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

Juliana Parsons¹, Friedrich Altmann², Manuela Graf¹, Johannes Stadlmann², Ralf Reski¹,3,4 & Eva L. Decker¹

¹Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestr. 1, 79104 Freiburg, Germany, ²Department of Chemistry, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria, ³BIOS Centre for Biological Signalling Studies, 79104 Freiburg, Germany, ⁴FRIAS 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 in¹) with the first product (Elelyso 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 β1,2-xylose and α1,3-fucose residues to the core N-glycan was abolished in different plant systems²-⁵. In addition, plant-produced recombinant human EPO (rhEPO) devoid of Lewis A epitopes on N-glycans was reported recently⁶. 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 β1,4 galactosyltransferase⁷,⁸ and additional heterologous enzymes necessary for engineering sialylation⁹,¹⁰. 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 by¹¹) and induces antibody formation in mammals¹²,¹³. Immunogenicity of biopharmaceuticals may result in reduced product efficacy and is a potential risk for the patients¹⁴,¹⁵. 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 in¹⁶,¹⁷) while no further modification of Hyp occurs in mammals¹⁸. Although Hyp is always synthesized post-translationally by prolyl-4-hydroxylases (P4Hs) via hydroxylation of the γ carbon of proline, recognition sequences on the target proteins differ between mammals and plants¹⁹. 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 yet¹⁹. 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 proteins²⁰,²¹. Prolyl-hydroxylation and subsequent glycosylation of plant cell wall proteins is of major importance for growth, differentiation, development and stress adaptation²²,²³.
The target motifs for Hyp-anchored O-glycosylation in plants, so-called glycomodules, were defined and validated\(^{27,28}\). From these, the consensus motif [A/S/T/T/V]-P(1,4)-X(0,10)-[A/S/T/T/V]-P(1,4) (where X can be any amino acid) was derived for predicting prolyl-hydroxylation in plants\(^{11}\). 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\(^{11}\). Indeed, undesired plant-typical prolyl-hydroxylation\(^{24–26}\) and in some cases subsequent arabinosylation of biopharmaceuticals was reported\(^{27–29}\). On the other hand, the artificial introduction of Hyp-O-glycosylation motifs was suggested as an alternative to PEylation (the attachment of polyethylene glycol-oligomers to proteins or peptide drugs) to increase the serum half-life of biopharmaceuticals\(^{30,31}\). However, non-human prolyl-hydroxylation does not only alter the native sequence of the protein, but also serves as 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.\(^{32}\)). Further, several recombinant proteins have been produced in the moss bioreactor, including rhEPO\(^{33}\), one of the top ten biopharmaceuticals world-wide\(^{34}\). EPO is a highly glycosylated peptide hormone stimulating erythropoiesis. Recombinant H.EPO 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 glycoengineered 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\(^{35,36}\). Production of correctly N-glycosylated asialo-EPO in the moss bioreactor was reported recently\(^{37}\). However, plant-derived rhEPO from moss and Nicotiana benthamiana was shown to be hydroxylated within the motif SPP (amino acids 147–149)\(^{37,38}\) 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) we were able to identify six sequences from the moss 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 Pp1s172_91V6.1 (P4H4), Pp1s12_247V6.1 (P4H5) and Pp1s328_29V6.1 (P4H6), we were able to identify six sequences from the Cosmoss database (www.cosmoss.org) we were able to identify six sequences from the moss 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). Accordingly, gene targeting constructs (Supplementary Fig. S3 online) were designed for the six P4H genes and transferred to the moss bioreactor. An ER-targeted GFP version (ASP-GFP-KDEL\(^{41}\)) as well as GFP without any signal peptide displaying GFP fluorescence in the cytoplasm as well as the nucleus 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 non-human prolyl-hydroxylation of moss-produced rhEPO we aimed to ablate the gene functions of each of the seven P4Hs. Accordingly, gene targeting constructs (Supplementary Fig. S3 online) were designed for the six P4H genes and transferred to the rhEPO-producing moss line 174.1624 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 (∆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...
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 the parental plant, rhEPO produced in moss lines with ablated P4H2, P4H3, P4H4, P4H5 or P4H6, respectively, was hydroxylated in

**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.
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 ΔP4H1 lines. To verify P4H1 enzymatic activity in prolyl-hydroxylation we ectopically expressed this gene in the ΔP4H1 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 (P4H1OE) 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 EASPPDAASAAPLR (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 28, we analysed the rhEPO N-terminal peptide APPRLICDSRVL 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 P4H1OE-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 EASPPDAASAAPLR (144–158) derived from rhEPO produced in moss lines 174.16 (control parental plant), ΔP4H1 #192, ΔP4H2 #6, ΔP4H3 #21, ΔP4H4 #95, ΔP4H5 #29 and ΔP4H6 #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 ΔP4H1 #192 and the presence of hydroxylated peptide (Hyp) in the line ΔP4H4 #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 ΔP4H1 #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.25). 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 recently26,28,29,42 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 biopharmaceuticals25. Several human proteins expressed in plants were shown to be hydroxylated24–29. Recombinant rhEPO produced both in moss and in N. benthamiana was shown to be hydroxylated in the consensus sequence SPP25,28 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 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 production29. This post-translational modification could be partially reduced by P4H inhibitors29,43 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 development22,23, or to the high number of isoforms of this enzyme in plants4 the engineering of plant genomes for mutation or deletion of P4H genes was suggested28,42 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 EAISPDDASAAPLR (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 specificities44, 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 plants46,47, might be conceivable to remove unwanted Hyp formation in both systems.

The consecutive prolines in the N-terminal sequence APPRIL-DDSRVL 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 straightforward 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 P4H1, 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 W46,47 (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://wolfpsort.org/).

Plant material and transformation procedure. Physcomitrella patens (Hedw.) Bruch & Schimp was cultivated as described previously49. 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 was used as the parental line for the P4H knockout generation and the line AP4H1 #192 was used for the generation of P4H1 overexpression lines. In these moss lines the st1.3 fucosyltransferase and the β1,2 xylosyltransferase genes are disrupted7. Wild-type moss was used for the subcellular localization experiments with P4H-GFP.

Plantlet isolation and PEG-mediated transformation was performed as described previously40,41. Mutant selection was performed with Zeocin™ (Invitrogen) or sulfadiazine (Sigma) as described before. For rhEPO production, P. patens was cultivated as described before.

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 pET12.1 (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 mAV4mcs43 using the Xhol and BglII sites giving rise to N-termol 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 taken41.

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®-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® 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 into 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® by HindIII digestion was inserted into the
and the nos terminator into the mAV4mcs vector. For this purpose the GFP gene was used as controls. 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). Sample Laemmli loading buffer (Biorad, Munich, Germany) and electrophoretic analysis via semi-quantitative RT-PCR. For this purpose, cDNA equivalent to 150 ng per sample was amplified with 26 and 28 cycles using the primers listed in Supplementary Table S2 online. The absence of the p4H1 transcript in the KO constructs digested with the same enzyme, pTOPO and 292–299 (2005).
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. Tainen, 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., Hadeler, 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. Bütter-Mainik, A. *et al.* Production of biologically active recombinant human factor H in *Physcomitrella*. *Plant Biotechnol* **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 Stifting 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/.

**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