to membrane surfaces and activation by factor XIa, *Biochemistry*, 1994. 33(47): 14136-43.
10. Kumar, S.R., Industrial production of clotting factors: Challenges of expression, and choice of host cells, *Biotechnol J*, 2015. 10(7): 995-1004.
11. Kovnir, S.V., Orlova, N.A., Shakhparonov, M.I., Skryabin, K.G., Gabibov, A.G., and Vorobiev, II, A Highly Productive CHO Cell Line Secreting Human Blood Clotting Factor IX, *Acta Naturae*, 2018. 10(1): 51-65.
12. Wajih, N., Hutson, S.M., Owen, J., and Wallin, R., Increased production of functional recombinant human clotting factor IX by baby hamster kidney cells engineered to overexpress VKORC1, the vitamin K 2,3-epoxide-reducing enzyme of the vitamin K cycle, *J Biol Chem*, 2005. 280(36): 31603-7.
13. Pegg, C.L., Zacchi, L.F., Roche Recinos, D., Howard, C.B., and Schulz, B.L., Identification of novel glycosylation events on human serum-derived Factor IX, *bioRxiv*, 2019.

14. Hansson, K. and Stenflo, J., Post-translational modifications in proteins involved in blood coagulation, *J Thromb Haemost*, 2005. 3(12): 2633-48.
15. Diuguid, D.L., Rabiet, M.J., Furie, B.C., Liebman, H.A., and Furie, B., Molecular basis of hemophilia B: a defective enzyme due to an unprocessed propeptide is caused by a point mutation in the factor IX precursor, *Proc Natl Acad Sci U S A*, 1986. 83(16): 5803-7.
16. Wasley, L.C., Rehemtulla, A., Bristol, J.A., and Kaufman, R.J., PACE/furin can process the vitamin K-dependent pro-factor IX precursor within the secretory pathway, *J Biol Chem*, 1993. 268(12): 8458-65.
17. Thomas, G., Furin at the cutting edge: from protein traffic to embryogenesis and disease, *Nat Rev Mol Cell Biol*, 2002. 3(10): 753-66.
18. Di Scipio, R.G., Kurachi, K., and Davie, E.W., Activation of human factor IX (Christmas factor), *J Clin Invest*, 1978. 61(6): 1528-38.
19. Katayama, K., Ericsson, L.H., Enfield, D.L., Walsh, K.A., Neurath, H., Davie, E.W., and Titani, K., Comparison of amino acid sequence of bovine coagulation Factor IX (Christmas Factor) with that of other vitamin K-dependent plasma proteins, *Proc Natl Acad Sci U S A*, 1979. 76(10): 4990-4.
20. Furie, B. and Furie, B.C., The molecular basis of blood coagulation, *Cell*, 1988. 53(4): 505-18.
21. Bond, M., Jankowski, M., Patel, H., Karnik, S., Strang, A., Xu, B., Rouse, J., Koza, S., Letwin, B., Steckert, J., Amphlett, G., and Scoble, H., Biochemical characterization of recombinant factor IX, *Semin Hematol*, 1998. 35(2 Suppl 2): 11-7.
22. Carlisle, T.L. and Suttie, J.W., Vitamin K dependent carboxylase: subcellular location of the carboxylase and enzymes involved in vitamin K metabolism in rat liver, *Biochemistry*, 1980. 19(6): 1161-7.
23. Sadowski, J.A., Esmon, C.T., and Suttie, J.W., Vitamin K-dependent carboxylase. Requirements of the rat liver microsomal enzyme system, *J Biol Chem*, 1976. 251(9): 2770-6.
24. Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., and Iwanaga, S., Identification of a disaccharide (Xyl-Glc) and a trisaccharide (Xyl2-Glc) O-glycosidically linked to a serine residue in the first epidermal growth factor-like domain of human factors VII and IX and protein Z and bovine protein Z, *J Biol Chem*, 1989. 264(34): 20320-5.
25. Huang, L.J., Lin, J.H., Tsai, J.H., Chu, Y.Y., Chen, Y.W., Chen, S.L., and Chen, S.H., Identification of protein O-glycosylation site and corresponding glycans using liquid chromatography-tandem mass spectrometry via mapping accurate mass and retention time shift, *J Chromatogr A*, 2014. 1371: 136-45.
26. Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., and Iwanaga, S., Human factor IX has a tetrasaccharide O-glycosidically linked to serine 61 through the fucose residue, *J Biol Chem*, 1992. 267(25): 17520-5.
27. McMullen, B.A., Fujikawa, K., and Kisiel, W., The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood coagulation zymogens, *Biochem Biophys Res Commun*, 1983. 115(1): 8-14.
28. Fernlund, P. and Stenflo, J., Beta-hydroxyaspartic acid in vitamin K-dependent proteins, *J Biol Chem*, 1983. 258(20): 12509-12.

29. Harris, R.J., Papac, D.I., Truong, L., and Smith, K.J. *Partial phosphorylation of serine-68 in EGF-1 of human factor IX*. in *Proceedings of XIth international conference on methods in protein structure analysis*. 1996. Annecy.
30. King, S.L., Joshi, H.J., Schjoldager, K.T., Halim, A., Madsen, T.D., Dziegiel, M.H., Woetmann, A., Vakhrushev, S.Y., and Wandall, H.H., Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells, *Blood Adv*, 2017. 1(7): 429-442.
31. Makino, Y., Omichi, K., Kuraya, N., Ogawa, H., Nishimura, H., Iwanaga, S., and Hase, S., Structural analysis of N-linked sugar chains of human blood clotting factor IX, *J Biochem*, 2000. 128(2): 175-80.
32. Gil, G.C., Velander, W.H., and Van Cott, K.E., Analysis of the N-glycans of recombinant human Factor IX purified from transgenic pig milk, *Glycobiology*, 2008. 18(7): 526-39.
33. Kurachi, K. and Davie, E.W., Isolation and characterization of a cDNA coding for human factor IX, *Proc Natl Acad Sci U S A*, 1982. 79(21): 6461-4.
34. Chevreux, G., Faid, V., Andre, M.H., Tellier, Z., and Bihoreau, N., Differential investigations from plasma-derived and recombinant Factor IX revealed major differences in post-translational modifications of activation peptides, *Vox Sang*, 2013. 104(2): 171-4.
35. Agarwala, K.L., Kawabata, S., Takao, T., Murata, H., Shimonishi, Y., Nishimura, H., and Iwanaga, S., Activation peptide of human factor IX has oligosaccharides O-glycosidically linked to threonine residues at 159 and 169, *Biochemistry*, 1994. 33(17): 5167-71.
36. Seo, Y., Park, G.M., Oh, M.J., and An, H.J., Investigation of O-glycosylation heterogeneity of recombinant coagulation factor IX using LC-MS/MS, *Bioanalysis*, 2017. 9(18): 1361-1372.
37. Yehuda, S. and Padler-Karavani, V., Glycosylated Biotherapeutics: Immunological Effects of N-Glycolylneuraminic Acid, *Front Immunol*, 2020. 11: 21.
38. Jefferis, R., Posttranslational Modifications and the Immunogenicity of Biotherapeutics, *J Immunol Res*, 2016. 2016: 5358272.
39. Jenkins, N., Murphy, L., and Tyther, R., Post-translational modifications of recombinant proteins: significance for biopharmaceuticals, *Mol Biotechnol*, 2008. 39(2): 113-8.
40. Lubiniecki, A., Volkin, D.B., Federici, M., Bond, M.D., Nedved, M.L., Hendricks, L., Mehndiratta, P., Bruner, M., Burman, S., Dalmonte, P., Kline, J., Ni, A., Panek, M.E., Pikounis, B., Powers, G., Vafa, O., and Siegel, R., Comparability assessments of process and product changes made during development of two different monoclonal antibodies, *Biologicals*, 2011. 39(1): 9-22.
41. Liu, H., Nowak, C., and Patel, R., Modifications of recombinant monoclonal antibodies in vivo, *Biologicals*, 2019. 59: 1-5.
42. Yan, S.C., Razzano, P., Chao, Y.B., Walls, J.D., Berg, D.T., McClure, D.B., and Grinnell, B.W., Characterization and novel purification of recombinant human protein C from three mammalian cell lines, *Biotechnology (N Y)*, 1990. 8(7): 655-61.
43. Charlton, A., Napoli, M., Smrdelj, P., Stowers, A., Pirzas, V., and Schroeder, M., *Method of purifying polypeptides*, W.I.P. Organization, Editor. 2011.
44. Rathore, D., Faustino, A., Schiel, J., Pang, E., Boyne, M., and Rogstad, S., The role of mass spectrometry in the characterization of biologic protein products, *Expert Rev Proteomics*, 2018. 15(5): 431-449.

45. Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol Cell Proteomics*, 2012. 11(6): O111.016717.
46. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial, *Mol Syst Biol*, 2018. 14(8): e8126.
47. Kerr, E.D., Caboche, C.H., and Schulz, B.L., Post-translational modifications drive protein stability to control the dynamic beer brewing proteome, *Mol Cell Proteomics*, 2019.
48. Zacchi, L.F. and Schulz, B.L., SWATH-MS Glycoproteomics Reveals Consequences of Defects in the Glycosylation Machinery, *Mol Cell Proteomics*, 2016. 15(7): 2435-47.
49. Xu, Y., Bailey, U.M., and Schulz, B.L., Automated measurement of site-specific N-glycosylation occupancy with SWATH-MS, *Proteomics*, 2015. 15(13): 2177-86.
50. Zacchi, L.F. and Schulz, B.L., Data-Independent Acquisition for Yeast Glycoproteomics, *Methods Mol Biol*, 2019. 2049: 191-202.
51. Phung, T.K., Zacchi, L.F., and Schulz, B.L., DIALib: an automated ion library generator for data independent acquisition mass spectrometry analysis of peptides and glycopeptides, *Mol Omics*, 2020.
52. Keller, A., Bader, S.L., Kusebauch, U., Shteynberg, D., Hood, L., and Moritz, R.L., Opening a SWATH Window on Posttranslational Modifications: Automated Pursuit of Modified Peptides, *Mol Cell Proteomics*, 2016. 15(3): 1151-63.
53. Fan, L., Frye, C.C., and Racher, A.J., The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes, *Pharm. Bioprocess.*, 2013. 1(5): 487–502.
54. Zacchi, L.F., Roche Recinos, D., Otte, E., Aitken, C., Hunt, T., Sandford, V., Lee, Y.Y., Schulz, B.L., and Howard, C.B., S-Trap eliminates cell culture media polymeric surfactants for effective proteomic analysis of mammalian cell bioreactor supernatants, *J Proteome Res*, 2020.
55. Enculescu, C., Kerr, E.D., Yeo, K.Y.B., Schenk, G., Fortes, M.R.S., and Schulz, B.L., Proteomics Reveals Profound Metabolic Changes in the Alcohol Use Disorder Brain, *ACS Chem Neurosci*, 2019. 10(5): 2364-2373.
56. Cairns, J.R., Williamson, M.K., and Price, P.A., Direct identification of gamma-carboxyglutamic acid in the sequencing of vitamin K-dependent proteins, *Anal Biochem*, 1991. 199(1): 93-7.
57. Hallgren, K.W., Zhang, D., Kinter, M., Willard, B., and Berkner, K.L., Methylation of gamma-carboxylated Glu (Gla) allows detection by liquid chromatography-mass spectrometry and the identification of Gla residues in the gamma-glutamyl carboxylase, *J Proteome Res*, 2013. 12(6): 2365-74.
58. Nguyen, L.T., Zacchi, L.F., Schulz, B.L., Moore, S.S., and Fortes, M.R.S., Adipose tissue proteomic analyses to study puberty in Brahman heifers, *J Anim Sci*, 2018. 96(6): 2392-2398.
59. Fortes, M.R.S., Zacchi, L.F., Nguyen, L.T., Raidan, F., Weller, M., Choo, J.J.Y., Reverter, A., Rego, J.P.A., Boe-Hansen, G.B., Porto-Neto, L.R., Lehnert, S.A., Canovas, A., Schulz, B.L., Islas-Trejo, A., Medrano, J.F., Thomas, M.G., and Moore, S.S., Pre- and

post-puberty expression of genes and proteins in the uterus of Bos indicus heifers: the luteal phase effect post-puberty, *Anim Genet*, 2018. 49(6): 539-549.

60. Anson, D.S., Choo, K.H., Rees, D.J., Giannelli, F., Gould, K., Huddleston, J.A., and Brownlee, G.G., The gene structure of human anti-haemophilic factor IX, *Embo j*, 1984. 3(5): 1053-60.
61. McGraw, R.A., Davis, L.M., Noyes, C.M., Lundblad, R.L., Roberts, H.R., Graham, J.B., and Stafford, D.W., Evidence for a prevalent dimorphism in the activation peptide of human coagulation factor IX, *Proc Natl Acad Sci U S A*, 1985. 82(9): 2847-51.
62. North, S.J., Huang, H.H., Sundaram, S., Jang-Lee, J., Etienne, A.T., Trollope, A., Chalabi, S., Dell, A., Stanley, P., and Haslam, S.M., Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity, *J Biol Chem*, 2010. 285(8): 5759-75.
63. Kerr, E.D., Phung, T.K., Caboche, C.H., Fox, G.P., Platz, G.J., and Schulz, B.L., The intrinsic and regulated proteomes of barley seeds in response to fungal infection, *Anal Biochem*, 2019. 580: 30-35.
64. Deutsch, E.W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D.S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R.L., Carver, J.J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H., and Vizcaino, J.A., The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition, *Nucleic Acids Res*, 2017. 45(D1): D1100-d1106.
65. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Perez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A., and Vizcaino, J.A., The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res*, 2019. 47(D1): D442-d450.
66. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, *Bioinformatics*, 2009. 25(8): 1091-3.
67. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res*, 2003. 13(11): 2498-504.
68. Larson, P.J., Stanfield-Oakley, S.A., VanDusen, W.J., Kasper, C.K., Smith, K.J., Monroe, D.M., and High, K.A., Structural integrity of the gamma-carboxyglutamic acid domain of human blood coagulation factor IXa Is required for its binding to cofactor VIIIa, *J Biol Chem*, 1996. 271(7): 3869-76.
69. Gillis, S., Furie, B.C., Furie, B., Patel, H., Huberty, M.C., Switzer, M., Foster, W.B., Scoble, H.A., and Bond, M.D., gamma-Carboxyglutamic acids 36 and 40 do not contribute to human factor IX function, *Protein Sci*, 1997. 6(1): 185-96.
70. Kaufman, R.J., Wasley, L.C., Furie, B.C., Furie, B., and Shoemaker, C.B., Expression, purification, and characterization of recombinant gamma-carboxylated factor IX synthesized in Chinese hamster ovary cells, *J Biol Chem*, 1986. 261(21): 9622-8.
71. Kelleher, N.L., Zubarev, R.A., Bush, K., Furie, B., Furie, B.C., McLafferty, F.W., and Walsh, C.T., Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid, *Anal Chem*, 1999. 71(19): 4250-3.

