

Stable Protein Sialylation in *Physcomitrella*

Lennard L. Bohlender¹, Juliana Parsons¹, Sebastian N. W. Hoernstein¹, Christine Rempfer^{1,2}, Natalia Ruiz-Molina¹, Timo Lorenz¹, Fernando Rodriguez-Jahnke^{1,2}, Rudolf Figl³, Benjamin Fode⁴, Friedrich Altmann³, Ralf Reski^{1,2,5,6} and Eva L. Decker^{1*}

¹Plant Biotechnology, Faculty of Biology, University of Freiburg, Freiburg, Germany

²Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Freiburg, Germany

³Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria

⁴Eleva GmbH, Freiburg, Germany

⁵Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, Germany

⁶Cluster of Excellence livMatS @ FIT – Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Freiburg, Germany

*Correspondence:

Eva L. Decker

eva.decker@biologie.uni-freiburg.de

Keywords: glyco-optimization, *N*-glycan sialylation, *N*-glycan humanization, glyco-engineering, plant-made recombinant biopharmaceuticals, plant-made pharmaceuticals, PMP

Abstract

Recombinantly produced proteins are indispensable tools for medical applications. Since the majority of them are glycoproteins, their *N*-glycosylation profiles are major determinants for their activity, structural properties and safety. For therapeutical applications, a glycosylation pattern adapted to product and treatment requirements is advantageous. *Physcomitrella* (*Physcomitrium patens*, moss) is able to perform highly homogeneous complex-type *N*-glycosylation. Additionally, it has been glyco-engineered to eliminate plant-specific sugar residues by knock-out of the β 1,2-xylosyltransferase and α 1,3-fucosyltransferase genes (Δ xt/ft). Furthermore, *P. patens* meets wide-ranging biopharmaceutical requirements such as GMP compliance, product safety, scalability and outstanding possibilities for precise genome engineering. However, all plants, in contrast to mammals, lack the capability to perform *N*-glycan sialylation. Since sialic acids are a common terminal modification on human *N*-glycans, the property to perform *N*-glycan sialylation is highly desired within the plant-based biopharmaceutical sector. In this study, we present the successful achievement of protein *N*-glycan sialylation in stably transformed *P. patens*. The sialylation ability was achieved in a Δ xt/ft moss line by stable expression of six mammalian coding sequences combined with targeted organelle-specific localization of the encoded enzymes responsible for synthesis, activation, transport and transfer of sialic acid. Production of free and (CMP)-activated sialic acid was proven. The glycosidic anchor for the attachment of terminal sialic acid was generated by the introduction of a chimeric human β 1,4-galactosyltransferase gene under the simultaneous knock-out of the gene encoding the endogenous β 1,3-galactosyltransferase. Functional complex-type *N*-glycan sialylation was confirmed via mass spectrometric analysis of a stably co-expressed recombinant human protein.

1 Introduction

The biopharmaceutical sector, which in 2017 was valued at USD 1.654 billion in global sales, is continuously increasing its significant share of the global pharmaceutical market (Moorkens et al., 2017). Recombinant proteins, which represent the largest proportion of biopharmaceutical products, are manufactured only by a limited number of production platforms. These are mainly based on bacteria, yeast, insect cell, plant cell or mammalian cell cultures, whereas the latter are the most represented ones (Merlin et al., 2014). Biopharmaceuticals produced in mammalian cell cultures are in general well accepted, since these are able to perform post-translational modifications in a similar manner as in the human body (Jenkins et al., 2008).

Within mammalian systems, Chinese hamster ovary (CHO) cells are by far the most common production host (Walsh, 2018; Nguyen et al., 2020), even though they are naturally not able to link sialic acid to *N*-glycans in the human-typical predominant α 2,6-manner but perform α 2,3-linkages instead, which may have a destabilizing effect on recombinant antibodies (Zhang et al., 2019). Furthermore, *N*-glycans of glycoproteins produced in some non-human mammalian cell lines can carry immunogenic residues, such as the terminal *N*-glycolylneuraminic acids (Neu5Gc), a sialic acid not produced in humans (Chou et al., 2002; Tangvoranuntakul et al., 2003; Ghaderi et al., 2012) or α 1,3-linked galactoses, which especially occur in SP2/O cells (Hamadeh et al., 1992; Chung et al., 2008). Moreover, mammalian cell lines show a high heterogeneity in terms of *N*-glycosylation as well as a limited batch-to-batch reproducibility (Jefferis, 2005; Kolarich et al., 2012; Komatsu et al., 2016), which has a negative impact on the quality of a biopharmaceutical (Schiestl et al., 2011).

Alternative production platforms can offer advantages in sectors neglected by the current narrowing range of systems. In this regard, plants combine the protein processing capabilities of eukaryotic cells with cultivation requirements comparable to prokaryotic production systems, leading to lower manufacturing costs (Rozov et al., 2018). Further, plants lack human pathogens, endotoxins and oncogenic DNA sequences (Commandeur et al., 2003), and hence are generally safer than microbial or animal production hosts (Twyman et al., 2003). Between plants and animals, both the protein biosynthesis as well as the secretory pathway are highly conserved. This enables plants, in contrast to bacterial hosts, to synthesize and fold complex human proteins correctly (Tschofen et al., 2016), as well as to perform the majority of posttranslational modifications needed for high-quality biopharmaceuticals.

Due to their smaller glycome, plants display a reduced *N*-glycosylation microheterogeneity in comparison to mammalian cells (Bosch et al., 2013; Margolin et al., 2020). The latter usually synthesize a wide mixture of *N*-glycans, whereas in plants often one *N*-glycan structure predominates (Koprivova et al., 2003; Castilho et al., 2011). This leads to a rigid batch-to-batch stability and an improvement of quality and kinetics of recombinantly plant-produced biopharmaceuticals (Bosch et al., 2013; Stoger et al., 2014; Sochaj et al., 2015; Shen et al., 2016). Like mammals, plants produce complex-type *N*-glycans, which share an identical $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ (GnGn) core but differ in some post-ER processed residues from their mammalian counterparts (Viëtör et al., 2003). The distal GlcNAc residues of mammalian *N*-glycans harbor β 1,4-linked galactoses, which are often capped by a α 2,6-linked sialic acid residue (Varki et al., 2015). In contrast, plant *N*-glycans are predominantly terminated with exposed distal GlcNAcs, which additionally can be decorated with an β 1,3-linked galactose and an α 1,4-linked fucose (Lerouge et al., 1998). This terminal structure ($\text{Gal}\beta(1,3)(\text{Fuc}\alpha(1,4))\text{GlcNAc}$), known as Lewis A (Le^a) epitope, can be found in some plant cell-wall associated proteins (Strasser et al., 2007). It is, however, a tumor-associated carbohydrate structure in humans (Zhang et al., 2018) and it has been associated with antibody formation (Fitchette et al., 1999; Wilson et al., 2001). Additionally, the Asn-linked GlcNAc is substituted with an α 1,3-attached fucose in plants, whereas in mammals this

residue is α 1,6-linked. Furthermore, in plants the proximal mannose harbors a bisecting β 1,2-linked xylose, a sugar which is not present in mammals. Considering that the majority of biopharmaceuticals are glycoproteins and that their *N*-glycosylation plays an important role in their efficacy (Lingg et al., 2012), attention should be paid to *N*-glycosylation quality. Therefore, several model plants have been glyco-engineered in the last decades to obtain a humanized *N*-glycosylation pattern devoid of these structures (van Ree et al., 2000; Bardor et al., 2003; Gomord et al., 2005; Decker and Reski, 2012; Decker et al., 2014). Plant-specific β 1,2-xylosylation and α 1,3-fucosylation of *N*-glycans were eliminated in *P. patens* by targeted knock-outs of the responsible glycosyltransferase (XT and FT) coding genes via homologous recombination (Koprivova et al., 2004). The same genes were knocked down by RNAi in *Arabidopsis thaliana* (Strasser et al., 2004), *Lemna minor* (Cox et al., 2006), *Nicotiana benthamiana* (Strasser et al., 2008), *Medicago sativa* (Sourrouille et al., 2008) and *Oryza sativa* (Shin et al., 2011). More recently, these genes were knocked out in *Nicotiana benthamiana* (Jansing et al., 2018) and *Nicotiana tabacum* BY-2 suspension cells by targeting two XT and five FT encoding genes via CRISPR/Cas9 genome editing (Hanania et al., 2017; Mercx et al., 2017). Furthermore, the Le^a epitope, which is a rare terminal modification of *P. patens* *N*-glycans, was abolished in this organism by the single knock-out of the gene encoding the responsible β 1,3-galactosyltransferase 1 (GalT3; Parsons et al., 2012). Recombinant glycoproteins produced in *P. patens* with the triple KO of XT, FT and GalT3 displayed an outstanding homogeneity in the *N*-glycans, with a sharply predominant GnGn pattern (Parsons et al., 2012), providing a suitable platform for further glyco-optimization approaches. These studies also revealed a sufficient plasticity of plants towards glyco-engineering approaches without observable phenotypic impairments (Webster and Thomas, 2012).

Plant systems are already being used for the production of biopharmaceuticals. β -glucocerebrosidase, an enzyme for replacement therapy in Morbus Gaucher treatment, is obtained from carrot-cell suspensions (van Dussen et al., 2013) or ZMapp, a combination of antibodies for treatment of Ebola infections, is produced in *N. benthamiana* lacking plant-typical *N*-glycan modifications (Margolin et al., 2018). Besides these approved plant-made biopharmaceuticals there are further promising candidates in clinical trials, such as virus-like particles bearing influenza hemagglutinin proteins as *Medicago sativa*-derived vaccine against influenza (D'Aoust et al., 2010) or the moss-produced α -galactosidase for enzyme replacement therapy in Morbus Fabry treatment (Shen et al., 2016; Hennermann et al., 2019). The majority of pharmaceutically interesting glycoproteins are terminally sialylated in their native form. *N*-glycan sialylation is highly desirable due to its role in half-life, solubility, stability and receptor binding (Varki et al., 2015). However, plants are unable to perform β 1,4-galactosylation, which in mammals serves as acceptor substrate for terminal *N*-glycan sialylation. Moreover, they are not able to produce, activate and link sialic acid (Zeleny et al., 2006; Bakker et al., 2008; Castilho et al., 2008), which in mammals requires the coordinated activity of six enzymes. To achieve *in planta* β 1,4-galactosylation, several trials have been undertaken in different systems with varying degrees of success, including the expression of the β 1,4-galactosyltransferase from humans (hGalT4) or animals or chimeric varieties (Palacpac et al., 1999; Bakker et al., 2001, 2006; Misaki et al., 2003; Huether et al., 2005; Fujiyama et al., 2007; Hesselink et al., 2014; Kittur et al., 2020; Kriechbaum et al., 2020). The efficiency and quality of galactosylation was shown to depend on the localization of the enzyme in the Golgi apparatus (Bakker et al., 2001; Strasser et al., 2009), on its expression level (Kallolimath et al., 2018) and on the investigated protein (Kriechbaum et al., 2020). *De novo* synthesis of sialic acid involves three enzymes in a four-step process within the cytosol. The biosynthesis starts with the synthesis of *N*-acetylmannosamine (ManNAc) out of its UDP-activated precursor substrate *N*-acetylglucosamine (UDP-GlcNAc), which is present in both, plants and mammals. The first reaction steps of the sialic acid production are catalyzed by the enzyme UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE), which is the key enzyme of sialic acid biosynthesis (Reinke et al., 2009). GNE bifunctionally accomplishes the cleavage of UDP and the subsequent epimerization of GlcNAc to

ManNAc, followed by ManNAc phosphorylation at C-6, resulting in ManNAc-6P. Subsequently, Neu5Ac 9-phosphate synthase (NANS) catalyzes the condensation with phosphoenolpyruvate (PEP) leading to Neu5Ac-9P (Roseman et al., 1961). Finally, Neu5Ac-9P is dephosphorylated by the Neu5Ac-9-P phosphatase (NANP) resulting in *N*-acetylneuraminic acid (Neu5Ac, sialic acid, Maliekal et al., 2006). Activation of Neu5Ac is then performed by the nuclear CMP-Neu5Ac synthetase (CMAS), forming CMP-Neu5Ac. This is translocated in exchange for CMP in an antiporter mechanism by the CMP-sialic acid transporter (CSAT) from the cytosol into the Golgi lumen (Aoki et al., 2003; Tiralongo et al., 2006; Zhao et al., 2006). In the final step of mammalian *N*-glycosylation, the sialyltransferase (ST) catalyzes the release of Neu5Ac from CMP and the covalent attachment of it to an β 1,4-galactosylated acceptor in the (trans) Golgi apparatus, resulting in the complex-type mammalian *N*-glycosylation. In recent years several approaches have been performed to establish functional sialylation in plants. Accordingly, the simultaneous expression of the murine GNE and human NANS and CMAS coding sequences (CDS) in *A. thaliana* lead to the production of activated sialic acid (Castilho et al., 2008). In *N. benthamiana* functional *N*-glycan sialylation could be achieved in a transient approach by co-expressing five CDSs of the sialylation pathway genes, together with a chimeric GalT4 containing *N*-terminally the sub-Golgi apparatus localization determining cytoplasmic, transmembrane, and stem (CTS) region of the rat ST. This resulted in a sialylation efficiency of about 80%, detected on an additionally transiently co-expressed recombinant monoclonal antibody (Castilho et al., 2010). The first stable genetically engineered sialylating *N. benthamiana* line was described by Kallolimath et al. (2016).

Particularly, with regard to recent achievements in the production of candidate biopharmaceuticals, combined with its Good Manufacturing Practice (GMP)-compliant production capabilities, the moss *P. patens* serves as a competitive production platform for biopharmaceuticals (Decker and Reski, 2020). Preclinical trials of moss-derived recombinant human complement factor H were recently effectively accomplished (Häffner et al., 2017; Michelfelder et al., 2017), along with the moss-GAA (acid alpha-1,4-glucosidase) against Pompe disease (Hintze et al., 2020). Furthermore, the clinical phase I study of the recombinantly produced candidate biopharmaceutical moss-aGal against Fabry disease was successfully completed (Reski et al., 2018; Hennermann et al., 2019). Moss-GAA and moss-aGal proved to have better overall performance compared to their variants produced in mammalian cell cultures (Hennermann et al., 2019; Hintze et al., 2020), demonstrating the potential of *P. patens* to produce biobetters. However, the last step in humanizing *N*-glycans, protein sialylation, still has to be accomplished.

Here, we report the successful establishment of stable protein sialylation in *P. patens*, after subsequently accomplishing the synthesis of free as well as activated sialic acid, and the β 1,4 linkage of the sialic acid anchor galactose to protein *N*-glycans. This will further increase the attractiveness of moss as plant-based biopharmaceutical production platform.

2 Materials and Methods

2.1 Construct generation

For cloning, coding sequences were amplified by PCR with Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA). All primers used in this study are compiled in Supplementary Table S1. After each PCR or restriction digest step, agarose gel purification of respective fragments was performed using QIAEX II Gel Extraction Kit (QIAGEN, Hilden, Germany) and ligations were performed using pJET 1.2 Cloning Kit or TOPO™ TA Cloning™ (Thermo Fisher Scientific), all according to the manufacturer's protocols. Assembled vectors were verified by sequencing.

For the heterologous expression of all six genes of the sialylation pathway two multi-gene constructs, containing three PCR-amplified CDS each, were assembled via restriction-site based cloning. GNE and CSAT were amplified from murine cDNA and ST from rat cDNA. NANS, a truncated CMAS-version lacking 120 bases at the 5' end (Castilho et al., 2010) and NANP were amplified from human cDNA. Additionally, a Blasticidin-S deaminase (BSD) expression cassette under the control of the CaMV 35S promoter and CaMV 35S terminator, provides resistance towards Blasticidin S (Kubo et al., 2013).

To assemble the multi-gene constructs, each CDS as well as the BSD cassette was amplified with restriction site-introducing primers (primers 1-14), and subcloned in the pJET 1.2 cloning vector. Expression of each glycosylation-related CDS was driven by the long CaMV 35S promoter (Horstmann et al., 2004) and the nos terminator. Promoter, multiple cloning site (MCS) and terminator were amplified via CDS-specific restriction site-introducing primers (primers 15-22) and subcloned into pJET 1.2, resulting in six different target vectors. To assemble the expression cassettes, the subcloned CDS and the multiple cloning site of the corresponding target vector were digested with the respective restriction enzymes and ligated, resulting in seven expression cassettes, containing GNE, NANS, NANP, CMAS, CSAT, ST or BSD, respectively. For generation of the two multi-gene constructs, two assembly vectors, either based on pJET 1.2 or pCR™4-TOPO® TA-vector harboring the corresponding homologous flanks for targeted genome integration and a designed MCS were created, respectively (Figures 1A and 1B). For the introduction of the first part of the sialylation pathway containing GNE, NANS and NANP (GNN) vector, homologous flanks were designed to target the integration of the construct into the *P. patens* adenine phosphoribosyltransferase (APT) gene (Pp3c8_16590V3.1). To create the multiple cloning site of the GNN-assembly vector, the APT-5' homologous flank was amplified with a primer pair introducing an LglI restriction site at the 5' end and AgeI, SgrDI and AvrII restriction sites at the 3' end of the PCR product (oligonucleotides 23 and 24). The APT-3' homologous flank was amplified with a primer pair introducing AvrII and AsclI restriction sites at the 5' end and an LglI restriction site at the 3' end of the PCR product (primers 25 and 26). The respective PCR-products were digested with AvrII, ligated and subsequently cloned into pJET 1.2, resulting in the GNN assembly vector (Figure 1A). GNE-, NANS- and NANP expression cassettes were subsequently introduced into the GNN-assembly vector using the AgeI + SgrDI, SgrDI + AvrII or AvrII + AsclI restriction sites, respectively, resulting in the GNN-construct (Supplementary Figure S1A).

For construction of the CMAS, CSAT, ST and BSD-cassette containing multi-gene construct (CCSB) a similar strategy was used. This construct was created based on the pTOPO vector, targeted to the endogenous prolyl-4-hydroxylase (P4H1) gene (Pp3c8_7140V3.1) and additionally harbored a Blasticidin S resistance (BSD) for selection purposes. To create the multiple cloning site of the CCSB-assembly vector, the P4H1-5' homologous flank was amplified with a primer pair introducing a NotI restriction site at the 5' end and SgrDI, AgeI and AvrII restriction sites at the 3' end of the PCR product (primers 27 + 28). The P4H1-3' homologous flank was amplified with a primer pair introducing AvrII, AgeI and AatII restriction sites at the 5' end and a NotI restriction site at the 3' end of the PCR product (primers 29 + 30). The respective PCR-products were digested with AvrII, ligated, polyadenylated with Taq-polymerase using primers 27 and 30 and cloned into the pTOPO vector to create the CCSB-assembly vector (Figure 1B). CMAS, CAST, ST and BSD expression cassettes were subsequently introduced into the CCSB-assembly vector using the SgrDI + AsclI, AsclI + AvrII, AvrII + AgeI and AgeI + AatII restriction sites, respectively, resulting in the CCSB-construct (Supplementary Figure S1B). The LglI or NotI-cut GNN and CCSB constructs used for transfections are schematically illustrated in Figures 1C and 1D. A schematic overview of all GNN and CCSB cloning steps is provided in Supplementary Figure S1.

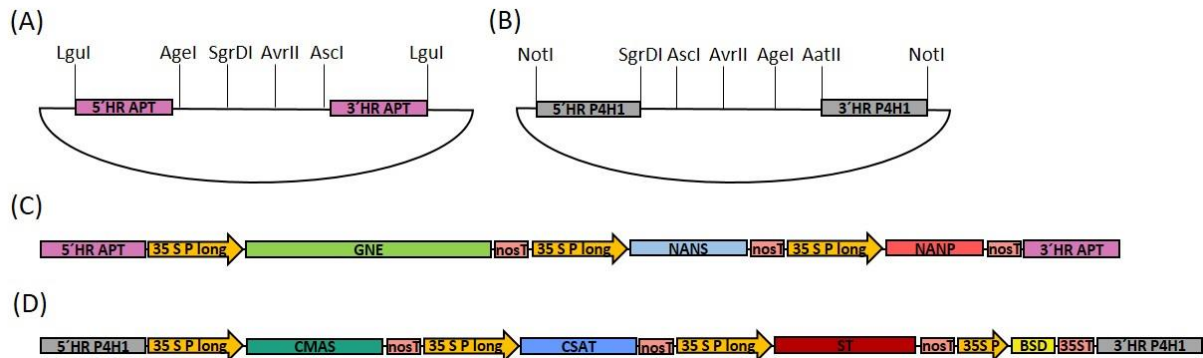


Figure 1. Schematic illustration of the cloned target vectors and the final GNN and CCSB constructs used for transfection. **(A)** pJET1.2-based assembly vector with a designed multiple cloning site for generation of the GNE, NANS and NANP-containing GNN construct with homologous flanks for integration into the adenine phosphoribosyltransferase (APT) locus. **(B)** pTOPO-based assembly vector with a designed multiple cloning site for generation of the CMAS, CSAT, ST and BSD-containing CCSB construct with homologous flanks for integration into the prolyl-4-hydroxylase 1 (P4H1) locus. Lgul or NotI were used to linearize GNN **(C)** and CCSB **(D)** constructs, respectively, before sequential transformation of *P. patens*. 5' HR: 5' homologous region, 3' HR: 3' homologous region, 35S P long: long CaMV 35S promoter, 35S P: CaMV 35S promoter, nosT: nos terminator, 35ST: CaMV 35S terminator.

2.1.1 GNE mutation

To circumvent CMP-Neu5Ac-triggered negative feedback regulation of GNE, a mutated version (GNEmut) with amino acid exchanges R263L and R266Q in the allosteric CMP-Neu5Ac binding site (Viswanathan et al., 2005; Son et al., 2011), was created. The corresponding GNE sequence alterations were performed on the previously described GNE expression construct via site-directed mutagenesis using the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For complete vector amplification a mismatching primer pair, which changes the codon at position 263 from CGA to CTA and replaces the codon CGG at position 266 with CAG, was used (primers 31 + 32). After successful mutagenesis, the native GNE of the GNN-construct was replaced by GNEmut via AgeI and SgrDI restriction sites, resulting in the GM-construct.

2.1.2 Cloning of a chimeric β 1,4-galactosyltransferase

For cloning of a chimeric β 1,4-galactosyltransferase (FTGT), the localization-determining cytoplasmic, transmembrane, and stem region (CTS) of the human β 1,4-galactosyltransferase (GalT4, NM_001497.4) was replaced by 390 bp encoding the 130 amino acids comprising the CTS of *P. patens* α 1,4-fucosyltransferase (FT4, Pp3c18_90V3.1), responsible for the last step in *N*-glycosylation. The CTS from the FT4 (FT-CTS) was amplified from moss cDNA with the primers 33 and 34, generating an overlapping region to the GalT4 sequence. In parallel, 927 bp comprising the catalytic domain from the GalT4 CDS were amplified with primers 35 and 36. Subsequently, the PCR products were assembled in a two-template PCR with primers 33 and 36, giving rise to the CDS of the chimeric FT-CTS: β 1,4-galactosyltransferase (FTGT). For the expression of the FTGT, a bicistronic vector was cloned, where the sequence of the chimeric FTGT and the *ble* gene (Parsons et al., 2013), conferring resistance to ZeocinTM (Zeocin), were separated by the sequence of the self-cleaving 2A peptide, P2A and driven by the long CaMV 35S promoter. This sequence triggers ribosomal skipping during translation, leading to the release of two individual proteins (Donnelly et al., 2001; Kim et al., 2011). The sequence for P2A was included at the 3' end of the chimeric FTGT CDS by two successive PCR reactions using primers 33 and 37, and 33 and 38. The CaMV 35S promoter from the Zeocin resistance cassette in the β 1,3-galactosyltransferase 1-KO construct (Parsons et al., 2012) was excised with BamHI and XhoI and

replaced by a long CaMV 35S promoter amplified with primers 39 and 40 using the same restriction sites. This construct bears homologous flanks targeting the endogenous β 1,3-galactosyltransferase 1 locus (GalT3, Pp3c22_470V3.1) to knock it out (Supplementary Figure S2). The FTGT-P2A sequence was then inserted between promoter and *ble* gene using the XhoI site. For transfection the construct was cut using EcoRI endonuclease.

2.1.3 Cloning of a chimeric α 2,6-sialyltransferase

For the introduction of a chimeric sialyltransferase (FTST), consisting of the endogenous FT-CTS and the catalytic domain of the rat α 2,6-sialyltransferase, a construct with homologous flanks for targeted integration into the P4H2 locus (Pp3c20_10350V3.1) (Supplementary Figure S3) was cloned via Gibson assembly. Amplification of the 5' and 3' homologous flanks was performed from moss genomic DNA using overhang introducing primers 41 and 42 or 43 and 44, respectively. The hygromycin selection cassette coding for a hygromycin B phosphotransferase (*hpt*) (Decker et al., 2015) was amplified using the primers 45 and 46. Gibson assembly was performed as described in Gibson et al. (2009) into a pJET 1.2 backbone. For introduction of the chimeric FTST variant this vector was linearized with BseRI at the 3' end of the 5' homologous flank and subsequently used as backbone for the introduction of the FTST anew via Gibson assembly. For this the FT-CTS together with the long CaMV 35S promoter was amplified with the primers 47 and 48 from the previously described FTGT construct and the catalytic domain of the rat ST including the nos terminator was amplified from the previously described CCSB-construct using primers 49 and 50. For transfection the construct was cut using NotI and XbaI endonucleases.

2.2 Plant material and cell culture

P. patens was cultivated as described previously (Frank et al., 2005) in Knop medium supplemented with microelements (KnopME, Horst et al. (2016)). The sialylating lines were obtained by stable transformation of the Δ xt/ft moss line (IMSC no.: 40828), in which the α 1,3-fucosyltransferase and the β 1,2-xylosyltransferase genes have been disrupted via homologous recombination (Koprivova et al., 2004). This line additionally produces a reporter glycoprotein.

2.2.1 Transfection of *P. patens* protoplasts and selection of transgenic lines

Protoplast isolation, transformation and regeneration were performed as described before (Decker et al., 2015). For each transfection event 50 μ g of linearized plasmid were used. Selection for the marker-free GNN or GM lines with targeted ATP locus disruption was performed on 0.3 mM 2-fluoroadenine (2-FA) containing solid KnopME plates. Protoplasts were regenerated in liquid medium for 5 days, then transferred to Knop solid medium covered with cellophane for 3 days. Subsequently, the cellophane sheets with the regenerating protoplasts were transferred onto 2-FA containing Knop plates for a further 3 weeks. Selection with Blasticidin S (Sigma Aldrich) or Zeocin (Invitrogen) were started in liquid medium on day 8 after transfection via the addition of 75 mg/L Blasticidin S or 50 mg/L Zeocin, respectively. After 4 days of selection in liquid medium the protoplasts were transferred to solid KnopME plates containing either 75 mg/L Blasticidin S or 100 mg/L Zeocin and 1% MES, respectively. The selection was done in two successive cycles of 3 weeks on selective plates interrupted by a 2-week release on non-selective KnopME plates. Hygromycin selection was performed as described before (Wiedemann et al., 2018).

