



OPEN

SUBJECT AREAS:

MOLECULAR  
ENGINEERING IN PLANTS

BIOLOGICS

PLANT MOLECULAR BIOLOGY

DRUG SAFETY

Received  
2 May 2013

Accepted  
4 October 2013

Published  
22 October 2013

Correspondence and  
requests for materials  
should be addressed to  
E.L.D. (Eva.Decker@  
biologie.uni-freiburg.  
de)

# A gene responsible for prolyl-hydroxylation of moss-produced recombinant human erythropoietin

Juliana Parsons<sup>1</sup>, Friedrich Altmann<sup>2</sup>, Manuela Graf<sup>1</sup>, Johannes Stadlmann<sup>2</sup>, Ralf Reski<sup>1,3,4</sup> & Eva L. Decker<sup>1</sup>

<sup>1</sup>Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestr. 1, 79104 Freiburg, Germany, <sup>2</sup>Department of Chemistry, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria, <sup>3</sup>BIOSS Centre for Biological Signalling Studies, 79104 Freiburg, Germany, <sup>4</sup>FRIAS Freiburg Institute for Advanced Studies, 79104 Freiburg, Germany.

**Recombinant production of pharmaceutical proteins is crucial, not only for personalized medicine. While most biopharmaceuticals are currently produced in mammalian cell culture, plant-made pharmaceuticals gain momentum. Post-translational modifications in plants are similar to those in humans, however, existing differences may affect quality, safety and efficacy of the products. A frequent modification in higher eukaryotes is prolyl-4-hydroxylase (P4H)-catalysed prolyl-hydroxylation. P4H sequence recognition sites on target proteins differ between humans and plants leading to non-human posttranslational modifications of recombinant human proteins produced in plants. The resulting hydroxyprolines display the anchor for plant-specific O-glycosylation, which bears immunogenic potential for patients. Here we describe the identification of a plant gene responsible for non-human prolyl-hydroxylation of human erythropoietin (hEPO) recombinantly produced in plant (moss) bioreactors. Targeted ablation of this gene abolished undesired prolyl-hydroxylation of hEPO and thus paves the way for plant-made pharmaceuticals humanized via glyco-engineering in moss bioreactors.**

Plant-based systems are gaining acceptance as alternative production platforms for recombinant biopharmaceuticals (reviewed in<sup>1</sup>) with the first product (Elyso by Protalix) released to the market. With regard to the differences existing in posttranslational modifications between humans and plants considerable progress was achieved in the humanization of Asparagin (N)-linked glycosylation of plant-made pharmaceuticals. The attachment of immunogenic plant-specific  $\beta$ 1,2-xylose and  $\alpha$ 1,3-fucose residues to the core N-glycan was abolished in different plant systems<sup>2-5</sup>. In addition, plant-produced recombinant human EPO (rhEPO) devoid of Lewis A epitopes on N-glycans was reported recently<sup>6</sup>. Lewis A is a trisaccharide structure which occurs only rarely on glycoproteins of healthy adult humans but is widespread on plants. Further humanization of the N-glycosylation on plant proteins was achieved by expression of the human  $\beta$ 1,4 galactosyltransferase<sup>7,8</sup> and additional heterologous enzymes necessary for engineering sialylation<sup>9,10</sup>. Despite this progress in engineering N-glycosylation, O-glycosylation, which means the attachment of glycans to the hydroxyl group of amino acids, can affect product quality. Plant O-glycosylation differs explicitly from the typical human mucin-type O-glycosylation (reviewed by<sup>11</sup>) and induces antibody formation in mammals<sup>12,13</sup>. Immunogenicity of biopharmaceuticals may result in reduced product efficacy and is a potential risk for the patients<sup>14,15</sup>. Such adverse effects hamper the broad use of plants as production hosts for biopharmaceuticals. In plants, the main anchor for O-glycosylation is 4-trans-hydroxyproline (Hyp) (reviewed in<sup>16,17</sup>) while no further modification of Hyp occurs in mammals<sup>18</sup>. Although Hyp is always synthesized post-translationally by prolyl-4-hydroxylases (P4Hs) via hydroxylation of the  $\gamma$  carbon of proline, recognition sequences on the target proteins differ between mammals and plants<sup>18</sup>. The action of both, mammalian and plant P4Hs leads to Hyp, while its diastereomer 4-cis-hydroxyproline has not been found in a natural protein yet<sup>19</sup>. Hyp is an important structural component of plant cell walls and of the extracellular matrix of animals. Here, Hyp plays a key role in stabilizing the structure of collagen, one of the most abundant proteins in mammals, in which the second proline of the tripeptide PPG is usually hydroxylated by collagen P4Hs. In plants, Hyp residues are the attachment sites for O-glycosylation of hydroxyproline-rich glycoproteins (HRGPs), the most abundant proteins in the plant extracellular matrix and cell wall. HRGPs include extensins, proline-rich glycoproteins and arabinogalactan proteins<sup>16,20,21</sup>. Prolyl-hydroxylation and subsequent glycosylation of plant cell wall proteins is of major importance for growth, differentiation, development and stress adaptation<sup>22,23</sup>.



The target motifs for Hyp-anchored O-glycosylation in plants, so-called glycomodules, were defined and validated<sup>20,21</sup>. From these, the consensus motif [A/S/T/V]-P<sub>(1,4)</sub>-X<sub>(0,10)</sub>-[A/S/T/V]-P<sub>(1,4)</sub> (where X can be any amino acid) was derived for predicting prolyl-hydroxylation in plants<sup>11</sup>. According to *in silico* analysis of the human proteome, approximately 30% of all proteins contain this motif, making them candidates for non-human prolyl-hydroxylation and subsequent O-glycosylation when expressed in plant systems<sup>11</sup>. Indeed, undesired plant-typical prolyl-hydroxylation<sup>24–26</sup> and in some cases subsequent arabinosylation of biopharmaceuticals was reported<sup>27–29</sup>. On the other hand, the artificial introduction of Hyp-O-glycosylation motifs was suggested as an alternative to PEGylation (the attachment of polyethylene glycol-oligomers to proteins or peptide drugs) to increase the serum half-life of biopharmaceuticals<sup>30,31</sup>. However, non-human prolyl-hydroxylation does not only alter the native sequence of the protein, but also serves as an anchor for O-glycans, which in turn may be immunogenic. Thus, the elimination of the anchor Hyp is the only safe way to avoid adverse O-glycosylation in PMPs.