72. Ramstrom, M. and Sandberg, H., Characterization of gamma-carboxylated tryptic peptides by collision-induced dissociation and electron transfer dissociation mass spectrometry, *Eur J Mass Spectrom (Chichester)*, 2011. 17(5): 497-506.
73. Charlton, A., Napoli, M., Smrdelj, P., Stowers, A., and Schroder, M., *Method of purifiying polypeptides*, W.I.P. Organization, Editor. 2010: United States.
74. Aghamohseni, H., Ohadi, K., Spearman, M., Krahn, N., Moo-Young, M., Scharer, J.M., Butler, M., and Budman, H.M., Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody, *J Biotechnol*, 2014. 186: 98-109.
75. Butler, M., Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals, *Appl Microbiol Biotechnol*, 2005. 68(3): 283-91.
76. Butler, M. and Meneses-Acosta, A., Recent advances in technology supporting biopharmaceutical production from mammalian cells, *Appl Microbiol Biotechnol*, 2012. 96(4): 885-94.
77. Sunley, K., Tharmalingam, T., and Butler, M., CHO cells adapted to hypothermic growth produce high yields of recombinant beta-interferon, *Biotechnol Prog*, 2008. 24(4): 898-906.
78. Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N., Katinger, H., and Muller, D., Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors, *Biotechnol Bioeng*, 2006. 94(6): 1033-44.
79. Enjolras, N., Dargaud, Y., Perot, E., Guillaume, F., Becchi, M., and Negrier, C., Human hepatoma cell line HuH-7 is an effective cellular system to produce recombinant factor IX with improved post-translational modifications, *Thromb Res*, 2012. 130(5): e266-73.
80. Jorgensen, M.J., Cantor, A.B., Furie, B.C., Brown, C.L., Shoemaker, C.B., and Furie, B., Recognition site directing vitamin K-dependent gamma-carboxylation resides on the propeptide of factor IX, *Cell*, 1987. 48(2): 185-91.
81. Bentley, A.K., Rees, D.J., Rizza, C., and Brownlee, G.G., Defective propeptide processing of blood clotting factor IX caused by mutation of arginine to glutamine at position -4, *Cell*, 1986. 45(3): 343-8.
82. Furie, B. and Furie, B.C., Molecular basis of gamma-carboxylation. Role of the propeptide in the vitamin K-dependent proteins, *Ann N Y Acad Sci*, 1991. 614: 1-10.
83. Chen, G., Zhang, Y., Trinidad, J.C., and Dann, C., 3rd, Distinguishing Sulfotyrosine Containing Peptides from their Phosphotyrosine Counterparts Using Mass Spectrometry, *J Am Soc Mass Spectrom*, 2018. 29(3): 455-462.
84. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M., Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode, *Anal Chem*, 2001. 73(7): 1440-8.
85. Edelson-Averbukh, M., Shevchenko, A., Pipkorn, R., and Lehmann, W.D., Discrimination between peptide O-sulfo- and O-phosphotyrosine residues by negative ion mode electrospray tandem mass spectrometry, *J Am Soc Mass Spectrom*, 2011. 22(12): 2256-68.
86. Valliere-Douglass, J.F., Brady, L.J., Farnsworth, C., Pace, D., Balland, A., Wallace, A., Wang, W., Treuheit, M.J., and Yan, B., O-fucosylation of an antibody light chain:

characterization of a modification occurring on an IgG1 molecule, *Glycobiology*, 2009. 19(2): 144-52.

87. Schneider, M., Al-Shareffi, E., and Haltiwanger, R.S., Biological functions of fucose in mammals, *Glycobiology*, 2017. 27(7): 601-618.
88. Wang, Y., Lee, G.F., Kelley, R.F., and Spellman, M.W., Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains, *Glycobiology*, 1996. 6(8): 837-42.
89. Rana, N.A., Nita-Lazar, A., Takeuchi, H., Kakuda, S., Luther, K.B., and Haltiwanger, R.S., O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1, *J Biol Chem*, 2011. 286(36): 31623-37.
90. Derian, C.K., VanDusen, W., Przysiecki, C.T., Walsh, P.N., Berkner, K.L., Kaufman, R.J., and Friedman, P.A., Inhibitors of 2-ketoglutarate-dependent dioxygenases block aspartyl beta-hydroxylation of recombinant human factor IX in several mammalian expression systems, *J Biol Chem*, 1989. 264(12): 6615-8.
91. Rabiet, M.J., Jorgensen, M.J., Furie, B., and Furie, B.C., Effect of propeptide mutations on post-translational processing of factor IX. Evidence that beta-hydroxylation and gamma-carboxylation are independent events, *J Biol Chem*, 1987. 262(31): 14895-8.
92. Yu, H. and Takeuchi, H., Protein O-glucosylation: another essential role of glucose in biology, *Curr Opin Struct Biol*, 2019. 56: 64-71.
93. Kuraya, N., Omichi, K., Nishimura, H., Iwanaga, S., and Hase, S., Structural analysis of O-linked sugar chains in human blood clotting factor IX, *J Biochem*, 1993. 114(6): 763-5.
94. Bolt, G., Bjelke, J.R., Hermit, M.B., Hansen, L., Karpf, D.M., and Kristensen, C., Hyperglycosylation prolongs the circulation of coagulation factor IX, *J Thromb Haemost*, 2012. 10(11): 2397-8.
95. Bolt, G., Kristensen, C., and Steenstrup, T.D., More than one intracellular processing bottleneck delays the secretion of coagulation factor VII, *Thromb Haemost*, 2008. 100(2): 204-10.
96. Bolt, G., Steenstrup, T.D., and Kristensen, C., All post-translational modifications except propeptide cleavage are required for optimal secretion of coagulation factor VII, *Thromb Haemost*, 2007. 98(5): 988-97.
97. Sinha, U. and Wolf, D.L., Carbohydrate residues modulate the activation of coagulation factor X, *J Biol Chem*, 1993. 268(5): 3048-51.
98. Walsh, G. and Jefferis, R., Post-translational modifications in the context of therapeutic proteins, *Nat Biotechnol*, 2006. 24(10): 1241-52.
99. Holdener, B.C. and Haltiwanger, R.S., Protein O-fucosylation: structure and function, *Curr Opin Struct Biol*, 2019. 56: 78-86.
100. Hansen, K., Szarka, S., Escoffier, E., Berthet, A., Venet, J., Collet-Brose, J., Hepburn, S., Wright, M., Wheller, R., Nelson, R., and Kay, R.G., Glu-C, an alternative digestive enzyme for the quantitative LC-MS/MS analysis of an IgG-based antibody biotherapeutic, *Bioanalysis*, 2018. 10(13): 997-1007.
101. Tait, A.S., Hogwood, C.E., Smales, C.M., and Bracewell, D.G., Host cell protein dynamics in the supernatant of a mAb producing CHO cell line, *Biotechnol Bioeng*, 2012. 109(4): 971-82.
102. Hogwood, C.E., Bracewell, D.G., and Smales, C.M., Host cell protein dynamics in recombinant CHO cells: impacts from harvest to purification and beyond, *Bioengineered*, 2013. 4(5): 288-91.