2.3 Molecular validation of transgenic moss lines

Plants surviving the selection were screened for targeted construct integration via direct PCR (Schween et al., 2003). Successful extraction of DNA, which was performed as described before (Parsons et al., 2012), was assayed by amplifying a part of the elongation factor 1 gene (EF1 α , Pp3c2_10310V3.1) with primer pair 51 and 52. Targeted integration of the GNN or GM constructs in the APT locus was confirmed using the primer pairs 53 and 54 for 5'- and 55 and 56 for 3'-integration. Targeted integration of the CCSB construct within the P4H1 locus was assayed with the primers 57 and 58 as well as 59 and 60 for the 5'- and 3'-integration, respectively. Homologous integration of the chimeric β 1,4-galactosyltransferase FTGT into the GalT3 locus was confirmed using the primers 61 and 62 (5'-integration), whereas correct 3'-integration was confirmed by the primers 63 and 64. Targeted integration of the FTST construct into the P4H2 locus was assayed with the primers 65 and 66 for 5'- and 67 and 68 for 3'-integration.

2.4 Expression analysis of transgenic moss lines

2.4.1 Real-time qPCR analysis of gene expression

To determine the expression levels of the transgenes, total RNA was isolated using Trizol (Thermo Fischer Scientific) according to the manufacturer's instructions. RNA concentrations were determined via UV-Vis spectrometric measurement (NanoDrop ND 1000, PEQLAB Biotechnologie GmbH, Erlangen) and quality was checked via agarose gel electrophoresis. Isolated RNA was subsequently digested with DNaseI (Thermo Fisher Scientific) and cDNA synthesis was performed with random hexamers using TaqMan[®] Reverse Transcription Reagents (Thermo Fisher Scientific), both according to the manufacturer's protocols. Completeness of DNaseI digestion was confirmed by a non-transcribed control, without the addition of MultiScribe[®] RT enzyme. Primer pairs were designed using the Universal Probe Library by Roche (https://lifescience.roche.com/en_de/brands/universal-probe-library.html#assay-design-center) and selected according to the lowest amount of off-target hits identified in a Phytozome (Goodstein et al., 2012) search against the *P. patens* transcriptome (V3.3; Lang et al., 2018). Oligonucleotide pair efficiencies of 2 were confirmed prior to analysis with a qPCR-run of a serial 1:2 dilution row of cDNA completed by a control without cDNA. Presence of off-targets was excluded via melting curve analysis. Measurements were prepared in white 96-multiwell plates and carried out in triplicate, using the SensiMix[™] Kit and SYBRGreen (Bioline, Luckenwalde, Germany). For each triplicate cDNA amounts corresponding to 50 ng RNA were used and measurements were conducted in a LightCycler[®] 480 (Roche) according to the manufacturer's instructions. Absence of DNA contamination in reagents was confirmed for each primer pair with a control reaction without template addition. Amplification was performed in 40 cycles with a melting temperature of 60°C. Computational analysis of the qRT-PCR was done with LightCycler[®] 480 software (Roche). The analysis of the expression levels of the transgenes was performed relatively against the expression values of the internal housekeeping genes coding for EF1 α and for the ribosomal protein L21 (Pp3c13_2360V3.1) (Beike et al., 2015). The relative expression of the gene of interest (GOI) compared to the controls (ctrl) EF1 α and L21 was calculated as $2^{-\Delta C_T}$, assuming an primer efficiency of 2, and where $\Delta C_T = C_{TGOI} - C_{Tctrl}$ and C_T is defined as the cycle number at which each sample reaches an arbitrary threshold (Livak and Schmittgen, 2001). Primer pairs used: GNE: 73 and 74; NANS: 75 and 76; NANP: 77 and 78; CMAS: 79 and 80; CSAT: 81 and 82; ST and FTST: 83 and 84; FTGT: 85 and 86; EF1 α : 69 and 70 and L21: 71 and 72.

2.4.2 RNAseq library preparation and data analysis

Total RNA was extracted from 100 mg fresh weight (FW) protonema tissue in biological triplicates for each line, using Direct-zol™ RNA MicroPrep Kit (Zymo Research, Freiburg) according to the manufacturer's protocol. RNAseq library preparations were performed by the Genomics Unit at Instituto Gulbenkian de Ciencia (Portugal) with conditions optimized according to Baym et al. (2015), Macaulay et al. (2016) and Picelli et al. (2014) using Smart-seq2. Sequencing was performed on an Illumina NextSeq 500 instrument producing 75 bp long single-end reads.

We assessed sequence quality with FastQC (Galaxy Version 0.72+galaxy1; Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> and multiQC; Galaxy Version 1.6; Ewels et al., 2016) on the public European Galaxy instance at <https://usegalaxy.org> (Afgan et al., 2018). For transcript quantification, reads of each library were pseudoaligned against the *P. patens* transcriptome (V3.3; Lang et al., 2018), obtained from Phytozome v12.1.5 (Goodstein et al., 2012) and complemented with the sequences of the introduced transgenes using Kallisto quant (Galaxy Version 0.43.1.4; Bray et al., 2016) for single-end reads in “unstranded” mode. According to the information obtained from the sequencing facility, the average fragment length of the libraries was specified as 340 with an estimated standard deviation of 34. For the quantification algorithm a number of 100 bootstrap samples was chosen.

2.5 Detection of sialic acid via periodate-resorcinol-assay

The detection of total sialic acid concentrations in moss extracts was performed according to Jourdian et al. (1971) with the following modifications: Protonema, the young filamentous tissue of *P. patens*, was harvested via vacuum filtration and frozen. Approximately 150 mg FW of frozen plant material was disrupted with a glass and a metal bead (\emptyset 3 mm; QIAGEN GmbH, Hilden, Germany) in a TissueLyser (MM400, Retsch GmbH, Haan, Germany) at 30 Hz for 1.5 min. For Neu5Ac extraction, the four-fold amount (v/w) of 50 mM Tris/HCl (pH 7.0) was added and the samples were vortexed for 10 min and afterwards sonicated (Bandelin Sonorex RK52, Bandelin electronic GmbH & Co. KG, Berlin, Germany) for 15 min at 4°C. Crude extracts were cleared via two subsequent centrifugation steps at 14,000 rpm and 4°C for 10 min and further 30 min centrifugation of the supernatant in a fresh reaction tube. Serial dilutions of the samples (1:2 to 1:64) and a standard row containing 1 to 40 nmol Neu5Ac (Sigma Aldrich) were prepared in extraction buffer, final volume 120 μ l. For sialic acid oxidation 30 μ l of a 0.032 M periodic acid solution were added, followed by 45 min incubation under gently shaking conditions at 4°C. Samples were additionally cleared via a 10 min centrifugation step at 14,000 rpm and 4°C. Subsequently, 100 μ l of freshly prepared resorcinol solution (0.06% w/v resorcinol (Sigma-Aldrich), 16.8% HCl, 0.25 μ M CuSO₄) and 50 μ l sample or standard were mixed in a 96-well plate (Greiner Bio-One, Frickenhausen, Germany), sealed and incubated for 45 min at 80°C. Color complexes were stabilized by the addition of 100 μ l tert-butanol (Sigma-Aldrich) and the absorbance measured at 595 nm (Sunrise absorbance reader, Tecan, Männedorf, Switzerland, software Magellan™ V 7.1). Calculation of Neu5Ac concentration in the samples was performed by linear regression.

2.6 Detection of sialic acid (Neu5Ac) via RP-HPLC-FLD

Detection of sialic acid in moss extracts was performed as described previously for *Arabidopsis thaliana* (Castilho et al., 2008) via reverse phase high performance liquid chromatography coupled with fluorescence detection (RP-HPLC-FLD). Moss protonema tissue was dispersed with an ULTRA-TURRAX® (IKA, Staufen, Germany), 1 ml of the crude extract was taken for dry weight (DW) determination and further 500 μ l were mixed with 50 μ l acetic acid and incubated for 10 min under shaking conditions. Extracts were cleared via 1 min of centrifugation at 16,100 x g and supernatants were further purified

via a C18 SPE column (Strata C18-E, 50 mg; Phenomenex), pre-equilibrated with 1% acetic acid. After loading to the column, the samples were washed with 200 μ l 1% acetic acid. The column flow-through was vacuum-dried and afterwards resuspended in 30 μ l ultra-pure water. Ten μ l of the C18 SPE purified extracts were DMB-labeled with 60 μ l of DMB labeling reagent (Sigma-Aldrich), at 50°C for 2.5 h in the dark under shaking conditions at 750 rpm (Hara et al., 1987). Prior to analysis, quenching of the reaction was performed by the addition of 730 μ l of ultra-pure water. For detection of sialic acid, 5 to 20 μ l of the labelled extracts were injected on a HPLC (Nexera X2 HPLC system) with a RF-20AxS Fluorescence Detector, equipped with a semimicro flow cell (Shimadzu, Korneuburg, Austria). Separation was performed on an Aquasil C18 column (250 cm x 3 mm, 5 μ m particle size; Thermo Fisher Scientific) at a flow rate of 1 ml/min, applying a linear gradient from 30% to 42% B (70% 100 mM ammonium acetate pH 5.5, 30% acetonitrile; Solvent A was water) over 12 min and the column thermostat was set to 35°C. Fluorescence was measured with wavelengths Excitation/Emission 373 nm and 448 nm. Identification of Neu5Ac in the moss extract was performed in comparison with a DMB-Neu5Ac standard sample and confirmed by its co-eluting fluorescent profile; its concentration was determined via peak area integration.

2.7 Detection of activated sialic acid (CMP-Neu5Ac) via mass spectrometry

Five ml of protonema suspension culture were supplemented to an ammonia concentration of 1% (v/v) and dispersed with an ULTRA-TURRAX®. Afterwards the extracts were cleared via 10 min of centrifugation at 16,100 x g and supernatants were applied on a 10 mg HyperSep™ Hypercarb™ SPE cartridge (Thermo Scientific). Washing of the column was conducted with 1 ml 1% ammonia, and CMP-Neu5Ac was eluted with 600 μ l 50% ACN in 1% ammonia. The elution fraction was vacuum-dried and resuspended in 10 μ l 80 mM formic acid, buffered to pH 9 with ammonia. Five μ l was injected on a Dionex Ultimate 3000 LC-system, using a Hypercarb™ Porous Graphitic Carbon LC Column (150x0.32 mm; Thermo Scientific). Solvent A consisted of 80 mM formic acid, buffered to pH 9, solvent B of 80% ACN in solvent A. At a flow of 6 μ l/min and column oven set to 31°C, initial conditions of 1.3% B were held for 5 min, went to 19% B over 27 min and finally 63% in one minute, which was held for 7 min. The LC was directly coupled to a Bruker amaZone speed ETD ion trap instrument (Bruker, Bremen, Germany) with standard ESI source settings (capillary voltage 4.5 kV, nebulizer gas pressure 0.5 bar, drying gas 5 L/min, 200°C). Spectra were recorded in negative ion data dependent acquisition mode with MS¹ set on m/z = 613.1 which corresponds to the [M-H]⁻-ion of CMP-Neu5Ac, and simulated selected ion monitoring of m/z = 322.0 ([M-H]⁻ of CMP) was performed with MS². Confirmation of CMP-Neu5Ac identity was performed via the measurement of a CMP-Neu5Ac standard demonstrating an identical fragment pattern on MS² level. Quantification of CMP-Neu5Ac amounts were performed via peak area integration and compared to peak areas gained from defined amounts of a CMP-Neu5Ac standard (Sigma-Aldrich).

2.8 Glycoprotein analysis

2.8.1 Protein extraction for mass spectrometric analyses

Six-days-old plant material (1-2 g of fresh weight), cultivated in KnopME at pH 4.5 with 2.5 mM ammonium tartrate (Sigma Aldrich) was harvested over a 100 μ m sieve and resuspended in 3-fold amount extraction buffer (408 mM NaCl, 60 mM Na₂HPO₄x2H₂O, 10.56 mM KH₂PO₄, 60 mM EDTA, 1% protease inhibitor (Sigma Aldrich), pH 7.4). The material was homogenized with an ULTRA-TURRAX® for 10 min at 10,000 rpm on ice. The crude cell lysate was cleared via two successive centrifugation steps for 5 min at 5,000 rpm and 4°C. Extracts were frozen in 150 μ l aliquots, in liquid nitrogen and stored at -80°C.

2.8.2 Glycopeptide analysis

For mass spectrometry (MS) protein extracts were supplemented with 2% SDS and 25 mM DTT (final concentration) and incubated for 10 min at 90°C. After cooling to room temperature, proteins were S-alkylated with 60 mM iodoacetamide for 20 min in darkness. Prior to SDS-PAGE the samples were mixed with 4x sample loading buffer (Bio-Rad, Munich, Germany). Separation of proteins was carried out via SDS-PAGE in 7.5% gels (Ready Gel Tris-HCl; BioRad) in TGS buffer (BioRad) at 120 V. After electrophoresis, the gel was washed three times for 10 min in water followed by a 1 h staining period with PageBlue® Protein Staining Solution (Thermo Fisher Scientific). For MS analysis gel bands corresponding to high molecular weight range proteins were excised. Gel band preparations, MS measurements on a QExactive Plus instrument and data analysis were performed as described previously (Top et al., 2019). Raw data were processed using Mascot Distiller V2.5.1.0 (Matrix Science, USA) and the peak lists obtained were searched with Mascot V2.6.0 against an in-house database containing all *P. patens* V3.3 protein models (Lang et al., 2018) and the sequence of the reporter protein. Glycopeptides were identified from the Mascot mgf files using a custom Perl script. Glycopeptide precursor masses were searched within these files with a precursor mass tolerance of ± 5 ppm and further validated by the presence of typical GlcNAc oxonium ions (m/z -values: $[\text{GlcNAc}]^+ = 204.087$, $[\text{GlcNAc} - \text{H}_2\text{O}]^+ = 186.076$, $[\text{GlcNAc} - 2\text{H}_2\text{O}]^+ = 168.066$, $[\text{GlcNAc} - \text{C}_2\text{H}_4\text{O}_2]^+ = 144.065$, $[\text{GlcNAc} - \text{CH}_6\text{O}_3]^+ = 138.055$, $[\text{GlcNAc} - \text{C}_2\text{H}_6\text{O}_3]^+ = 126.055$), glycan fragment ions ($[\text{GlcNAcHex}]^+ = 366.139$, $[\text{GlcNAcHex}_2]^+ = 528.191$) or in the case of sialic acid containing glycopeptides the presence of Neu5Ac-specific oxonium ions (m/z -values: $[\text{Neu5Ac}]^+ = 292.103$ and $[\text{Neu5Ac} - \text{H}_2\text{O}]^+ = 274.092$) (Halim et al., 2014). A fragment mass tolerance of 0.02 Da was used. A list of the glycopeptides searched and their calculated precursor masses is provided in Supplementary Table S2. Quantification of identified glycopeptides was done using a default MaxQuant search (V1.6.0.16) on the raw data. For each glycopeptide identified from Mascot mgf files, the intensity value (peak area) was extracted from the MaxQuant allPeptides.txt file by matching raw file name, precursor mass and MS² scan number using another custom Perl script.

3 Results

3.1 Generation of stable sialic acid-producing lines

To establish protein sialylation in *P. patens*, six sequences encoding the enzymes required to synthesize, activate, transport and finally transfer sialic acid onto β 1,4-galactosylated *N*-glycans needed to be introduced into the moss genome. These CDSs were distributed to two multi-gene cassettes, with homologous flanks for targeted integration into the *P. patens* genome (Figures 1C and 1D), of a Δ xt/ft (β 1,2-xylosyltransferase and the α 1,3-fucosyltransferase) knock-out line expressing a reporter glycoprotein to verify the *N*-glycan sialylation of recombinant proteins.

To enable the production of sialic acid, the enzymes responsible for its biosynthesis, GNE, NANS and NANP were introduced via two versions of a multi-gene expression construct only differing in the GNE CDS used. One expression construct (GNN) contained the native GNE and in the other the GNE CDS was altered with two point mutations, R263L and R266Q, in the CMP-Neu5Ac binding site (GM). Both constructs bear homologous regions targeting the adenine phosphoribosyltransferase locus, whose disruption leads to interruption of the adenine-salvage-pathway resulting in lines resistant towards the toxic adenine analogue 2-fluoro adenine (2-FA) (Trouiller et al., 2007). The selection on 2-FA resulted in four GNN-lines with the native GNE (GNN1-4) and 64 GM-lines with the mutated GNE, which were directly screened by PCR for targeted integration of the respective construct within the genome. Homologous 5' - and 3' -construct integration was confirmed for two lines harboring the GNN-construct

(GNN2 and GNN4) and 3' homologous integration was confirmed for two further GNN-lines (GNN 1 and GNN3) as well as 26 GM-construct containing lines (GM9, 10-12, 16, 21, 22, 25-28, 36, 41, 44-47, 49, 51, 52, 57 and 61-65), respectively (Supplementary Figures S4 and S5). These lines were chosen for further experiments.

3.1.1 GNN-lines produce sialic acid

In mammals, the coordinated activity of GNE, NANS and NANP leads to the synthesis of free sialic acid. To prove the activity of these enzymes in *P. patens*, sialic acid was quantified in plant extracts via a colorimetric periodate-resorcinol assay, which measures total sialic acid. Three out of four analyzed GNN-lines and 20 out of 22 analyzed GM-lines could be identified as sialic acid producers, while, as expected, the parental line does not show any signal indicating sialic acid production. Sialic acid levels between 9.8 ± 0.5 and 16.5 ± 0.7 $\mu\text{mol/g}$ FW were detected in the GNN lines, whereas similar amounts ranging from 4.1 ± 0.1 up to 15.9 ± 0.5 $\mu\text{mol/g}$ FW were obtained in the GM lines (Figure 2A), indicating that the epimerase and kinase activities of the enzyme with the double mutation are not affected, as previously shown in CHO cells (Son et al., 2011). For the overall best producing line GNN2 presence of sialic acid was confirmed via fluorescence detection of DMB-labeled moss extracts separated over reverse phase HPLC (RP-HPLC-FLD) compared to a DMB-Neu5Ac standard (Figure 2B). According to the peaks' area ratio, 85 μmol Neu5Ac/g DW could be detected in GNN2 extracts, which is in agreement with the result of the colorimetric assay.

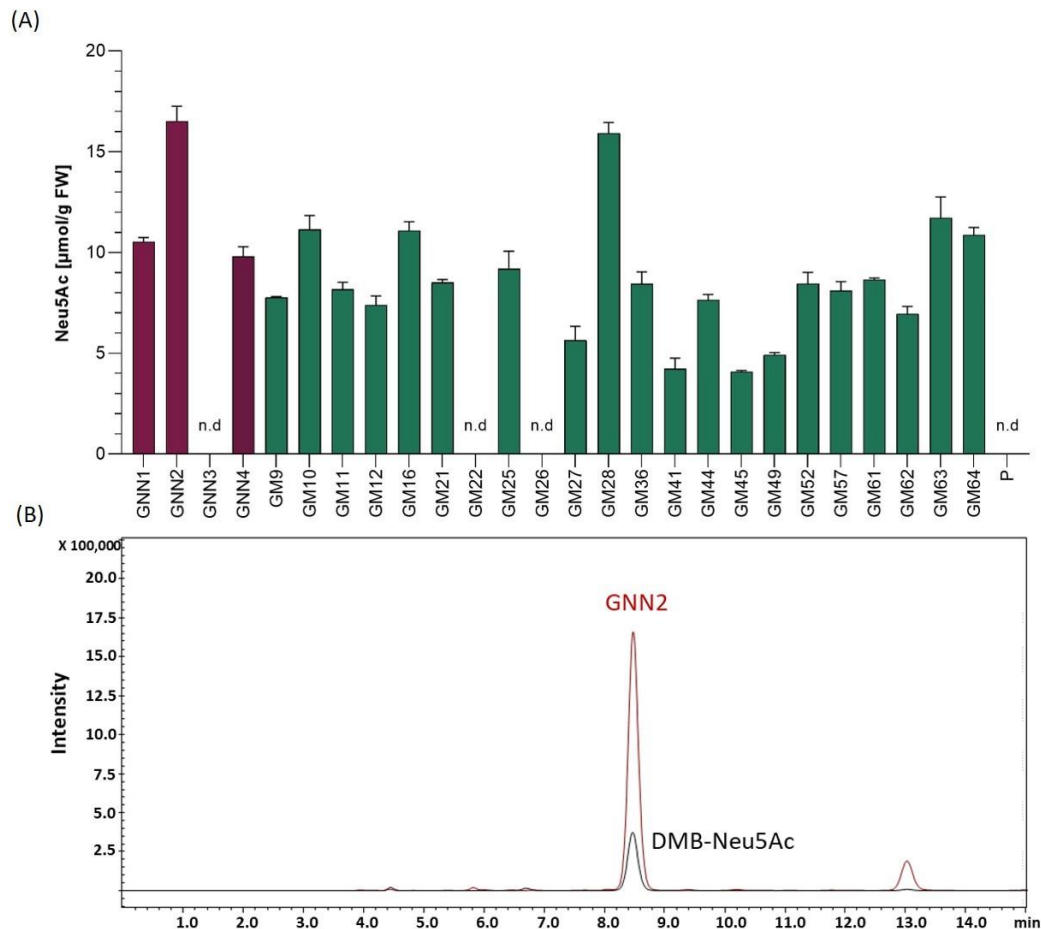


Figure 2. Detection of sialic acid in moss extracts of lines expressing NANS, NANP and either the native GNE (GNN lines) or the mutated version GNE_{mut} (GM lines). **(A)** Detection of sialic acid via periodate-resorcinol assay in protonema extracts of four GNN (purple), 22 GM (green) lines and the parental line (P). Neu5Ac quantification was performed in comparison to a Neu5Ac standard. Extracts were measured in duplicates in serial dilutions ranging from 1:2 to 1:32 and standard deviations (SDs) are given. Absorption was measured at 595 nm. FW: fresh weight, n.d: not detectable. **(B)** Detection of sialic acid in line GNN2 via reverse phase high performance liquid chromatography coupled with fluorescence detection (RP-HPLC-FLD). For fluorescent detection, sialic acid in the protonema extract was derivatized with 1,2-diamino-4,5-methylenedioxybenzene (DMB), resulting in DMB-Neu5Ac. Identity of the DMB-Neu5Ac in the moss extract (red line) was confirmed by comparison to DMB-Neu5Ac standard (black line).

3.1.2 Expression analysis of GNE, NANS and NANP in GNN and GM-lines

Four Neu5Ac-producing lines (GNN1, GNN2, GNN4 and GM28) were characterized via Real-Time Quantitative Reverse Transcription PCR (qRT-PCR) regarding their transgene expression levels. The results are summarized in Figure 3 via the respective $2^{-(\Delta\text{CT})}$ -values which represent the ratios between the expression level of the respective transgene compared to the expression levels of the housekeeping genes EF1 α and L21. All analyzed lines strongly express the introduced transgenes, whereas the values obtained for the expression of the transgenes in the parental line are considered as background signal of the method (Figure 3). The expression level of the transgenes varied in the different lines, notably the expression level of NANP seems to be related to the amount of free sialic acid found in these lines (Figure 2A and Figure 3), indicating that the expression of this enzyme might play an important role in the efficiency of sialic acid synthesis.

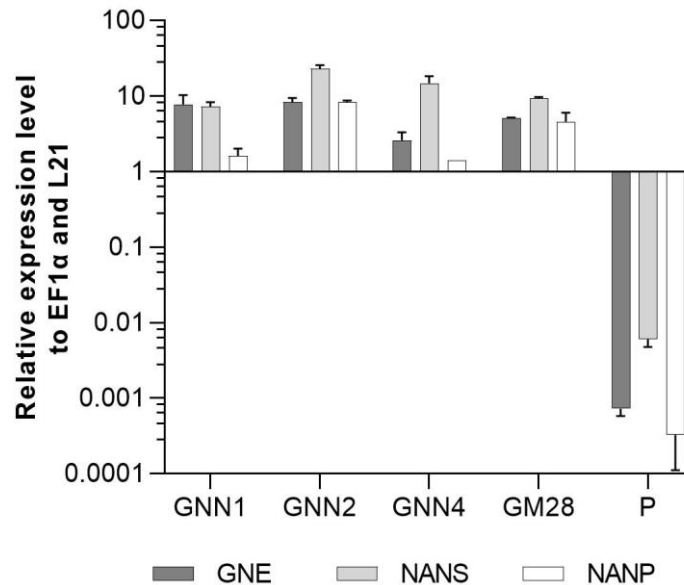


Figure 3. qRT-PCR based expression analysis of GNE, NANS and NANP relative to the expression of the internal housekeeping genes EF1 α and L21. Expression levels of the transgenes were normalized against the expression levels of the two housekeeping genes EF1 α and L21. The relative expression levels are expressed as $2^{-(\Delta C_T)}$, where $\Delta C_T = C_T$ transgene - C_T housekeeping genes. Bars represent mean values \pm SD from $n = 3$. The parental plant (P) was used as negative control.

3.2 Generation of the complete sialylation pathway

After confirming the synthesis of sialic acid, the second part of the sialylation pathway catalyzing the activation, transport and glycosidic transfer of sialic acid was introduced. For this, the best Neu5Ac producing lines GNN2 and GM28 were transfected with the second multi-gene construct, called CCSB, containing the CDSs of CMAS, CSAT, ST as well as a Blastidicin S resistance cassette targeted to the moss P4H1 gene, to prevent plant typical prolyl hydroxylation of recombinant proteins (Parsons et al., 2013). After selection, surviving lines were directly screened by PCR for targeted integration of the CCSB-construct into the P4H1 locus. This resulted in two lines in the GNN2-background (GNC lines 4 and 7, Supplementary Figure S6) and four lines with confirmed 3'-integration in the GM28 background (GMC lines 5, 22, 23 and 46, Supplementary Figure S7).

3.2.1 Expression analysis of CMAS, CSAT and ST

The expression levels of CMAS, CSAT and ST were analyzed via qRT-PCR in all seven lines with targeted integration of the CCSB-construct alongside their respective parental lines GNN2 for GNC and GM28 for GMC as negative controls. The expression levels of previously introduced GNE, NANS and NANP were analyzed anew. In four of these lines, GNC7, GMC5, GMC23 and GMC46, all six transgenes are strongly expressed, ranging from 1.4 to 100 times the expression level of the housekeeping genes EF1 α and L21 (Figure 4). The other three lines showed a low expression level for at least one gene, mainly the CMAS. Unexpectedly, in GNC4 a marked decrease in the expression levels of GNE, NANS and NANP was observed.

