Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants

Alexandra Castilho1, Laura Neumann2, Pia Gattinger1, Richard Strasser1, Karola Vorauer-Uhl3, Thomas Sterovsky4, Friedrich Altmann2, Herta Steinkellner1*

1 Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria, 2 Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria, 3 Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, 4 Polymun Scientific GmbH, Klosterneuburg, Austria

Abstract

Hyperglycosylated proteins are more stable, show increased serum half-life and less sensitivity to proteolysis compared to non-sialylated forms. This applies particularly to recombinant human erythropoietin (rhEPO). Recent progress in N-glycoengineering of non-mammalian expression hosts resulted in in vivo protein sialylation at great homogeneity. However the synthesis of multi-sialylated N-glycans is so far restricted to mammalian cells. Here we used a plant based expression system to accomplish multi-antennary protein sialylation. A human erythropoietin fusion protein (EPOFc) was transiently expressed in Nicotiana benthamiana :XTFT, a glycosylation mutant that lacks plant specific N-glycan residues. cDNA of the hormone was co-delivered into plants with the necessary genes for (i) branching (ii) β1,4-galactosylation as well as for the (iii) synthesis, transport and transfer of sialic acid. This resulted in the production of recombinant EPOFc carrying bi-, tri- and tetra-sialylated complex N-glycans. The formation of this highly complex oligosaccharide structure required the coordinated expression of 11 human proteins acting in different subcellular compartments at different stages of the glycosylation pathway. In vitro receptor binding assays demonstrate the generation of biologically active molecules. We demonstrate the in planta synthesis of one of the most complex mammalian glycoforms pointing to an outstanding high degree of tolerance to changes in the glycosylation pathway in plants.

Introduction

Recombinant human erythropoietin (rhEPO) was the first hematopoietic growth factor approved to treat anemia associated with kidney failure, cancer and other pathological conditions [1]. Mature EPO is a 30 kDa glycoprotein with 166 amino acids carrying three N-linked (Asn-24, -38 and -83) and one O-linked (Ser-126) carbohydrate chains which account for 40% of the total molecular weight [2,3]. Glycosylation has a profound effect in maintaining the overall stability and in vivo hematopoietic activity of hEPO [4–6]. Several studies report that terminal sialic acid increases the circulatory half-life of rhEPO, moreover a positive correlation between the in vivo biological activity and the ratio of tetra- to bi-antennary sialylated oligosaccharides was shown [7,8]. Due to the complexity of the glycosylation pattern, therapeutic rhEPO is exclusively produced in mammalian cell cultures, mainly in Chinese hamster ovary (CHO) [9–11]. Many efforts have been made to improve the sialylation content of the hormone [12–14]. Indeed, hyper-sialylated rhEPOs with prolonged half-life and subsequent enhanced drug efficacy were produced [6]. Another strategy to improve drug efficacy of rhEPO is its fusion to stabilizing peptides/proteins. The application of immunoglobulin Fc-fusions to therapeutic proteins has become very popular since the Fc fragment can extend the conjugated protein serum half-life by being recycled via the neonatal Fc receptor (FcRn). EPOFc fusions have been successfully explored in this direction [15].

The limited production capacity and expensive mammalian cell based production facilities make the recombinant hormone very costly. A viable alternative for the large-scale and low cost production of biopharmaceuticals is the use of plants [16,17]. Recent progress in expression levels, production speed and up-scaling, have placed this expression system into an encouraging position. Another important feature of using plants as production platform is their ability to carry out human-like complex N-glycosylation. Due to their comparable small repertoire of glycosylation reactions, plants carry out complex N-glycosylation with remarkable homogeneity, which makes them especially amenable for N-glycoengineering. Indeed, over the past years many research groups have concentrated their efforts on modulating plant N-glycosylation to enable the production of recombinant proteins with human-like structures [Review [18,19]]. One of the most impressive results is the introduction of the mammalian biosynthetic pathway for in planta protein sialylation [20].

Previous attempts to produce rhEPO in plants resulted in the generation of a recombinant hormone that shows in vitro activity
However, plant-derived rEPO was not active in vivo most probably due to the lack of sialylation [24]. Regrettably, most of these studies did not consider the glycosylation status of the recombinant hormone. rhEPO and rhEPOFc produced in glycoengineered moss and

*N. benthamiana* carried mainly human type complex GlcNAc2Man3GlcNAc2 (GnGn) structures, lacking plant specific xylose and fucose [25,26]. Moreover the production of rhEPOFc with tetra-antennary bisected complex N-glycans was achieved upon overexpression of mammalian N-acetylglucosaminyltransferases (GnTIII, GnTIV and GnTV) [26,27]. Overall, the results demonstrate the feasibility of plants to generate active rhEPOFc with a targeted N-glycosylation profile, however plant derived (multi-) sialylated rhEPOFc remains elusive.

In this investigation we set out to express in plants rhEPOFc carrying tri- and tetra-sialylated N-glycans. Agrobacterium containing rhEPOFc cDNA was delivered to *N. benthamiana* ΔXTFT mutants (lacking the plant specific N-glycan residues β1,2-xylose and core α1,3-fucose) together with the mammalian genes required for in planta protein sialylation [20]. N-glycosylation profiling of the recombinant protein using LC-ESI-MS exhibited the synthesis of mainly complex bi-antennary sialylated N-glycans, i.e. NeuAc2Gal2GlcNAc2Man2GlcNAc2 (NaNa). Transient co-expression of rhEPOFc with mammalian genes necessary for the branching and sialylation of N-glycans (in total 11 genes) resulted in the synthesis of rhEPOFc decorated with tri- and tetra-sialylated oligosaccharides. All glycoforms exhibit biological activities comparable to the CHO derived rhEPOFc, as determined by cell-based receptor binding assays.