Among plants, the moss *Physcomitrella patens* offers the unique possibility for precise targeted genetic engineering via homologous recombination (e.g.<sup>3,32</sup>). Further, several recombinant proteins have been produced in the moss bioreactor, including rhEPO<sup>33</sup>, one of the top-ten biopharmaceuticals world-wide<sup>34</sup>. EPO is a highly glycosylated peptide hormone stimulating erythropoiesis. Recombinant hEPO produced in CHO (Chinese hamster ovary) cells is used for prevention or treatment of anaemia in nephrology and oncology patients, and can be abused for illegal doping activities. A glyco-engineered version of EPO (asialo-EPO) has no hematopoietic activity but can serve as a safe drug with neuro- and tissue-protective functions after stroke and additional hypoxia stress<sup>35,36</sup>. Production of correctly N-glycosylated asialo-EPO in the moss bioreactor was reported recently<sup>6</sup>. However, plant-derived rhEPO from moss and *Nicotiana benthamiana* was shown to be hydroxylated within the motif SPP (amino acids 147–149)<sup>24,25</sup> which may have adverse effects on patients.

In this study, we aimed to identify and destroy the genes responsible for undesired non-human prolyl-hydroxylation of rhEPO produced in the moss bioreactor.

## Results

As the proline hydroxylation on moss-produced rhEPO occurred within a consensus motif recognized by plant P4Hs, we searched for homologues of these enzymes in the moss genome. Via BLAST (basic local alignment search tool) searches in the Cosmoss database ([www.cosmoss.org](http://www.cosmoss.org)) we were able to identify six sequences from the *P. patens* genome with homology to P4H enzymes: Pp1s8\_114V6.1 (P4H1), Pp1s192\_51V6.1 (P4H2), Pp1s19\_322V6.1 (P4H3), Pp1s172\_91V6.1 (P4H4), Pp1s12\_247V6.1 (P4H5) and Pp1s328\_29V6.1 (P4H6). As sequence information was not complete for P4H2, 3 and 6 mRNA, 5' RACE-PCR was employed to obtain full length sequences (Supplementary Fig. S1 online). Two different cDNAs were amplified for the P4H6 gene, corresponding to alternative splice forms of the mRNA, from which two protein variants with different N-termini could be predicted (P4H6a and P4H6b, with P4H6a containing an N-terminal extension not present in P4H6b). All deduced protein sequences had a prolyl-4-hydroxylase alpha subunit catalytic domain (SMART 0702). N-terminal transmembrane domains were predicted for all homologues except P4H2 (TMHMM server v.2.0, <http://www.cbs.dtu.dk/services/TMHMM/>). As none of the sequences possessed CXXXC motifs typical for prolyl-3-hydroxylases<sup>37</sup>, we exclude 3-hydroxylation activity of the predicted proteins.

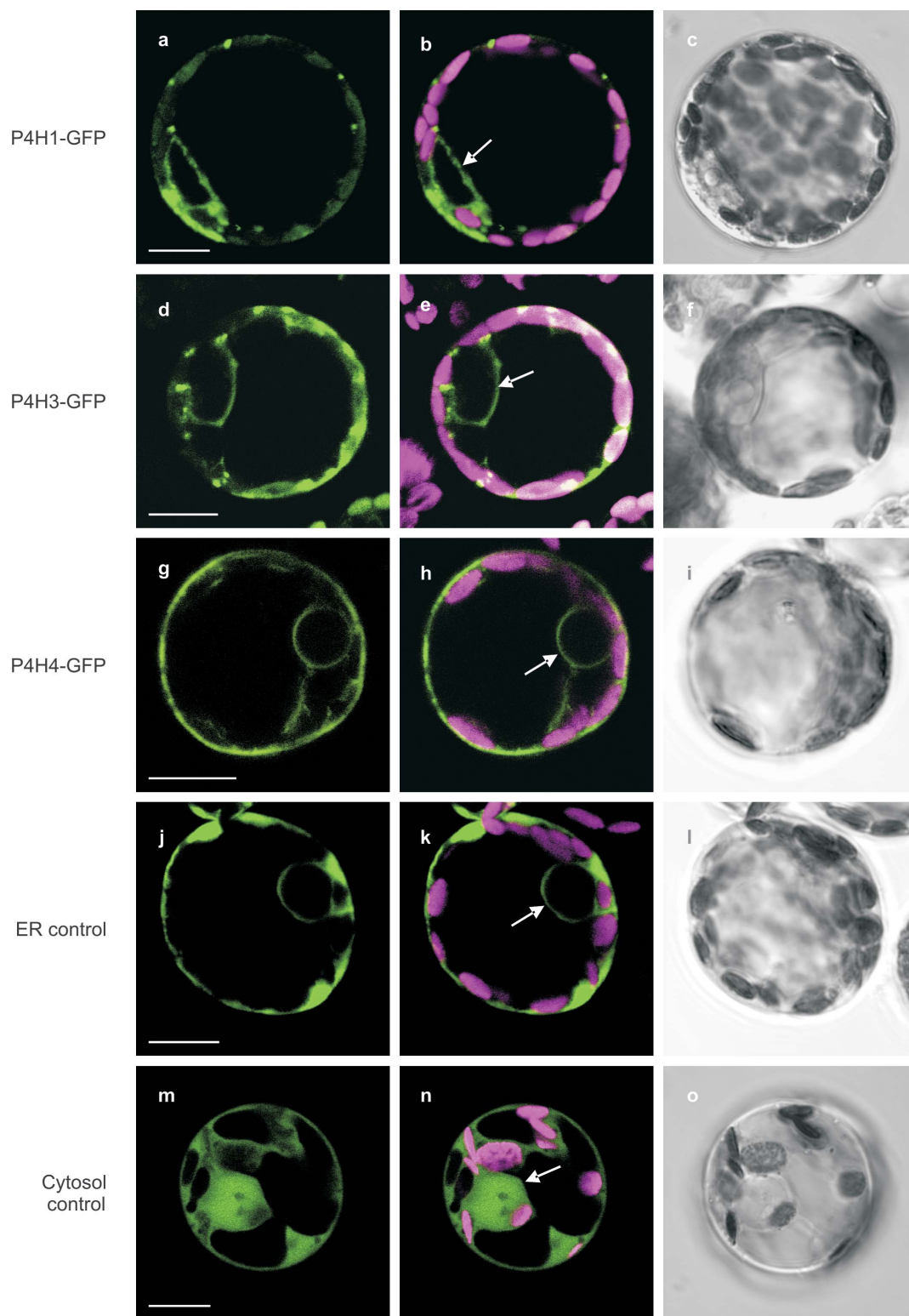
In order to gain more information about the predicted P4H enzymes, the deduced amino acid sequences were aligned with sequences of already characterized P4Hs from human, *Arabidopsis*

and *Nicotiana tabacum*. The catalytic domain in the C-terminal end of the protein is highly conserved in all seven *P. patens* homologues (Supplementary Fig. S2 online). The seven putative P4Hs share 16–24% identity with the human catalytic  $\alpha$  (I) subunit and 30–63% identity with ATP4H1. Among the moss sequences the degree of identity is between 30 and 81%. All sequences contain the motif HXD and a distal histidine, which are necessary to bind the cofactor Fe<sup>2+</sup>. Further, they contain the basic residue lysine which binds the C-5 carboxyl group of 2-oxoglutarate (Supplementary Fig. S2 online). These residues are indispensable for the activity of collagen P4Hs<sup>38</sup> and of P4H1 from *A. thaliana*<sup>39</sup>, indicating that all seven sequences from *P. patens* are functional prolyl-4-hydroxylases.