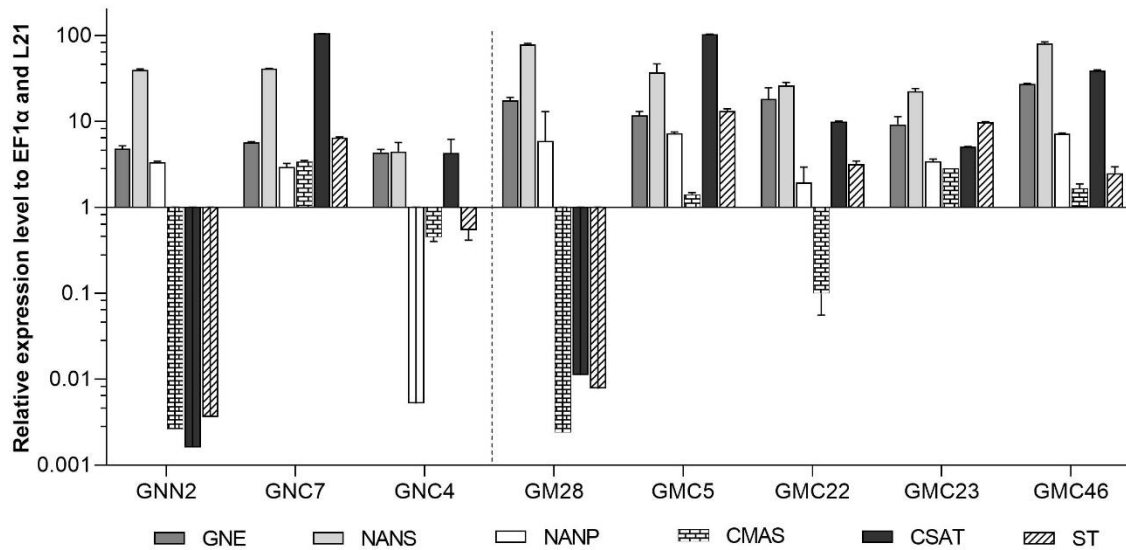


Figure 4. qRT-PCR based expression analysis of all six heterologous sialic acid pathway genes in two GNC lines and their parental line GNN2 and five GMC lines and their parental line GM28 relative to the expression of the internal housekeeping genes EF1 α and L21. Expression levels of the transgenes were normalized against the expression levels of the two housekeeping genes EF1 α and L21. The relative expression levels are expressed as $2^{-\Delta\Delta CT}$, where $\Delta\Delta CT = CT_{transgene} - CT_{housekeeping\ genes}$. Bars represent mean values \pm SD from $n = 3$. Lines expressing the native GNE are displayed on the left side of the dotted line, while the ones harboring the mutated GNE are depicted on the right side.

3.2.2 Mass spectrometric determination of CMP-Neu5Ac in lines expressing the complete sialylation pathway

Activation of the free sialic acid by addition of CMP is a prerequisite for later transfer to the *N*-glycan structure. Therefore, the four lines robustly expressing all six transgenes (GNC7, GMC5, GMC23 and GMC46) were tested regarding their ability to activate sialic acid by measuring the CMP-Neu5Ac content via mass spectrometry in negative ion mode. The expected [CMP-Neu5Ac - H]⁻ at $m/z = 613.1$ on MS¹ level and the CMP-fragment ion [CMP - H]⁻ at $m/z = 322.0$ on MS² level could be detected in all analyzed lines, confirming the presence of CMP-Neu5Ac exemplarily shown for GMC5 in Figure 5. Verification of CMP-Neu5Ac presence in the other lines is depicted in Supplementary Figure S8. Peak area integration of the respective extracted ion chromatograms (EICs) corresponding to the m/z -value of 613.1 in comparison to the standard yielded in CMP-Neu5Ac amounts of only 2 nmol/g DW in line GNC7, in contrast 14 nmol/g DW in the lines GMC23 and GMC46 and 58 nmol/g DW in the line GMC5. This confirmed the activity of the CMAS version used in moss leading to the production of activated sialic acid. Additionally, CMP-Neu5Ac-levels in GNE mutated lines were up to 25-fold higher than those in the native GNE-harboring line GNC7.

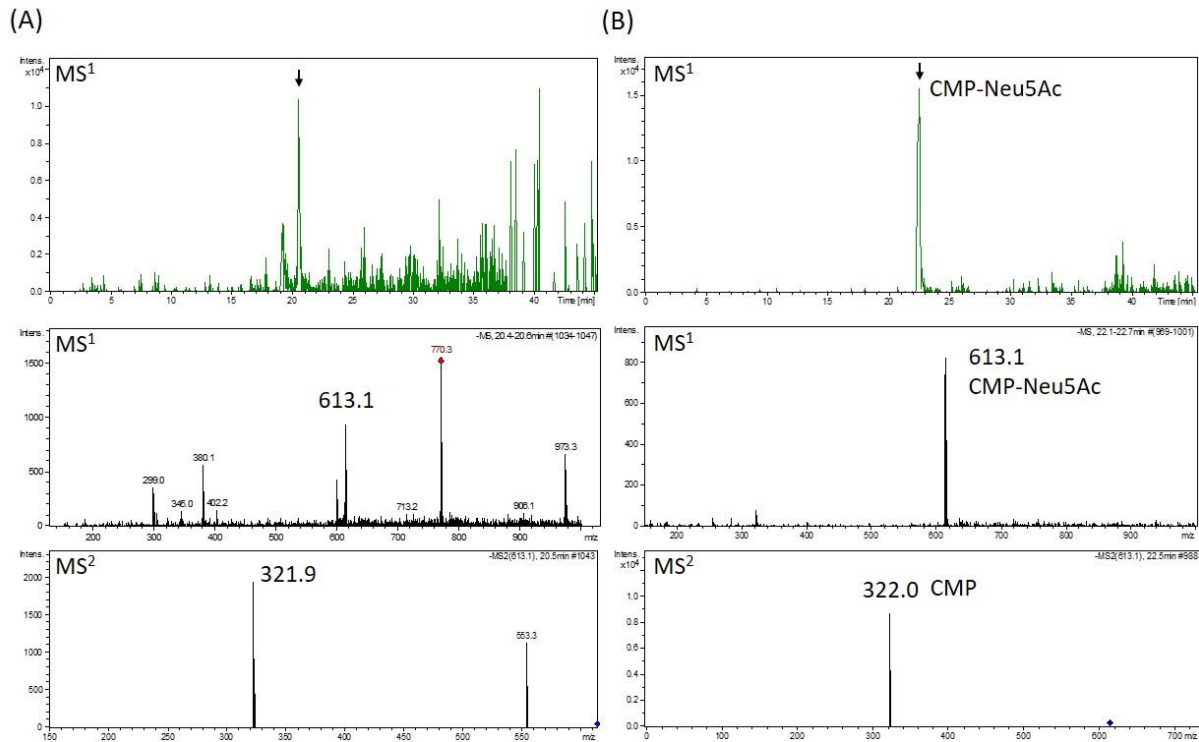


Figure 5. Mass spectrometric detection of CMP-Neu5Ac in a moss extract (A) compared to the measurement of a CMP-Neu5Ac standard (B). Extracted ion chromatogram (m/z of 613.1 ± 0.1) of LC-MS measured extract of line GMC5 (A) or 70 pmol of a CMP-Neu5Ac standard (B), respectively (upper panels). In both measurements, peaks corresponding to the $[M-H]^-$ -ion of CMP-Neu5Ac of $m/z = 613.1$ could be detected on MS^1 level (middle panels). Confirmation of corresponding MS^1 -peak identities (indicated by the black arrows) was performed on MS^2 level via the identification of the $[M-H]^-$ -CMP-fragment ion of $m/z = 322.0$ (lower panels). CMP-Neu5Ac-concentration in GMC5 was determined via peak area integration in comparison to defined standard values and resulted in 58 nmol/g DW.

3.3 CMP-Neu5Ac triggers GNE feedback inhibition, which can be overcome by mutating GNE

Due to the difference in CMP-Neu5Ac amounts between plants harboring the two different versions of GNE, sialic acid content of all four selected lines expressing the six transgenes was quantified again via periodate resorcinol assay in comparison to their respective parental lines. This analysis revealed that with the introduction of the second half of the sialylation pathway the Neu5Ac content declined over 60% in the native GNE-expressing line GNC7 compared to the GNN2 parental line (Figure 6), despite no changes of the GNE, NANS and NANP expression levels being detected (Figure 4). In contrast, in the three analyzed GMC lines 5, 23 and 46, expressing the mutated GNE-version, the sialic acid content remained stable compared to the GM28 parental line (Figure 6). The higher content of CMP-Neu5Ac in lines carrying the GNE_{mut} in comparison to the native GNE containing line GNC7 indicates that the negative feedback of the activated sialic acid on the activity of GNE could be overcome with the double point mutation, which was previously described for CHO cells (Son et al., 2011). Thus, the mutated version of the enzyme is the correct choice to establish protein sialylation in moss. According to these results the GNE-mutated lines were chosen for further experiments.

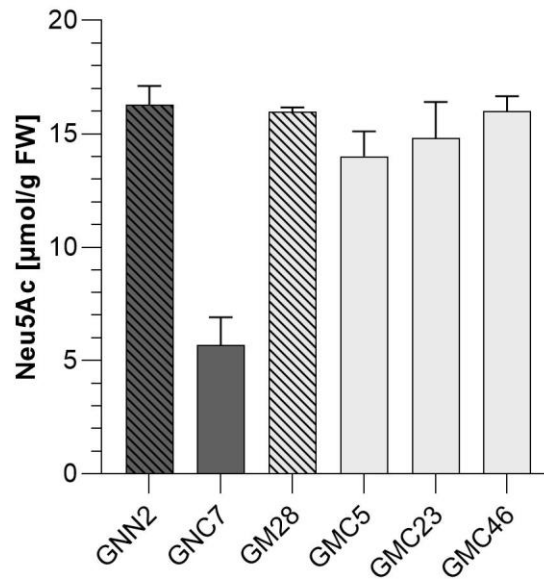


Figure 6. Colorimetric detection of total sialic acid content via periodate-resorcinol assay. Total sialic acid content was measured in protonema extracts of four lines expressing all six transgenes of the sialic acid pathway (GNC7, GMC5, GMC23 and GMC46, solid filled bars) and their respective parental lines GNN2 and GM28 (bars with pattern), both lacking the CMAS responsible for the activation of sialic acid. The lines expressing the native GNE are displayed in dark gray, while lines harboring the mutated GNE-version are depicted in light gray. Neu5Ac quantification was performed in comparison to a Neu5Ac standard. Extracts were measured in duplicates in serial dilutions ranging from 1:2 to 1:32 and SDs are given. Absorption was measured at 595 nm. FW: fresh weight.

3.4 Generation of *N*-glycan sialylation by introduction of the β 1,4-galactosyltransferase

After the introduction of the whole sialylation pathway, we aimed to provide terminal β 1,4-linked galactoses for the anchoring of sialic acid to *N*-glycans. For this, a construct to simultaneously introduce the human β 1,4-galactosyltransferase (GalT4) and to knock out the undesired activity of the endogenous β 1,3-galactosyltransferase 1 (GalT3), responsible for the formation of Le^a epitopes (Parsons et al., 2012), was created. Moreover, as the glycosyltransferases in the Golgi apparatus should act in a sequential fashion, for appropriate localization of the GalT4 in the late plant Golgi apparatus (Bakker et al., 2006), a chimeric version of the enzyme was designed, bearing the CTS of the last acting enzyme in plant *N*-glycosylation, the α 1,4-fucosyltransferase. This CTS region was N-terminally fused to the catalytic domain of the GalT4, resulting in the chimeric FT-CTS-GalT4 (FTGT). FTGT expression was driven by the long CaMV 35S promoter, which is according to Horstmann et al. (2004) four times stronger than the CaMV 35S promoter in *P. patens*. Line GMC23, with a high expression level of all six genes introduced previously, with the GNE_{mut} version, was transformed with the FTGT-encoding construct. After Zeocin selection, resulting lines were screened for homologous integration in the GalT3 locus, which resulted in four lines with targeted FTGT-construct integration (GMC_GT19, GMC_GT25, GMC_GT28 and GMC_GT80, Supplementary Figure S9). These four lines as well as the line GM28 as a negative control, were analyzed for the respective expression level via qRT-PCR. All lines expressed the chimeric β 1,4-galactosyltransferase strongly, between 29 and 155 times higher than the housekeeping genes EF1 α and L21 (Figure 7).

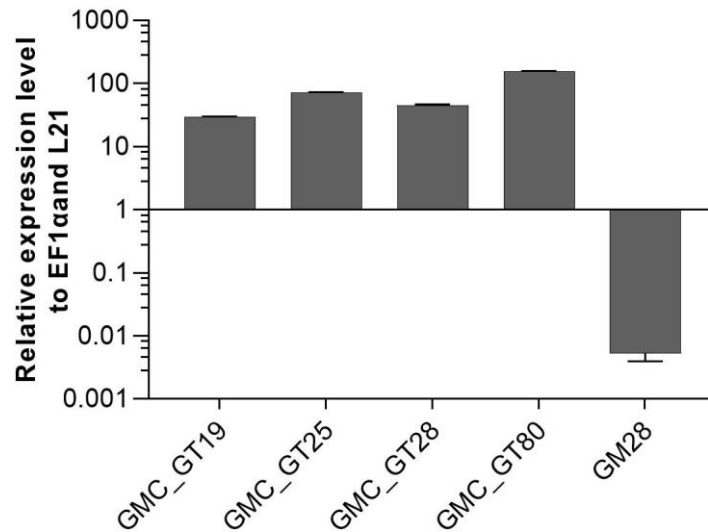


Figure 7. qRT-PCR based expression analysis of chimeric FTGT relative to the expression of the internal housekeeping genes EF1 α and L21. Expression levels of the chimeric β 1,4-galactosyltransferase (FTGT) were determined with a primer pair targeting the catalytic domain of the human GalT4 and normalized internally against the expression levels of the two housekeeping genes EF1 α and L21. The relative expression levels are expressed as $2^{-(\Delta CT)}$, where $\Delta CT = CT \text{ transgene} - CT \text{ housekeeping genes}$. Bars represent mean values \pm SD from $n=3$.

3.5 Mass spectrometric analysis of the *N*-glycosylation pattern of complete sialylation lines

The efficiency of galactosylation and subsequent sialylation of *N*-glycans in the FTGT-expressing lines GMC_GT19, 25, 28 and 80, was assessed on glycopeptides of the stably co-expressed reporter protein via mass spectrometry. Evaluation of the *N*-glycosylation pattern was performed via integration of MS²-confirmed glycopeptide elution profiles on MS¹ level. Our present glycopeptide analysis does not allow us to distinguish conformational isomers of *N*-glycans but enables us to make conclusions about their composition. For simplification in the description of the results, we present all possible isomers of an *N*-glycan combined under one structure, e.g. AM, MA or a mixture of both, are all displayed together as AM. The MS-analyses revealed that the investigated lines displayed a galactosylation efficiency between 45% and 60% within the analyzed glycopeptides (Figure 8). Among galactosylated glycopeptides, 80% corresponded to AM-structures. On average, 8% of the *N*-glycans were biantennary galactosylated (indicated by AA, Figure 8). Further, our data indicate that the degree of bi-galactosylated peptides is higher in plants with lower expression level of the FTGT (Figure 7 and Figure 8). On glycopeptides harboring galactoses mass shifts of 132.0423 Da or 264.0846 Da were observed in up to 50% of all detected galactosylated *N*-glycans. This suggests the linkage of one or two pentoses to the newly introduced β 1,4-linked galactoses (Figure 8 and Supplementary Figure S10). Non-galactosylated glycopeptides carried almost exclusively the complex-type GnGn glycoform. *N*-glycans bearing plant specific xylose, α 1,3-linked fucose, Le^a epitopes or the corresponding aglycons were not detected for any of the analyzed glycopeptides. Despite the fact that with the introduction of a β 1,4-galactosyltransferase into a CMP-Neu5Ac-producing and CSAT and ST-expressing moss line all requirements for sialylation are addressed, in none of the MS-analyzed lines sialylation of *N*-glycans was detectable. A corresponding result was reported for *N. benthamiana* by Kittur et al. (2020).

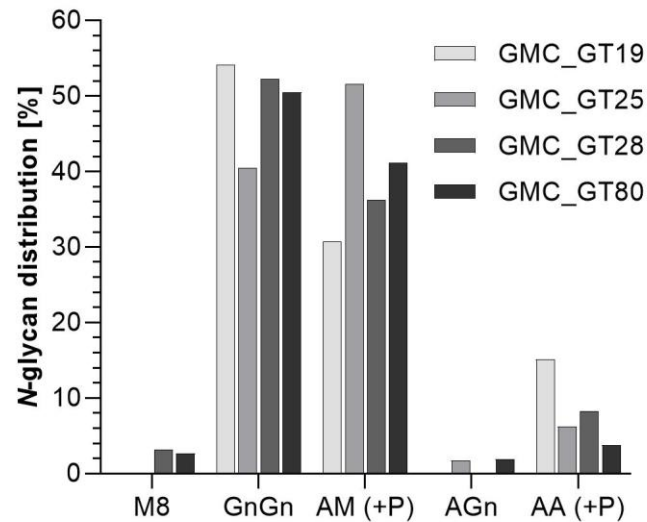


Figure 8. Mass spectrometric analysis of glycosylation patterns in GMC_GT lines. MS-based relative quantification of glycopeptides identified on the tryptically digested reporter glycoprotein of complete sialylation lines. Relative quantification was based on peak area integration of extracted ion chromatograms (EICs) on MS¹ level, for which peak identities were confirmed on MS² level. For quantification, areas of all confirmed peaks per measurement were summed up and the relative percentages are given for each identified glycan structure. M: mannose, Gn: *N*-acetylglucosamine, A: galactose, (+P): indicates that a variable proportion of the corresponding *N*-glycan is decorated with pentoses.

3.6 RNAseq-analysis revealed the drop of ST expression after FTGT introduction

To address the question why no sialic acid was attached to *N*-glycans in the GMC_GT lines harboring all genes necessary for sialylation, we analyzed the changes in gene expression in the line GMC_GT25, and its parental line GMC23, without the chimeric β 1,4-galactosyltransferase. RNAseq analysis revealed that the introduction of FTGT led to a decrease of CMAS and CSAT expression of approximately 95% and nearly abolished the ST expression, whereas the expression levels of the first three pathway CDSs remained stable (Figure 9). Even if the reason for the striking drop in ST expression level could not be explained, we assume that this lack of expression is responsible for the absence of sialylation in the GMC_GT lines.

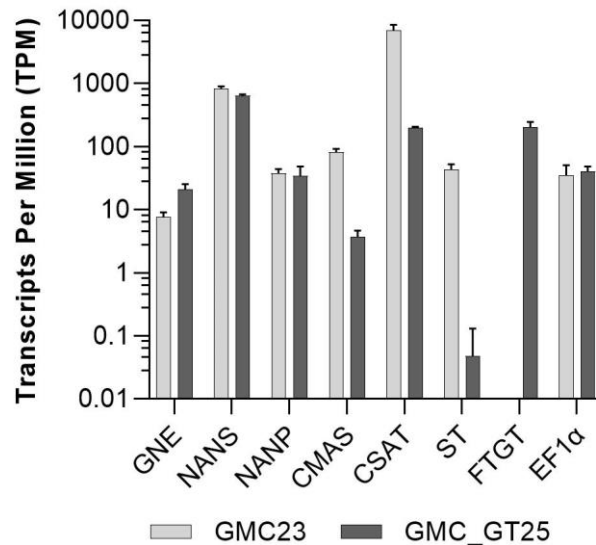


Figure 9. RNAseq expression analysis. Expression levels of all seven pathway related heterologous sequences and the strongly expressed housekeeping gene EF1 α in the lines GMC23 and GMC_GT25. Bars represent mean values \pm SD from n = 3 biological triplicates.

3.7 Lines strongly expressing the FTST are able to perform stable *N*-glycan sialylation

To overcome the absence of ST expression in GMC_GT25, the line was transformed with a chimeric ST version. In this variant, the original CTS of the rat ST was replaced by the CTS of the endogenous α 1,4-fucosyltransferase already used for the FTGT, giving rise to the chimeric variant FTST. The integration of the FTST was targeted to the P4H2 locus, which is part of the plant-prolylhydroxylases family (Parsons et al., 2013). Plants surviving the hygromycin-based selection were analyzed for targeted construct integration within the P4H2 locus, which could be confirmed for eight GMC_GT_FTST lines (GMC_GT_FTST2, 12, 13, 24, 64, 69, 78 and 79; Supplementary Figure S11). Although these plants express several transgenes and display multiple modifications, no obvious impairment in growth could be assessed.

FTST expression was quantified via qRT-PCR in all GMC_GT_FTST lines with targeted genome integration. The used primers were targeted to the catalytic domain of the sialyltransferase, thus amplifying both ST variants. Therefore, the lines GMC_GT_FTST, carrying both the native and the chimeric ST were compared to their parental line GMC_GT25, only expressing the native ST version. As previously detected, the GMC_GT25 displayed only a very weak ST expression, while six out of the eight tested GMC_GT_FTST lines strongly express the chimeric sialyltransferase, in a range between 23 and 104 times the expression level of the housekeeping genes EF1 α and L21 (Figure 10).

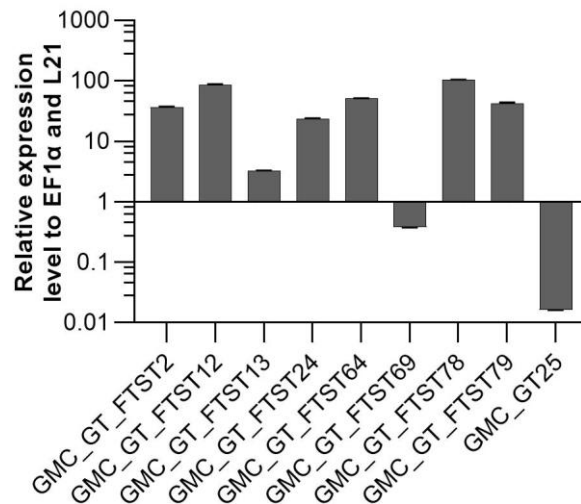


Figure 10. qRT-PCR based expression analysis of chimeric FTST variant relative to the expression of the internal housekeeping genes EF1 α and L21. Expression level of the FTGT variant was determined with a primer pair targeting the catalytic domain of the rat ST. FTST-expression levels were normalized line internally against the expression levels of the two housekeeping genes EF1 α and L21. The relative expression levels are expressed as $2^{(-\Delta CT)}$, where $\Delta CT = CT \text{ transgene} - CT \text{ housekeeping genes}$. Bars represent mean values \pm SD from $n = 3$.

To investigate FTST enzyme activity, three lines with FTST expression levels ranging from low to high (GMC_GT_FTST13, 64 and 78) were chosen for MS-based glycopeptide analysis. This analysis revealed that the galactosylation efficiency increased from 60% in the GMC_GT25 parental line to 85-89% in FTST expressing lines (Supplementary Figure S12). Moreover, a higher share of biantennary galactosylated *N*-glycans could be detected in all analyzed lines after the introduction of the FTST, increasing from 6.3% in GMC23_GT25 to e.g. 33.7% in GMC_GT_FTST64 (Supplementary Figure S12). Further, in the two moderate and strong FTST expressing lines GMC_GT_FTST64 and 78 stable *N*-glycan sialylation could be confirmed, whereas in the weak-expressing line GMC_GT_FTST13 no *N*-glycan sialylation was detectable (Figure 11A and Supplementary Figure S12). Verification of *N*-glycan sialylation was performed via the detection of defined glycopeptide *m/z*-ratios on MS¹ level (Supplementary Figures S13-S16) as well as the detection of sialic acid reporter ions [Neu5Ac - H₂O]⁺ with *m/z*=274.092 and [Neu5Ac]⁺ with *m/z*=292.103 and fragments of the corresponding peptide backbone on MS² level (Figure 11B-11D and Supplementary Figures S13-S16). In GMC_GT_FTST78, the line with the highest FTST expression level, sialylation could be detected in all three analyzed glycopeptides (Figure 11B-11D). Altogether, sialic acid was linked to 6.3% of all detected glycopeptides, of which 6.0% were NaM, while the remaining 0.3% displayed NaGn structures. Further, 70.8% AM and 0.8% AGn structures were detected, but almost half of them were decorated with one or two pentoses. Interestingly, no pentoses were found when these *N*-glycans were sialylated. Biantennary galactosylation was present in 11.3% of the glycopeptides, but no sialylation of this structure could be detected. GMC_GT_FTST64, the analyzed line with the intermediate FTST-expression level, displayed an overall sialylation efficiency of 1% NaM structures (Supplementary Figure S12), indicating that the efficiency of sialylation is dependent on the expression level of the sialyltransferase. These results verified the correct synthesis, localization and sequential activity of seven heterologous mammalian enzymes and confirmed that stable *N*-glycan sialylation is possible in moss.