**Materials and Methods**

**Vectors for single gene expression**

MagnICON pro-vector system was used for the expression of rhEPOFc chimeric proteins as described before [26]. For modulation of rhEPOFc N-glycosylation profile we used the previously described binary vectors each carrying a single gene necessary to produce multi-antennary N-glycans (FUT11GnTIV and FUT11GnTV, [26]) and to assemble in planta the metabolic pathway for protein sialylation (GNE, NANS and CMAS, CST, STGalT and ST, [20]).

**Binary vectors for multiple gene expression**

We combined six expression cassettes in two different binary plasmids: one for the expression of the genes necessary for the synthesis of sugar-activated sialic acid, CMP-Neu5Ac (GNE, NANS and CMAS) and another for the expression of genes
necessary for synthesis of the acceptor substrate (β1,4-galactosylation), Golgi transport and transfer of sialic acid (CST, STGalT and ST). To this intent we used the versatile pSAT family that allows target genes to be cloned under a large choice of promoters and terminators that are easily interchangeable [Fig 1A, [17]]. cDNA from each gene were amplified from the correspondent binary vector described in Castilho et al [20]. Appropriate rare-cutting enzymes flanking the expression cassettes in each pSAT vector were used to assemble several cassettes into plant transformation RCS2-based vectors carrying the same rare-cutting enzymes [17]. The pSAT auxiliary vectors and the pPZP-RCS2 binary vectors were purchased from University of Michigan, USA.

Construction of vector for the expression of GNE, NANS and CMAS

The cDNA of GNE was amplified with primers GNE R1/F1 digested with XhoI/BglII and cloned into pSAT1A digested XhoI/ BamHI (pSAT1A-GNE). NANS and CMAS cDNAs were amplified with primers NANS F1/R1 and CMAS F1/R1 respectively, digested with XhoI/BamHI and cloned into pSAT1A (pSAT1A-NANS and pSAT1A-CMAS). The expression cassette of pSAT1A-CMAS was transferred into the AgdI-Xel sites of the pSAT6A. To obtain the construct for simultaneous expression of the three proteins the expression cassette of pSAT1A-GNE was removed by AgdI digestion and cloned into the Xel site of pPZP-RCS2, the expression cassette from pSAT1A-NANS was removed by Xel digestion and cloned into the I-SceI site of pPZP-RCS2 and finally the CMAS expression cassette was inserted into the site PI-Pol of pPZP-RCS2 (pG144, Figure 1B).

Construction of vector for the expression of CST, STGalT and ST

The cDNA from CMP-Neu5Ac transporter was amplified from the correspondent binary vector with the primer pair CST F1/R1, digested with XhoI/BamHI and cloned into pSAT3A digested XhoI/ BamHI (pSAT3A-ST). STGalT R1 primers digested and cloned into the XhoI site of pPZP-RCS2 the expression cassette of pSAT3A-ST was removed by XhoI digestion and cloned into the XhoI site of pPZP-RCS2 and finally the CMAS expression cassette was inserted into the site PI-Pol of pPZP-RCS2 (pG371, Figure 1C).

All binary vectors were transformed into the Agrobacterium tumefaciens strain U1A 143 and magnICON constructs were transformed into strain GV3101 pMP90. All primers used in this investigation are listed in Table S1.

Plant material and transient protein expression

Nicotiana benthamiana AXTFT plants [29] were grown in a growth chamber at 22°C with a 16 h light/8 h dark photoperiod.

Transient expression of rEPOFc was done in four-to-five-week old plants by agroinfiltration. The magnICON 3′-vector containing cDNA was co-infiltrated with the corresponding 5′-vector carrying the signal peptide for secretion in combination.

Table 1. Expression of rEPOFc in N. benthamiana.

| rEPOFc | mg/kg | % TSP |
|--------|-------|-------|
| AXTFT  | 9.13  | 0.18  |
| Sia    | 6.14  | 0.12  |
| TriSia | 9.12  | 0.18  |
| TetraSia | 9.11 | 0.18  |

Concentration of transiently expressed rEPOFc was determined using a commercially available immunoassay. For each sample the concentration is given in mg/kg of fresh leaf. The percentage of the total soluble protein (TSP) was also calculated. rEPOFc was expressed in N. benthamiana AXTFT mutants (AXTFT), co-expressed in AXTFT with mammalian genes for protein sialylation (Sia), co-expressed in AXTFT with mammalian genes for synthesis of tri-antennary sialylated N-glycans (TriSia) and co-expressed in AXTFT with mammalian genes for synthesis of tetra-sialylated N-glycans (TetraSia).

doi:10.1371/journal.pone.0054836.t001
with the binary vector for the expression of the recombinase [30].

For modulation of the N-glycosylation profiles, binary vectors containing the cDNA of the different mammalian genes were co-infiltrated with the magnICON viral-based vectors. Agrobacteria carrying the magnICON constructs were infiltrated using optical density (OD600) 0.1 and 0.05 for agrobacteria carrying binary constructs.