Non-human prolyl-hydroxylation occurred on moss-derived rhEPO which has been secreted to the medium of the moss bioreactor culture. Therefore, we concluded that the P4H enzyme responsible for post-translational rhEPO modification is located in the secretory compartments, i.e. the endoplasmic reticulum (ER) or the Golgi apparatus. All sequences display a hydrophobic segment near the N-terminus following a positively charged residue, suggesting a localization of the proteins in the secretory compartments<sup>40</sup>. However, we examined the subcellular localization of the seven moss P4Hs *in silico* with four different programs based on different algorithms. As no consistent prediction was obtained by this approach (Supplementary Table S1 online) we subsequently studied the *in vivo* intracellular localization of each of the seven moss P4Hs by expressing them as GFP (green fluorescent protein) fusions (P4H-GFP) in moss cells. Transfected cells were analysed via Confocal Laser Scanning Microscopy. In optical sections recorded 3–14 days after transfection, GFP signals from all seven different P4H fusion proteins were predominantly detected as defined circular structures around the nucleus, indicating labelling of the nuclear membrane (Fig. 1). As the nuclear membrane is part of the endomembrane continuum of eukaryotic cells, these signals reveal that all seven moss P4Hs were targeted to the secretory compartments. An ER-targeted GFP version (ASP-GFP-KDEL<sup>41</sup>) as well as GFP without any signal peptide displaying GFP fluorescence in the cytoplasm as well as the nucleus<sup>41</sup> served as controls. Thus, these experiments provided no clear indication of a specific P4H responsible for generation of Hyp on secreted rhEPO but left all of them as candidates.

In order to definitely identify those homologues responsible for plant-typical prolyl-hydroxylation of moss-produced rhEPO we aimed to ablate the gene functions of each of the moss P4Hs. Accordingly, gene targeting constructs (Supplementary Fig. S3 online) were designed for the six P4H genes and transferred to the rhEPO-producing moss line 174.16<sup>24</sup> to generate specific deletion (knockout, KO) lines by allele replacement via homologous recombination for each of the P4H genes. After antibiotic selection, surviving plants were screened for homologous integration of the KO construct into the correct genomic locus. Loss of the respective transcript was proven by RT-PCR (Supplementary Fig. S4 online), confirming successful gene ablation. Even if truncated N-terminal P4H fragments might exist, we can exclude any residual enzymatic activity as critical catalytic residues are located in the C-terminus of the P4H family (Supplementary Fig. S2 online). One line for each genetic modification was chosen for further analysis, and stored in the International Moss Stock Center (<http://www.moss-stock-center.org>; Supplementary Table S3 online).

To investigate the effect of each of the P4H ablations on the prolyl-hydroxylation observed for moss-produced rhEPO, the recombinant protein from each of the KO lines ( $\Delta$ P4H) was analysed via mass spectrometry. For this purpose, total soluble proteins were precipitated from the culture supernatant of the parental plant and one knockout line from each P4H homologue, and separated by SDS-PAGE. Subsequently, the main rhEPO-containing band was cut from the Coomassie-stained gel, digested with trypsin and subjected to mass spectrometry for an analysis of the tryptic peptide



**Figure 1 | Subcellular localization of *P. patens* P4H homologues.** Fluorescence of P4H-GFP fusion proteins in *P. patens* protoplasts was observed by confocal microscopy 3 to 14 days after transfection. The images obtained for P4H1-GFP, P4H3-GFP and P4H4-GFP are taken as example of the fluorescence pattern which was observed for all homologues. (a–c) P4H1-GFP, (d–f) P4H3-GFP, (g–i) P4H4-GFP, (j–l) ASP-GFP-KDEL as control for ER localization, (m–o) GFP without any signal peptide as control for cytosolic localization. (a, d, g, j and m) single optical sections emitting GFP fluorescence (494–558 nm), (b, e, h, k and n) merge of chlorophyll autofluorescence (601–719 nm) and GFP fluorescence, (c, f, i, l and o) transmitted light images. The arrows indicate the cell nucleus membrane.

EAISPPDAASAAPLR (144–158). In the parental plant 174.16, almost half of the rhEPO was hydroxylated (Fig. 2), mainly in the second proline from the SPP motif, as shown by MS/MS (Supplementary Fig. S5 online). The stereochemistry of the observed

hydroxyproline - assumably 4-trans-hydroxyproline - was not experimentally verified.

Surprisingly, while rhEPO produced in moss lines with ablated *P4H2*, *P4H3*, *P4H4*, *P4H5* or *P4H6*, respectively, was hydroxylated in