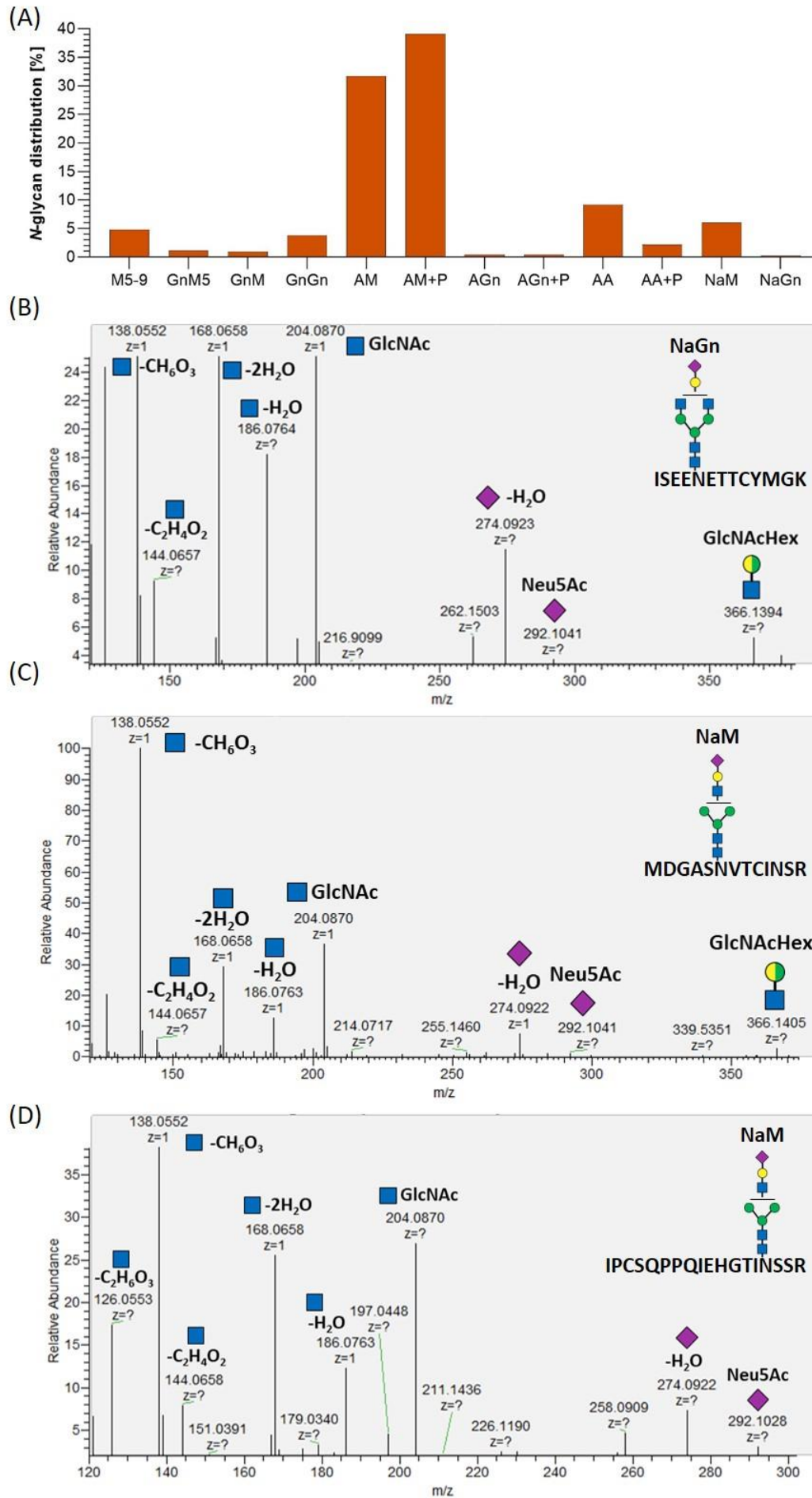


Figure 11. Mass spectrometric analysis of the glycosylation pattern in GMC_GT_FTST78 and MS²-based verification of sialylated glycopeptides. **(A)** MS-based relative quantification of glycopeptides identified on the tryptic digested reporter glycoprotein of the complete sialylation line GMC_GT_FTST78 with a chimeric sialyltransferase. Relative quantification was based on peak area integration of extracted ion chromatograms (EICs) on MS¹ level, for which glycopeptide identities were confirmed on MS² level. For quantification, areas of all confirmed peaks per measurement were summed up and the relative percentages are given for each identified glycan structure. **(B-D)** MS²-based verification of *N*-glycan sialylation on all three analyzed tryptic glycopeptides by the identification of *N*-acetylglucosamine (GlcNAc) and sialic acid (Neu5Ac) reporter ions with the following *m/z*-values: [GlcNAc]⁺ = 204.087, [GlcNAc - H₂O]⁺ = 186.076, [GlcNAc - 2H₂O]⁺ = 168.066, [GlcNAc - C₂H₄O₂]⁺ = 144.065, [GlcNAc - CH₆O₃]⁺ = 138.055, [GlcNAc - C₂H₆O₃]⁺ = 126.055), [Neu5Ac]⁺ = 292.103, [Neu5Ac - H₂O]⁺ = 274.092 and the detection of the glycan fragment ion [GlcNAcHex]⁺ = 366.139. M: mannose (green circle), Gn: *N*-acetylglucosamine (blue square), A: galactose (yellow circle), Na: sialic acid (purple rhombos), Hex : Hexose (yellow and green circle: stands for the presence of either mannose or galactose).

4 Discussion

Since many therapeutics are sialylated glycoproteins, the ability to perform sialylation on complex-type *N*-glycans is highly required for biopharmaceutical production platforms. *N*-glycan sialylation plays an important role for the plasma half-life of glycoproteins, as it protects them from clearance (reviewed in Varki, 2008), which is highly favorable to achieve long lasting therapeutic effects.

To achieve *N*-glycan sialylation in *P. patens* several challenges need to be addressed, namely the production of free sialic acid (Neu5Ac) from the precursor UDP-*N*-acetylglucosamine, which is also present in plants (Stanley et al., 2015), its activation to CMP-Neu5Ac, its transport to the Golgi apparatus and finally its transfer onto a β 1,4-galactosylated *N*-glycan. This requires the coordinated activity of seven mammalian enzymes localized in three different plant subcellular compartments including the nucleus, cytosol and Golgi apparatus. Stable integration of the respective genes was directed via gene targeting, which is a frequently used tool for genome engineering in *P. patens* (e.g. Khraiwesh et al., 2010; Sakakibara et al., 2013; Decker et al., 2015; Horst et al., 2016; Wiedemann et al., 2018). These knock-in approaches were aimed at selection or to avoid undesired plant-typical posttranslational modifications. The adenine phosphoribosyltransferase is an enzyme involved in the purine salvage pathway and recycles adenine into AMP, but can also recognize adenine analogues as substrates, which are toxic when metabolized (Schaff, 1994). Knock-out of the single coding gene for this enzyme in *P. patens* results in resistance to adenine analogues as 2-fluoro adenine; an effect that has been employed as a marker for homologous recombination frequency (Trouiller et al., 2007; Kamisugi et al., 2012). We used this strategy to positively select GNN-transformed plants avoiding the introduction of an additional foreign gene for selection. The integration of the CCSB- and FTST-expression cassettes was targeted to the genes coding for the prolyl-4-hydroxylase 1 (P4H1) and the prolyl-4-hydroxylase 2 (P4H2), respectively, which are members of the plant P4H-family, responsible for hydroxylation of prolines in consensus sequences different to those recognized in humans (Parsons et al., 2013). Moreover, as hydroxyproline serves as anchor for plant-typical O-glycosylation, the formation of these structures should be avoided in biopharmaceuticals. Besides, the generation of the potential immunogenic Le^a-epitope was abolished in *P. patens* by targeting the FTGT-expression cassette to the gene encoding β 1,3-galactosyltransferase 1 (Parsons et al., 2013).

The production of free sialic acid was achieved by the simultaneous expression of the sequences encoding the first three cytosolic active enzymes of the sialic acid biosynthesis pathway: murine GNE, human NANS and human NANP, resulting in the production of Neu5Ac in 10 times higher amounts than previously reported for plants (Castilho et al., 2008). In *A. thaliana* and *N. benthamiana* transformed with GNE and NANS, endogenous activity of a putative plant NANP was sufficient for the

synthesis of Neu5Ac (Castilho et al., 2008). However, our data indicates that in moss additional NANP-expression levels lead to higher Neu5Ac production.

As the full-length CMAS-CDS could not be amplified from human cDNA, a shorter version, where 120 bp at the 5' end are missing (Castilho et al., 2008), was cloned and introduced in sialic acid-producing lines. Moss plants expressing the truncated version of CMAS were able to activate sialic acid, confirming its activity in *P. patens*.

The activity of GNE, the first enzyme in the biosynthesis of sialic acid, is known to be inhibited by the CMAS product CMP-Neu5Ac (Kornfeld et al., 1964). Mutated versions of this enzyme (R263L and/or R266Q/R266W) were already used in *N. benthamiana*, insect and CHO cells to overcome this negative feedback and increase sialylation of recombinant proteins (Viswanathan et al., 2005; Bork et al., 2007; Son et al., 2011; Kallolimath et al., 2016). We expressed in moss the GNE_{mut} version carrying the double mutation R263L, R266Q which was previously described to be more active *in vitro* than the single mutated versions (Son et al., 2011). Sialic acid-synthesizing moss plants carrying GNE_{mut} produced similar amounts of sialic acid as plants expressing the native GNE-version, confirming that the R263L, R266Q mutations in the allosteric site of GNE do not affect the epimerase/kinase activities of the enzyme. In contrast to the native GNE-expressing lines, no decrease in the production of sialic acid was observed in GNE_{mut} expressing lines after the introduction of CMAS, revealing that the negative feedback loop of GNE was successfully prevented in moss. Consequently, up to 25 times more CMP-Neu5Ac was achieved in moss plants expressing the GNE_{mut} compared to lines with the native GNE.

The 2A peptide sequences derived from the *Picornaviridae* family are known to cause a ribosomal skipping event during the process of translation (Szymczak et al., 2004), resulting in isolated proteins from a single open reading frame. The ability of *P. patens* to recognize the 2A peptide sequence from the foot-and-mouth disease virus has already been reported (Pan et al., 2015). We used the P2A sequence (Kim et al., 2011), originated from the porcine teschovirus-1, another virus belonging to this family, to separate the transgene of interest, in our case the chimeric β 1,4-galactosyltransferase, from the *ble* resistance gene. This resulted in the synthesis of independent FTGT and the Bleomycin resistance protein, verified by their corresponding enzymatic activities.

Different qualities and degrees of galactosylation have been reported so far after the introduction of the β 1,4-galactosyltransferase in plants. Enzymes involved in *N*-glycosylation are localized in the ER and/or in the Golgi apparatus according to their sequential manner of action in the *N*-glycan maturation, and this localization is determined by their N-terminal CTS region (Saint-Jore-Dupas et al., 2006). Early incorporation of galactose into the maturing *N*-glycan before the action of mannosidase II leads to hybrid-type glycans (Schachter et al., 1983; Hesselink et al., 2014). Therefore, we aimed to target the β 1,4-galactosyltransferase activity to the trans Golgi apparatus by creating a chimeric enzyme consisting of the catalytic domain of the human β 1,4-galactosyltransferase and the localization-determining CTS region of the moss endogenous α 1,4-fucosyltransferase (FT4), which is the last enzyme known to act on the plant *N*-glycans maturation (Saint-Jore-Dupas et al., 2006). Lines expressing this synthetic GalT4 variant (FTGT) displayed up to 60% of galactosylation on the reporter glycopeptides. Mono- and biantennary galactosylated glycans were detected, up to 50% and 15%, respectively. In addition, in some of the galactosylated *N*-glycans the appearance of mass shifts corresponding to the molecular weight of one or two pentoses occurred, indicating that β 1,4-galactosylated *N*-glycans were decorated with additional sugars of thus far unknown identity. The characterization of these residues remains a future task. A very similar galactosylation pattern, including the appearance of pentoses, was described by Kittur et al. (2020) on erythropoietin in transgenic *N. tabacum* lines expressing a chimeric GalT4 with the CTS region of the rat α 2,6-sialyltransferase (STGT). Interestingly, in transgenic Δ xt/ft *N. benthamiana* lines the same STGT variant led to predominantly bi-galactosylated *N*-glycan structures on a transiently co-expressed antibody

(Strasser et al., 2009). However, the extent of biantennary galactosylation also seems to depend on the investigated glycoprotein, as in contrast to the high galactosylation efficiencies obtained on antibodies, poor levels were achieved for other proteins in the same system (Kriechbaum et al., 2020). Further, our results indicate that the expression level of the enzyme plays a role in the quality of galactosylation. Earlier studies performed in *N. benthamiana* indicated that there is an optimal GalT expression level and levels beyond it resulted in higher amounts of immature *N*-glycans (Kallolimath et al., 2018). This is in agreement with our results, since the amount of bi-galactosylated *N*-glycans was higher in lines with a lower FTGT expression level. Therefore, an improvement of the glycosylation pattern towards a higher share of bi-galactosylated *N*-glycan structures via modulation of promoter strength or number of inserted copies in stably transformed moss plants is a future task.

Achieving stable sialylation in *P. patens* requires not just the correct subcellular localization and coordinated activity of seven mammalian enzymes involved in this process, but also their optimal expression levels to produce fully processed glycans with terminal sialic acid. Since the same promoter was used to express all genes, the differences in the expression levels observed between moss transgenic lines should mainly be determined by the number of constructs integrated into the genome or the loci of integration, respectively. The use of homologous flanks to target the integration in *P. patens* do not always limit this recombination process to the intended locus alone. In contrast to transient systems, the process of plant selection according to the expression level of the gene of interest is a necessary step. According to RNAseq data, the introduction of the FTGT resulted additionally in a drop of CMAS and CSAT and a dramatically decrease of ST expression levels. It has been reported that during moss transfection rearrangements may appear, such as small deletions, insertions or concatenation of the constructs, which are characteristic of non-homologous end-joining or homologous recombination pathways (Kamisugi et al., 2006; Murén et al., 2009). These events might be an explanation for the marked decrease in the expression level of these genes, since some copies of the transgenes might be damaged by recombination events. However, we currently cannot explain why ST expression was especially affected.

To circumvent the loss of ST expression, we created and inserted the chimeric gene for FTST into the genome of the production line. The efficient expression of this chimeric ST variant led to stable sialylation of the reporter glycoprotein. The observed structures were NaM and NaGn *N*-glycans, or their isomers, appearing on all analyzed glycopeptides. Furthermore, FTST expression led to an increased galactosylation efficiency from 60 up to 89% combined with an increase in the share of biantennary galactosylated *N*-glycans. Although the sialylation levels achieved in this study are still low (around 6.3%), this is an important step towards the production of sialylated, high-quality biopharmaceuticals in *P. patens*. Different strategies need to be addressed to further optimize the expression levels of the genes involved in the sialylation pathway. It is also well known that culture conditions affect the glycosylation pattern in CHO cells, such as culture pH, temperature, media and supplements, operation mode of the bioreactor and feeding strategies (Hossler et al., 2009; Lewis et al., 2016; Wang et al., 2018; Ehret et al., 2019). Therefore, evaluating the influence of cell culture factors might be a further strategy to increase sialylation levels in *P. patens*.

We established stable *N*-glycan sialylation in *P. patens* by the introduction of seven mammalian enzymes, some of them as chimeric versions, resulting in a sialylated co-expressed human glycoprotein. This demonstrates the coordinated activity and correct subcellular localization of the corresponding enzymes within the cytosol, the nucleus and the Golgi apparatus. Stable expression of all transgenes involved in both the production of the reporter protein as well as the glycosylation, enables the complete characterization of the production line and ensures consistent product quality. The now established ability of stably transformed *P. patens* to sialylate proteins widens the range of *N*-glycosylation patterns attainable in this system considerably. This facilitates the customized production of recombinant biopharmaceuticals with various degrees of post-translational

modifications to ensure maximal activity. This makes moss a versatile and attractive production system within the biopharmaceutical industry.

5 Author Contributions

LLB performed most of the experiments, SNWH performed the MS data analysis, CR performed the RNAseq data analysis, TL cloned the GNN, GM and CCSB constructs and FRJ cloned the FTGT construct, RF performed the mass-spectrometric based Neu5Ac and CMP-Neu5Ac analyses, LLB, JP, NRM, FA, BF, RR and ELD designed the study and wrote the manuscript.

6 Funding

This work was supported by the Excellence Initiative and Strategy of the German Federal and State Governments (GSC-4 SGBM to FRJ and EXC 294 BIOSS to RR), Eleva GmbH (to RR) and the Deutscher Akademischer Austauschdienst DAAD (to NRM).

7 Acknowledgements

We thank Agnes Novakovic, Ingrid Heger, Kai Clemens Carsten Gaa and Maxime Karrer for their support to this work, Prof. Dr. Mitsuyasu Hasebe for kindly providing the BSD plasmid, Dr. Karsten Häffner for providing human and rat cDNA, Dr. Asifa Akhtar for providing murine cDNA, and the Genomics Unit of Instituto Gulbenkian de Ciencia for RNAseq analyses. We thank Prof. Dr. Bettina Warscheid for the possibility to use the QExactive Plus instrument and Anne Katrin Prowse for proof-reading of the article.

8 Conflict of Interests

The authors are inventors of patents and patent applications related to the production of recombinant proteins in moss. RR is an inventor of the moss bioreactor and a founder of Greenovation Biotech, now Eleva GmbH. He currently serves as advisory board member of this company.

9 References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi:10.1093/nar/gky379.
- Aoki, K., Ishida, N., and Kawakita, M. (2003). Substrate recognition by nucleotide sugar transporters further characterization of substrate recognition regions by analyses of UDP-galactose/CMP-sialic acid transporter chimeras and biochemical analysis of the substrate specificity of parental and chimeric transporters. *J. Biol. Chem.* 278, 22887–22893. doi:10.1074/jbc.M302620200.
- Bakker, H., Bardor, M., Molthoff, J. W., Gomord, V., Elbers, I., Stevens, L. H., et al. (2001). Galactose-extended glycans of antibodies produced by transgenic plants. *PNAS* 98, 2899–2904. doi:10.1073/pnas.031419998.

- Bakker, H., Routier, F., Ashikov, A., Neumann, D., Bosch, D. and Gerardy-Schahn, R. (2008). A CMP-sialic acid transporter cloned from *Arabidopsis thaliana*. *Carbohydrate Research* 343, 2148–2152. doi:10.1016/j.carres.2008.01.010.
- Bakker, H., Rouwendal, G. J. A., Karnoup, A. S., Florack, D. E. A., Stopen, G. M., Helsper, J. P. F. G., et al. (2006). An antibody produced in tobacco expressing a hybrid β -1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes. *PNAS* 103, 7577–7582. doi:10.1073/pnas.0600879103.
- Bardor, M., Faveeuw, C., Fitchette, A.-C., Gilbert, D., Galas, L., Trottein, F., et al. (2003). Immunoreactivity in mammals of two typical plant glyco-epitopes, core α (1,3)-fucose and core xylose. *Glycobiology* 13, 427–434. doi:10.1093/glycob/cwg024.
- Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLOS ONE* 10, e0128036. doi:10.1371/journal.pone.0128036.
- Beike, A. K., Lang, D., Zimmer, A. D., Wüst, F., Trautmann, D., Wiedemann, G., et al. (2015). Insights from the cold transcriptome of *Physcomitrella patens* : global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytologist* 205, 869–881. doi:10.1111/nph.13004.
- Bork, K., Reutter, W., Weidemann, W., and Horstkorte, R. (2007). Enhanced sialylation of EPO by overexpression of UDP-GlcNAc 2-epimerase/ManAc kinase containing a sialuria mutation in CHO cells. *FEBS Lett.* 581, 4195–4198. doi:10.1016/j.febslet.2007.07.060.
- Bosch, D., Castilho, A., Loos, A., Schots, A., and Steinkellner, H. (2013). N-glycosylation of plant-produced recombinant proteins. *Curr. Pharm. Des.* 19, 5503–5512.
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34, 525–527. doi:10.1038/nbt.3519.
- Castilho, A., Gattinger, P., Grass, J., Jez, J., Pabst, M., Altmann, F., et al. (2011). N-glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex N-glycans. *Glycobiology* 21, 813–823. doi:10.1093/glycob/cwr009.
- Castilho, A., Pabst, M., Leonard, R., Veit, C., Altmann, F., Mach, L., et al. (2008). Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant Physiol* 147, 331–339. doi:10.1104/pp.108.117572.
- Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., et al. (2010). In planta protein sialylation through overexpression of the respective mammalian pathway. *J. Biol. Chem.* 285, 15923–15930. doi:10.1074/jbc.M109.088401.
- Chou, H.-H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., et al. (2002). Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11736–11741. doi:10.1073/pnas.182257399.

- Chung, C. H., Mirakhur, B., Chan, E., Le, Q.-T., Berlin, J., Morse, M., et al. (2008). Cetuximab-induced anaphylaxis and IgE specific for galactose- α -1,3-galactose. *N. Engl. J. Med.* 358, 1109–1117. doi:10.1056/NEJMoa074943.
- Commandeur, U., Twyman, R. M., and Fischer, R. (2003). The biosafety of molecular farming in plants. *AgBiotechNet. The online service for agricultural biotechnology* 5, 9.
- Cox, K. M., Sterling, J. D., Regan, J. T., Gasdaska, J. R., Frantz, K. K., Peele, C. G., et al. (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat. Biotechnol.* 24, 1591–1597. doi:10.1038/nbt1260.
- D'Aoust, M.-A., Couture, M. M.-J., Charland, N., Trépanier, S., Landry, N., Ors, F., et al. (2010). The production of hemagglutinin-based virus-like particles in plants: a rapid, efficient and safe response to pandemic influenza. *Plant Biotechnol. J.* 8, 607–619. doi:10.1111/j.1467-7652.2009.00496.x.
- Decker, E. L., Parsons, J., and Reski, R. (2014). Glyco-engineering for biopharmaceutical production in moss bioreactors. *Frontiers in Plant Science* 5. doi:10.3389/fpls.2014.00346.
- Decker, E. L., and Reski, R. (2012). Glycoprotein production in moss bioreactors. *Plant Cell Rep* 31, 453–460. doi:10.1007/s00299-011-1152-5.
- Decker, E. L., and Reski, R. (2020). Mosses in biotechnology. *Current Opinion in Biotechnology* 61, 21–27. doi:10.1016/j.copbio.2019.09.021.
- Decker, E. L., Wiedemann, G., and Reski, R. (2015). Gene targeting for Precision glyco-engineering: Production of biopharmaceuticals devoid of plant-typical glycosylation in moss bioreactors in *Glyco-Engineering: Methods and Protocols* Methods in Molecular Biology., ed. A. Castilho (New York, NY: Springer), 213–224. doi:10.1007/978-1-4939-2760-9_15.
- Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., et al. (2001). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* 82, 1013–1025. doi:10.1099/0022-1317-82-5-1013.
- Ehret, J., Zimmermann, M., Eichhorn, T., and Zimmer, A. (2019). Impact of cell culture media additives on IgG glycosylation produced in Chinese hamster ovary cells. *Biotechnol Bioeng* 116, 816–830. doi:10.1002/bit.26904.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi:10.1093/bioinformatics/btw354.
- Fitchette, A.-C., Cabanes-Macheteau, M., Marvin, L., Martin, B., Satiat-Jeunemaitre, B., Gomord, V., et al. (1999). Biosynthesis and immunolocalization of Lewis a-containing N-Glycans in the plant cell. *Plant Physiology* 121, 333–344. doi:10.1104/pp.121.2.333.

- Frank, W., Decker, E. L., and Reski, R. (2005). Molecular tools to study *Physcomitrella patens*. *Plant Biology* 7, 220–227. doi:10.1055/s-2005-865645.
- Fujiyama, K., Furukawa, A., Katsura, A., Misaki, R., Omasa, T., and Seki, T. (2007). Production of mouse monoclonal antibody with galactose-extended sugar chain by suspension cultured tobacco BY2 cells expressing human $\beta(1,4)$ -galactosyltransferase. *Biochemical and Biophysical Research Communications* 358, 85–91. doi:10.1016/j.bbrc.2007.04.054.
- Ghaderi, D., Zhang, M., Hurtado-Ziola, N., and Varki, A. (2012). Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol. Genet. Eng. Rev.* 28, 147–175. doi:10.5661/bger-28-147.
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* 6, 343–345. doi:10.1038/nmeth.1318.
- Gomord, V., Chamberlain, P., Jefferis, R., and Faye, L. (2005). Biopharmaceutical production in plants: problems, solutions and opportunities. *Trends in Biotechnology* 23, 559–565. doi:10.1016/j.tibtech.2005.09.003.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–1186. doi:10.1093/nar/gkr944.
- Häffner, K., Parsons, J., Bohlender, L. L., Hoernstein, S., Niederkrüger, H., Fode, B., et al. (2017). Treatment of experimental C3 Glomerulopathy by human complement factor H produced in glycosylation-optimized *Physcomitrella patens*. *Molecular Immunology* 89, 120. doi:10.1016/j.molimm.2017.06.050.
- Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., et al. (2014). Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC–MS/MS of glycopeptides. *J. Proteome Res.* 13, 6024–6032. doi:10.1021/pr500898r.
- Hamadeh, R. M., Jarvis, G. A., Galili, U., Mandrell, R. E., Zhou, P., and Griffiss, J. M. (1992). Human natural anti-Gal IgG regulates alternative complement pathway activation on bacterial surfaces. *J. Clin. Invest.* 89, 1223–1235. doi:10.1172/JCI115706.
- Hanania, U., Ariel, T., Tekoah, Y., Fux, L., Sheva, M., Gubbay, Y., et al. (2017). Establishment of a tobacco BY2 cell line devoid of plant-specific xylose and fucose as a platform for the production of biotherapeutic proteins. *Plant Biotechnol. J.* 15, 1120–1129. doi:10.1111/pbi.12702.
- Hara, S., Takemori, Y., Yamaguchi, M., Nakamura, M., and Ohkura, Y. (1987). Fluorometric high-performance liquid chromatography of *N*-acetyl- and *N*-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. *Analytical Biochemistry* 164, 138–145. doi:10.1016/0003-2697(87)90377-0.

- Hennermann, J. B., Arash-Kaps, L., Fekete, G., Schaaf, A., Busch, A., and Frischmuth, T. (2019). Pharmacokinetics, pharmacodynamics, and safety of moss- α -Galactosidase A in patients with Fabry disease. *J. Inherit. Metab. Dis.* 42, 527–533. doi:10.1002/jimd.12052.
- Hesselink, T., Rouwendal, G. J. A., Henquet, M. G. L., Florack, D. E. A., Helsper, J. P. F. G., and Bosch, D. (2014). Expression of natural human β 1,4-GalT1 variants and of non-mammalian homologues in plants leads to differences in galactosylation of *N*-glycans. *Transgenic Res* 23, 717–728. doi:10.1007/s11248-014-9806-z.
- Hintze, S., Limmer, S., Dabrowska-Schlepp, P., Berg, B., Krieghoff, N., Busch, A., et al. (2020). Moss-Derived Human Recombinant GAA Provides an Optimized Enzyme Uptake in Differentiated Human Muscle Cells of Pompe Disease. *Int J Mol Sci* 21. doi:10.3390/ijms21072642.
- Horst, N. A., Katz, A., Pereman, I., Decker, E. L., Ohad, N., and Reski, R. (2016). A single homeobox gene triggers phase transition, embryogenesis and asexual reproduction. *Nature Plants* 2, 15209. doi:10.1038/nplants.2015.209.
- Horstmann, V., Huether, C. M., Jost, W., Reski, R., and Decker, E. L. (2004). Quantitative promoter analysis in *Physcomitrella patens*: a set of plant vectors activating gene expression within three orders of magnitude. *BMC Biotechnology* 4, 13. doi:10.1186/1472-6750-4-13.
- Hossler, P., Khattak, S. F., and Li, Z. J. (2009). Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19, 936–949. doi:10.1093/glycob/cwp079.
- Huether, C. M., Lienhart, O., Baur, A., Stemmer, C., Gorr, G., Reski, R., et al. (2005). Glyco-engineering of moss lacking plant-specific sugar residues. *Plant Biol (Stuttg)* 7, 292–299. doi:10.1055/s-2005-837653.
- Jansing, J., Sack, M., Augustine, S. M., Fischer, R., and Bortesi, L. (2018). CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in *Nicotiana benthamiana* for the production of recombinant proteins lacking β -1,2-xylose and core α -1,3-fucose. *Plant Biotechnol. J.* doi:10.1111/pbi.12981.
- Jefferis, R. (2005). glycosylation of recombinant antibody therapeutics. *Biotechnology Progress* 21, 11–16. doi:10.1021/bp040016j.
- Jenkins, N., Murphy, L., and Tyther, R. (2008). Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. *Mol. Biotechnol.* 39, 113–118. doi:10.1007/s12033-008-9049-4.
- Jourdain, G. W., Dean, L., and Roseman, S. (1971). The sialic acids XI. A periodate-resorcinol method for the quantitative estimation of free sialic acids and their glycosides. *J. Biol. Chem.* 246, 430–435.
- Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., et al. (2016). Engineering of complex protein sialylation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9498–9503. doi:10.1073/pnas.1604371113.