Protein purification, N-glycan analysis and peptide mapping

rhEPOFc was purified from agroinfiltrated leaves (200–300 mg) with rProteinA Sepharose Fast Flow (GE Healthcare) as described previously [26]. For glycopeptide analysis, purified rhEPOFc were resolved by SDS-PAGE and bands corresponding to 55 kDa were cut out, S-alkylated and double-digested with trypsin and endoproteinase Glu-C. This double digestion allows site-specific analysis of all four N-glycosylation sites (GPs): EPO GP1: E/A22ENITTGCAE, EPO GP2: E/H32CSLNENITVPD TK45, EPO GP3: R/G77QALLNNSQPWEPLQHLVDK and Fc glycopeptide: R/EEQYNSTYR. Subsequently samples were analysed by liquid chromatography electrospray ionization-mass spectrometry, LC-ESI-MS [31,32]. Briefly, a BioBasic C18 column (150×0.32 mm, 5 μm; Thermo Scientific) was eluted with 0.3% formic acid buffered to pH 3.0 with ammonia as the aqueous solvent and a gradient from 10% to 55% acetonitrile developed over 40 min of 1.5 mL/min. The glycoforms of a given peptide co-eluted due to the use of buffered eluent [32]. The elution zone of each peak was summed and the spectra were deconvoluted using MaxEnt3 (Waters Micromass). Peak heights were taken as indicators of the molar ratios of glycoforms, which was recently shown to give meaningful results for Fc-glycopeptides [31].

The 30 kDa protein band corresponding to free Fc was analysed by LC ESI MS/MS for peptide mapping in order to identify the N-terminus. The data was analyzed using the X! Tandem open source software to match tandem mass spectra with the EPO-Fc protein sequence. The N-terminal peptide was identified by the GPM (Global Protein Machine) search engine.

Immunoblot Analysis

Five micrograms of total soluble protein and Protein A purified rhEPOFc were subjected to 12% SDS-PAGE under reducing conditions and blotted onto Hybond Enhanced Chemiluminescence nitrocellulose membranes (GE Healthcare). The blots were blocked in 1xPBS containing 0.1% (v/v) Tween 20 and 3% (w/v) BSA for 1 h and the protein bands were analysed by immuno blotting using either anti-hEPO (1:3000 dilution MAB2871, R&D Systems, Minneapolis, MN), anti-human IgG (1:5000 dilution anti-Fc, W4031 Promega, Mannheim, Germany) or anti-Lewis-A (1:40 dilution JIM84, kindly provided by Paul Knox, University of Leeds, UK) antibodies.

rhEPOFc quantification and in vitro assay

The expression level of plant-derived rhEPOFc was measured in total soluble proteins using the Quantikine IVD ELISA for human EPO (DEPOO, R&D Systems) according to manufacturer’s instructions. The biological activity of protein A purified rhEPOFc was measured in a UT-7 cell based proliferation assay. Briefly, the UT-7 cell line [33] was maintained in RPMI 1640 (Biochrome AG) supplemented with 10% fetal calf serum (PAN Biotech.), 4 mM L-glutamine and 5 ng/mL EPO. The cells were washed with EPO free culture medium and incubated for 4 h at 37°C and 7% CO2. In a 96-well culture plate increasing amounts of CHO-derived rhEPOFc (0.009–60 ng/ml) and plant-derived rhEPOFc (ranging from 0.003–20 ng/mL) were added to 100 μL of medium containing about 105 cells. After 4 days at 37°C and 7% CO2, 10 μL of a MTT (Thiazolyl Blue Tetrazolium Bromide;
Figure 4. Generation of bi-sialylated structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in N. benthamiana ΔXTFT with mammalian genes for protein sialylation (GNE, NANS, CMAS, CST, α2-6GalT and ST) (rhEPOFcSia; Figure 2B, lane 2). Glycosylation patterns of rhEPO Gp1: E/A22ENITTGCAE31; Gp2: E/H32CSLNEITVPDTK45 and Gp3: R/G77QALLVNSQPWEPLQHLVDK97 are shown. N-glycosylation profile of the Fc glycopeptide is shown in Figure S1. Peak labels were made according to the ProGlycAn system (www.proglycan.com). Illustrations display N-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5.
doi:10.1371/journal.pone.0054836.g004
Table 2. Relative abundance of different complex glycoforms detected in rhEPOFc. (oligomannosidic structures that are present in all samples are not included).

| Glycoform (%) | EpofFcSia | EpofFcTriSia | EpofFcTetraSia |
|--------------|-----------|-------------|---------------|
| Glycoform (%) | Gp1 | Gp2 | Gp3 | Gp1 | Gp2 | Gp3 | Gp1 | Gp2 | Gp3 |
| GnGn | 9.1 | 10.1 | 10.9 | 4.8 | 3.7 | 6.6 | 5.6 | 4.4 | 3.4 |
| MNAiso | 12.1 | 17.6 | 22.5 | - | - | - | 12.3 | - | 11.2 |
| NaaNa | 62 | 49.7 | 66.6 | 4 | 12.4 | 15.3 | 13.5 | 29.4 | 26.9 |
| [GnGn]Gn | 18.8 | 10.1 | 10.4 | 10.7 | 2 | 2.1 |
| [GnGn]GnF | - | 2.2 | - | - | - | - |
| [AGn]GnF | - | 2.1 | - | - | - | - |
| [NaNa]Na | 72.4 | 68.1 | 67.7 | 56.3 | 39.6 | 21.5 |
| [GnGn]Gn | 1.6 | 7.7 | 10.4 |
| [GnGn]GnAiso | - | 5.3 | 9.7 |
| [GnA]GnGn | - | 1.2 | 1.3 |
| [NaNa][NaNa] | - | 10.4 | 13.5 |
| Glycoform (%) | 90.9 | 89.9 | 89.1 | 76.4 | 81.5 | 83 | 82.1 | 79.4 | 73.1 |

Relative abundance of complex N-glycans determined by LC-ESI-MS. rhEPOFcSia: rhEPOFc co-expressed with mammalian genes for protein sialylation; rhEPOFcTriSia: rhEPOFc co-expressed with mammalian genes for synthesis of tri-antennary sialylated N-glycans; rhEPOFcTetraSia: rhEPOFc co-expressed with mammalian genes for synthesis of tetra-sialylated N-glycans. ΔXTFT was used as expression host. Values are in percentages. Quantifications were done for complex N-glycans (oligomannosidic structures were not included in calculations). Gp1: glycopeptide 1; Gp2: Glycopeptide 2; Gp3: Glycopeptide 3.