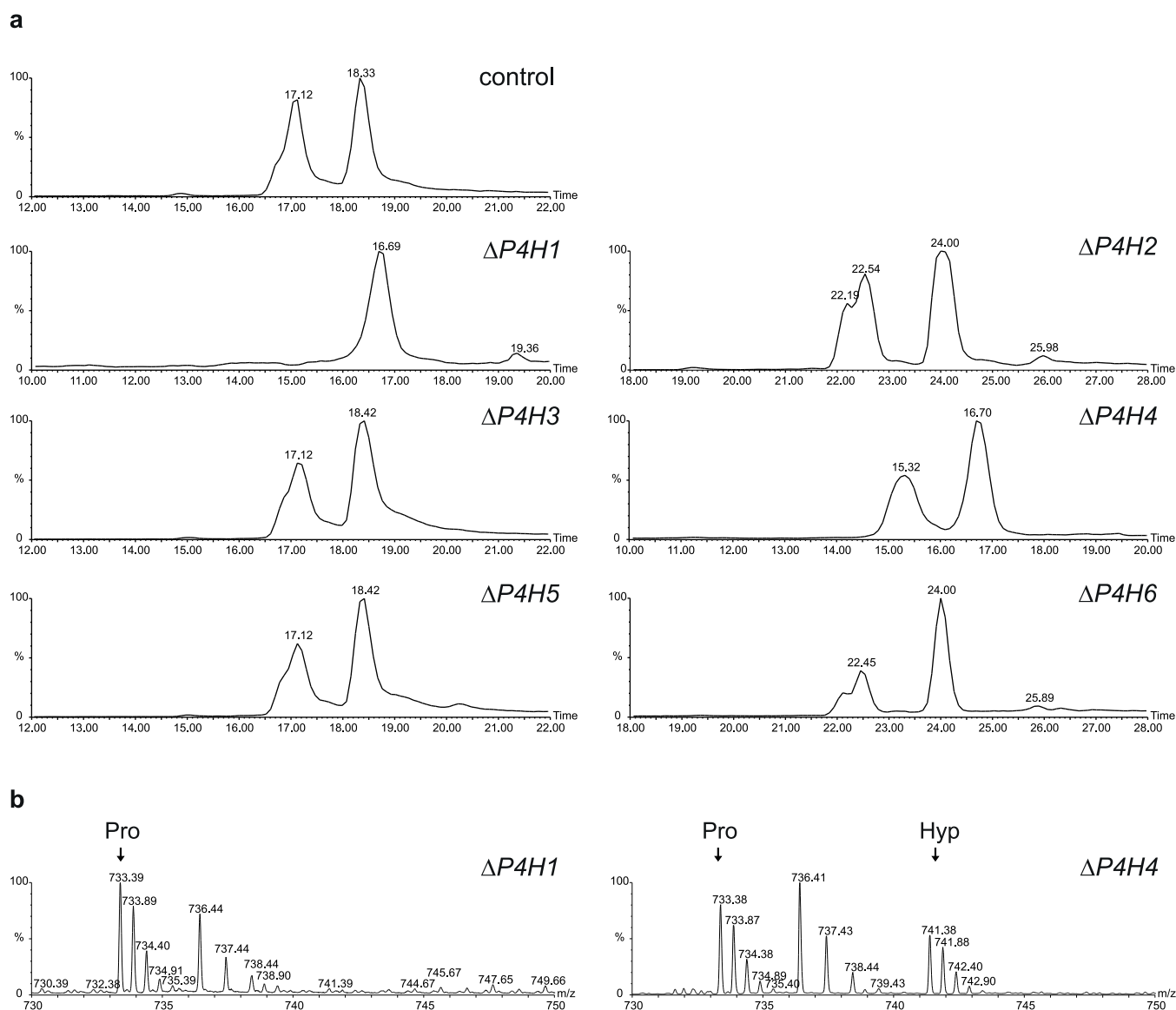


similar levels to those found on the parental plant, the ablation of exclusively the *P4H1* gene was sufficient to completely abolish the prolyl-hydroxylation of the SPP motif (Fig. 2). Growth rate, rhEPO productivity and secretion of the protein to the culture medium were not impaired in these knockout plants compared to the parental line (data not shown).

We showed the complete lack of Hyp on rhEPO produced by the  $\Delta P4H1$  lines. To verify P4H1 enzymatic activity in prolyl-hydroxylation we ectopically expressed this gene in the  $\Delta P4H1$  knockout line #192. Strong overexpression of the *P4H1* transcript was confirmed in the resulting lines via semi-quantitative RT-PCR (Supplementary Fig. S4 online). Five *P4H1* overexpression lines (*P4H1*OE) were analysed for rhEPO-Pro-hydroxylation. LC-ESI-MS measurements revealed that *P4H1* overexpression restored prolyl-hydroxylation of the moss-produced rhEPO (Fig. 3). The proportion of hydroxylated rhEPO, as well as the hydroxylation pattern, was altered by the

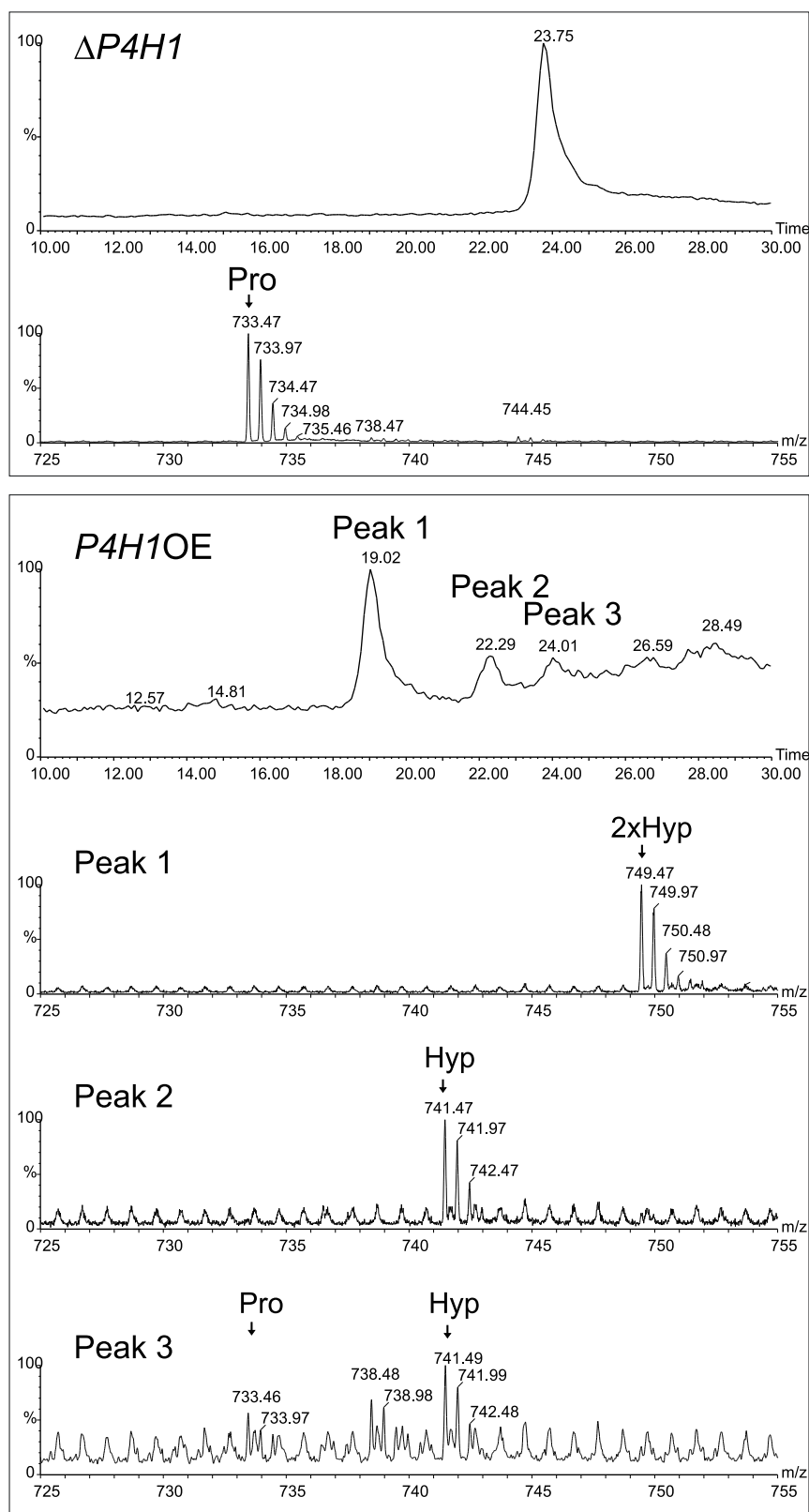
elevated expression levels of the gene. In the parental plant 174.16, with native P4H1 activity, approximately half of rhEPO displayed Hyp (Fig. 2), whereas nearly all rhEPO was oxidized in the *P4H1* overexpressors (Fig. 3). Furthermore, in the overexpressors not only Pro149 in the peptide EAISPPDAASAAPLR (144–158) was hydroxylated as also seen in the parental plant, but a second proline of this peptide was converted to Hyp (Fig. 3).