- Kallolimath, S., Gruber, C., Steinkellner, H., and Castilho, A. (2018). Promoter choice impacts the efficiency of plant glyco-engineering. *Biotechnol J* 13. doi:10.1002/biot.201700380.
- Kamisugi, Y., Schaefer, D. G., Kozak, J., Charlot, F., Vrielynck, N., Holá, M., et al. (2012). MRE11 and RAD50, but not NBS1, are essential for gene targeting in the moss *Physcomitrella patens*. *Nucleic Acids Res* 40, 3496–3510. doi:10.1093/nar/gkr1272.
- Kamisugi, Y., Schlink, K., Rensing, S. A., Schween, G., von Stackelberg, M., Cuming, A. C., et al. (2006). The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Res.* 34, 6205–6214. doi:10.1093/nar/gkl832.
- Khraiwesh, B., Arif, M. A., Seumel, G. I., Ossowski, S., Weigel, D., Reski, R., et al. (2010). Transcriptional control of gene expression by MicroRNAs. *Cell* 140, 111–122. doi:10.1016/j.cell.2009.12.023.
- Kim, J. H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K. Y., et al. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLOS ONE* 6, e18556. doi:10.1371/journal.pone.0018556.
- Kittur, F. S., Hung, C.-Y., Zhu, C., Shajahan, A., Azadi, P., Thomas, M. D., et al. (2020). Glycoengineering tobacco plants to stably express recombinant human erythropoietin with different *N*-glycan profiles. *International Journal of Biological Macromolecules* 157, 158–169. doi:10.1016/j.ijbiomac.2020.04.199.
- Kolarich, D., Jensen, P. H., Altmann, F., and Packer, N. H. (2012). Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protocols* 7, 1285–1298. doi:10.1038/nprot.2012.062.
- Komatsu, E., Buist, M., Roy, R., Gomes de Oliveira, A. G., Bodnar, E., Salama, A., et al. (2016). Characterization of immunoglobulins through analysis of *N*-glycopeptides by MALDI-TOF MS. *Methods* 104, 170–181. doi:10.1016/j.ymeth.2016.01.005.
- Koprivova, A., Altmann, F., Gorr, G., Kopriva, S., Reski, R., and Decker, E. L. (2003). *N*-glycosylation in the moss *Physcomitrella patens* is organized similarly to that in higher plants. *Plant Biology* 5, 582–591. doi:10.1055/s-2003-44721.
- Koprivova, A., Stemmer, C., Altmann, F., Hoffmann, A., Kopriva, S., Gorr, G., et al. (2004). Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic *N*-glycans. *Plant Biotechnology Journal* 2, 517–523. doi:10.1111/j.1467-7652.2004.00100.x.
- Kornfeld, S., Kornfeld, R., Neufeld, E. F., and O'Brien, P. J. (1964). The feedback control of sugar nucleotide biosynthesis in liver. *Proc Natl Acad Sci U S A* 52, 371–379.
- Kriechbaum, R., Ziaee, E., Grünwald-Gruber, C., Buscaill, P., van der Hoorn, R. A. L., and Castilho, A. (2020). BGAL1 depletion boosts the level of β -galactosylation of *N*- and *O*-glycans in *N. benthamiana*. *Plant Biotechnol. J.* 18, 1537–1549. doi:10.1111/pbi.13316.

- Kubo, M., Imai, A., Nishiyama, T., Ishikawa, M., Sato, Y., Kurata, T., et al. (2013). System for stable β -estradiol-inducible gene expression in the moss *Physcomitrella patens*. *PLOS ONE* 8, e77356. doi:10.1371/journal.pone.0077356.
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi:10.1111/tpj.13801.
- Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.-C., Gomord, V., and Faye, L. (1998). N-Glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol Biol* 38, 31–48. doi:10.1023/A:1006012005654.
- Lewis, A. M., Croughan, W. D., Aranibar, N., Lee, A. G., Warrack, B., Abu-Absi, N. R., et al. (2016). Understanding and controlling sialylation in a CHO Fc-fusion process. *PLOS ONE* 11, e0157111. doi:10.1371/journal.pone.0157111.
- Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012). The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis. *Biotechnology Journal* 7, 1462–1472. doi:10.1002/biot.201200078.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta C(T))}$ Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262.
- Macaulay, I. C., Teng, M. J., Haerty, W., Kumar, P., Ponting, C. P., and Voet, T. (2016). Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nature Protocols* 11, 2081–2103. doi:10.1038/nprot.2016.138.
- Maliekal, P., Vertommen, D., Delpierre, G., and Van Schaftingen, E. (2006). Identification of the sequence encoding N-acetylneuraminic acid-9-phosphate phosphatase. *Glycobiology* 16, 165–172. doi:10.1093/glycob/cwj050.
- Margolin, E. A., Strasser, R., Chapman, R., Williamson, A.-L., Rybicki, E. P., and Meyers, A. E. (2020). Engineering the plant secretory pathway for the production of next-generation pharmaceuticals. *Trends in Biotechnology* 38, 1034–1044. doi:10.1016/j.tibtech.2020.03.004.
- Margolin, E., Chapman, R., Williamson, A., Rybicki, E. P., and Meyers, A. E. (2018). Production of complex viral glycoproteins in plants as vaccine immunogens. *Plant Biotechnol J* 16, 1531–1545. doi:10.1111/pbi.12963.
- Mercx, S., Smargiasso, N., Chaumont, F., De Pauw, E., Boutry, M., and Navarre, C. (2017). Inactivation of the $\beta(1,2)$ -xylosyltransferase and the $\alpha(1,3)$ -fucosyltransferase genes in *Nicotiana tabacum* BY-2 Cells by a multiplex CRISPR/Cas9 strategy results in glycoproteins without plant-specific glycans. *Front Plant Sci* 8. doi:10.3389/fpls.2017.00403.
- Merlin, M., Gecchele, E., Capaldi, S., Pezzotti, M., and Avesani, L. (2014). Comparative evaluation of recombinant protein production in different biofactories: The green perspective. *BioMed Research International*. doi:10.1155/2014/136419.

- Michelfelder, S., Parsons, J., Bohlender, L. L., Hoernstein, S. N. W., Niederkrüger, H., Busch, A., et al. (2017). Moss-produced, glycosylation-optimized human factor H for therapeutic application in complement disorders. *Journal of the American Society of Nephrology : JASN* 28, 1462. doi:10.1681/ASN.2015070745.
- Misaki, R., Kimura, Y., Palacpac, N. Q., Yoshida, S., Fujiyama, K., and Seki, T. (2003). Plant cultured cells expressing human β 1,4-galactosyltransferase secrete glycoproteins with galactose-extended N-linked glycans. *Glycobiology* 13, 199–205. doi:10.1093/glycob/cwg021.
- Moorkens, E., Meuwissen, N., Huys, I., Declerck, P., Vulto, A. G., and Simoens, S. (2017). The market of biopharmaceutical medicines: A snapshot of a diverse industrial landscape. *Front Pharmacol* 8. doi:10.3389/fphar.2017.00314.
- Murén, E., Nilsson, A., Ulfstedt, M., Johansson, M., and Ronne, H. (2009). Rescue and characterization of episomally replicating DNA from the moss *Physcomitrella*. *PNAS* 106, 19444–19449. doi:10.1073/pnas.0908037106.
- Nguyen, T. S., Misaki, R., Ohashi, T., and Fujiyama, K. (2020). Enhancement of sialylation in rlgG in glyco-engineered Chinese hamster ovary cells. *Cytotechnology* 72, 343–355. doi:10.1007/s10616-020-00381-z.
- Palacpac, N. Q., Yoshida, S., Sakai, H., Kimura, Y., Fujiyama, K., Yoshida, T., et al. (1999). Stable expression of human β 1,4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns. *PNAS* 96, 4692–4697. doi:10.1073/pnas.96.8.4692.
- Pan, X.-W., Han, L., Zhang, Y.-H., Chen, D.-F., and Simonsen, H. T. (2015). Sclareol production in the moss *Physcomitrella patens* and observations on growth and terpenoid biosynthesis. *Plant Biotechnol Rep* 9, 149–159. doi:10.1007/s11816-015-0353-8.
- Parsons, J., Altmann, F., Arrenberg, C. K., Koprivova, A., Beike, A. K., Stemmer, C., et al. (2012). Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* 10, 851–861. doi:10.1111/j.1467-7652.2012.00704.x.
- Parsons, J., Altmann, F., Graf, M., Stadlmann, J., Reski, R., and Decker, E. L. (2013). A gene responsible for prolyl-hydroxylation of moss-produced recombinant human erythropoietin. *Scientific Reports* 3, 3019. doi:10.1038/srep03019.
- Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* 9, 171–181. doi:10.1038/nprot.2014.006.
- Reinke, S. O., Lehmer, G., Hinderlich, S., and Reutter, W. (2009). Regulation and pathophysiological implications of UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE) as the key enzyme of sialic acid biosynthesis. *Biol. Chem.* 390, 591–599. doi:10.1515/BC.2009.073.
- Reski, R., Bae, H., and Simonsen, H. T. (2018). *Physcomitrella patens*, a versatile synthetic biology chassis. *Plant Cell Reports* 37, 1409–1417. doi:10.1007/s00299-018-2293-6.

- Roseman, S., Jourdian, G. W., Watson, D., and Rood, R. (1961). Enzymatic synthesis of sialic acid 9-phosphates *. *Proc Natl Acad Sci U S A* 47, 958–961.
- Rozov, S. M., Permyakova, N. V., and Deineko, E. V. (2018). Main strategies of plant expression system glycoengineering for producing humanized recombinant pharmaceutical proteins. *Biochemistry Moscow* 83, 215–232. doi:10.1134/S0006297918030033.
- Saint-Jore-Dupas, C., Nebenführ, A., Boulaflous, A., Follet-Gueye, M.-L., Plasson, C., Hawes, C., et al. (2006). Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. *Plant Cell* 18, 3182–3200. doi:10.1105/tpc.105.036400.
- Sakakibara, K., Ando, S., Yip, H. K., Tamada, Y., Hiwatashi, Y., Murata, T., et al. (2013). KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. *Science* 339, 1067–1070. doi:10.1126/science.1230082.
- Schachter, H., Narasimhan, S., Gleeson, P., and Vella, G. (1983). Control of branching during the biosynthesis of asparagine-linked oligosaccharides. *Can. J. Biochem. Cell Biol.* 61, 1049–1066. doi:10.1139/o83-134.
- Schaff, D. A. (1994). The adenine phosphoribosyltransferase (APRT) selectable marker system. *Plant Science* 101, 3–9. doi:10.1016/0168-9452(94)90159-7.
- Schiestl, M., Stangler, T., Torella, C., Čepeljnik, T., Toll, H., and Grau, R. (2011). Acceptable changes in quality attributes of glycosylated biopharmaceuticals. *Nature Biotechnology* 29, 310–312. doi:10.1038/nbt.1839.
- Schween, G., Hohe, A., Koprivova, A., and Reski, R. (2003). Effects of nutrients, cell density and culture techniques on protoplast regeneration and early protonema development in a moss, *Physcomitrella patens*. *Journal of Plant Physiology* 160, 209–212. doi:10.1078/0176-1617-00855.
- Shen, J.-S., Busch, A., Day, T. S., Meng, X.-L., Yu, C. I., Dabrowska-Schlepp, P., et al. (2016). Mannose receptor-mediated delivery of moss-made α -galactosidase A efficiently corrects enzyme deficiency in Fabry mice. *J Inherit Metab Dis* 39, 293–303. doi:10.1007/s10545-015-9886-9.
- Shin, Y.-J., Chong, Y.-J., Yang, M.-S., and Kwon, T.-H. (2011). Production of recombinant human granulocyte macrophage-colony stimulating factor in rice cell suspension culture with a human-like N-glycan structure. *Plant Biotechnol. J.* 9, 1109–1119. doi:10.1111/j.1467-7652.2011.00636.x.
- Sochaj, A. M., Świdarska, K. W., and Otlewski, J. (2015). Current methods for the synthesis of homogeneous antibody–drug conjugates. *Biotechnology Advances* 33, 775–784. doi:10.1016/j.biotechadv.2015.05.001.
- Son, Y.-D., Jeong, Y. T., Park, S.-Y., and Kim, J. H. (2011). Enhanced sialylation of recombinant human erythropoietin in Chinese hamster ovary cells by combinatorial engineering of selected genes. *Glycobiology* 21, 1019–1028. doi:10.1093/glycob/cwr034.

- Sourrouille, C., Marquet-Blouin, E., D'Aoust, M.-A., Kiefer-Meyer, M.-C., Seveno, M., Pagny-Salehabadi, S., et al. (2008). Down-regulated expression of plant-specific glycoepitopes in alfalfa. *Plant Biotechnol. J.* 6, 702–721. doi:10.1111/j.1467-7652.2008.00353.x.
- Stanley, P., Taniguchi, N., and Aebi, M. (2015). “N-Glycans,” in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, et al. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press). Available at: <http://www.ncbi.nlm.nih.gov/books/NBK453020/> [Accessed September 21, 2020].
- Stoger, E., Fischer, R., Moloney, M., and Ma, J. K.-C. (2014). Plant molecular pharming for the treatment of chronic and infectious diseases. *Annual Review of Plant Biology* 65, 743–768. doi:10.1146/annurev-arplant-050213-035850.
- Strasser, R., Altmann, F., Mach, L., Glössl, J. and Steinkellner, H. (2004). Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett.* 561, 132–136. doi:10.1016/S0014-5793(04)00150-4.
- Strasser, R., Bondili, J. S., Vavra, U., Schoberer, J., Svoboda, B., Glössl, J., et al. (2007). A unique β 1,3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in *Arabidopsis thaliana*. *The Plant Cell* 19, 2278–2292. doi:10.1105/tpc.107.052985.
- Strasser, R., Castilho, A., Stadlmann, J., Kunert, R., Quendler, H., Gattinger, P., et al. (2009). Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous β 1,4-galactosylated N-glycan profile. *J. Biol. Chem.* 284, 20479–20485. doi:10.1074/jbc.M109.014126.
- Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., et al. (2008). Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol. J.* 6, 392–402. doi:10.1111/j.1467-7652.2008.00330.x.
- Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., et al. (2004). Correction of multi-gene deficiency in vivo using a single “self-cleaving” 2A peptide-based retroviral vector. *Nat. Biotechnol.* 22, 589–594. doi:10.1038/nbt957.
- Tangvoranuntakul, P., Gagneux, P., Diaz, S., Bardor, M., Varki, N., Varki, A., et al. (2003). Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *PNAS* 100, 12045–12050. doi:10.1073/pnas.2131556100.
- Tiralongo, J., Ashikov, A., Routier, F., Eckhardt, M., Bakker, H., Gerardy-Schahn, R., et al. (2006). Functional expression of the CMP-sialic acid transporter in *Escherichia coli* and its identification as a simple mobile carrier. *Glycobiology* 16, 73–81. doi:10.1093/glycob/cwj029.
- Top, O., Parsons, J., Bohlender, L. L., Michelfelder, S., Kopp, P., Busch-Steenberg, C., et al. (2019). Recombinant production of MFHR1, a novel synthetic multitarget complement inhibitor, in moss bioreactors. *Frontiers in Plant Science* 10. doi:10.3389/fpls.2019.00260.

- Trouiller, B., Charlot, F., Choinard, S., Schaefer, D. G., and Nogu e, F. (2007). Comparison of gene targeting efficiencies in two mosses suggests that it is a conserved feature of Bryophyte transformation. *Biotechnol Lett* 29, 1591–1598. doi:10.1007/s10529-007-9423-5.
- Tschofen, M., Knopp, D., Hood, E., and St oger, E. (2016). Plant molecular farming: Much more than medicines. *Annual Review of Analytical Chemistry* 9, 271–294. doi:10.1146/annurev-anchem-071015-041706.
- Twyman, R. M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. *Trends in Biotechnology* 21, 570–578. doi:10.1016/j.tibtech.2003.10.002.
- van Dussen, L., Zimran, A., Akkerman, E. M., Aerts, J. M. F. G., Petakov, M., Elstein, D., et al. (2013). Taliglucerase alfa leads to favorable bone marrow responses in patients with type I Gaucher disease. *Blood Cells, Molecules, and Diseases* 50, 206–211. doi:10.1016/j.bcmd.2012.11.001.
- van Ree, R., Cabanes-Macheteau, M., Akkerdaas, J., Milazzo, J. P., Loutelier-Bourhis, C., Rayon, C., et al. (2000). Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J. Biol. Chem.* 275, 11451–11458.
- Varki, A. (2008). Sialic acids in human health and disease. *Trends Mol Med* 14, 351–360. doi:10.1016/j.molmed.2008.06.002.
- Varki, A., Schnaar, R. L., and Schauer, R. (2015). Sialic Acids and Other Nonulosonic Acids, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, et al. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press). Available at: <http://www.ncbi.nlm.nih.gov/books/NBK453082/> [Accessed August 30, 2020].
- Vi etor, R., Loutelier-Bourhis, C., Fitchette, A.-C., Margerie, P., Gonneau, M., Faye, L., et al. (2003). Protein N-glycosylation is similar in the moss *Physcomitrella patens* and in higher plants. *Planta* 218, 269–275. doi:10.1007/s00425-003-1099-z.
- Viswanathan, K., Narang, S., Hinderlich, S., Lee, Y. C., and Betenbaugh, M. J. (2005). Engineering intracellular CMP-sialic acid metabolism into insect cells and methods to enhance its generation. *Biochemistry* 44, 7526–7534. doi:10.1021/bi047477y.
- Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology* 36, 1136–1145. doi:10.1038/nbt.4305.
- Wang, Q., Chung, C.-Y., Chough, S., and Betenbaugh, M. J. (2018). Antibody glycoengineering strategies in mammalian cells. *Biotechnol. Bioeng.* 115, 1378–1393. doi:10.1002/bit.26567.
- Webster, D. E., and Thomas, M. C. (2012). Post-translational modification of plant-made foreign proteins; glycosylation and beyond. *Biotechnology Advances* 30, 410–418. doi:10.1016/j.biotechadv.2011.07.015.

- Wiedemann, G., Gessel, N. van, Köchl, F., Hunn, L., Schulze, K., Maloukh, L., et al. (2018). RecQ helicases function in development, DNA repair, and gene targeting in *Physcomitrella patens*. *The Plant Cell* 30, 717–736. doi:10.1105/tpc.17.00632.
- Wilson, I. B. H., Zeleny, R., Kolarich, D., Staudacher, E., Stroop, C. J. M., Kamerling, J. P., et al. (2001). Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core α 1,3-linked fucose and xylose substitutions. *Glycobiology* 11, 261–274. doi:10.1093/glycob/11.4.261.
- Zeleny, R., Kolarich, D., Strasser, R., and Altmann, F. (2006). Sialic acid concentrations in plants are in the range of inadvertent contamination. *Planta* 224, 222–227. doi:10.1007/s00425-005-0206-8.
- Zhang, Z., Shah, B., and Richardson, J. (2019). Impact of Fc *N*-glycan sialylation on IgG structure. *mAbs* 11, 1381–1390. doi:10.1080/19420862.2019.1655377.
- Zhang, Z., Wuhrer, M., and Holst, S. (2018). Serum sialylation changes in cancer. *Glycoconj. J.* 35, 139–160. doi:10.1007/s10719-018-9820-0.
- Zhao, W., Chen, T.-L. L., Vertel, B. M., and Colley, K. J. (2006). The CMP-sialic acid transporter is localized in the medial-trans Golgi and possesses two specific endoplasmic reticulum export motifs in its carboxyl-terminal cytoplasmic tail. *J. Biol. Chem.* 281, 31106–31118. doi:10.1074/jbc.M605564200.

Supplementary Material

Supplementary Table S1. Primers used. Introduced restriction sites are written in *italic*. Lower-case letters indicate sequences of primer overhangs. Bolt letters indicate sequence mismatches for site-directed mutagenesis.

1	GNE_BamHI_fwd	TTTGGATCCATGGAGAAGAATGGAATAACCGA
2	GNE_Sall_rev	TTTGTGACCTAGTAGATCCTGCGTGTGT
3	NANS_BamHI_fwd	TTTGGATCCATGCCGCTGGAGCTG
4	NANS_Sall_rev	TTTGTGACTTAAGACTTGATTTTTTGGCATGATTATCTAC
5	NANP_BamHI_fwd	TTTGGATCCATGGGGCTGAGCCGC
6	NANP_Sall_rev	TTTGTGACTTAAGTGGACATACTGACTTTGCAGTCTA
7	CMAS_trc_XhoI_fwd	TTTCTGAGATGAAGCCCCCGCA
8	CMAS_KpnI_rev	TTTGGTACCCTATTTTTGGCATGAATTATTAACCTTTTCC
9	CSAT_BamHI_fwd	TTTGGATCCATGGCTGCCCGAGA
10	CSAT_Sall_rev	TTTGTGACTCACACACCAATAACTCTCTCTTT
11	ST6Gal1_BamHI_fwd	TTTGGATCCATGATTCACACCAACCTGAAGAAA
12	ST6Gal1_Sall_rev	TTTGTGACTTAGCAGTGAATGGTCCGG
13	BSD_35S_P_AgeI_fwd	TTTACCGGTCTACTCCAAAAATGTCAAAGATACAG
14	BSD_35S_T_AatII_rev	AGTTTGACGTCGGATTTTAGTACTGGATTTTGTT
15	GNE_AgeI_35S_P_fwd	TTTACCGGTATGCCTGCAGGTCCC
16	GNE_NosT_SgrDI_rev	TTTTCGTCGACGCGATCTAGTAACATAGATGACACCG
17	NANS/CMAS_SgrDI_35S_P_f	TTTTCGTCGACGATGCCTGCAGGTCCC
18	NANS/CSAT_NosT_AvrII_rev	TTTCCTAGGCGATCTAGTAACATAGATGACACCG
19	NANP/ST_AvrII_35S_fwd	TTTCCTAGGATGCCTGCAGGTCCC
20	NANP/CMAS_NosT_AscI_rev	TTTGGCGCGCCGATCTAGTAACATAGATGACACCG
21	CSAT_AscI_35S_P_fwd	TTTGGCGCGCCATGCCTGCAGGTCCC
22	ST_NosT_AgeI_rev	TTTACCGGTGATCTAGTAACATAGATGACACCG
23	APT_5HR_MCS_fwd	TTTGCTCTTCGAAAGCTTATGCCAATAGAAAAGATCA
24	APT_5HR_MCS_rev	AAACCTAGGAAAACGTCGACGTTTACCGGTAAATTAATCCA GTCTGTGCAAGT
25	APT_3HR_MCS_fwd	TTCTAGGTTTGGCGCGCCTGGTACCGATTGCATTGAGATG
26	APT_3HR_MCS_rev	AAAGCTCTTCTAAAAGCTTACTCGCCCTCATATTC
27	P4H1_5HR_NotI_fwd	TTTGCGGCCGCTACACCATAGAAGGTAGCAAC
28	P4H1_5HR_MCS_rev	AACCTAGGTTTGGCGCGCCAAAACGTCGACGTTTGAACCTT ACTCTCGATTC
29	P4H1_3HR_MCS_fwd	TTCTAGGTTTACCGGTAGCTAGACGTCGAAGCTTGAGCTCCT CCAAGTT
30	P4H1_3HR_NotI_rev	GTTGCGGCCGCGAATTCAGAGGGTAGGATTGTGTG
31	SDM_GNEmut_fwd	ATGC AGA AAGAAGGGCATCGAGCATC
32	SDM_GNEmut_rev	CACT AGA ACCATCTCCTTGCTGCC
33	XhoI_1,4FT5'UTR_fwd	CTCGAGGGTCTCCTGCTCACAAACAGTC
34	FT_CTS_GT4_J_R	gcgcggtgggaggaggcACTGTGGGGTTTTGAAGACTGCGT
35	GT4_FT_CTS_J_F	ttcaaaccacagtgGCTCCTCCAGCCGCGC
36	GT4_J_R	GCTCGGTGTCGGATGTCCAC

37	GT4_P2A_1_J_R	ccagcctgtttcagaaggctgaagtggtagcGCTCGGTGTCCCGATGT CCACTG
38	GT4_P2A_2_J_R	tctcgagagggccagggttctcttcaacatctCCAGCCTGTTTCAGAAG GCTGAAGTTGG
39	BamHI_35S_long_fwd	GGATCCAAGCTTGCATGCCTGC
40	XhoI_35S_long_rev	CTCGAGTTATATAGAGGAAGGGTCTTGCG
41	P4H2-5'-fwd	ccgccagatcttccggatggctcgagttttcagcaagatatacGCTCAGTTT CGTACGAGCGATTTG
42	P4H2-5'-rev	ccctgcccatcagatccttggcggAACCCACGCTCAGCAGCACA
43	P4H2-3'-fwd	cggtgtcatctatgttactagatcgCTGCTCTTCTCAGCCTACA
44	P4H2-3'-rev	atggcagctgagaatattgtaggagatcttctagaagatatacGTCTTCGTT AATTAATCCACTTG
45	P4H2-hpt-fwd	gatcctgtgctgctgagcgtgggtgggtgttCCGCAAGGATCTGATG
46	P4H2-hpt-rev	agggtgtaggctgaagaagagcagCGATCTAGTAACATAGATGAC ACC
47	P4H2_35Slong_FTCTS_fw	ggaacgccacggtttcacattggttacacAGATTAGCCTTTTCAATTT CAGAAAG
48	P4H2_35Slong_FTCTS_rev	tgagaacacaacctgggaagcagaccccatCACTGTGGGGTTTTGAA GAC
49	P4H2_rST_nosT_fw	actttacgcagtcttcaaaccacagtgATGGGGTCTGCTCCCAG
50	P4H2_rST_nosT_rev	cgaaagcggagacagtgaggagcctgagacCGATCTAGTAACATAGA TGACACCGC
51	EF1a_fw	AGCGTGGTATCACAATTGAC
52	EF1a_rev	GATCGCTCGATCATGTTATC
53	Leaf_APT_HR5_fwd	GCCTTTTGAACAAGATGACAAAGTG
54	Leaf_GNE_mus_rev	AACAATCGTGTGGAGCCTGG
55	Leaf_NANP_fwd	CTGGGGACTGTGTGATGGTC
56	Leaf_APT_HR3_rev	GCAGTACGATCTGCTTTTGCC
57	P4Hsc8KO_1547F	GGCTAATGATGAAGATGCGAGA
58	CMAS_leaf_rev	GGAAGGCCCTGAATCCA
59	full length mrna BSD Stop	GTTATGTGTGGGAGGGCTAA
60	P4Hsc8KO_4023r	AGCATCCCCTCGTTTAGGTT
61	RT1	TGTCGTGCTCCACCATGTTG
62	B1,3galT3_RT_879F	ACTTGGACAGTAAGCAGAGATTGG
63	3HR_zeo_fwd	CTTCGTGGCCGAGGAGCAGG
64	3genome_GalT3_rev	CGAAACTTACCGGAGCCGGCG
65	Leaf_P4H2_5HR_f	GGCATTGTGCCGTTTATGC
66	Leaf_P4H2_5HR_rev	GGAAGGAAGGCTGTGGTTTG
67	qhpt_f	ATACGAGGTCGCCAACATCT
68	P4H2_3HR_rev	CACAACTTCATCGACTTGATCC
69	EF1a_qf	CGACGCCCTGGACATC
70	EF1a_qr	CCTGCGAGGTTCCCGTAA
71	c45-tg2F	ACGCACCGGCATCGT
72	c45-tg2R	TGCTTGTTTCATCACGACACCA
73	GNE_mus_qPCR_fwd	CTACCGGACGTCCTCAATCG
74	GNE_mus_qPCR_rev	CGTCAAATCGGTCTCCGTGA
75	NANS_qPCR_fwd	TTGCAGTGTACCAGCGCATA

76	NANS_qPCR_rev	TATTCCGAGATGACCCGCAG
77	NANP_qPCR_fwd	GCTTGTGCCTGTCAGTCCTA
78	NANP_qPCR_rev	ATGGACGGTGCTGGTTTCTC
79	qPCR CMAS fwd	CTTGAAGAGAGTGGGCCTAAGT
80	qPCR CMAS rev	GCCTTCTGGGCAGTAGAACA
81	CSAT_mus_qPCR_fwd	TTCCGACACTTCCCTTTGGG
82	CSAT_mus_qPCR_rev	ACGTCACAACGATCCCTGAC
83	ST6_rat_qPCR_fwd	AAACTCCCAGCTTGGTCGAG
84	ST6_rat_qPCR_rev	GGTAGGGGCCCCATTAACC
85	qPCR chGalT4 fw	TTGACCTCGGTCCCAGTG
86	qPCR chGalT4 rev	GGCATGTTAAACTCAATCAGCA

Supplementary Table S2. Monoisotopic masses of glycopeptides used for mass spectrometric analysis. All combinations of the searched peptides with possible glycan structures are given in three calculated charge stages ($[M+nH]^{n+}$). cam: carbamidomethylation (C, + 57.021464 Da, fixed); ox: oxidation (M, +15.994915 Da, optional); M: mannose, Gn: *N*-acetylglucosamine, A: galactose, Na: sialic acid, P: pentose.