Results

Transient expression of EPOFc in N. benthamiana ΔXTFT

We used N. benthamiana ΔXTFT, a glycosylation mutant that synthesizes complex N-glycans devoid of plant specific B1,2-xylene and core a1,3-fucose, as expression platform [29]. In previous studies we have shown the versatility of these plants for the modulation of plant N-glycosylation towards mammalian-like structures (recently reviewed [19]). Using the potent viral-based expression system magnICON, [30] appropriate agrobacteria carrying hEPOFc cDNA were delivered to ΔXTFT leaves. 4–5 days post infiltration (dpi) expression was monitored by Western blotting. Antibodies against EPO and Fc enabled the detection of a soluble protein (Table 1). rhEPO was purified via protein A-based chromatography and separated by SDS PAGE. Coomassie staining revealed the presence of two bands as already detected by immunoblotting (Figure 2B). Peptide mapping and MS analyses demonstrated that the 55 kDa band corresponds to the intact rhEPOFc, while the 30 kDa band refers to free Fc (data not shown). Similar observations of rhEPOFc fragmentation have been reported in earlier studies in transgenic chickens [34] and in plants [26,27]. In our attempts to enhance the expression of full-length rhEPOFc, different fusion constructs were generated. These included plant codon-optimization of the hEPO fragment using the GeneArt® Gene Synthesis and GeneOptimizer® process (www.liftechnologies.com, GenBank accession No. KC329647), amino acid variations in the hinge region of Fc, and exchange of the hinge-Fc fragment from IgG1 for the IgGD hinge and the IgG4-Fc regions [35]. Another concern is the post translational elimination of the arginyl (Arg166) amino acid residue. Analysis of the C-terminus of CHO-rhEPO and human EPO purified from the urine demonstrates that the Arg166 predicted to be at the C-terminus is missing. This is presumably due to the enzymatic activity of endogenous carboxypeptidases [36]. Since plants contain several types of carboxypeptidases the trimming of Arg166 and the consequent loss of tags fused to the C-terminus cannot be excluded. To possibly prevent this eventual cleavage we generated a hEPOFc fusion lacking this amino acid. Unfortunately none of the strategies led to improved expression of full length EPOFc (data not shown). Moreover, the identification of the N-terminus on the free Fc fraction by LC ESI MS/MS was not clear and the results showed that the ~30 kDa band consist of a mixture of Fc fragment fused to varying sizes of EPO sequence. It was therefore not possible to identify an exact cleavage site between the hEPO and the Fc.

LC-ESI-MS analysis was performed to determine the N-glycosylation profile of purified rhEPOFc expressed in ΔXTFT (rhEPOFcΔXTFT, Figure 2B, lane 1). MS data revealed that all three N-glycosylation sites of rhEPO carry a similar glycosylation pattern (Figure 3), with a major glycoform, GlnGn. In addition significant amounts of structures compatible to Gln[FA]iso were present, a carbohydrate formation already detected previously on plant derived rhEPO and rhEPOFc [25,26]. The presence of the terminal trisaccharide consisting of α1,4-fucose and B1,3-galactose linked to N-acetylgalcosamine also known as Lewis-a epitope can...
rhEPOFc\textsubscript{Trisia}

\textbf{Gp1: AENITTGCAE}

\begin{itemize}
  \item Man8: 2914.08
  \item Man9: 3078.26
  \item NaNa: 3399.24
\end{itemize}

\textbf{Gp2: HCSLNENITVPDTK}

\begin{itemize}
  \item GnGn: 2823.16
  \item Man8: 3330.36
  \item Man9: 3462.40
  \item NaNa: 3632.28
\end{itemize}

\textbf{Gp3: GQALLVNSSQPWEPLQLVVDK}

\begin{itemize}
  \item Man8: 4601.70
  \item Man9: 4223.02
  \item NaNa: 4563.37
\end{itemize}
be detected by immunoreaction to the monoclonal antibody, JIM04 [23,37]. Total soluble proteins (TSP) and protein A purified rhEPOFcEXTT analysed by Western blotting showed that the 55 kDa band corresponding to the intact rhEPOFc reacts with anti-Lewis-x antibodies, while the free Fc 30 kDa band does not (Figure 2C). In fact, the glycosylation profile of Fc exhibits exclusively GnGn structures (Figure S1).

**Multiple gene expression vectors**

In previous studies we have shown that in planta sialylation can be accomplished by co-infiltration of 6 agrobacteria cultures into a plant leaf (each carries a binary vector with a mammalian glycognition gene) [20]. To achieve this, all recombinant proteins including the target protein, which is also co-delivered, must work in the same cell in a highly coordinated fashion. However, the infection of a single cell via agro-infiltration is a random procedure, thus the delivery of single constructs might lead to inefficiencies. To facilitate the simultaneous delivery of all cDNAs to the same cell, two multi gene vectors were generated, each carrying three mammalian glycosylation genes. The pSAT-family vectors allow target genes to be cloned under a large choice of promoters and terminators and the expression cassettes are easily interchangeable (Figure 1A [17]). The six different cDNAs were initially cloned into pSAT vector and subsequently groups of three expression cassettes were assembled in two binary vectors: (i) pC144, carries the genes necessary for the synthesis of nucleotide sugar activated sialic acid, CMP-Neu5Ac (GNE, NANS and CAMS, Figure 1B); (ii) pG371, carries the genes necessary for the synthesis of the β1,4-galactosyl acceptor substrate, Golgi transport and transfer of sialic acid (CST, STGal and ST, Figure 1C). For detailed description of the vectors see Experimental Procedures.