As hydroxylation and arabinosylation of the human epithelial mucin MUC1 at the sequence APP was reported upon expression in *N. benthamiana*<sup>28</sup>, we analysed the rhEPO N-terminal peptide APPRLCDSRVL for prolyl-hydroxylation in moss. After chymotryptic digestion of rhEPO derived from the parental plant 174.16, the knockout plant *P4H1* #192 and the overexpressor *P4H1*OE-45, LC-ESI-MS analysis revealed that this peptide was not hydroxylated in any of these cases (Supplementary Fig. S6 online). The other Pro containing peptides, the two EPO glycopeptides, have been analysed



**Figure 2 | Mass spectrometric analysis of the hydroxylation of moss-produced rhEPO.** (a) Reversed-phase liquid chromatogram of tryptic peptides showing peaks of oxidized and non-oxidized peptide EAISPPDAASAAPLR (144–158) derived from rhEPO produced in moss lines 174.16 (control parental plant),  $\Delta P4H1$  #192,  $\Delta P4H2$  #6,  $\Delta P4H3$  #21,  $\Delta P4H4$  #95,  $\Delta P4H5$  #29 and  $\Delta P4H6$  #8. Selected ion chromatograms for the doubly charged ions of non-oxidized ( $m/z = 733.4$ ) and oxidized peptide ( $m/z = 741.4$ ) are shown. (b) Broad band sum spectra for peptide 144–158 showing the absence of prolyl-hydroxylation (Pro) in the line  $\Delta P4H1$  #192 and the presence of hydroxylated peptide (Hyp) in the line  $\Delta P4H4$  #95, as an example. The singly charged peak between “Pro” and “Hyp” is caused by the incidentally co-eluting peptide YLLEAK. Retention time deviations are technical artefacts.





**Figure 3** | Effect of overexpression of the prolyl-hydroxylase gene *P4H1*. Comparison of reversed-phase chromatograms showing the retention time for the moss-produced rhEPO peptide EAISPPDAASAAPLR (144–158) and its hydroxylated versions in the knockout moss line  $\Delta P4H1$  #192 (upper panel) and in the overexpressing line  $P4H1OE$  #32 (lower panel). The spectra of each peak are shown below the chromatograms. In the overexpressing line, the doubly hydroxylated peptide and two singly hydroxylated isomers – one coeluting with the parent peptide - were found.



by mass spectrometry before in the parental plant 174.16 and no Hyp residues were detected (e.g.<sup>6,24</sup>). Thus we can exclude any additional prolyl hydroxylation on moss-produced rhEPO.

## Discussion

One of the most common posttranslational modifications in higher eukaryotes is prolyl-4-hydroxylase (P4H)-catalysed formation of hydroxyproline (Hyp) residues, though sequence recognition sites on target proteins differ between animals and plants. Moreover, Hyp in plants is the main anchor for O-glycosylation, which again diverges from mammalian O-glycosylation. The engineering of the human O-glycosylation machinery in plants was tackled recently<sup>25,26,29,42</sup> leading to plant proteins with so-called mucin-type O-glycosylation. Nevertheless, the absence of plant-specific prolyl-hydroxylation and subsequent O-glycosylation should be guaranteed for the production of safe biopharmaceuticals<sup>26</sup>. Several human proteins expressed in plants were shown to be hydroxylated<sup>24–29</sup>. Recombinant hEPO produced both in moss and in *N. benthamiana* was shown to be hydroxylated in the consensus sequence SPP<sup>24,25</sup> but O-glycosylation was not observed. The Hyp proportion in the moss line 174.16 was higher than that found on rhEPO from *N. benthamiana*<sup>25</sup> which might be due to different production systems, transient expression in *N. benthamiana* vs. stable production in moss. It was reported before that less Hyp formation occurs in transient systems than in stable production<sup>29</sup>. This post-translational modification could be partially reduced by P4H inhibitors<sup>29,43</sup>, however, the complete elimination of the anchor Hyp on the biopharmaceutical by abolishment of P4H activity is the only reliable way to avoid this adverse modification. Due to the importance of P4Hs in plant growth and development<sup>22,23</sup>, or to the high number of isoforms of this enzyme in plants<sup>44</sup> the engineering of plant genomes for mutation or deletion of *P4H* genes was suggested<sup>26,42</sup> but to our knowledge not yet conducted.

In this work, we identified 6 *P4H* homologue genes in *P. patens*, from which 7 protein sequences were deduced containing the essential motifs for functionality. All of them were shown to be localized in the secretory compartments. By means of precise gene targeting via homologous recombination, knockout lines for each of these genes were generated in order to identify the *P4H* homologues involved in the prolyl-hydroxylation on secreted rhEPO in *P. patens*. As proven by MS analysis, the ablation of exclusively *P4H1* leads to moss-produced rhEPO free from non-human Hyp. Thus, we demonstrated that the expression of *P4H1* is essential and sufficient for the prolyl-hydroxylation of the moss-produced rhEPO.