103. Hogwood, C.E., Tait, A.S., Koloteva-Levine, N., Bracewell, D.G., and Smales, C.M., The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process, *Biotechnol Bioeng*, 2013. 110(1): 240-51.
104. Goey, C.H., Alhuthali, S., and Kontoravdi, C., Host cell protein removal from biopharmaceutical preparations: Towards the implementation of quality by design, *Biotechnol Adv*, 2018. 36(4): 1223-1237.
105. Wang, X., Song, X., Zhuo, W., Fu, Y., Shi, H., Liang, Y., Tong, M., Chang, G., and Luo, Y., The regulatory mechanism of Hsp90alpha secretion and its function in tumor malignancy, *Proc Natl Acad Sci U S A*, 2009. 106(50): 21288-93.
106. Mullen, L., Hanschmann, E.M., Lillig, C.H., Herzenberg, L.A., and Ghezzi, P., Cysteine Oxidation Targets Peroxiredoxins 1 and 2 for Exosomal Release through a Novel Mechanism of Redox-Dependent Secretion, *Mol Med*, 2015. 21: 98-108.
107. Rubartelli, A., Bajetto, A., Allavena, G., Wollman, E., and Sitia, R., Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway, *J Biol Chem*, 1992. 267(34): 24161-4.
108. Gilgunn, S., El-Sabbahy, H., Albrecht, S., Gaikwad, M., Corrigan, K., Deakin, L., Jellum, G., and Bones, J., Identification and tracking of problematic host cell proteins removed by a synthetic, highly functionalized nonwoven media in downstream bioprocessing of monoclonal antibodies, *J Chromatogr A*, 2019. 1595: 28-38.
109. de Castilho Fernandes, A., Fontes, A., Gonsales, N., Swiech, K., Picanco-Castro, V., Faca, S., and Covas, D., Stable and high-level production of recombinant Factor IX in human hepatic cell line, *Biotechnol Appl Biochem*, 2011. 58(4): 243-9.
110. de Sousa Bomfim, A., Cristina Correa de Freitas, M., Picanco-Castro, V., de Abreu Soares Neto, M., Swiech, K., Tadeu Covas, D., and Maria de Sousa Russo, E., Human cell lines: A promising alternative for recombinant FIX production, *Protein Expr Purif*, 2016. 121: 149-56.
111. do Amaral, R.L., de Sousa Bomfim, A., de Abreu-Neto, M.S., Picanco-Castro, V., de Sousa Russo, E.M., Covas, D.T., and Swiech, K., Approaches for recombinant human factor IX production in serum-free suspension cultures, *Biotechnol Lett*, 2016. 38(3): 385-94.
112. Hallgren, K.W., Hommema, E.L., McNally, B.A., and Berkner, K.L., Carboxylase overexpression effects full carboxylation but poor release and secretion of factor IX: implications for the release of vitamin K-dependent proteins, *Biochemistry*, 2002. 41(50): 15045-55.
113. Perot, E., Enjolras, N., Le Quellec, S., Indalecio, A., Girard, J., Negrier, C., and Dargaud, Y., Expression and characterization of a novel human recombinant factor IX molecule with enhanced in vitro and in vivo clotting activity, *Thromb Res*, 2015. 135(5): 1017-24.
114. Bohm, E., Seyfried, B.K., Dockal, M., Graninger, M., Hasslacher, M., Neurath, M., Konetschny, C., Matthiessen, P., Mitterer, A., and Scheiflinger, F., Differences in N-glycosylation of recombinant human coagulation factor VII derived from BHK, CHO, and HEK293 cells, *BMC Biotechnol*, 2015. 15: 87.
115. Pereira, S., Kildegaard, H.F., and Andersen, M.R., Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients, *Biotechnol J*, 2018. 13(3): e1700499.

116. Brooks, A.R., Sim, D., Gritzan, U., Patel, C., Blasko, E., Feldman, R.I., Tang, L., Ho, E., Zhao, X.Y., Apeler, H., and Murphy, J.E., Glycoengineered factor IX variants with improved pharmacokinetics and subcutaneous efficacy, *J Thromb Haemost*, 2013. 11(9): 1699-706.
117. Begbie, M.E., Mamdani, A., Gataiance, S., Eltringham-Smith, L.J., Bhakta, V., Hortelano, G., and Sheffield, W.P., An important role for the activation peptide domain in controlling factor IX levels in the blood of haemophilia B mice, *Thromb Haemost*, 2005. 94(6): 1138-47.
118. Griffith, M.J., Monroe, D.M., K.E., v.C., Walker, A., Waugh, S., Kumar, A., and Drohan, W.N., N-glycan sialylation is important for in vivo recovery of recombinant factor IX., *J Thromb Haemost*, 2007. 5 (Suppl. 2): P-M-043.
119. Chang, R.Y., Brock, J., Griffith, M.J., and Monroe, D.M., Glycosylation of the activation peptide of FIX determines plasma half-life., *J Thromb Haemost* 2007. 5 (Suppl. 2): O-M-088.
120. Soriano-Garcia, M., Park, C.H., Tulinsky, A., Ravichandran, K.G., and Skrzypczak-Jankun, E., Structure of Ca2+ prothrombin fragment 1 including the conformation of the Gla domain, *Biochemistry*, 1989. 28(17): 6805-10.
121. Baeuerle, P.A. and Huttner, W.B., Tyrosine sulfation is a trans-Golgi-specific protein modification, *J Cell Biol*, 1987. 105(6 Pt 1): 2655-64.
122. Farhan, H., Tyrosine kinase signaling in and on the endoplasmic reticulum, *Biochem Soc Trans*, 2020. 48(1): 199-205.
123. Perdivara, I., Deterding, L.J., Przybylski, M., and Tomer, K.B., Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification?, *J Am Soc Mass Spectrom*, 2010. 21(7): 1114-7.
124. Morand, K., Talbo, G., and Mann, M., Oxidation of peptides during electrospray ionization, *Rapid Commun Mass Spectrom*, 1993. 7(8): 738-43.
125. Auclair, J.R., Salisbury, J.P., Johnson, J.L., Petsko, G.A., Ringe, D., Bosco, D.A., Agar, N.Y., Santagata, S., Durham, H.D., and Agar, J.N., Artifacts to avoid while taking advantage of top-down mass spectrometry based detection of protein S-thiolation, *Proteomics*, 2014. 14(10): 1152-7.
126. Raftery, M.J., Determination of oxidative protein modifications using mass spectrometry, *Redox Rep*, 2014. 19(4): 140-7.
127. Ramallo Guevara, C., Philipp, O., Hamann, A., Werner, A., Osiewacz, H.D., Rexroth, S., Rogner, M., and Poetsch, A., Global Protein Oxidation Profiling Suggests Efficient Mitochondrial Proteome Homeostasis During Aging, *Mol Cell Proteomics*, 2016. 15(5): 1692-709.
128. Takeuchi, H., Kantharia, J., Sethi, M.K., Bakker, H., and Haltiwanger, R.S., Site-specific O-glucosylation of the epidermal growth factor-like (EGF) repeats of notch: efficiency of glycosylation is affected by proper folding and amino acid sequence of individual EGF repeats, *J Biol Chem*, 2012. 287(41): 33934-44.
129. Shao, L., Moloney, D.J., and Haltiwanger, R., Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the Abruptex region, *J Biol Chem*, 2003. 278(10): 7775-82.
130. Harris, R.J. and Spellman, M.W., O-linked fucose and other post-translational modifications unique to EGF modules, *Glycobiology*, 1993. 3(3): 219-24.