**Generation of bi-sialylated N-glycans on rhEPOFc**

In order to elongate the GnGn glycosides present on rhEPOFcEXTT with β1,4-galactose and α2,6-linked sialic acid, the hormone was co-expressed with the multi gene vectors, pC144 and pG371, allowing a total of 9 genes to be simultaneously delivered to ΔXTFT. Site specific N-glycosylation-profiling of the purified recombinant hormone (rhEPOFcEXTT, Figure 2B, lane 2) showed that all N-glycosylation sites on the rhEPO are similarly occupied and were efficiently modulated (Figure 4). MS analysis of the 55 kDa band revealed that about 90% of complex N-glycans was sialylated (Table 2). Notably, we observed a dominant N-linked glycoform, i.e. bi-antennary sialylated structures (NaNa), which accounts for more than 60% of all complex structures. In addition fucosylated (NaNaF) and incompletely sialylated (MNα) glycosomes were detected and about 10–15% of rhEPOFcEXTT carried oligomannosidic structures (not included in Table 2). Surprisingly, no Lewis-a structures were detected. In contrast, the N-glycan profile of Fc exhibited a largely heterogeneous glycosylation profile, including GnGn, mono and bi-galactosylated structures (GaNA, AA), incompletely processed structures (MNα) and oligomannosidic glycoforms (Figure S1). Notably, the procedure worked in a similar way when single binary vectors were used [20].

**Generation of multi-sialylated N-glycans on rhEPOFc**

The generation of plant derived rhEPOFc carrying branched (tri- and tetra-antennary), N-glycans has been reported previously [26,27]. This was achieved by the co-expression of rhEPOFc with mammalian N-acetylgalcosaminyltransferases IV and V targeted to medial Golgi compartment (FUT11GnTIV or FUT11GnTV; [26]). Here we set out to generate multi-antennary sialylated rhEPOFc. To approach this issue, we first co-expressed rhEPOFc with pC144 and pG371 in combination with either FUT11GnTIV or FUT11GnTV. SDS-PAGE analysis of purified rhEPOFc (rhEPOFcTriSia, Figure 2B lane 3) showed that the 55 kDa band corresponding to the fusion protein appears as a “smeary” band compared to rhEPOFcEXTT and rhEPOFcSia (Fig 2B, lane 1 and 2, respectively). The relative occurrence of the different complex glycosylomers present in rhEPOFcTriSia is shown in Table 2. In total about 80% of all glycans were sialylated, with the dominant N-glycan, being tri-sialylated oligosaccharide. Low amounts of tri-antennary non-sialylated structures are also detected and oligomannosidic structures account for ca. 10–12% of the total N-glycans. In addition the “smeary” 55 kDa band was separated into two fractions (A and B, Figure 2B) and they were individually analysed. The N-glycosylation profile of fraction A (which corresponds to a size slightly larger than 55 kDa) exhibits almost exclusively tri-antennary sialylated carbohydrates in all three glycosites ([NaNa][NaNa] (Figure 5). In contrast, fraction B (which corresponds to the lower part of the 55 kDa band) was decorated mainly with tri-antennary non-sialylated N-glycans with or without galactosylation ([GnGn][Gn, [AGn][Gn], accompanied by oligomannosidic N-glycans (Figure S2).

Finally rhEPOFc was co-expressed with pC144 and pG371 in combination with both FUT11GnTIV and FUT11GnTV. This procedure encompasses a coordinated action of eleven heterologous proteins. The purified product (rhEPOFcTetraSia) exhibited on Coomassie stained SDS-PAGE a “smeary” 55 kDa band as observed for rhEPOFcTriSia.

LC-ESI-MS analysis revealed that rhEPOFcTetraSia glycopeptides carried about 80% sialylated structures including tri- and tetra-sialylation (Table 2). Tri-antennary sialylated structures were the major glycoform in all three rhEPO glycosites (up to 56%). While GP 2 and 3 carried about 10-13% tetra-sialylated structures, surprisingly, this complex carbohydrate was not present on GP 1. Moreover ~15% of rhEPOFcTetraSia are decorated with oligomannosidic structures (not included in Table 2). As before, N-glycosylation analysis was individually performed on the two fractions A and B. (Figure 2B, lane 4). The main glycoform of rhEPOFc in fraction A is tri-sialylated with significant amounts of bi- and tetra-sialylated N-glycans (NaNa and [NaNa][NaNa]) on glycopeptide 2 (Gp2, Asn-38) and Gp3 (Asn-83) (Figure 6). Interestingly, on Gp1 (Asn-24) a single dominant peak corresponding to tri-sialylated structures is detected as well as smaller fractions of bi-sialylated glycans (NaNa) but no tetra-sialylated N-glycans were detected (Figure 6). Fraction B exhibited a variety of non-sialylated branched N-glycans some carrying one or two galactose residues ([GnGn][GnGn], GnGnGn[Gal], Gp2, GnGn][GnA] and [GnA][GnA] alike. Consistently with fraction A, Gp1 carries only GnGn and tri-antennary N-glycans (Figure S3).
Figure 6. Generation of tetra-sialylated structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in N. benthamiana ΔXFFT with mammalian genes for synthesis of tetra-sialylated N-glycans (rhEPOC_tetraSia). The analysis was performed on rhEPOC_tetraSia present on fraction A of the 55kDa band (Figure 2B, lane 4). Glycosylation patterns of rhEPO Gp1: E/A22E NITTGCAE31; Gp2: E/H32CSLNE NITVPDTK45 and Gp3: R/G77QALLV NSSQPWEPLQHLVDK97 are shown. N-glycosylation profile of the Fc glycopeptide is shown in Figure S1.
Glycosylation profile of rhEPOFc present on fraction B of the 55kDa band is shown in Figure S3. Peak labels were made according to the ProGlycAn system (www.proplycan.com illustrations display N-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5).