By overexpressing the enzyme, we could also demonstrate that a higher expression level of *P4H1* influences its enzyme activity, not only in the proportion of hydroxylated protein molecules but also in the pattern of hydroxylation. As opposed to moss lines with native *P4H1* activity, which hydroxylate only one proline (mainly Pro149) in the peptide EAISPPDAASAAPLR (144–158) of rhEPO, moss lines overexpressing *P4H1* produce rhEPO with a second proline hydroxylated in this motif. We demonstrated that of all the moss P4H proteins only *P4H1* was active on rhEPO. Of the biopharmaceuticals expressed in *P. patens* so far, rhEPO was the only one on which Hyp formation was detected. However, considering that different P4H homologues may possess distinct substrate specificities<sup>45</sup>, it is possible that recombinant proteins bearing a different hydroxylation sequence as the one presented here, could be substrate for other P4H homologues. In that case, to completely exclude non-human Hyp formation on moss-produced biopharmaceuticals, multiple knockouts had to be performed in parallel. Due to high sequence identities among the P4H homologues, also knockdown strategies, which are feasible in both moss and higher plants<sup>46,47</sup>, might be conceivable to remove unwanted Hyp formation in both systems.

The consecutive prolines in the N-terminal sequence APPRLIC-DSRVL of rhEPO could be assumed to be a putative hydroxylation

site. Therefore, we analysed the rhEPO N-terminal peptide, both in moss plants with endogenous activity of *P4H1*, and also in plants overexpressing this enzyme. No prolyl-hydroxylation of this peptide could be detected, indicating that the mere presence of contiguous proline residues preceded by an alanine in the protein of interest is not sufficient to be recognized by moss prolyl-hydroxylases. Thus we confirmed that in moss-produced rhEPO only the SPP motif was hydroxylated.

Analysing the effects of plant-typical O-glycosylated biopharmaceuticals in the human body would require cost-intensive clinical trials. Furthermore, even slight differences between PMPs and their native counterparts will hamper the approval of a drug by the relevant authorities. Thus, in our opinion the straight-forward approach is to precisely eliminate the attachment sites for plant-specific O-glycosylation, the hydroxylated proline residues, on the recombinant protein. As demonstrated here for the production of rhEPO, this can be achieved in the moss production system by the ablation, and most likely also by a down-regulation, of a single *P4H* gene, thus paving the way to a further humanization of plant-made biopharmaceuticals in the moss bioreactor.

## Methods

**Identification of prolyl-4-hydroxylases in *P. patens*.** For the identification of prolyl-4-hydroxylase homologues in *P. patens*, the amino acid sequence of the *Arabidopsis thaliana* P4H1 (AT2G43080.1) was used to perform a BLAST (basic local alignment search tool) search against the gene models in the *Physcomitrella patens* resource (www.cosmos.org). The 5' complete sequences of the *P4H2*, 3 and 6 cDNAs were obtained via RACE (rapid amplification of cDNA-ends)-PCR (GeneRacer™, Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol.

Protein sequence alignments were performed with the program CLUSTAL W<sup>48</sup> (www.ebi.ac.uk/Tools/msa/clustalw2/) and visualized with Jalview (www.jalview.org/).

**In silico prediction of intracellular localization.** The *in silico* predictions for intracellular localization of *P. patens* P4H homologues were performed with four different programs: TargetP (http://www.cbs.dtu.dk/services/TargetP/), MultiLoc (https://abi.inf.uni-tuebingen.de/Services/MultiLoc), SherLoc (https://abi.inf.uni-tuebingen.de/Services/SherLoc2) and Wolf PSORT (http://wolffpsort.org/).

**Plant material and transformation procedure.** *Physcomitrella patens* (Hedw.) Bruch & Schimp was cultivated as described previously<sup>49</sup>. Moss-produced rhEPO was shown to be hydroxylated at the prolyl-hydroxylation consensus motif SPP (amino acids 147–149), therefore the rhEPO-producing *P. patens* line 174.16<sup>24</sup> was used as the parental line for the *P4H* knockout generation and the line *ΔP4H1* #192 was used for the generation of *P4H1* overexpression lines. In these moss lines the  $\alpha$ 1,3 fucosyltransferase and the  $\beta$ 1,2 xylosyltransferase genes are disrupted<sup>3</sup>. Wild-type moss was used for the subcellular localization experiments with P4H-GFP.

Protoplast isolation and PEG-mediated transfection was performed as described previously<sup>49,50</sup>. Mutant selection was performed with Zeocin™ (Invitrogen) or sulfadiazine (Sigma) as described before<sup>6</sup>.

For rhEPO production, *P. patens* was cultivated as described before<sup>6</sup>.

**Generation of plasmid constructs.** The cDNAs corresponding to the seven P4H homologues identified in *Physcomitrella patens* were amplified using the primers listed in Supplementary Table S2 online, and cloned into pJET 1.2 (CloneJET™ PCR CloningKit, Fermentas, St Leon-Rot, Germany). Subsequently, the *P4H* coding sequences including a portion of the 5' UTR were cloned into the plasmid mAV4mcs<sup>41</sup> using the XhoI and BglII sites giving rise to N-terminal fusion P4H-GFP proteins under the control of the cauliflower mosaic virus (CaMV) 35S promoter. Unmodified mAV4mcs was used as a control for cytoplasmic and nuclear localization. As positive control for ER localization, pASP-GFP-KDEL was taken<sup>41</sup>.

To generate the *P4H* knockout constructs, *P. patens* genomic DNA fragments corresponding to the prolyl-4-hydroxylases were amplified using the primers listed in Supplementary Table S2 online and cloned either into pCR®4-TOPO® (Invitrogen, Karlsruhe, Germany) or into pETBlue-1 AccepTor™ (Novagen, Merck KGaA, Darmstadt, Germany). The pTOPO\_ *P4H1* genomic fragment was first linearized using BstBI and SacI, thus deleting a 273 bp fragment, and recircularized by ligating double-stranded oligonucleotide containing restriction sites for BamHI and HindIII. These sites were used for the insertion of a zeomycin resistance cassette (zeo-cassette). The zeo-cassette was obtained from pUC-zeo<sup>6</sup> by digestion with HindIII and BamHI. For the *P4H5* KO construct, a 1487 bp fragment was cut out from the pTOPO\_ *P4H5* using SalI and BglII sites and replaced by double-stranded oligonucleotide containing restriction sites for BamHI and HindIII. These restriction sites were used for the insertion of the zeo-cassette obtained from the pUC-Zeo plasmid. The *P4H2* KO construct was cloned into the pETBlue-1 AccepTor™, and the zeo-cassette replaced a 270 bp genomic fragment deleted by digestion with KpnI and HindIII. The zeo-cassette obtained from pRT101-zeo<sup>6</sup> by HindIII digestion was inserted into the



pET\_4PH3 and the pTOPO\_4PH4 KO constructs digested with the same enzyme, replacing a 990 bp and a 1183 bp genomic fragment, respectively. For the 4PH6 KO construct, the zeo-cassette was obtained from the pUC-zeo via digestion with HindIII and SacI and inserted into pTOPO\_4PH6, replacing a 1326 bp genomic fragment. In all KO constructs the regions homologous to the target gene had approximately the same size at both ends of the selection cassette, comprising between 500 and 1,000 bp.