131. Takeuchi, H., Schneider, M., Williamson, D.B., Ito, A., Takeuchi, M., Handford, P.A., and Haltiwanger, R.S., Two novel protein O-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling, *Proc Natl Acad Sci U S A*, 2018. 115(36): E8395-e8402.
132. Chevreux, G., Tilly, N., Faid, V., and Bihoreau, N., Mass spectrometry based analysis of human plasma-derived factor X revealed novel post-translational modifications, *Protein Sci*, 2015. 24(10): 1640-8.
133. Li, Z., Fischer, M., Satkunarajah, M., Zhou, D., Withers, S.G., and Rini, J.M., Structural basis of Notch O-glucosylation and O-xylosylation by mammalian protein-O-glucosyltransferase 1 (POGLUT1), *Nat Commun*, 2017. 8(1): 185.

Supplementary Material

Supplementary Tables

S1: Output of the Byonic searches for rFIX characterization of PTMs using Trypsin, GluC, AspN, or Chymotrypsin, and +/- PNGase F.
S2: Glycan databases for searching in Byonic: *N*-glycan database 50 and 57 common biantennary glycans from Byonic combined with 14 additional *N*-glycoforms with LacNAc extensions, and *O*-glycan database with 11 *O*-glycans.
S3: ProteinPilot output of combined searches for peptides and proteins in the fractionated and unfractionated supernatant and purified samples from bioreactors H1 and H2.
S4: Ion library used to measure the abundance of rFIX and host-cell proteins during bioreactor operation and in the purified material.
S5: PeakView output of the quantification of rFIX and host-cell proteins abundance during bioreactor operation and in the purified material. FDR measurements, fragment ion intensity measurements, peptide intensity measurements, protein intensity measurements, protein intensity measurement recalculated based on FDR, and FDR-filtered protein intensity measurement normalized to trypsin.
S6: Comparison of the relative abundance of proteins before (supernatant) and after purification.
S7. Output of the Byonic searches of DDA files of pdFIX and H1 and H2 rFIX digested with trypsin and selected PSMs used for rFIX PTM quantification by DIA-MS.
S8: Ion library used to measure select FIX PTMs abundance during bioreactor operation and in the purified material.
S9: PeakView output of the quantification of FIX PTMs abundance during bioreactor operation and in the purified material. FDR measurements, fragment ion intensity measurements, peptide intensity measurements, protein intensity measurement, and peptide intensity measurement filtered based on FDR.
S10. Output of the Byonic searches of DDA files of methylated H1 and H2 rFIX digested with trypsin, and selected PSMs used for rFIX GLA γ-carboxylation quantification by DIA-MS.
S11: PeakView output of the quantification of methylated rFIX GLA domain. A) Byonic searches for methyl-carboxy rFIX characterization, focused on the GLA domain. B) Ion library used to measure methyl-carboxy rFIX GLA domain during bioreactor operation and in the purified material. C-F) PeakView output of the quantification of methyl-carboxy rFIX abundance during bioreactor operation and in the purified material, using either 1. XIC 6 min or 2. XIC 2 min, C) FDR measurements, D) Fragment ion intensity measurements, E) Peptide intensity measurements, F) Peptide intensity measurement filtered based on FDR. G) Statistical comparisons.
S12: Viability and metabolic data for CHO cells expressing rFIX from bioreactors H1 and H2.
S13: GO and network analyses of the proteins that were differentially abundant in the supernatant of bioreactor H1 *vs* H2 at day 13.
S14: ProteinPilot output of search for peptides and proteins in a purified sample from bioreactor H1.
S15: ProteinPilot output of search for peptides and proteins in a purified sample from bioreactor H2.

S16. CHO host cell proteins co-purifying with rFIX in samples from bioreactors H1 and/or H2 identified by DDA.

Names	Entry	Protein names	Organism	Calcium Binding	Subcellular location
	P00740	**Coagulation factor IX**	**Human**	**Yes**	**S**
	G3H8V5	**Carboxypeptidase**	**CHO**	**Yes**	**S**
	G3HHV4	**Thrombospondin-1**	**CHO**	**Yes**	**S**
Common in H1 and H2	*G3I8R9*	*Endoplasmic reticulum chaperone BiP*	*CHO*		*IC(S)*
	G3H354	Heat shock protein HSP 90-alpha	CHO		IC
	G3INX0	*Histone H2B*	*CHO*		*IC*
	G3H6V7	Lipoprotein lipase	CHO	Yes	S
	G3GYP9	*Peroxiredoxin-1*	*CHO*	*Yes*	*IC*
	G3I664	Procollagen C-endopeptidase enhancer 1	CHO		S
	G3IBF4	Serine protease HTRA1	CHO		S
	G3HQL6	Thioredoxin reductase 1, cytoplasmic	CHO		IC
	G3HNJ3	*Clusterin*	*CHO*		*S*
	G3GYG0	Vitamin K-dependent protein S	CHO	Yes	S
Only in H1	G3HQY6	Lipase	CHO		IC
Only in H2	G3HG36	Glutamine synthetase	CHO		IC
	G3HLB3	*Glutamine synthetase*	*CHO*		*IC*
	G3I027	Heat shock protein 75 kDa, mitochondrial	CHO		M
	G3IF52	Nucleobindin-2	CHO	Yes	S
	G3I3H2	60S acidic ribosomal protein P2	CHO		IC
	G3IFL2	*Multifunctional protein ADE2*	*CHO*		*N/F*
	G3HUC4	Sushi, von Willebrand factor type A	CHO	Yes	S
	G3GRV0	Kelch domain-containing protein 4	CHO		N/F
	G3INW9	Histone H2A	CHO		IC
	G3HMD1	Glyceraldehyde-3-phosphate dehydrogenase	CHO		IC
	G3I3U5	Nidogen-1	CHO	Yes	S
	G3HMI3	Semaphorin-3C	CHO		S
	G3HSL4	Elongation factor 2	CHO		IC
	G3HQM6	Endoplasmin	CHO		S
	G3GY17	Cullin-associated NEDD8-dissociated protein 1	CHO		IC
	G3HB04	Protein disulfide-isomerase A6	CHO		IC(S)
	G3H609	Glutathione reductase, mitochondrial	CHO		M
	G3IDS2	F-actin-capping protein subunit alpha	CHO		IC
	G3IH63	Myosin-9	CHO		IC
	G3HBI1	Peroxidasin-like	CHO		S
	G3ID82	Beta-actin-like protein 2	CHO		IC
	G3H0C2	Proteasome subunit alpha type	CHO		IC
	G3HG95	Lamin-A/C	CHO		IC

	G3IBK2	Filamin-A	CHO		IC
	G3I5H3	Elongation factor 1-delta	CHO		IC
	G3HXV5	*Nucleolar phosphoprotein p130*	*CHO*		*IC*
	G3H3Q1	Pyruvate kinase	CHO		IC
	G3GWR8	Proteasome endopeptidase complex	CHO		IC
	G3I486	Heterochromatin protein 1-binding protein 3	CHO		IC
	G3HCW9	PRDX2	CHO		IC
	G3IK13	*Eukaryotic translation initiation factor 3 subunit C eIF3c*	*CHO*		*IC*
	G3GZ90	Calumenin	CHO	Yes	IC(S)
	G3H1W4	Tubulointerstitial nephritis antigen-like	CHO		IC
	G3HSF3	Proteasome subunit alpha type	CHO		IC
	G3I9G7	Proteasome subunit alpha type	CHO		IC