doi:10.1371/journal.pone.0054836.g006

Analysis of Fc glycosylation in rhEPOFCTriSia and rhEPOFCTetraSia shows a largely heterogeneous N-glycosylation profile with a mixture of GnGn and oligomannosidic glycoforms, but also minor amounts of tri-antennary, galactosylated and sialylated structures (Figure S1). Notably expression levels of all glycoforms were in the same range (Table 1) indicating that co-infiltration of human glycosylation enzymes did not alter expression level of the recombinant fusion protein.

In vitro activity of different EPOFc glycoforms

Finally the biological activity of the plant-derived rhEPOFc variants was assessed using an erythropoietin-dependent human leukemia cell line, UT-7. Proliferation of the UT-7 cells is induced by the presence of EPO. The proliferation of UT-7 cells was measured and half maximal effective dose (ED$_{50}$) values were compared. All plant-derived rhEPOFc glycoforms had similar ED$_{50}$ values ranging 0.26–0.54 ng/mL. Comparably a slightly reduced receptor binding was obtained for the CHO derived counterpart (ED$_{50}$ 1.7 ng/mL) (Table 3). This might be due to different downstream procedures of plant and CHO derived recombinant hormones, e.g. CHO derived rhEPOFc but not the plant derived counterparts was subjected to a virus inactivation test.

Discussion

With the recognition of the N-glycan nature of the ABO blood group types, glycoconjugates were accepted to elicit specific reactions [38]. Since then numerous studies have highlighted the impact of this important posttranslational modification on the function of proteins. A well-known example is rhEPO, one of the leading biopharmaceutical products. The human EPO is a highly glycosylated molecule with three N- and one O-linked glycans. The relevance of O-glycosylation for the biological activity of EPO is unclear, and implications for a role in secretion are not conclusive [39]. On the contrary the biological implications of N-glycosylation are well characterized [4–6]. In the course of enhancing drug efficacy in anemia treatment, increased in vivo half-life via enhanced terminal sialylation was achieved. Moreover, fusing the hormone to an IgG-Fc domain resulted in a significant extension of the serum half-life of the recombinant hormone [40]. Here we report the transient expression of rhEPOFc in plants. Using the N. benthamiana ΔXTFT in combination with the magnICON based expression system, we achieved expression levels of rhEPOFc of up to 9 mg/kg leaves, which accounts for 0.2% of TSP. This is a relatively modest expression level in comparison to amounts reported for other recombinant proteins with the magnICON systems [30,41], however they are in agreement with rhEPOFc expressed previously in plants [27]. Low expression could result from the fact that a large portion of the recombinantly expressed protein (about 30–50 times) refers to Fc lacking the hEPO fragment. We designed different rhEPOFc chimeras to address this issue. However neither manipulation on the Fc-hinge region nor the presence/absence of the EPO Arginine$^{166}$ residue had a significant influence on expression of the full length fusion protein. The generation of free Fc has been already reported previously upon expression of hEPOFc in chicken [34], however it is not present when produced in mammalian cells [9,42]. The reason for this phenomenon has not been investigated in detail, although degradation of the fusion protein by plant proteases is a plausible explanation. Several studies refer to the proteolytic degradation of heterologous proteins in plants [43] and the outcome indicated that this occurs preferentially in the apoplast [44]. Importantly human proteins like EPOFc have not evolved in the context of plant proteases and thereby they represent novel targets. The apoplastic fluid of N. benthamiana is enriched of acidic proteases, e.g. the presence of papain-like cystein family was reported [45]. Papain is a non-specific protease that cleaves monoclonal antibodies preferentially in the N-terminal side of the hinge region and was effectively used to separate the rhEPO from the Fc fragment during N-glycan profiling of rhEPOFc produced in CHO cells [9]. Proteolysis is a major issue of recombinant proteins affecting the product yield, not only in plants but also in other expression systems. Different strategies are being considered to avoid or minimize proteolysis of heterologous proteins expressed in plants [46]. And the outcome hopefully will allow enhanced expression of full length rhEPOFc in plants.

Here we report the generation of rhEPOFc glyco-variants which largely resembles that of the CHO derived counterparts (Figure S4). Expression of rhEPOFc in ΔXTFT mutants results in the formation of almost exclusively GnGn structures on all glycosylation sites. Interestingly, although present in total soluble proteins extracted from ΔXTFT mutants and in some recombinantly expressed proteins [47], no truncated paucimannosidic structures, i.e. MM, were detected. Co-expression of hEPOFc with the mammalian genes involved in protein sialylation permitted the production of a hormone largely decorated with bi-antennary sialylated complex N-glycans. These structures are one of the major glycoforms of EPO present human serum [48], however accounts only for about 15% on the CHO derived counterpart [49]. Moreover, we report the synthesis of plant-derived rhEPOFc carrying multi-antennary sialylated N-glycans, the major structures of mammalian cell derived therapeutic rhEPO. Co-expression of rhEPOFc with the genes necessary for N-glycan branching and sialylation resulted in a mixture of neutral and charged oligosaccharides. In total approximately 10–16 glycoforms, with different relative amounts, are distributed by the four glycopeptides, similar to the observation for CHO-derived rhEPOFc [9]. In summary, rhEPOFc carried 80–90% sialylated structures upon co-expression with the sialylation pathway, Bi- and tri-sialylation were the major