For the overexpression construct, the 4PH1 coding sequence and 79 bp of the 5' UTR were amplified from moss WT cDNA with the primers listed in Supplementary Table S2 online, and cloned under the control of the 35S promoter and the nos terminator into the mAV4mcs vector<sup>41</sup>. For this purpose the GFP gene was deleted from the vector by digestion with Ecl136II and SmaI and subsequent religation of the vector. The 4PH1 cDNA was inserted into the vector via XhoI and BglII restriction sites. The 4PH1 overexpression construct was linearized via digestion with EcoRI and PstI and transferred into the line  $\Delta$ 4PH1 #192 together with pUC 18 sul<sup>6</sup> for sulfadiazine selection.

**Screening of transformed plants.** Screening of stable transformed plants was performed via direct PCR<sup>51</sup> with genomic DNA extracted as described before<sup>6</sup>. From these extracts, 2  $\mu$ l each were used as template for PCR, using the primers listed in Supplementary Table S2 online to check the 5' and 3' integration of the knockout construct in the correct genomic locus and to check the integration of the overexpression construct into the moss genome, respectively. Plants which showed the expected PCR products were considered as putative knockouts or overexpression lines, respectively, and subsequently analysed. The absence of the 4PH1 transcripts in the KO lines was analysed via RT-PCR as described before<sup>6</sup> using the primers listed in Supplementary Table S2 online. Expression of 4PH1 in the overexpression lines was analysed via semi-quantitative RT-PCR. For this purpose, cDNA equivalent to 150 ng RNA was amplified with 24, 26 and 28 cycles using the 4PH1 primers listed in Supplementary Table S2 online. The primers for the constitutively expressed TATA box-binding protein, TBP fwd and TBP rev (Supplementary Table S2 online) were used as controls.

**Protein analysis.** Total soluble proteins were recovered from 160 ml of a 16-days-old culture supernatant by precipitation with 10% (w/v) trichloroacetic acid (TCA, Sigma-Aldrich, Deisenhofen, Germany) as described<sup>52</sup>. The pellet was resuspended in sample Laemmli loading buffer (Biorad, Munich, Germany) and electrophoretic separation of proteins was carried out in 12% SDS-polyacrylamide gels (Ready Gel Tris-HCl, BioRad) at 150 V for 1 h under non-reducing conditions.

For peptide analysis, the proteins in the gels were stained with PageBlue<sup>®</sup> Protein Staining Solution (Fermentas) and the bands corresponding to 25 kDa were cut out, S-alkylated and digested with trypsin or chymotrypsin<sup>53</sup>. Analysis by reversed-phase liquid chromatography coupled to electrospray ionization mass spectrometry on a Q-TOF instrument (LC-ESI-MS and MS/MS) was performed as described previously<sup>53</sup>.

Quantification of the moss-produced rhEPO was performed using a hEPO Quantikine IVD ELISA kit (cat. no DEP00, R&D Systems) according to the manufacturer's protocol.

**Subcellular localization of P4H homologues.** Subcellular localization of the seven different P4H-GFP fusion proteins was analysed 3 to 14 days after transfection by Confocal Laser Scanning Microscopy (510 META; Carl Zeiss MicroImaging, Jena, Germany) and the corresponding software (version 3.5). Excitation at 488-nm was achieved with an argon laser and emission was measured with a META detector at 494–558 nm for GFP and at 601–719 nm for the chlorophyll. Cells were examined with a C-Apochromat 63x/1.2 W corr water immersion objective (Carl Zeiss MicroImaging). Confocal planes were exported from the ZEN2010 software (Carl Zeiss MicroImaging).

- Paul, M. & Ma, J. K. Plant-made pharmaceuticals: leading products and production platforms. *Biotechnol Appl Biochem* **58**, 58–67 (2011).
- Cox, K. M. *et al.* Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* **24**, 1591–1597 (2006).
- Koprivova, A. *et al.* Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic N-glycans. *Plant Biotechnol J*, **2**, 517–523 (2004).
- Strasser, R., Altmann, F., Mach, L., Glössl, J. & Steinkellner, H. 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 (2004).
- Sourrouille, C. *et al.* Down-regulated expression of plant-specific glycoepitopes in alfalfa. *Plant Biotechnol J* **6**, 702–721 (2008).
- Parsons, J. *et al.* Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol J* **10**, 851–861 (2012).
- Bakker, H. *et al.* An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes. *Proc Natl Acad Sci U S A* **103**, 7577–7582 (2006).
- Huether, C. M. *et al.* Glyco-engineering of moss lacking plant-specific sugar residues. *Plant Biol* **7**, 292–299 (2005).
- Castilho, A. *et al.* In planta protein sialylation through overexpression of the respective mammalian pathway. *J Biol Chem* **285**, 15923–15930 (2010).
- Jez, J. *et al.* Expression of functionally active sialylated human erythropoietin in plants. *Biotechnol J* **8**, 371–82 (2013).