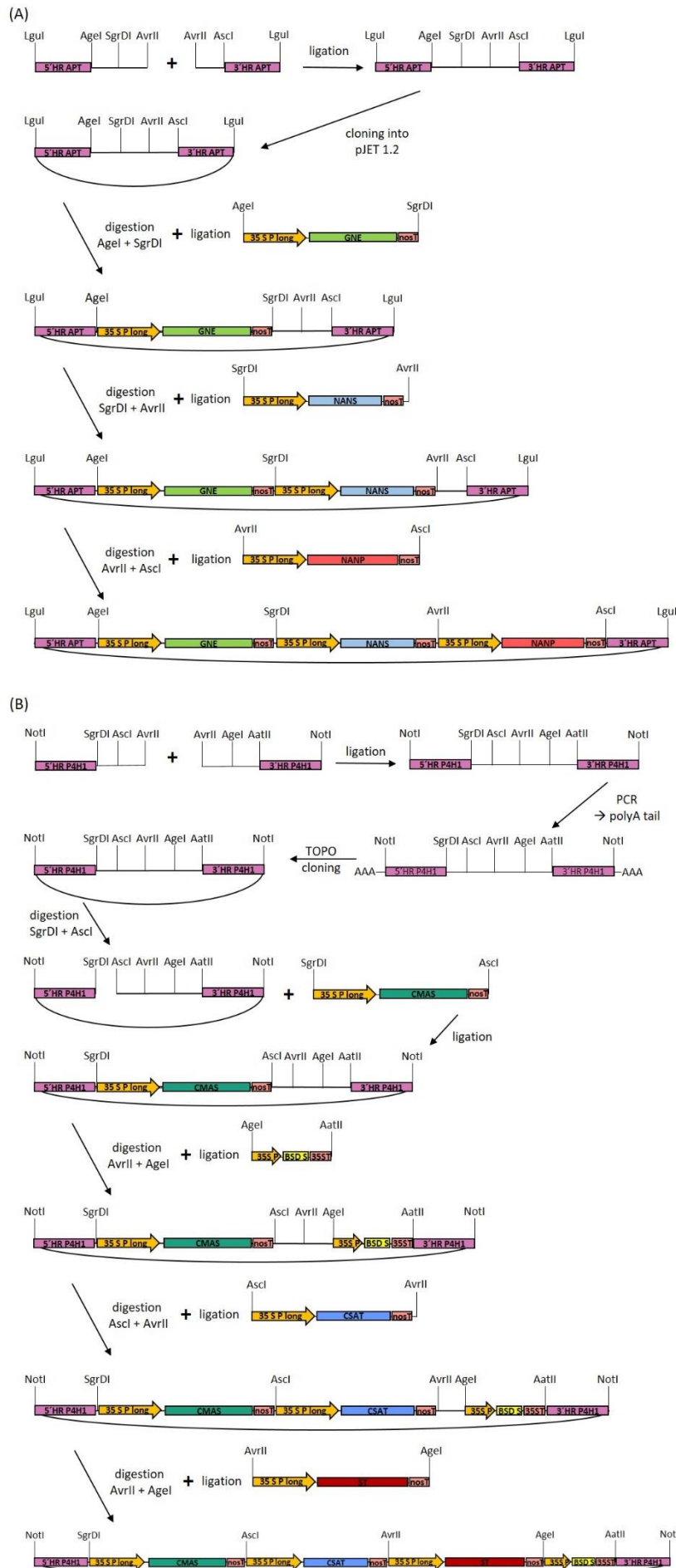
Glycopeptide	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
IPCcamSQPPQIEHGTINSSR_AA	1822.2879 M2H	1215.1944 M3H	911.6476 M4H
IPCcamSQPPQIEHGTINSSR_AA_2P	1954.3302 M2H	1303.2226 M3H	977.6688 M4H
IPCcamSQPPQIEHGTINSSR_AA_3P	2020.3514 M2H	1347.2367 M3H	1010.6793 M4H
IPCcamSQPPQIEHGTINSSR_AA_4P	2086.3725 M2H	1391.2508 M3H	1043.6899 M4H
IPCcamSQPPQIEHGTINSSR_AA_5P	2152.3936 M2H	1435.2649 M3H	1076.7005 M4H
IPCcamSQPPQIEHGTINSSR_AA_6P	2218.4148 M2H	1479.2790 M3H	1109.7110 M4H
IPCcamSQPPQIEHGTINSSR_AA_P	1888.3091 M2H	1259.2085 M3H	944.6582 M4H
IPCcamSQPPQIEHGTINSSR_AGn	1741.2615 M2H	1161.1768 M3H	871.1344 M4H
IPCcamSQPPQIEHGTINSSR_AGn_2P	1873.3038 M2H	1249.2050 M3H	937.1556 M4H
IPCcamSQPPQIEHGTINSSR_AGn_3P	1939.3250 M2H	1293.2191 M3H	970.1661 M4H
IPCcamSQPPQIEHGTINSSR_AGn_4P	2005.3461 M2H	1337.2332 M3H	1003.1767 M4H
IPCcamSQPPQIEHGTINSSR_AGn_5P	2071.3673 M2H	1381.2473 M3H	1036.1873 M4H
IPCcamSQPPQIEHGTINSSR_AGn_6P	2137.3884 M2H	1425.2614 M3H	1069.1979 M4H
IPCcamSQPPQIEHGTINSSR_AGn_P	1807.2827 M2H	1205.1909 M3H	904.1450 M4H
IPCcamSQPPQIEHGTINSSR_AM	1639.7218 M2H	1093.4836 M3H	820.3646 M4H
IPCcamSQPPQIEHGTINSSR_AM_2P	1771.7641 M2H	1181.5118 M3H	886.3857 M4H
IPCcamSQPPQIEHGTINSSR_AM_3P	1837.7853 M2H	1225.5259 M3H	919.3963 M4H
IPCcamSQPPQIEHGTINSSR_AM_4P	1903.8064 M2H	1269.5400 M3H	952.4069 M4H
IPCcamSQPPQIEHGTINSSR_AM_5P	1969.8276 M2H	1313.5541 M3H	985.4174 M4H
IPCcamSQPPQIEHGTINSSR_AM_6P	2035.8487 M2H	1357.5682 M3H	1018.4280 M4H
IPCcamSQPPQIEHGTINSSR_AM_P	1705.7430 M2H	1137.4977 M3H	853.3751 M4H
IPCcamSQPPQIEHGTINSSR_GnGn	1660.2351 M2H	1107.1592 M3H	830.6212 M4H
IPCcamSQPPQIEHGTINSSR_GnM	1558.6954 M2H	1039.4660 M3H	779.8514 M4H
IPCcamSQPPQIEHGTINSSR_GnM5	1720.7482 M2H	1147.5012 M3H	860.8777 M4H
IPCcamSQPPQIEHGTINSSR_M4	1538.1821 M2H	1025.7905 M3H	769.5947 M4H
IPCcamSQPPQIEHGTINSSR_M5	1619.2085 M2H	1079.8081 M3H	810.1079 M4H
IPCcamSQPPQIEHGTINSSR_M6	1700.2349 M2H	1133.8257 M3H	850.6211 M4H
IPCcamSQPPQIEHGTINSSR_M7	1781.2613 M2H	1187.8433 M3H	891.1343 M4H
IPCcamSQPPQIEHGTINSSR_M8	1862.2877 M2H	1241.8609 M3H	931.6475 M4H
IPCcamSQPPQIEHGTINSSR_M9	1943.3141 M2H	1295.8785 M3H	972.1607 M4H
IPCcamSQPPQIEHGTINSSR_MM	1457.1557 M2H	971.7729 M3H	729.0815 M4H
IPCcamSQPPQIEHGTINSSR_NaA_2P	2033.8568 M2H	1356.2403 M3H	1017.4320 M4H
IPCcamSQPPQIEHGTINSSR_NaA_3P	2099.8779 M2H	1400.2544 M3H	1050.4426 M4H
IPCcamSQPPQIEHGTINSSR_NaA_4P	2165.8990 M2H	1444.2685 M3H	1083.4532 M4H
IPCcamSQPPQIEHGTINSSR_NaA_5P	2231.9202 M2H	1488.2826 M3H	1116.4637 M4H
IPCcamSQPPQIEHGTINSSR_NaA_6P	2297.9414 M2H	1532.2967 M3H	1149.4743 M4H
IPCcamSQPPQIEHGTINSSR_NaA_P	1967.8356 M2H	1312.2262 M3H	984.4215 M4H
IPCcamSQPPQIEHGTINSSR_NaGn	1886.8092 M2H	1258.2086 M3H	943.9083 M4H
IPCcamSQPPQIEHGTINSSR_NaGn_P	1952.8303 M2H	1302.2227 M3H	976.9188 M4H
IPCcamSQPPQIEHGTINSSR_NaM	1785.2695 M2H	1190.5154 M3H	893.1384 M4H

IPCcamSQPPQIEHGTINSSR_NaNa	2113.3833 M2H	1409.2580 M3H	1057.1953 M4H
IPCcamSQPPQIEHGTINSSR_NaNa_P	2245.4256 M2H	1497.2861 M3H	1123.2164 M4H
ISEENETTCcamYMGK_AA	1592.6199 M2H	1062.0824 M3H	796.8136 M4H
ISEENETTCcamYMGK_AA_2P	1724.6622 M2H	1150.1106 M3H	862.8348 M4H
ISEENETTCcamYMGK_AA_3P	1790.6834 M2H	1194.1247 M3H	895.8453 M4H
ISEENETTCcamYMGK_AA_4P	1856.7045 M2H	1238.1388 M3H	928.8559 M4H
ISEENETTCcamYMGK_AA_5P	1922.7257 M2H	1282.1529 M3H	961.8665 M4H
ISEENETTCcamYMGK_AA_6P	1988.7468 M2H	1326.1670 M3H	994.8771 M4H
ISEENETTCcamYMGK_AA_P	1658.6411 M2H	1106.0965 M3H	829.8242 M4H
ISEENETTCcamYMGK_AGn	1511.5935 M2H	1008.0648 M3H	756.3004 M4H
ISEENETTCcamYMGK_AGn_2P	1643.6358 M2H	1096.0930 M3H	822.3216 M4H
ISEENETTCcamYMGK_AGn_3P	1709.6570 M2H	1140.1071 M3H	855.3321 M4H
ISEENETTCcamYMGK_AGn_4P	1775.6781 M2H	1184.1212 M3H	888.3427 M4H
ISEENETTCcamYMGK_AGn_5P	1841.6993 M2H	1228.1353 M3H	921.3533 M4H
ISEENETTCcamYMGK_AGn_6P	1907.7204 M2H	1272.1494 M3H	954.3639 M4H
ISEENETTCcamYMGK_AGn_P	1577.6146 M2H	1052.0789 M3H	789.3110 M4H
ISEENETTCcamYMGK_AM	1410.0538 M2H	940.3716 M3H	705.5306 M4H
ISEENETTCcamYMGK_AM_2P	1542.0961 M2H	1028.3998 M3H	771.5517 M4H
ISEENETTCcamYMGK_AM_3P	1608.1173 M2H	1072.4139 M3H	804.5623 M4H
ISEENETTCcamYMGK_AM_4P	1674.1384 M2H	1116.4280 M3H	837.5729 M4H
ISEENETTCcamYMGK_AM_5P	1740.1596 M2H	1160.4421 M3H	870.5834 M4H
ISEENETTCcamYMGK_AM_6P	1806.1807 M2H	1204.4562 M3H	903.5940 M4H
ISEENETTCcamYMGK_AM_P	1476.0750 M2H	984.3857 M3H	738.5411 M4H
ISEENETTCcamYMGK_GnGn	1430.5671 M2H	954.0472 M3H	715.7872 M4H
ISEENETTCcamYMGK_GnM	1329.0274 M2H	886.3540 M3H	665.0174 M4H
ISEENETTCcamYMGK_GnM5	1491.0802 M2H	994.3892 M3H	746.0438 M4H
ISEENETTCcamYMGK_M4	1308.5141 M2H	872.6785 M3H	654.7607 M4H
ISEENETTCcamYMGK_M5	1389.5405 M2H	926.6961 M3H	695.2739 M4H
ISEENETTCcamYMGK_M6	1470.5669 M2H	980.7137 M3H	735.7871 M4H
ISEENETTCcamYMGK_M7	1551.5933 M2H	1034.7313 M3H	776.3003 M4H
ISEENETTCcamYMGK_M8	1632.6197 M2H	1088.7489 M3H	816.8135 M4H
ISEENETTCcamYMGK_M9	1713.6461 M2H	1142.7665 M3H	857.3267 M4H
ISEENETTCcamYMGK_MM	1227.4877 M2H	818.6609 M3H	614.2475 M4H
ISEENETTCcamYMGK_NaA_2P	1804.1888 M2H	1203.1283 M3H	902.5980 M4H
ISEENETTCcamYMGK_NaA_3P	1870.2099 M2H	1247.1424 M3H	935.6086 M4H
ISEENETTCcamYMGK_NaA_4P	1936.2311 M2H	1291.1565 M3H	968.6192 M4H
ISEENETTCcamYMGK_NaA_5P	2002.2522 M2H	1335.1706 M3H	1001.6298 M4H
ISEENETTCcamYMGK_NaA_6P	2068.2733 M2H	1379.1847 M3H	1034.6403 M4H
ISEENETTCcamYMGK_NaA_P	1738.1676 M2H	1159.1142 M3H	869.5875 M4H
ISEENETTCcamYMGK_NaGn	1657.1412 M2H	1105.0966 M3H	829.0743 M4H
ISEENETTCcamYMGK_NaGn_P	1723.1623 M2H	1149.1107 M3H	862.0848 M4H
ISEENETTCcamYMGK_NaM	1555.6015 M2H	1037.4034 M3H	778.3044 M4H
ISEENETTCcamYMGK_NaNa	1883.7153 M2H	1256.1460 M3H	942.3613 M4H
ISEENETTCcamYMGK_NaNa_P	2015.7576 M2H	1344.1741 M3H	1008.3824 M4H
ISEENETTCcamYMoxGK_AA	1600.6174 M2H	1067.4140 M3H	800.8123 M4H
ISEENETTCcamYMoxGK_AA_2P	1732.6597 M2H	1155.4422 M3H	866.8335 M4H

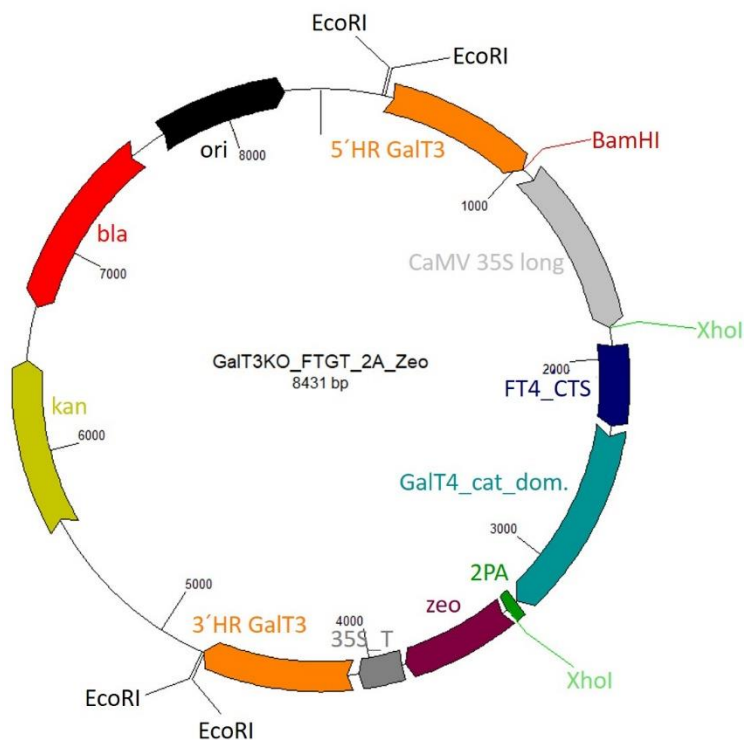
ISEENETTcamYMoxGK_AA_3P	1798.6808 M2H	1199.4563 M3H	899.8441 M4H
ISEENETTcamYMoxGK_AA_4P	1864.7020 M2H	1243.4704 M3H	932.8546 M4H
ISEENETTcamYMoxGK_AA_5P	1930.7231 M2H	1287.4845 M3H	965.8652 M4H
ISEENETTcamYMoxGK_AA_6P	1996.7443 M2H	1331.4986 M3H	998.8758 M4H
ISEENETTcamYMoxGK_AA_P	1666.6385 M2H	1111.4281 M3H	833.8229 M4H
ISEENETTcamYMoxGK_AGn	1519.5910 M2H	1013.3964 M3H	760.2991 M4H
ISEENETTcamYMoxGK_AGn_2P	1651.6333 M2H	1101.4246 M3H	826.3203 M4H
ISEENETTcamYMoxGK_AGn_3P	1717.6544 M2H	1145.4387 M3H	859.3309 M4H
ISEENETTcamYMoxGK_AGn_4P	1783.6756 M2H	1189.4528 M3H	892.3414 M4H
ISEENETTcamYMoxGK_AGn_5P	1849.6967 M2H	1233.4669 M3H	925.3520 M4H
ISEENETTcamYMoxGK_AGn_6P	1915.7179 M2H	1277.4810 M3H	958.3626 M4H
ISEENETTcamYMoxGK_AGn_P	1585.6121 M2H	1057.4105 M3H	793.3097 M4H
ISEENETTcamYMoxGK_AM	1418.0513 M2H	945.7033 M3H	709.5293 M4H
ISEENETTcamYMoxGK_AM_2P	1550.0936 M2H	1033.7315 M3H	775.5504 M4H
ISEENETTcamYMoxGK_AM_3P	1616.1147 M2H	1077.7456 M3H	808.5610 M4H
ISEENETTcamYMoxGK_AM_4P	1682.1359 M2H	1121.7597 M3H	841.5716 M4H
ISEENETTcamYMoxGK_AM_5P	1748.1570 M2H	1165.7738 M3H	874.5822 M4H
ISEENETTcamYMoxGK_AM_6P	1814.1782 M2H	1209.7879 M3H	907.5927 M4H
ISEENETTcamYMoxGK_AM_P	1484.0724 M2H	989.7174 M3H	742.5399 M4H
ISEENETTcamYMoxGK_GnGn	1438.5646 M2H	959.3788 M3H	719.7859 M4H
ISEENETTcamYMoxGK_GnM	1337.0249 M2H	891.6857 M3H	669.0161 M4H
ISEENETTcamYMoxGK_GnM5	1499.0777 M2H	999.7209 M3H	750.0425 M4H
ISEENETTcamYMoxGK_M4	1316.5116 M2H	878.0101 M3H	658.7594 M4H
ISEENETTcamYMoxGK_M5	1397.5380 M2H	932.0277 M3H	699.2726 M4H
ISEENETTcamYMoxGK_M6	1478.5644 M2H	986.0453 M3H	739.7858 M4H
ISEENETTcamYMoxGK_M7	1559.5908 M2H	1040.0629 M3H	780.2990 M4H
ISEENETTcamYMoxGK_M8	1640.6172 M2H	1094.0805 M3H	820.8122 M4H
ISEENETTcamYMoxGK_M9	1721.6436 M2H	1148.0981 M3H	861.3254 M4H
ISEENETTcamYMoxGK_MM	1235.4852 M2H	823.9925 M3H	618.2462 M4H
ISEENETTcamYMoxGK_NaA_2P	1812.1862 M2H	1208.4599 M3H	906.5968 M4H
ISEENETTcamYMoxGK_NaA_3P	1878.2074 M2H	1252.4740 M3H	939.6073 M4H
ISEENETTcamYMoxGK_NaA_4P	1944.2285 M2H	1296.4881 M3H	972.6179 M4H
ISEENETTcamYMoxGK_NaA_5P	2010.2497 M2H	1340.5022 M3H	1005.6285 M4H
ISEENETTcamYMoxGK_NaA_6P	2076.2708 M2H	1384.5163 M3H	1038.6391 M4H
ISEENETTcamYMoxGK_NaA_P	1746.1651 M2H	1164.4458 M3H	873.5862 M4H
ISEENETTcamYMoxGK_NaGn	1665.1387 M2H	1110.4282 M3H	833.0730 M4H
ISEENETTcamYMoxGK_NaGn_P	1731.1598 M2H	1154.4423 M3H	866.0835 M4H
ISEENETTcamYMoxGK_NaM	1563.5990 M2H	1042.7351 M3H	782.3031 M4H
ISEENETTcamYMoxGK_NaNa	1891.7128 M2H	1261.4776 M3H	946.3600 M4H
ISEENETTcamYMoxGK_NaNa_P	2023.7550 M2H	1349.5058 M3H	1012.3812 M4H
MDGASNVTcamINSR_AA	1524.1073 M2H	1016.4073 M3H	762.5573 M4H
MDGASNVTcamINSR_AA_2P	1656.1496 M2H	1104.4355 M3H	828.5785 M4H
MDGASNVTcamINSR_AA_3P	1722.1708 M2H	1148.4496 M3H	861.5890 M4H
MDGASNVTcamINSR_AA_4P	1788.1919 M2H	1192.4637 M3H	894.5996 M4H
MDGASNVTcamINSR_AA_5P	1854.2131 M2H	1236.4778 M3H	927.6102 M4H
MDGASNVTcamINSR_AA_6P	1920.2342 M2H	1280.4919 M3H	960.6208 M4H

MDGASNVTccamINSR_AA_P	1590.1285 M2H	1060.4214 M3H	795.5679 M4H
MDGASNVTccamINSR_AGn	1443.0809 M2H	962.3897 M3H	722.0441 M4H
MDGASNVTccamINSR_AGn_2P	1575.1232 M2H	1050.4179 M3H	788.0652 M4H
MDGASNVTccamINSR_AGn_3P	1641.1444 M2H	1094.4320 M3H	821.0758 M4H
MDGASNVTccamINSR_AGn_4P	1707.1655 M2H	1138.4461 M3H	854.0864 M4H
MDGASNVTccamINSR_AGn_5P	1773.1867 M2H	1182.4602 M3H	887.0970 M4H
MDGASNVTccamINSR_AGn_6P	1839.2078 M2H	1226.4743 M3H	920.1076 M4H
MDGASNVTccamINSR_AGn_P	1509.1020 M2H	1006.4038 M3H	755.0547 M4H
MDGASNVTccamINSR_AM	1341.5412 M2H	894.6966 M3H	671.2743 M4H
MDGASNVTccamINSR_AM_2P	1473.5835 M2H	982.7248 M3H	737.2954 M4H
MDGASNVTccamINSR_AM_3P	1539.6047 M2H	1026.7389 M3H	770.3060 M4H
MDGASNVTccamINSR_AM_4P	1605.6258 M2H	1070.7530 M3H	803.3166 M4H
MDGASNVTccamINSR_AM_5P	1671.6470 M2H	1114.7671 M3H	836.3271 M4H
MDGASNVTccamINSR_AM_6P	1737.6681 M2H	1158.7812 M3H	869.3377 M4H
MDGASNVTccamINSR_AM_P	1407.5624 M2H	938.7107 M3H	704.2848 M4H
MDGASNVTccamINSR_GnGn	1362.0545 M2H	908.3721 M3H	681.5309 M4H
MDGASNVTccamINSR_GnM	1260.5148 M2H	840.6790 M3H	630.7611 M4H
MDGASNVTccamINSR_GnM5	1422.5676 M2H	948.7142 M3H	711.7875 M4H
MDGASNVTccamINSR_M4	1240.0015 M2H	827.0034 M3H	620.5044 M4H
MDGASNVTccamINSR_M5	1321.0279 M2H	881.0210 M3H	661.0176 M4H
MDGASNVTccamINSR_M6	1402.0543 M2H	935.0386 M3H	701.5308 M4H
MDGASNVTccamINSR_M7	1483.0807 M2H	989.0562 M3H	742.0440 M4H
MDGASNVTccamINSR_M8	1564.1071 M2H	1043.0738 M3H	782.5572 M4H
MDGASNVTccamINSR_M9	1645.1335 M2H	1097.0914 M3H	823.0704 M4H
MDGASNVTccamINSR_MM	1158.9751 M2H	772.9858 M3H	579.9912 M4H
MDGASNVTccamINSR_NaA_2P	1735.6762 M2H	1157.4532 M3H	868.3417 M4H
MDGASNVTccamINSR_NaA_3P	1801.6973 M2H	1201.4673 M3H	901.3523 M4H
MDGASNVTccamINSR_NaA_4P	1867.7185 M2H	1245.4814 M3H	934.3629 M4H
MDGASNVTccamINSR_NaA_5P	1933.7396 M2H	1289.4955 M3H	967.3735 M4H
MDGASNVTccamINSR_NaA_6P	1999.7608 M2H	1333.5096 M3H	1000.3840 M4H
MDGASNVTccamINSR_NaA_P	1669.6550 M2H	1113.4391 M3H	835.3312 M4H
MDGASNVTccamINSR_NaGn	1588.6286 M2H	1059.4215 M3H	794.8179 M4H
MDGASNVTccamINSR_NaGn_P	1654.6497 M2H	1103.4356 M3H	827.8285 M4H
MDGASNVTccamINSR_NaM	1487.0889 M2H	991.7284 M3H	744.0481 M4H
MDGASNVTccamINSR_NaNa	1815.2027 M2H	1210.4709 M3H	908.1050 M4H
MDGASNVTccamINSR_NaNa_P	1947.2450 M2H	1298.4991 M3H	974.1261 M4H
MoxDGASNVTccamINSR_AA	1532.1048 M2H	1021.7389 M3H	766.5560 M4H
MoxDGASNVTccamINSR_AA_2P	1664.1471 M2H	1109.7671 M3H	832.5772 M4H
MoxDGASNVTccamINSR_AA_3P	1730.1682 M2H	1153.7812 M3H	865.5878 M4H
MoxDGASNVTccamINSR_AA_4P	1796.1894 M2H	1197.7953 M3H	898.5983 M4H
MoxDGASNVTccamINSR_AA_5P	1862.2105 M2H	1241.8094 M3H	931.6089 M4H
MoxDGASNVTccamINSR_AA_6P	1928.2317 M2H	1285.8235 M3H	964.6195 M4H
MoxDGASNVTccamINSR_AA_P	1598.1259 M2H	1065.7530 M3H	799.5666 M4H
MoxDGASNVTccamINSR_AGn	1451.0784 M2H	967.7213 M3H	726.0428 M4H
MoxDGASNVTccamINSR_AGn_2P	1583.1207 M2H	1055.7495 M3H	792.0640 M4H
MoxDGASNVTccamINSR_AGn_3P	1649.1418 M2H	1099.7636 M3H	825.0746 M4H