Table 3. in vitro activity of CHO- and plant-derived rhEPOFC.

| rhEPOFc   | ED$_{50}$ (ng/mL) |
|-----------|------------------|
| CHO       | 1.7              |
| ΔXTFT     | 0.45             |
| Sia       | 0.54             |
| TriSia    | 0.25             |
| TetraSia  | 0.26             |

In vitro activity assay of plant- and CHO- derived rhEPOFc. Half maximal effective doses (ED$_{50}$) are displayed. rhEPOFc was expressed in CHO cells (CHO); in N. benthamiana ΔXTFT mutants (ΔXTFT); co-expressed in ΔXTFT with mammalian genes for protein sialylation (Sia); co-expressed in ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated N-glycans (TriSia) and co-expressed in ΔXTFT with mammalian genes for synthesis of tetra-sialylated N-glycans (TetraSia).
glycoform in recombinant products rhEPOFcSia and rhEPOFc-TriSia, respectively. In contrast to mammalian cell derived rhEPO, tetra-sialylation is inefficiently synthesized in plants and rhEPOFc-TetraSia carries only about 10–14% of this highly complex carbohydrate formation. Beside complex N-glycans, plant derived rhEPOFc carries about 10–15% oligomannosidic structures. Notably, and as for mammalian cells, there are significant differences on the N-glycan profile of rhEPO and of Fe domain. While efficient modulation towards mammalian-like structures was observed on all rhEPO glycosites, Fe glycosylation exhibited unusual structures. A homogenous glycosylation profile (namely, GnGn oligosaccharides) was obtained for Fe produced in ΔXTFT; all further modification steps (branching, sialylation) led to the synthesis of a largely heterogeneous glycosylation profile with unusual incompletely processed structures. In particular, sialylation was modest. The unusual Fe glycan-modulation was already observed previously for the synthesis of tetra-antennary N-glycans in rhEPOFc [26]. The reason for this different performance in glyco-modulation of rhEPO and Fe is currently not understood. One explanation could be different accessibility of the N-glycosylation sites, EPO glycosites are considered very exposed and as a consequence they have restricted accessibility to N-glycan processing enzymes.

Multi-sialylation of rhEPOFc requires the coordinated expression of 11 exogenous genes in a single cell. To reduce the number of agrobacteria cultures and to facilitate the simultaneous delivery of glycosylation genes into the same cell, two multi gene vectors carrying the six genes for in planta sialylation were constructed (pC144 and pG371). With a future intention of using these vectors to stably introduce the sialic acid pathway into plants, different plant selection markers have been placed to the vector backbones [50] and as a consequence they have restricted accessibility to N-glycan processing enzymes.

Transgenic N. benthamiana plants stable expressing mammalian glycosyltransferases can be extremely useful for the production of recombinant proteins with a highly homogenous human-like glycosylation profile as recently demonstrated [27,28].

Importantly all plant-derived rhEPOFc glycoforms are biologically active as seen in receptor binding assays. These results are a good starting point for follow up advanced structure-function studies, with the aim to determine the most suitable glycoforms. These will be the focus of future experiments.

With the generation of multi-sialylated glycans we display in vivo engineering of one of the most complex human N-glycan structures in planta and thereby demonstrate the enormous plasticity of plants to tolerate modifications on their protein N-glycosylation. The results presented here together with other achievements in plant N-glycoengineering (reviewed by [19]) provide the know-how for the generation of recombinant proteins with targeted N-glycosylation profiles. This allows advanced protein-carbohydrate structure-function studies to better understand the impact of N-glycans and to develop next generation drugs, where patients would benefit from optimally glycosylated drugs.

Supporting Information

Figure S1 N-glycosylation profile observed in the Fc glycopeptide (R/EEQYNSTYR) of rhEPOFcΔXTFT: rhEPOFc expressed in N. benthamiana ΔXTFT mutants; rhEPOFcSia: rhEPOFc co-expressed in ΔXTFT with mammalian genes for protein sialylation; rhEPOFcTria: rhEPOFc co-expressed in ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated N-glycans; rhEPOFc-TetraSia: rhEPOFc co-expressed in ΔXTFT with mammalian genes for synthesis of tetra-sialylated N-glycans. For interpretation of glycoforms present in assigned peaks see Figure S5.

(TIF)

Figure S2 N-glycosylation profile of rhEPOFcSia present in fraction B of the 55kDa band (Figure 2B, lane 3). Glycosylation patterns of rhEPO Gp1: E/A22ENITTGCAE; Gp2: E/H36CSSLNENITVPDK and Gp3: R/G77QALLVNSSQPWEPLQHLVDR are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.

(TIF)

Figure S3 N-glycosylation profile of rhEPOFcTetraSia present in fraction B of the 55kDa band (Figure 2B, lane 4). Glycosylation patterns of rhEPO Gp1: E/A22ENITTGCAE; Gp2: E/H36CSSLNENITVPDK and Gp3: R/G77QALLVNSSQPWEPLQHLVDR are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.

(TIF)

Figure S4 N-glycosylation profile of rhEPOFc expressed in CHO cells. Glycosylation patterns of rhEPO Gp1: E/A22ENITTGCAE; Gp2: E/H36CSSLNENITVPDK and Gp3: R/G77QALLVNSSQPWEPLQHLVDR and the Fc glycopeptide (R/EEQYNSTYR) are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.