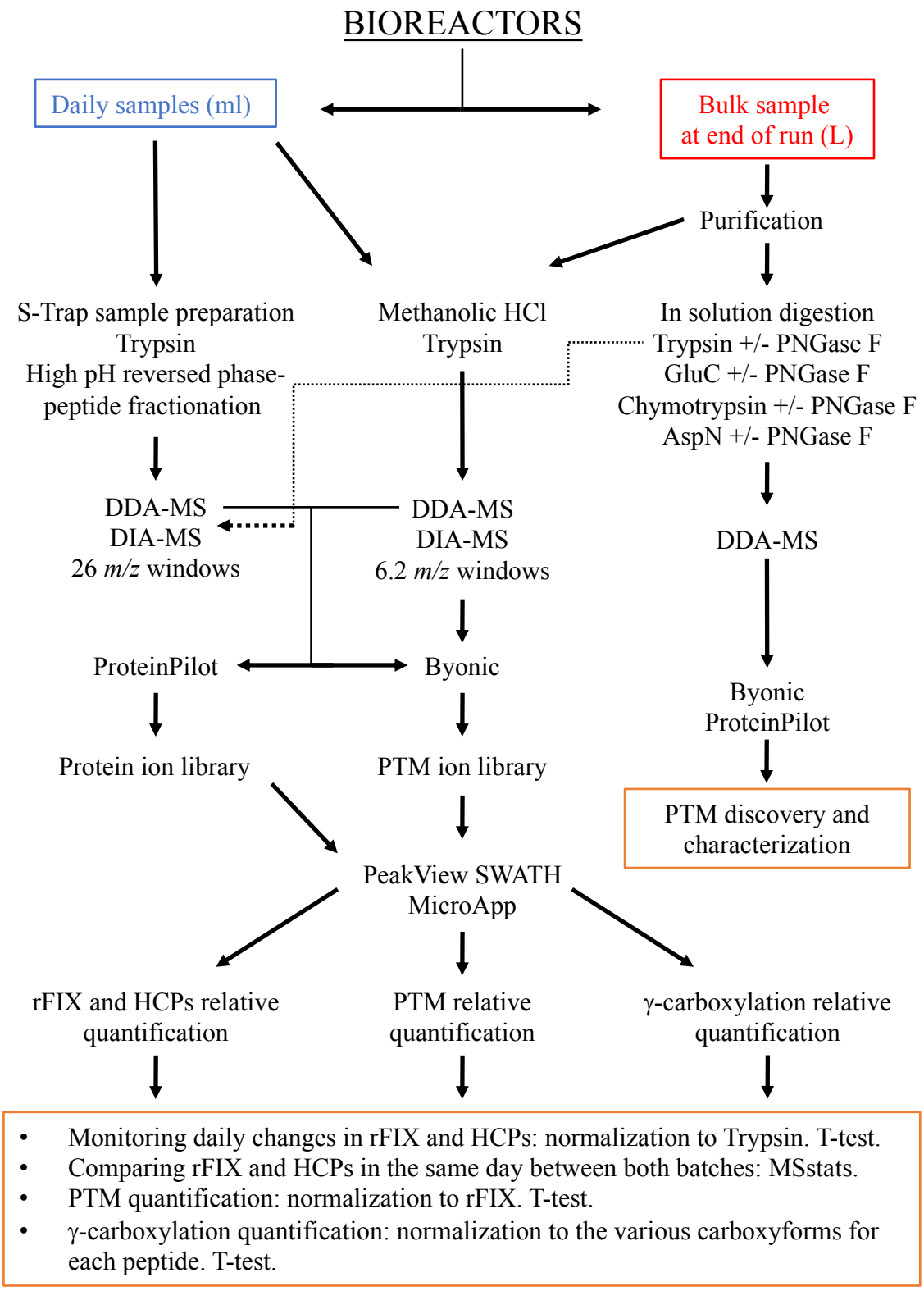
S=secretory; IC=intracellular; IC(S)=intracellular and possibly secreted; M=mitochondrial; N/F not found. Data was obtained from Uniprot. When no information was available on GO or calcium binding for the CHO proteins, the table was completed using data from the human homologs.

Bold: proteins significantly more abundant in H1 rFIX sample compared to H2 rFIX sample (DIA-MS and MSstats, $P < 10^{-5}$).

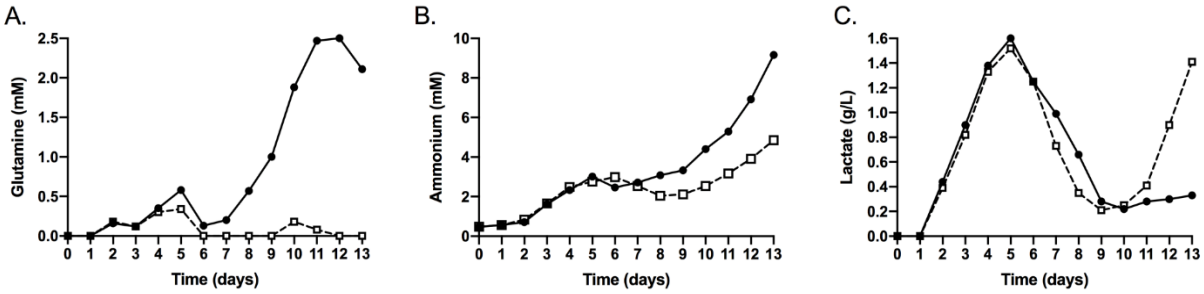
Italics: proteins significantly more abundant in H2 rFIX sample compared to H1 rFIX sample (DIA-MS and MSstats, $P < 10^{-5}$).

S17. Description of DDA and DIA-MS files uploaded in ProteomeXchange with the dataset identifier PXD018229.