- Gomord, V. *et al.* Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J* **8**, 564–87 (2010).
- Leonard, R. *et al.* Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. *J Biol Chem* **280**, 7932–40 (2005).
- Yates, E. A. *et al.* Characterization of carbohydrate structural features recognized by anti-arabinogalactan-protein monoclonal antibodies. *Glycobiology* **6**, 131–9 (1996).
- Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **24**, 1241–52 (2006).
- Schellekens, H. Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat Rev Drug Discov* **1**, 457–62 (2002).
- Kieliszewski, M. J. The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**, 319–23 (2001).
- Taylor, C. M., Karunaratne, C. V. & Xie, N. Glycosides of hydroxyproline: some recent, unusual discoveries. *Glycobiology* **22**, 757–67 (2012).
- Gorres, K. L. & Raines, R. T. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* **45**, 106–24 (2010).
- Shoulders, M. D., Kotch, F. W., Choudhary, A., Guzei, I. A. & Raines, R. T. The aberrance of the 4S diastereomer of 4-hydroxyproline. *J Am Chem Soc* **132**, 10857–65 (2010).
- Kieliszewski, M. J. & Lampert, D. T. Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* **5**, 157–72 (1994).
- Shpak, E., Barbar, E., Leykam, J. F. & Kieliszewski, M. J. Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J Biol Chem* **276**, 11272–8 (2001).
- Velasquez, S. M. *et al.* O-glycosylated cell wall proteins are essential in root hair growth. *Science* **332**, 1401–3 (2011).
- Lampert, D. T., Kieliszewski, M. J. & Showalter, A. M. Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. *New Phytol* **169**, 479–92 (2006).
- Weise, A. *et al.* High-level expression of secreted complex glycosylated recombinant human erythropoietin in the *Physcomitrella* Delta-fuc-t Delta-xyl-t mutant. *Plant Biotechnol J* **5**, 389–401 (2007).
- Castilho, A. *et al.* Engineering of sialylated mucin-type O-glycosylation in plants. *J Biol Chem* **287**, 36518–26 (2012).
- Yang, Z. *et al.* Engineering mammalian mucin-type O-glycosylation in plants. *J Biol Chem* **287**, 11911–23 (2012).
- Karnoup, A. S., Turkelson, V. & Anderson, W. H. O-linked glycosylation in maize-expressed human IgA1. *Glycobiology* **15**, 965–81 (2005).
- Pinkhasov, J. *et al.* Recombinant plant-expressed tumour-associated MUC1 peptide is immunogenic and capable of breaking tolerance in MUC1.Tg mice. *Plant Biotechnol J* **9**, 991–1001 (2011).
- Yang, Z. *et al.* Toward stable genetic engineering of human O-glycosylation in plants. *Plant Physiol* **160**, 450–63 (2012).
- Xu, J., Tan, L., Goodrum, K. J. & Kieliszewski, M. J. High-yields and extended serum half-life of human interferon alpha2b expressed in tobacco cells as arabinogalactan-protein fusions. *Biotechnol Bioeng* **97**, 997–1008 (2007).
- Kieliszewski, M. J., Xu, J. & Meyer, G. Methods of producing peptides/proteins in plants and peptides/proteins produced thereby, United States Patent 20060026719.
- Strepp, R., Scholz, S., Kruse, S., Speth, V. & Reski, R. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc Natl Acad Sci U S A* **95**, 4368–73 (1998).
- Decker, E. L. & Reski, R. Glycoprotein production in moss bioreactors. *Plant Cell Rep* **31**, 453–60 (2012).
- Walsh, G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **28**, 917–24 (2010).
- Erbayraktar, S. *et al.* Asialoerythropoietin is a nonerythropoietic cytokine with broad neuroprotective activity in vivo. *Proc Natl Acad Sci U S A* **100**, 6741–6 (2003).
- Ishii, T. *et al.* A single injection of liposomal asialo-erythropoietin improves motor function deficit caused by cerebral ischemia/reperfusion. *Int J Pharm* **439**, 269–74 (2012).
- Vranka, J. A., Sakai, L. Y. & Bachinger, H. P. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem* **279**, 23615–21 (2004).
- Kivirikko, K. I. & Myllyharju, J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol* **16**, 357–68 (1998).
- Hieta, R. & Myllyharju, J. Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J Biol Chem* **277**, 23965–71 (2002).
- Gafvelin, G., Sakaguchi, M., Andersson, H. & von Heijne, G. Topological rules for membrane protein assembly in eukaryotic cells. *J Biol Chem* **272**, 6119–27 (1997).
- Schaaf, A., Reski, R. & Decker, E. L. A novel aspartic proteinase is targeted to the secretory pathway and to the vacuole in the moss *Physcomitrella patens*. *Eur J Cell Biol* **83**, 145–52 (2004).
- Daskalova, S. M. *et al.* Engineering of *N. benthamiana* L. plants for production of N-acetylgalactosamine-glycosylated proteins—towards development of a plant-based platform for production of protein therapeutics with mucin type O-glycosylation. *BMC Biotechnol* **10**, 62 (2010).



43. Moriguchi, R., Matsuoka, C., Suyama, A. & Matsuoka, K. Reduction of plant-specific arabinogalactan-type O-glycosylation by treating tobacco plants with ferrous chelator 2,2'-dipyridyl. *Biosci Biotechnol Biochem* **75**, 994–6 (2011).
44. Liang, Y. *et al.* Identification and characterization of in vitro galactosyltransferase activities involved in arabinogalactan-protein glycosylation in tobacco and *Arabidopsis*. *Plant Physiol* **154**, 632–42 (2010).
45. Tiainen, P., Myllyharju, J. & Koivunen, P. Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J Biol Chem* **280**, 1142–8 (2005).
46. Khraiwesh, B., Ossowski, S., Weigel, D., Reski, R. & Frank, W. Specific gene silencing by artificial MicroRNAs in *Physcomitrella patens*: an alternative to targeted gene knockouts. *Plant Physiol* **148**, 684–93 (2008).
47. Ossowski, S., Schwab, R. & Weigel, D. Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J* **53**, 674–90 (2008).
48. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8 (2007).
49. Frank, W., Decker, E. L. & Reski, R. Molecular tools to study *Physcomitrella patens*. *Plant Biol* **7**, 220–227 (2005).
50. Rother, S., Haderl, B., Orsini, J. M., Abel, W. O. & Reski, R. Fate of a mutant macrochloroplast in somatic hybrids. *J Plant Physiol* **143**, 72–77 (1994).
51. Schween, G., Fleig, S. & Reski, R. High-Throughput-PCR screen of 15,000 transgenic *Physcomitrella* plants. *Plant Molecular Biology Reporter* **20**, 43–47 (2002).
52. Büttner-Mainik, A. *et al.* Production of biologically active recombinant human factor H in *Physcomitrella*. *Plant Biotechnol J* **9**, 373–83 (2011).
53. Grass, J., Pabst, M., Chang, M., Wozny, M. & Altmann, F. Analysis of recombinant human follicle-stimulating hormone (FSH) by mass spectrometric approaches. *Anal Bioanal Chem* **400**, 2427–38 (2011).

## Acknowledgements

This work was supported by a grant from the German Federal Ministry of Education and Research (BMBF, No. 0313852C), by contract research 'Glykobiologie/Glykomik' of the Baden-Württemberg Stiftung and the Excellence Initiative of the German Federal and State Governments (EXC294 to R.R.). We are grateful to greenovation Biotech GmbH for the moss strain 174.16. We thank Dagmar Krischke and Ingrid Heger for excellent technical assistance and Anne Katrin Prowse for proof-reading of the manuscript. The article processing charge was funded by the open access publication fund of the Albert Ludwigs University Freiburg.

## Author contributions

J.P., F.A., R.R. and E.L.D. designed the research, analysed results and wrote the manuscript. M.G. did the experimental work for the intracellular localization. J.P. produced the knockout and overexpression lines. J.S. and F.A. performed the MS analysis and interpretation of the spectra.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Parsons, J. *et al.* A gene responsible for prolyl-hydroxylation of moss-produced recombinant human erythropoietin. *Sci. Rep.* **3**, 3019; DOI:10.1038/srep03019 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>