(TIF)

Figure S5 Illustration of N-glycan structures on transiently expressed rhEPOFc. Oligomannosidic structures (Man3–Man9) and complex N-glycans typical of plant-derived proteins (GnGnXf) and glycans carrying Lewis-a epitopes (GnFAx) are also illustrated. Schematic representations are based on the nomenclature proposed by the consortium for Functional Glycomics.

(TIF)

Table S1 List of primers as cited in Material and Methods.

(DOCX)

Acknowledgments

We thank Thomas Hackl and Hanna Weindorfer, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria for excellent technical support, also Dr. Eva Decker, Plant Biotechnology University of Freiburg, Freiburg, Germany for providing the hEPO quantification kit.

Footnote: Nucleotide sequence data for EPOFc codon optimized for plants is available in the Genbank under the accession number KC329647.

Author Contributions

Conceived and designed the experiments: AC HS. Performed the experiments: AC LN PG. Analyzed the data: AC LN FA TS KVU RS.
References

1. Maiese K, Chung ZZ, Li F, Shang YC (2000) Erythropoietin: elucidating new cellular targets that broaden therapeutic strategies. Prog Neurobiol 65: 194-213.

2. Sasaki H, Ochi N, Della A, Fukasawa M (1988) Site-specific glycosylation of human recombinant erythropoietin: analysis of glycopeptides or peptides at each glycosylation site by fast atom bombardment mass spectrometry. Biochemistry 27: 8610-8612.

3. Takahashi M, Takasaki S, Miyazaki H, Kato T, Hoshi S, et al. (1980) Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells. J Biol Chem 255: 3657-3663.

4. Dube S, Fisher JW, Possel JS (1980) Glycosylation at specific sites of human recombinant erythropoietin is critical for bioactivity, secretion, and biological function. J Biol Chem 256: 7703-7709.

5. Higuchi M, Obeda M, Kubonouchi H, Tomonou H, Shimonaka Y, et al. (1992) Role of sugar chains in the expression of the biological activity of human erythropoietin. J Biol Chem 267: 7703-7709.

6. Egrie JC, Browne JK (2001) Development and characterization of novel erythropoiesis stimulating protein (NESP), Nephol Dial Transplant 16 Suppl 3: 1-33.

7. Hocke CH, Bergwerff AA, Van Dedem GW, Kamerling JP, Vliegenthart JF (1995) Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units. J Biol Chem 270: 10801-10808.

8. Yuan CT, Storring PL, Tippley RJ, Izquierdo M, Wait R, et al. (2003) Relationships between the Nglycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures. Bi J. Biochem 228: 1011-1016.

9. Schriebl K, Trummer E, Lattenmayer C, Weik R, Kunert R, et al. (2006) Sialylation patterns and branch location of dimeric N-acetyllactosamine units. J Biol Chem 281: 1052-1058.

10. Wang MD, Yang M, Huzel N, Butler M (2002) Erythropoietin production from CHO cells grown by continuous culture in a fluidized-bed bioreactor. Biotechnol Bioeng 77: 194-203.

11. Elliott S, Lorenzini T, Asher S, Aoki K, Brankow D, et al. (2003) Enhancement of therapeutic protein expression in vivo activities through glycoengineering. Nat Biotechnol 21: 414-421.

12. Braconi A, Distefano G, Buckberry LD, Aebihs G, Foglioni C, et al. (2000) A new Chinese hamster ovary cell line expressing alpha2,6-sialyltransferase used as universal host for the production of human sialylated recombinant glycoproteins. Biochim Biophys Acta 1474: 273-282.

13. Wang Z, Park JH, Park HH, Tan W, Park TH (2011) Enhancement of recombinant human EPO production and sialylation in Chinese hamster ovary cells through Bombus mori 30Kc19 gene expression. Biotechnol Bioeng 108: 1634-1642.

14. Bork K, Reutter W, Weidemann W, Horstkorte R (2007) Enhanced sialylation through overexpression of the respective mammalian pathway. J Biol Chem 282: 9768-9779.

15. Bitonti AJ, Dumont JA, Low SC, Peters RT, Kropp KE, et al. (2004) Pulmonary delivery of an erythropoietin Fc fusion protein in non-human primates through a Bombyx mori 30Kc19 gene expression. Biotechnol Bioeng 77: 194-203.

16. Novak TJ, Hamelin WR, Brunkhorst B, Kong SM, Qi A, et al. (2005) Improvement of Fc-erythropoietin structure and pharmacokinetics by modification at a disulfide bond. Protein Eng Des Sel 18: 111-118.

17. Schlie Assembly M, Gebler JC, Wu J (2000) Papain digestion of different mouse IgG subclasses as studied by electrospray mass spectrometry. J Immunol Methods 238: 95-104.

18. Birnbacher D, Steinherz H, Abendfenn L, Remy JJ, van de Wiel D (2011) Secretion of biologically active glycoforms of bovine follicle stimulating hormone in plants. Eur J Biochem 268: 4576-4579.

19. Schiødt V, Nissen-Lie G, Torjesen P (2003) Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. Blood 98: 3626-3634.
49. Shahrokh Z, Royle L, Saldova R, Bones J, Abrahams JL, et al. (2011) Erythropoietin produced in a human cell line (Dynepo) has significant differences in glycosylation compared with erythropoietins produced in CHO cell lines. Mol Pharm 8: 286–296.

50. Sondermann P, Huber R, Oosthuizen V, Jacob U (2000) The 3.2-Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex. Nature 406: 267–273.