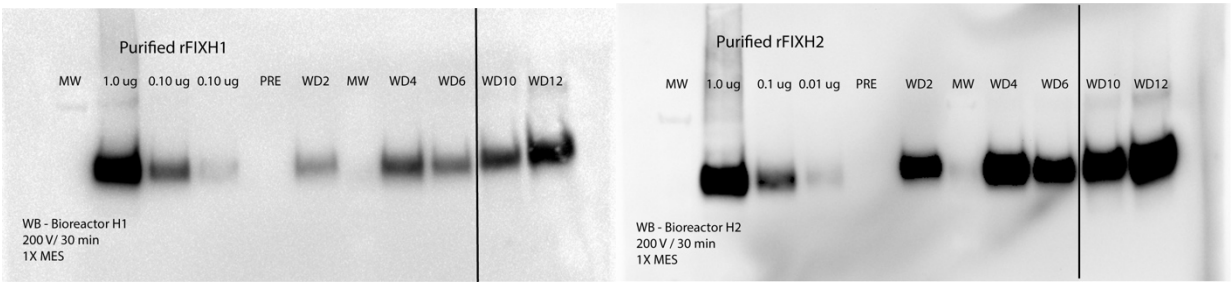
Supplementary Figures



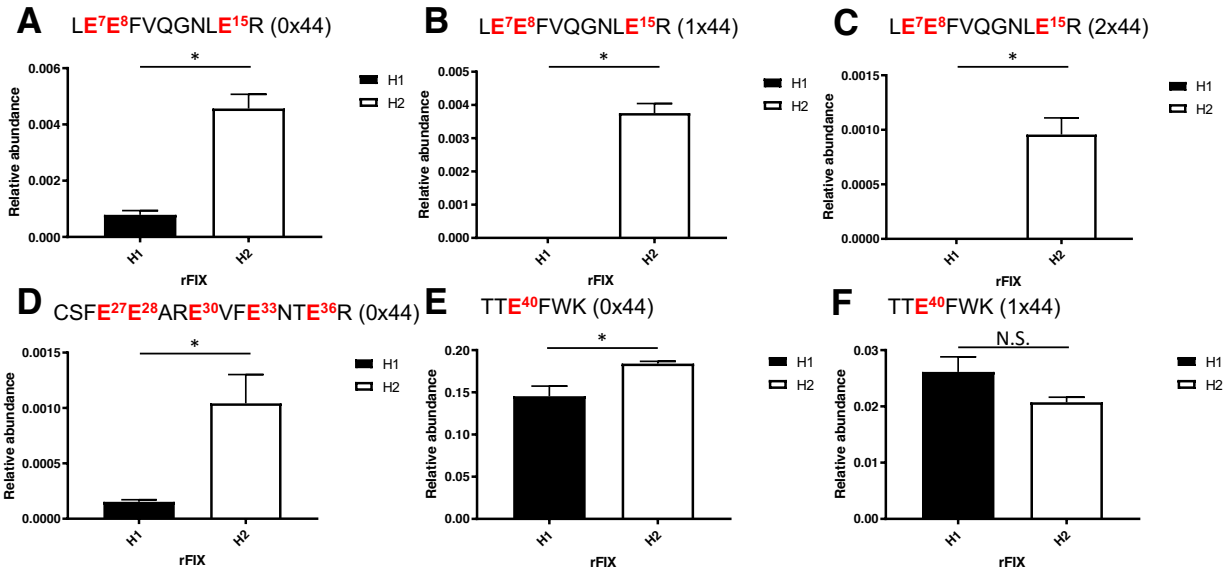
Supplementary Figure S1. Overview of the proteomic workflows used in this study.



Supplementary Figure S2. Metabolic profile of CHO cells expressing rFIX in both fed batch conditions. CHO cells expressing rFIX and PACE/Furin were grown in fed batch bioreactor mode with either EfficientyFeed A (H1, solid line and black circles) or EfficientFeed B (H2, dotted line and open square). The following metabolites were measured: (**A**) glutamine, (**B**) ammonium, and (**C**) lactate.



Supplementary Figure S3. Western blot of rFIX during bioreactor operation. Bioreactor H1 (left panel) and bioreactor H2 (right panel). Equal volumes of supernatant from H1 and H2 bioreactors at days 2, 4, 6, 10, and 12 were loaded onto the gel. Ten-fold dilutions of purified rFIX (1 mg/ml) from each bioreactor was used as a standard. PRE, medium before inoculation. rFIX showed the expected MW of ~ 57 kDa.



Supplementary Figure S4. Relative abundance of γ-carboxylated peptides from rFIX after purification. The abundance of select rFIX γ-carboxylated peptides (underivatized) was measured by DIA-MS in H1 and H2 bioreactors after rFIX purification. Graphs depict the mean +/- SEM of the abundance of each carboxypeptide relative to rFIX in purified samples from bioreactor H1 (black) or H2 (white). Shown are: **(A-C)** different carboxyforms of LE^{7}E^{8}FVQGNLE^{15}R, **(A)** uncarboxylated (0x44), **(B)** mono γ-carboxylated (1x44), or **(C)** di γ-carboxylated (2x44) at E$^{7/8/15}$; **(D)** uncarboxylated CSFE^{27}E^{28}ARE^{30}VFE^{33}NTE^{36}R (0x44), **(E-F)** different carboxyforms of TTE^{40}FWK, **(E)** uncarboxylated (0x44) or **(F)** mono γ-carboxylated (1x44) at E^{40}. One-tailed t-Test: * $p < 0.05$, N.S. = not significant (TTE(+44)FWK H1 *vs* H2, $p = 0.0632$). N = 2 - 3.

	% of coverage			
	Trypsin +/- PNGase F	**Glu-C +/- PNGase F**	**Chymotrypsin +/- PNGase F**	**AspN +/- PNGase F**
H1 rFIX	83.95	84.6	77.01	58.53
H2 rFIX	76.14	81.78	77.87	43.17
Plasma derived FIX	72.45	69.85	65.94	N/A

rFIX H1 digested with trypsin or trypsin + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

rFIX H2 digested with trypsin or trypsin + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

Plasma derived FIX digested with trypsin or trypsin + PNGase F

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

rFIX H1 digested with Glu-C or Glu-C + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

rFIX H2 digested with Glu-C or Glu-C + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

Plasma derived FIX digested with Glu-C or Glu-C + PNGase F

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

rFIX H1 digested with Chymotrypsin or Chymotrypsin + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

rFIX H2 digested with Chymotrypsin or Chymotrypsin + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

Plasma derived FIX digested with Chymotrypsin or Chymotrypsin + PNGase F

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

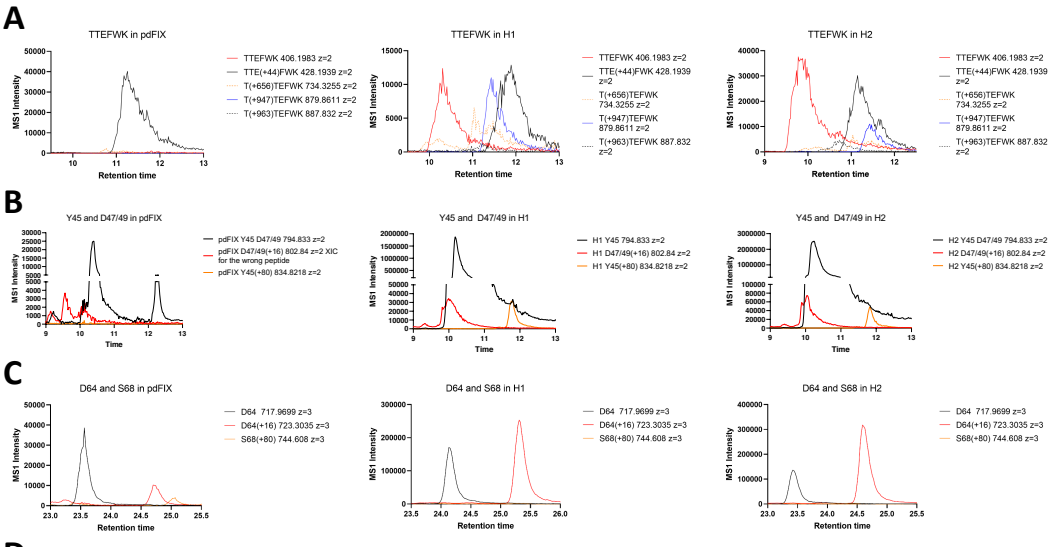
rFIX from H1 digested with AspN or AspN + PNGase F

MGRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRVKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

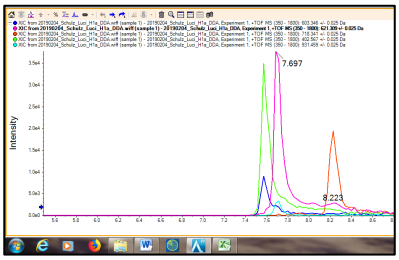
rFIX from H2 digested with AspN or AspN + PNGase F

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREV
FENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVV
CSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGE
DAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAA
INKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRS
TKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