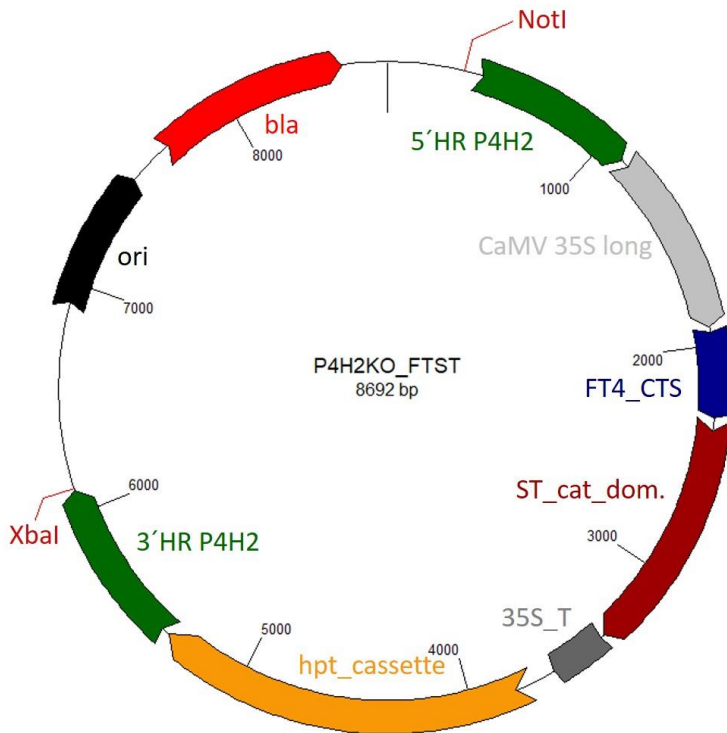
MoxDGASNVTCcamINSR_AGn_4P	1715.1630 M2H	1143.7777 M3H	858.0851 M4H
MoxDGASNVTCcamINSR_AGn_5P	1781.1841 M2H	1187.7918 M3H	891.0957 M4H
MoxDGASNVTCcamINSR_AGn_6P	1847.2053 M2H	1231.8059 M3H	924.1063 M4H
MoxDGASNVTCcamINSR_AGn_P	1517.0995 M2H	1011.7354 M3H	759.0534 M4H
MoxDGASNVTCcamINSR_AM	1349.5387 M2H	900.0282 M3H	675.2730 M4H
MoxDGASNVTCcamINSR_AM_2P	1481.5810 M2H	988.0564 M3H	741.2941 M4H
MoxDGASNVTCcamINSR_AM_3P	1547.6021 M2H	1032.0705 M3H	774.3047 M4H
MoxDGASNVTCcamINSR_AM_4P	1613.6233 M2H	1076.0846 M3H	807.3153 M4H
MoxDGASNVTCcamINSR_AM_5P	1679.6444 M2H	1120.0987 M3H	840.3259 M4H
MoxDGASNVTCcamINSR_AM_6P	1745.6656 M2H	1164.1128 M3H	873.3364 M4H
MoxDGASNVTCcamINSR_AM_P	1415.5598 M2H	944.0423 M3H	708.2836 M4H
MoxDGASNVTCcamINSR_GnGn	1370.0520 M2H	913.7037 M3H	685.5296 M4H
MoxDGASNVTCcamINSR_GnM	1268.5123 M2H	846.0106 M3H	634.7598 M4H
MoxDGASNVTCcamINSR_GnM5	1430.5651 M2H	954.0458 M3H	715.7862 M4H
MoxDGASNVTCcamINSR_M4	1247.9990 M2H	832.3351 M3H	624.5031 M4H
MoxDGASNVTCcamINSR_M5	1329.0254 M2H	886.3527 M3H	665.0163 M4H
MoxDGASNVTCcamINSR_M6	1410.0518 M2H	940.3703 M3H	705.5295 M4H
MoxDGASNVTCcamINSR_M7	1491.0782 M2H	994.3879 M3H	746.0427 M4H
MoxDGASNVTCcamINSR_M8	1572.1046 M2H	1048.4055 M3H	786.5559 M4H
MoxDGASNVTCcamINSR_M9	1653.1310 M2H	1102.4231 M3H	827.0691 M4H
MoxDGASNVTCcamINSR_MM	1166.9726 M2H	778.3175 M3H	583.9899 M4H
MoxDGASNVTCcamINSR_NaA_2P	1743.6736 M2H	1162.7848 M3H	872.3405 M4H
MoxDGASNVTCcamINSR_NaA_3P	1809.6948 M2H	1206.7989 M3H	905.3510 M4H
MoxDGASNVTCcamINSR_NaA_4P	1875.7159 M2H	1250.8130 M3H	938.3616 M4H
MoxDGASNVTCcamINSR_NaA_5P	1941.7371 M2H	1294.8271 M3H	971.3722 M4H
MoxDGASNVTCcamINSR_NaA_6P	2007.7582 M2H	1338.8412 M3H	1004.3828 M4H
MoxDGASNVTCcamINSR_NaA_P	1677.6525 M2H	1118.7707 M3H	839.3299 M4H
MoxDGASNVTCcamINSR_NaGn	1596.6261 M2H	1064.7531 M3H	798.8167 M4H
MoxDGASNVTCcamINSR_NaGn_P	1662.6472 M2H	1108.7672 M3H	831.8272 M4H
MoxDGASNVTCcamINSR_NaM	1495.0864 M2H	997.0600 M3H	748.0468 M4H
MoxDGASNVTCcamINSR_NaNa	1823.2002 M2H	1215.8025 M3H	912.1037 M4H
MoxDGASNVTCcamINSR_NaNa_P	1955.2424 M2H	1303.8307 M3H	978.1249 M4H



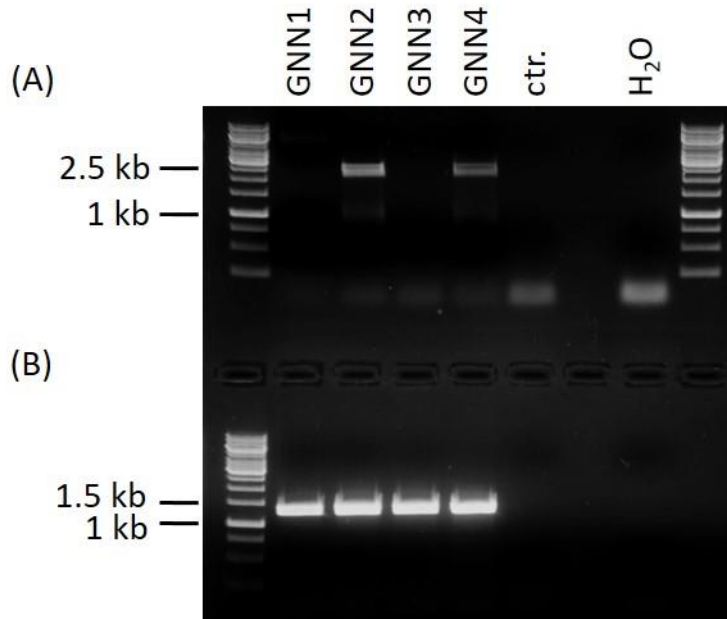
Supplementary Figure S1. Schematic overview of the cloning procedure for the GNN and CCSB multi-gene constructs. **(A)** Successive cloning procedure of the GNN-construct starting with the generation of the pJET1.2-based assembly vector with a designed multiple cloning site and homologous flanks for integration into the adenine phosphoribosyltransferase (APT, Pp3c8_16590V3.1) locus. This vector was subsequently linearized with two restriction enzymes cutting within the designed MCS and ligated with the respective transgene expression cassette digested with the same enzymes from the expression vector. This was performed subsequently for the GNE-, NANS, and NANP-expression cassettes, resulting in the assembled GNN-construct. **(B)** Successive cloning procedure of the CCSB-construct starting with the generation of the pTOPO-based assembly vector with a designed multiple cloning site and homologous flanks for integration into the prolyl-4-hydroxylase 1 (P4H1, Pp3c8_7140V3.1) locus. This vector was subsequently linearized with two restriction enzymes cutting within the designed MCS and ligated with the respective transgene expression cassette digested with the same enzymes from the expression vector. This was performed subsequently for the CMAS, CSAT, ST and BSD-expression cassettes, resulting in the assembled CCSB-construct. The restriction enzymes used are indicated above the constructs. 5'HR: 5'-homologous region, 3'HR: 3'-homologous region, 35SP_long: long CaMV 35S promoter, 35SP: CaMV 35S promoter, nosT: nos terminator, 35ST: CaMV 35S terminator.



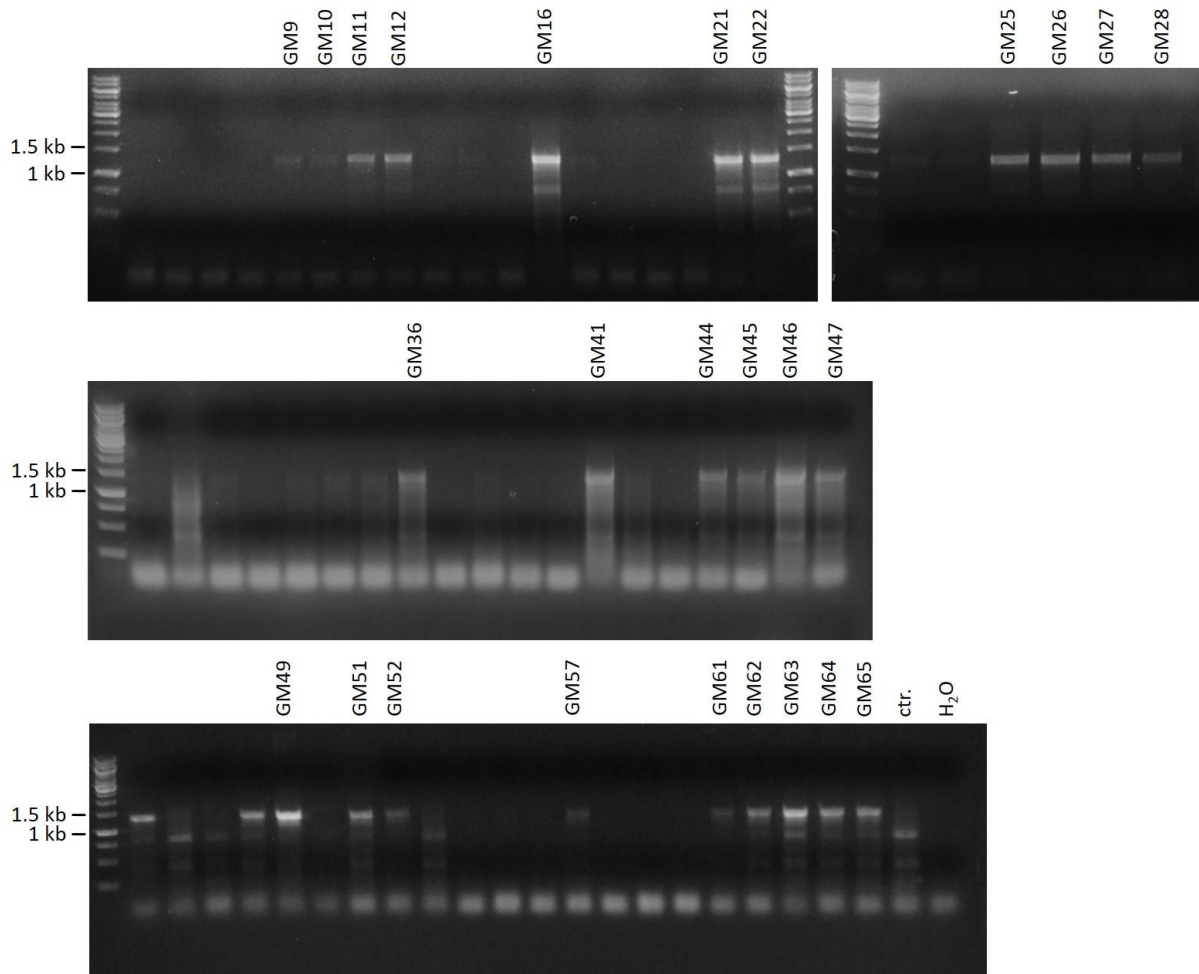
Supplementary Figure S2. Schematic illustration of the plasmid containing the FTGT expression construct for targeted integration into the GalT3 locus. The homologous 5'- and 3'-flanks (5'HR GalT3, 3'HR GalT3) for targeted integration into the β 1,3-galactosyltransferase 1 (GalT3) locus are displayed in orange. CaMV 35S long: long CaMV 35S promoter, FT4_CTS: sequence encoding the N-terminus of moss α 1,4-fucosyltransferase including the CTS domain; GalT4_cat_dom.: sequence encoding the catalytic domain of the human GalT4, P2A: 2A peptide sequence from the porcine teschovirus-1, zeo: zeocine resistance coding for bleomycin resistance protein, 35S_T: CaMV 35S terminator, kan: kanamycin selection cassette, bla: ampicillin resistance cassette coding for beta-lactamase.



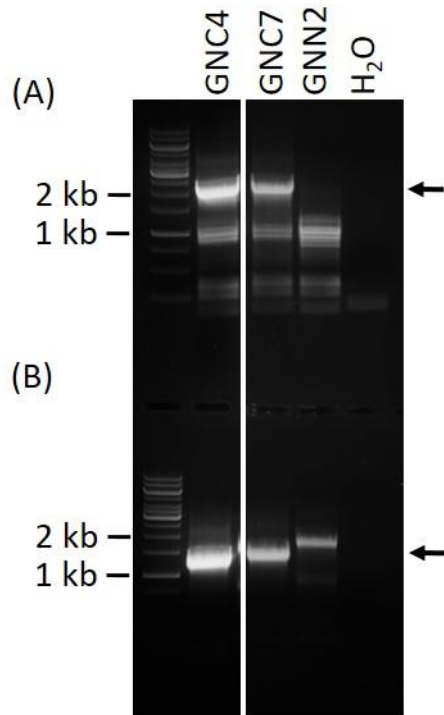
Supplementary Figure S3. Schematic illustration of the plasmid containing the FTST expression construct for targeted integration into the P4H2 locus. The homologous 5'- and 3'-flanks (5'HR_P4H2, 3'HR_P4H2) for targeted integration into the prolyl-4-hydroxylase 2 (P4H2) locus are displayed in green. CaMV 35S long: long CaMV 35S promoter, FT4_CTS: sequence encoding the N-terminus of moss α 1,4-fucosyltransferase including the CTS domain; ST_cat_dom.: sequence encoding the catalytic domain of the rat α 2,6-sialyltransferase, 35S_T: CaMV 35S terminator, hpt_cassette: hygromycin B phosphotransferase encoding resistance gene driven by the nos promoter and nos terminator, bla: ampicillin resistance cassette coding for beta-lactamase.



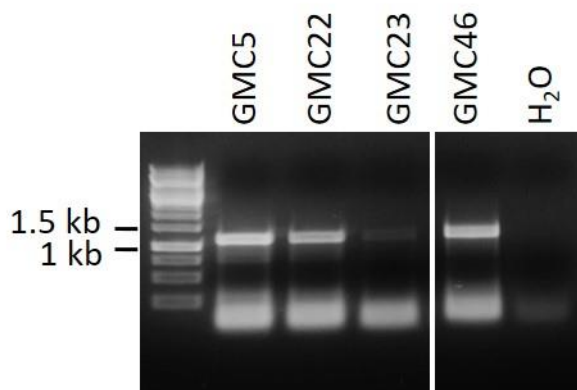
Supplementary Figure S4. PCR-based analysis of 5'- (A) and 3'-integration (B) of the GNN-construct within the APT locus. PCR reactions were performed using primer pairs 53 and 54 for 5'- and 55 and 56 for 3'-integration on genomic DNA of lines GNN1-4 as well as of parental line as negative control (ctr.). To demonstrate the absence of contamination a control without DNA (H₂O) was included. Amplicons of 2244 bp for the 5'- ((A), GNN2 and GNN4) and 1311 bp for the 3'-integration (B) confirmed the targeted integration.



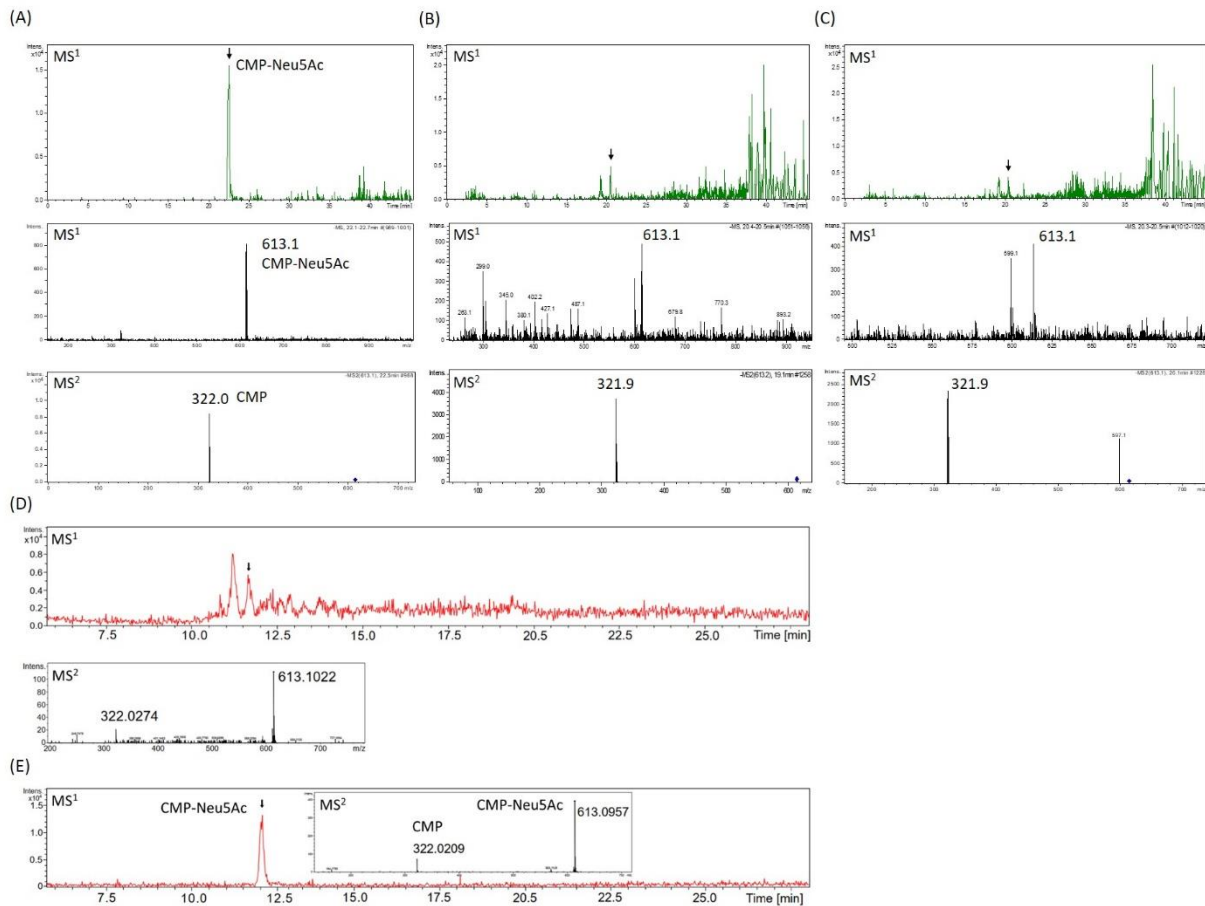
Supplementary Figure S5. PCR-based analysis of 3'-integration of the GM-construct within the APT locus. PCR reactions were performed using the primer pairs 55 and 56 for 3'-integration on genomic DNA of lines GM1-65 as well as of a negative control (ctr.). To demonstrate the absence of contamination a control without DNA (H₂O) was included. Amplicons of 1311 bp confirmed targeted 3' construct integration.



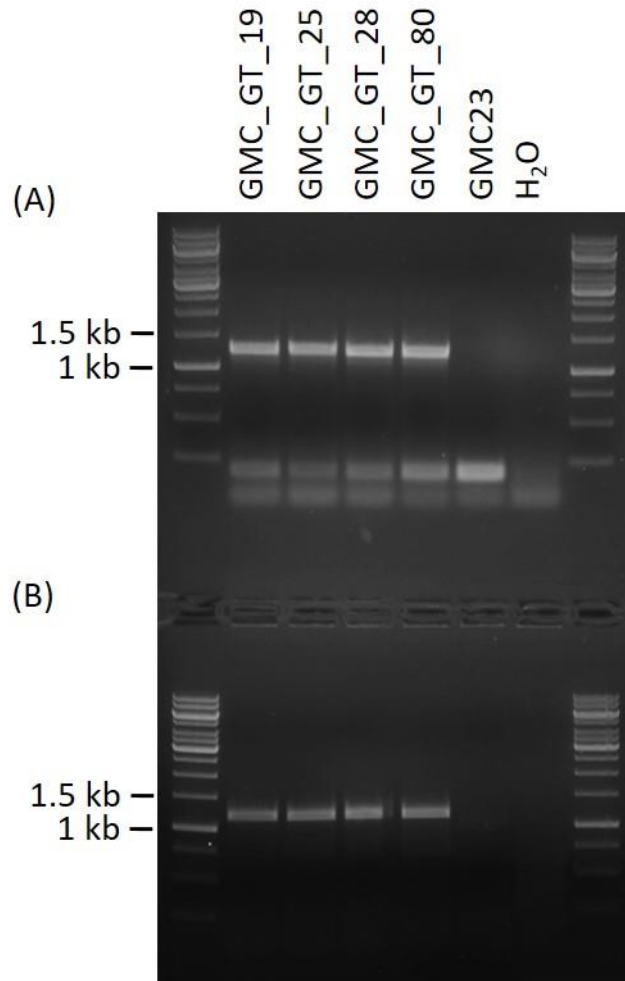
Supplementary Figure S6. PCR-based analysis of 5'- and 3'-integration of the CCSB-construct in the P4H1 locus. PCR reactions were performed using primer pairs 57 and 58 for 5'- and 59 and 60 for 3'-integration on genomic DNA of lines GNC4 and GNC7 as well as of the GNN2-parental line as negative control. To demonstrate the absence of contamination a control without DNA (H₂O) was included. Amplicons of 2222 bp for the 5'- (**A**) or 1406 bp for the 3'-integration (**B**) confirmed the targeted integration in the lines GNC4 and GNC7 (indicated by the black arrows).



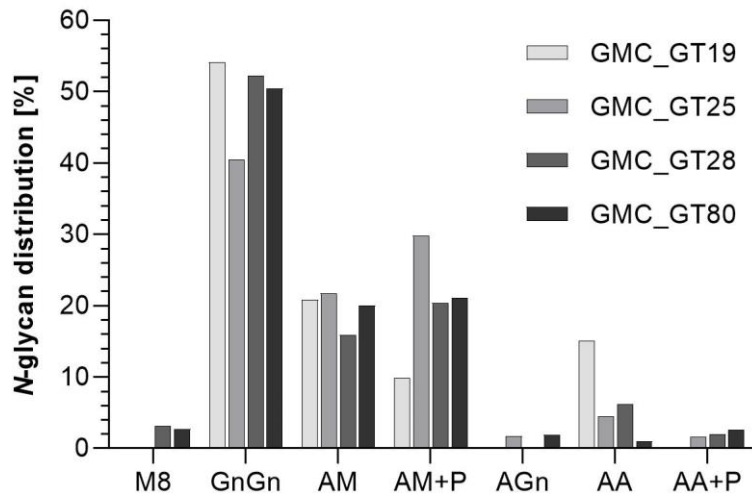
Supplementary Figure S7. PCR-based analysis of 3'-integration of the CCSB-construct in the P4H1 locus. PCR reactions were performed using primer pair 59 and 60 for 3'-integration on genomic DNA of lines GMC5, 22, 23, and 46, resulting from CCSB transfections of line GM28. To demonstrate the absence of contamination a control without DNA (H₂O) was included. Amplicons of 1406 bp confirmed targeted 3' construct integration in all analyzed lines.