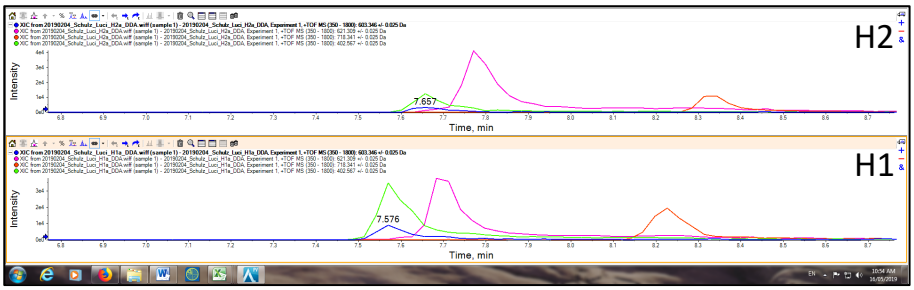
Supplementary Figure S5. FIX sequence coverage from DDA searches in ProteinPilot (SCIEX) of purified rFIX from bioreactors H1 and H2 or pdFIX digested with several single proteases +/- PNGase F. The table indicates the percentage of coverage. The sequences below show the coverage using the following color coding that indicates confidence assignment: Green: high; yellow: medium; red: low.



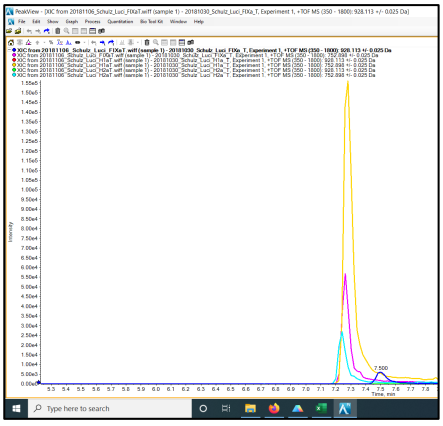
D



m/z	RT	Glycoform	z
402.5665	7.7461	VSVSQTSKLTR	3
603.3461	7.7811	VSVSQTSKLTR	2
621.309	7.7191	VSVSQTS[+656.22761]KLTR	3
931.459	7.7191	VSVSQTS[+656.22761]KLTR	2
718.3408	8.2172	VSVSQTS[+947.32303]KLTR	3

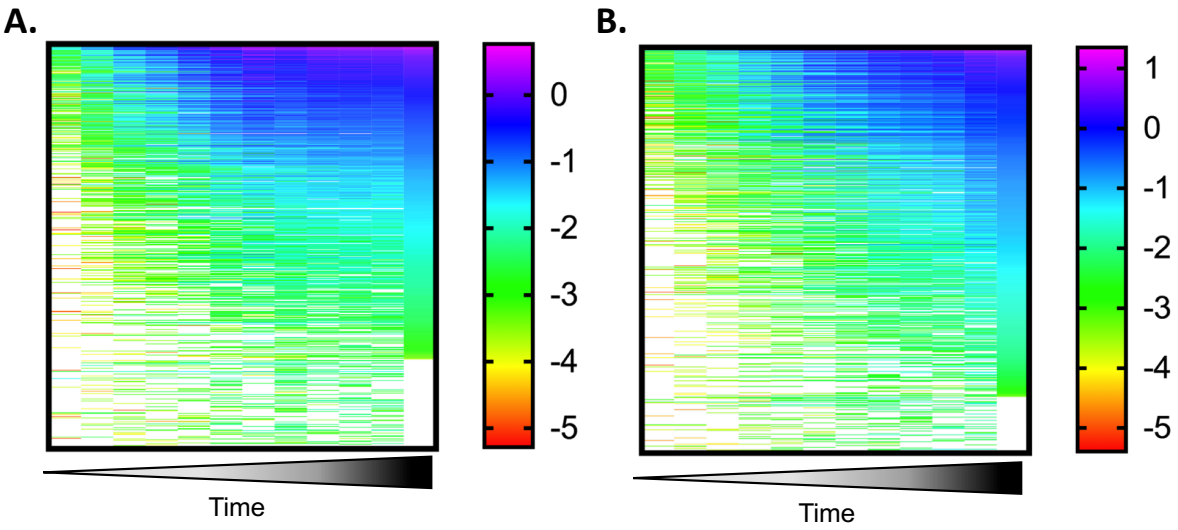


E

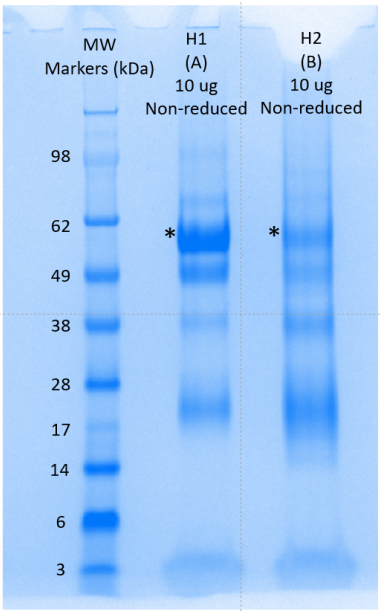


m/z	Glycoform	z	FIX
752.898	IIPHHNYNAAINK	2	H1
928.113	IIPHHN(+2204)YNAAINK	4	H1
752.898	IIPHHNYNAAINK	2	H2
928.113	IIPHHN(+2204)YNAAINK	4	H2
752.898	IIPHHNYNAAINK	2	pdFIX
928.113	IIPHHN(+2204)YNAAINK	4	pdFIX

Supplementary Figure S6. MS1 precursor intensity for select peptides and glycopeptides of interest. (A) $T^{38}TE^{40}FWK$. (B) $FWKQY^{45}VD^{47}GDQCE$. (C) $D^{64}DINS^{68}YECWCPFGFEGK$. (D) $VSVSQTS^{141}KLTR$. (E) $IIPHHN^{258}YNAAINK$.



Supplementary Figure S7. Host cell proteins (HCP) changes in relative abundance during culture. Heatmaps depicting the change in normalized relative abundance of HCPs through time (Day 1 to 13) in bioreactors H1 **(A)** and H2 **(B)**. Secreted HCPs were quantified by DIA-MS, normalized to trypsin, and log10 transformed. Each line represents each quantified HCP.



Supplementary Figure S8: Coomassie Blue stained non-reducing SDS-PAGE of purified rFIX from H1 and H2 fed-batch bioreactors. 10 µg total purified protein from H1 and H2 bioreactors were loaded in the gel. The asterisks show rFIX (estimated MW ~ 57 kDa).

Supplementary Information

1. Python script to recalculate peptide and protein abundance with peptide level 1% FDR cut-off.
2. Annotated MS/MS spectra from Byonic of all the post-translationally modified peptides used for the characterization of H1 rFIX digested with multiple single proteases and +/- PNGase F.
3. Annotated MS/MS spectra from Byonic of all the post-translationally modified peptides used for the characterization of H2 rFIX digested with multiple single proteases and +/- PNGase F.
4. Annotated MS/MS spectra from Byonic of all the post-translationally modified peptides used for the characterization of pdFIX digested with multiple single proteases and +/- PNGase F.
5. Annotated MS/MS spectra from Byonic of all the post-translationally modified peptides used for the quantification of PTMs in rFIX by DIA-MS.
6. Annotated MS/MS spectra from Byonic of all the post-translationally modified peptides used for the quantification of methylated GLA peptides in rFIX by DIA-MS.