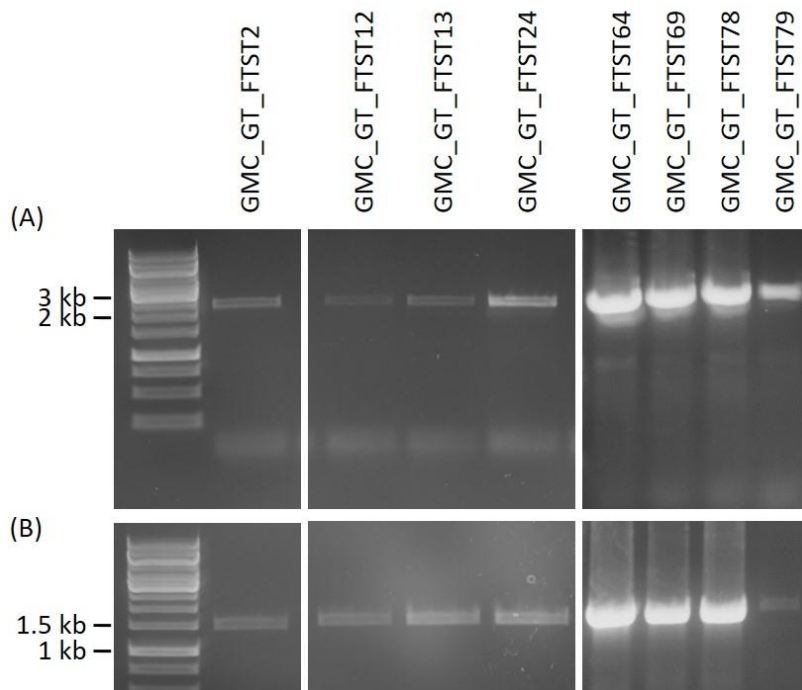
Supplementary Figure S8. Mass-spectrometric detection of CMP-Neu5Ac in moss extracts (**B-D**) compared to measurements of a CMP-Neu5Ac standard (**A+E**). Extracted ion chromatogram (m/z of 613.1 ± 0.1) of LC-MS measured extracts of the lines GMC23 (**B**), GMC46 (**C**), GNC7 (**D**) and 70 pmol of a CMP-Neu5Ac standard (**A** (standard¹ measurement for B + C) and **E** (standard measurement for D)), respectively (upper panels A-D and E). In all measurements peaks corresponding to the $[M-H]^-$ -ion of CMP-Neu5Ac of $m/z = 613.1$ could be detected on MS¹ level (middle panels A-C, E and lower panel of D). Confirmation of corresponding MS¹-peak identities (indicated by the black arrows) was performed on MS² level via the identification of the $[M-H]^-$ CMP fragment ion of $m/z = 322.0$ (lower panels A-D and E). CMP-Neu5Ac concentration of the moss extracts was determined via peak area integration in comparison to defined standard values and resulted in 14 nmol CMP-Neu5Ac/g DW in the lines GMC23 and 46 and 2.3 nmol CMP-Neu5Ac/g DW in GNC7.



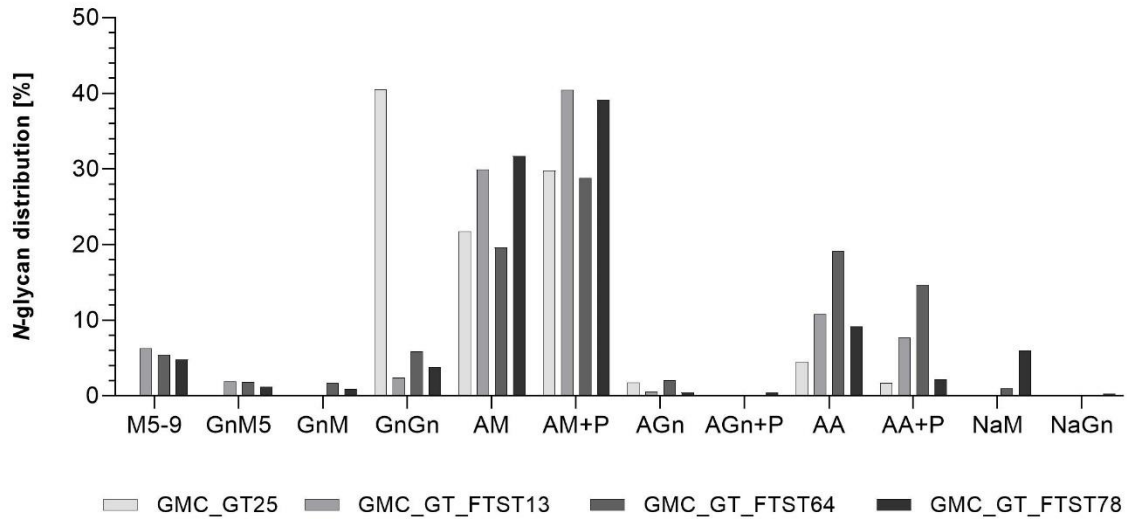
Supplementary Figure S9. PCR-based analysis of 5'- and 3'-integration of the FTGT-construct in the GalT3 locus. PCR reactions were performed on genomic DNA of the lines GMC_GT_19, 25, 28 and 80 as well as the GMC23 parental line using the primer pairs 61 and 62 for 5'- as well as 63 and 64 for 3'-integration, respectively. To demonstrate the absence of contamination a control without DNA (H₂O) was included. These analyses resulted in integration-confirming amplicons of 1253 for the 5'- (**A**) and 11171 bp for the 3'-integration (**B**) for all transgenic lines, whereas no integration-indicating signals were obtained in the parental line and the H₂O control.



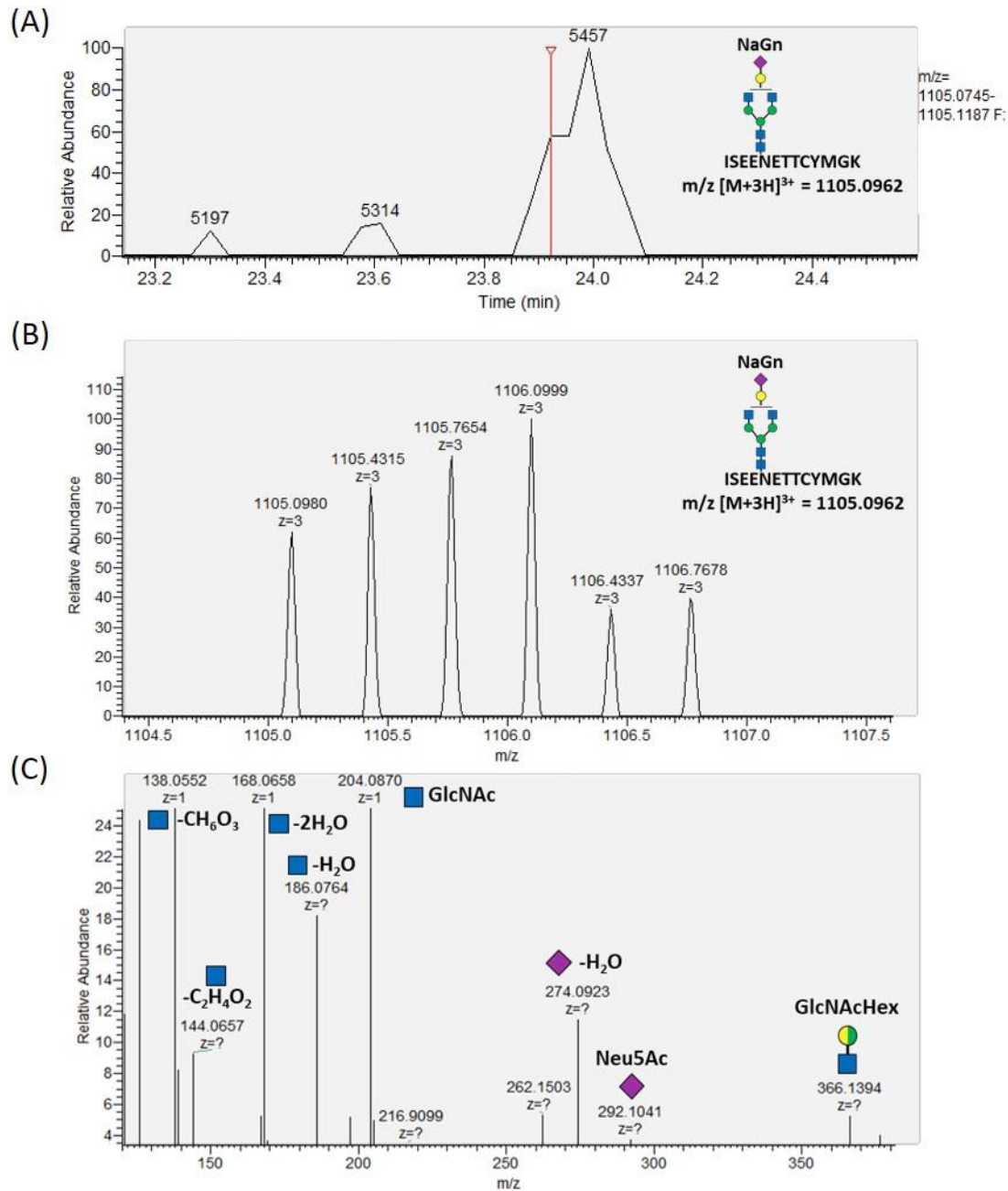
Supplementary Figure S10. Mass-spectrometric analysis of *N*-glycosylation patterns in GMC_GT lines. MS-based relative quantification of identified tryptic glycopeptides from the reporter glycoprotein of complete sialylation lines. Relative quantification was based on peak area integration of extracted ion chromatograms on MS¹ level, for which peak identities were confirmed on MS² level. For quantification, areas of all confirmed peaks per measurement were summed up and the relative percentages are given for each identified glycan structure. M: mannose, Gn: *N*-acetylglucosamine, A: galactose, P: represents the presence of one or two pentoses attached to the *N*-glycan.



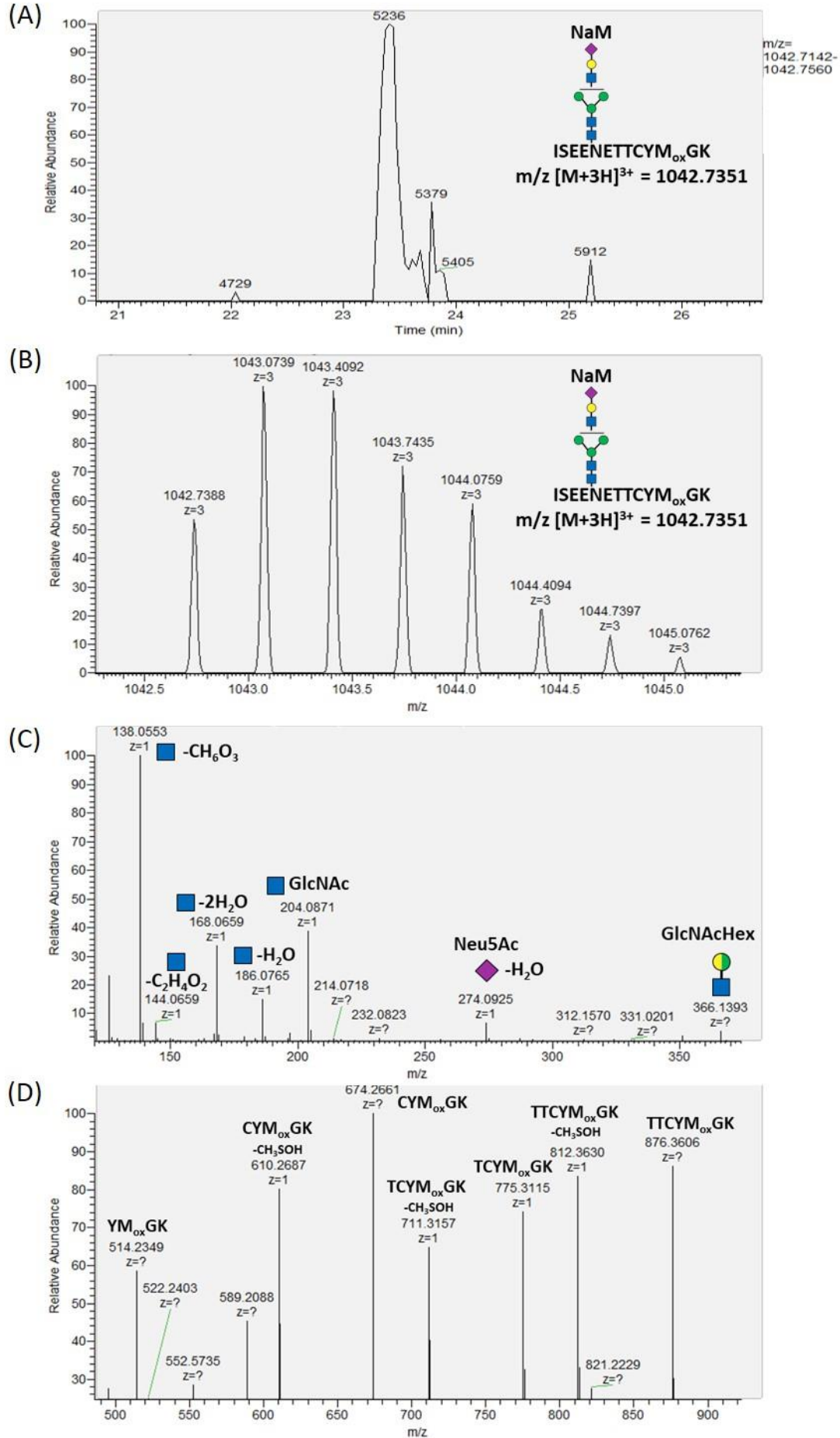
Supplementary Figure S11. PCR signals confirming targeted 5'- and 3'-integration of the FTST-construct within the P4H2 locus. PCR reactions were performed on genomic DNA using the primer pairs 65 and 66 for 5'- as well as 67 and 68 for 3'- integration, respectively. These analyses resulted for all analyzed lines in integration-confirming amplicons of 2169 for the 5'- (**A**) and 1464 bp for the 3'-integration (**B**), respectively.



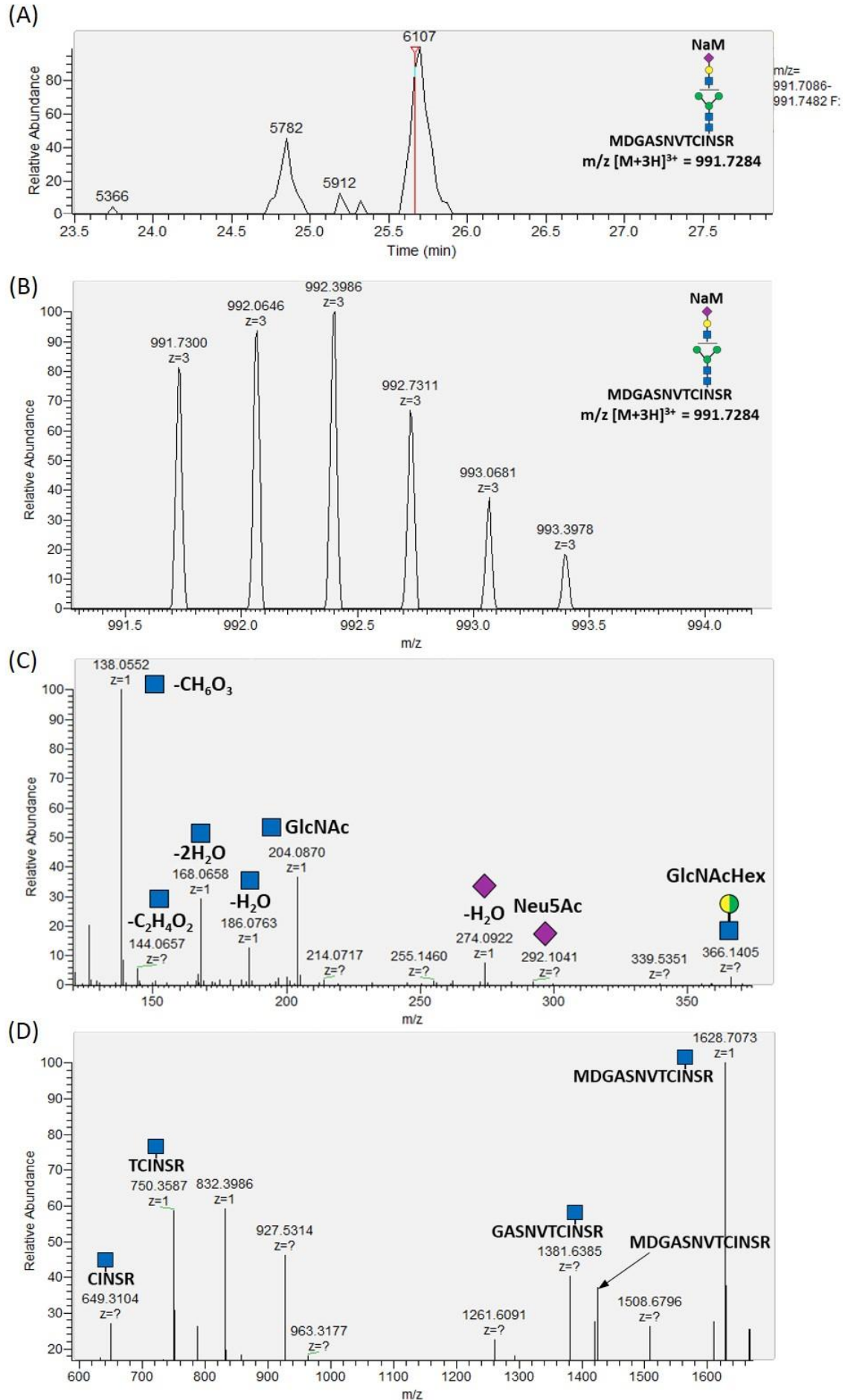
Supplementary Figure S12. Mass-spectrometric analysis of glycosylation patterns in FTST-expressing lines compared to the GMC_GT25 parental line. MS-based relative quantification of glycopeptides identified on the tryptically digested reporter glycoprotein. Relative quantification was based on peak area integration of extracted ion chromatograms on MS¹ level, for which peak identities were confirmed on MS² level. For quantification, areas of all confirmed peaks per measurement were summed up and the relative percentages are given for each identified glycan structure. M: mannose, Gn: *N*-acetylglucosamine, A: galactose, Na: sialic acid, P: indicates the presence of one or two pentoses on the corresponding *N*-glycan structure.



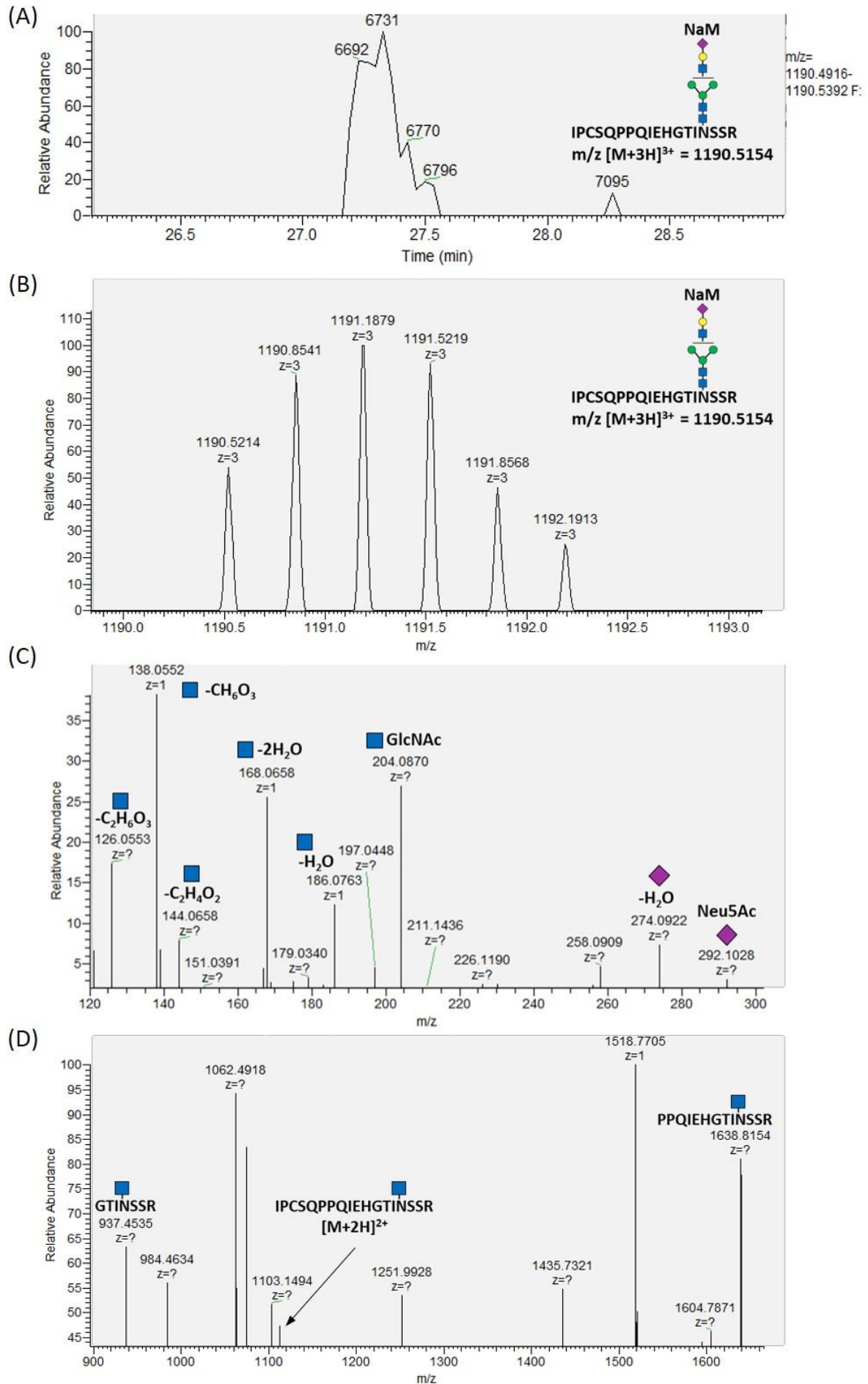
Supplementary Figure S13. Mass-spectrometric verification of the identified NaGn-glycosylation of the ISEENETTCYMGK glycopeptide on MS¹ (A+B) and MS² level (C). (A) Extracted ion chromatogram of the expected [M+3H]³⁺ m/z-ratio of 1105.0962 of the tryptic NaGn-ISEENETTCYMGK glycopeptide on MS¹ level. (B) Verification of the charge state (z=3) of the detected glycopeptide via investigation from the isotope pattern. (C) MS²-based verification of *N*-glycan sialylation via the detection of *N*-acetylglucosamine (GlcNAc) and sialic acid (Neu5Ac) reporter ions with the following m/z-values: [GlcNAc]⁺ = 204.087, [GlcNAc - H₂O]⁺ = 186.076, [GlcNAc - 2H₂O]⁺ = 168.066, [GlcNAc - C₂H₄O₂]⁺ = 144.065, [GlcNAc - CH₆O₃]⁺ = 138.055, [GlcNAc - C₂H₆O₃]⁺ = 126.055), [Neu5Ac]⁺ = 292.103, [Neu5Ac - H₂O]⁺ = 274.092 and the detection of the glycan fragment ion [GlcNAcHex]⁺ = 366.139. Blue square: *N*-acetylglucosamine (Gn, GlcNAc), green circle: mannose (M), yellow circle: galactose (A), purple rhombus: sialic acid (Na, Neu5Ac), Hex: hexose (yellow and green circle: stands for the presence of either mannose or galactose).



Supplementary Figure S14. Mass spectrometric verification of the identified NaM-glycosylation of the ISEENETTCYM_{ox}GK glycopeptide on MS¹ (**A+B**) and MS² level (**C+D**). (**A**) Extracted ion chromatogram of the expected [M+3H]³⁺ m/z-ratio of 1042.7351 of the tryptic NaM-ISEENETTCYM_{ox}GK glycopeptide on MS¹ level. (**B**) Verification of the charge state (z=3) of the detected glycopeptide via investigation from the isotope pattern. (**C**) MS²-based verification of *N*-glycan sialylation via the detection of *N*-acetylglucosamine (GlcNAc) and sialic acid (Neu5Ac) reporter ions with the following m/z-values: [GlcNAc]⁺ = 204.087, [GlcNAc - H₂O]⁺ = 186.076, [GlcNAc - 2H₂O]⁺ = 168.066, [GlcNAc - C₂H₄O₂]⁺ = 144.065, [GlcNAc - CH₆O₃]⁺ = 138.055, [GlcNAc - C₂H₆O₃]⁺ = 126.055), [Neu5Ac]⁺ = 292.103, [Neu5Ac - H₂O]⁺ = 274.092 and the detection of the glycan fragment ion [GlcNAcHex]⁺ = 366.139. (**D**) Fragmentation of the peptide backbone. Blue square: *N*-acetylglucosamine (Gn, GlcNAc), green circle: mannose (M), yellow circle: galactose (A), purple rhombus: sialic acid (Na, Neu5Ac), Hex: hexose (yellow and green circle: stands for the presence of either mannose or galactose).



Supplementary Figure S15. Mass-spectrometric verification of the identified NaM-glycosylation of the MDGASNVTCINSR glycopeptide on MS¹ (**A+B**) and MS² level (**C+D**). **(A)** Extracted ion chromatogram of the expected [M+3H]³⁺ m/z-ratio of 991.7284 of the tryptic NaM-MDGASNVTCINSR glycopeptide on MS¹ level. **(B)** Verification of the charge state (z=3) of the detected glycopeptide via investigation from the isotope pattern. **(C)** MS²-based verification of *N*-glycan sialylation via the detection of *N*-acetylglucosamine (GlcNAc) and sialic acid (Neu5Ac) reporter ions with the following m/z-values: [GlcNAc]⁺ = 204.087, [GlcNAc - H₂O]⁺ = 186.076, [GlcNAc - 2H₂O]⁺ = 168.066, [GlcNAc - C₂H₄O₂]⁺ = 144.065, [GlcNAc - CH₆O₃]⁺ = 138.055, [GlcNAc - C₂H₆O₃]⁺ = 126.055), [Neu5Ac]⁺ = 292.103, [Neu5Ac - H₂O]⁺ = 274.092 and the detection of the glycan fragment ion [GlcNAcHex]⁺ = 366.139. **(D)** Fragment ions of the glycopeptide. Blue square: *N*-acetylglucosamine (Gn, GlcNAc), green circle: mannose (M), yellow circle: galactose (A), purple rhombus: sialic acid (Na, Neu5Ac), Hex: hexose (yellow and green circle: stands for the presence of either mannose or galactose).



Supplementary Figure S16. Mass-spectrometric verification of the identified NaM-glycosylation of the IPCSQPPQIEHGTINSSR glycopeptide on MS¹ (**A+B**) and MS² level (**C+D**). **(A)** Extracted ion chromatogram of the expected [M+3H]³⁺ m/z-ratio of 1090.5154 of the tryptic MNa-IPCSQPPQIEHGTINSSR glycopeptide on MS¹ level. **(B)** Verification of the charge state (z=3) of the detected glycopeptide via investigation from the isotope pattern. **(C)** MS²-based verification of *N*-glycan sialylation via the detection of *N*-acetylglucosamine (GlcNAc) and sialic acid (Neu5Ac) reporter ions with the following m/z-values: [GlcNAc]⁺ = 204.087, [GlcNAc - H₂O]⁺ = 186.076, [GlcNAc - 2H₂O]⁺ = 168.066, [GlcNAc - C₂H₄O₂]⁺ = 144.065, [GlcNAc - CH₆O₃]⁺ = 138.055, [GlcNAc - C₂H₆O₃]⁺ = 126.055, [Neu5Ac]⁺ = 292.103 and [Neu5Ac - H₂O]⁺ = 274.092. **(D)** Fragment ions of the glycopeptide. Blue square: *N*-acetylglucosamine (Gn, GlcNAc), green circle: mannose (M), yellow circle: galactose (A), purple rhombus: sialic acid (Na, Neu5Ac), Hex: hexose (yellow and green circle: stands for the presence of either mannose or galactose).