Prolyl Hydroxylase Paralogs in *Nicotiana benthamiana* Show High Similarity With Regard to Substrate Specificity

Réka Mócsai¹, Kathrin Göritzer², David Stenitzer¹, Daniel Maresch¹, Richard Strasser² and Friedrich Altmann¹*

¹ Department of Chemistry, Institute of Biochemistry, University of Natural Resources and Life Sciences, Vienna, Austria, ² Department of Applied Genetics and Cell Biology, Institute of Plant Biotechnology and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria

Plant glycoproteins display a characteristic type of O-glycosylation where short arabinans or larger arabinogalactans are linked to hydroxyproline. The conversion of proline to 4-hydroxyproline is accomplished by prolyl-hydroxylases (P4Hs). Eleven putative *Nicotiana benthamiana* P4Hs, which fall in four homology groups, have been identified by homology searches using known *Arabidopsis thaliana* P4H sequences. One member of each of these groups has been expressed in insect cells using the baculovirus expression system and applied to synthetic peptides representing the O-glycosylated region of erythropoietin (EPO), IgA1, Art v 1 and the *Arabidopsis thaliana* glycoprotein STRUBBELIG. Unlike the situation in the moss *Physcomitrella patens*, where one particular P4H was mainly responsible for the oxidation of erythropoietin, the tobacco P4Hs exhibited rather similar activities, albeit with biased substrate preferences and preferred sites of oxidation. From a biotechnological viewpoint, this result means that silencing/knockout of a single P4H in *N. benthamiana* cannot be expected to result in the abolishment of the plant-specific oxidation of prolyl residues in a recombinant protein.

**Keywords:** prolyl-hydroxylase, *Nicotiana benthamiana*, substrate specificity, 4-hydroxyproline, plant-specific modification

**INTRODUCTION**

Plant-made pharmaceuticals are gaining more and more attention, therefore the difference between human post-translational modifications (PTMs) and plant PTMs should be addressed. One of the most prominent modifications to deal with is glycosylation of proteins. Glycoengineering of the N-glycosylation pathway of certain plants is already in a very advanced stage. Abolishment of the immunogenic plant-specific β1,2-xylene and α1,3-fucose residues of the core N-glycan was achieved in different plant systems (Koprivova et al., 2004; Strasser et al., 2004; Cox et al., 2006; Strasser et al., 2008; Sourrouille et al., 2008; Jansing et al., 2019) and production of erythropoietin (EPO) without Lewis-A epitopes was achieved in moss (Parsons et al., 2012). O-glycan engineering in plants on the other hand is still in its infancy and plant-produced pharmaceuticals with
human-like structures devoid of plant-specific modifications have not been reported (Castilho et al., 2012; Yang et al., 2012). In mammals, O-glycan biosynthesis occurs in a stepwise fashion involving the sequential transfer of single monosaccharide residues to serine and threonine residues forming mucin-type O-glycans. A similar mucin-type O-glycan biosynthesis pathway does not exist in plants (Strasser, 2012). The first step in the plant-specific O-glycosylation is the generation of Hyp residues. Many Hyp residues are further glycosylated by arabinogalactan, extensin type oligo-arabinosyl glycan chains or solanaceous lectins (Clarke et al., 1979; Lamport, 1980; Showalter and Varner, 1989; José and Puigdomènech, 1993). Hyp residues are generated by peptidyl-proline 4-dioxygenase, also called prolyl 4-hydroxylase (P4H). P4Hs belong to the large class of non-heme iron dioxygenases that use α-ketoglutarate (2-oxoglutarate) and O2 as cosubstrates to activate the metal center and form succinate and a high-valent iron(IV)-oxo species (Price et al., 2003; Proshlyakov et al., 2004). These plant-specific modifications are undesired as concerns have been raised for non-human PTMs on recombinant proteins (Price et al., 2003; Proshlyakov et al., 2004). These plant-specific modifications are undesired as concerns have been raised for the presence of non-human PTMs on recombinant proteins for therapeutic applications. Therefore, the responsible enzymes must be discovered and characterized in order to engineer Nicotiana benthamiana into a competitive expression system for recombinant O-glycosylated proteins.

In mammals two P4H families have been identified, Hypoxia Inducible Factor (HIF)-P4Hs and collagen-P4Hs. Hydroxylation of collagen proline residues results in the formation of the stable triple helical structure of collagen (Mylllyharju, 2003). HIF-P4Hs regulate the stability of a transcription factor (HIF1α) (Epstein et al., 2001). Prolyl 4-hydroxylation can also be detected on other proteins like elastin (Müller et al., 1978) and even on the nuclear protein Argonaute 2 (Qi et al., 2008). Mammalian P4Hs have clear preference for the sequence X-Pro-Gly, whereby good substrate affinity is only provided by strings of more than five such triplets (Kivirikko et al., 1989). In plants, the motifs recognized by P4Hs are more ambiguous. Typical substrates are proline-rich glycoproteins and the subgroup of arabinogalactan proteins (AGPs), which occur in the extracellular space or are linked to the outer membrane of plant cells. An example from a tomato AGP is the sequence TGQTPAAAAXVGAKAGTTPPAAP, where all prolines are converted to Hyp (Zhao et al., 2002). The synthetic sequences (SP)n, (SPP)n, and (SPPP)n (fused to green fluorescent protein) converted to Hyp (Zhao et al., 2002). The synthetic sequences

### MATERIALS AND METHODS

#### Plant Material and Growth Conditions

Wild-type Nicotiana benthamiana plants were grown under long-day conditions (16-h-light/8-h-dark photoperiod) on soil at 24°C.

#### Selection of P4H Paralogs

To discover putative P4Hs in N. benthamiana, Arabidopsis thaliana DNA sequences of known P4H genes were used to perform BLAST (basic local alignment search tool) searches against the gene models in the N. benthamiana database. Protein sequence alignments were performed using the MAFFT algorithm L-INS-i (Kato and Standley, 2013). A maximum likelihood phylogenetic tree was produced using PhyML software (Guindon et al., 2010) at the web server NGPhylogeny.fr (Lemoine et al., 2019) using SPR moves to optimize tree topology and 10 random starting trees. Statistical branch support was calculated using 500 bootstrap replications and the optimal substitution model was assessed by Smart Model Selection (Lefort et al., 2017) to be GTR + G + I + F under the Akaike information criterion (AIC). Multiple sequence alignments and phylogenetic trees were visualized in the MEGA7 software (Kumar et al., 2016).

#### P4H Cloning and Expression in Insect Cells

Total RNA was isolated from leaves of 5-week-old wild type N. benthamiana using the Monarch Total RNA Miniprep kit (New England Biolabs, Frankfurt am Main, Germany) and cDNA was prepared using the LunaScript RT SuperMix

---

1https://benthgenome.qut.edu.au/
kit (New England Biolabs). Four selected candidates of each phylogenetically relevant groups were subcloned into NEB pMiniT 2.0 vectors provided with the NEB PCR mini kit (New England Biolabs). All primers are shown in Supplementary Table 1. Sequences were then verified by Sanger sequencing (Microsynth AG, Balgach, Switzerland) and are deposited in Supplementary Data 1. GenBank accession numbers were assigned to each of these nucleotides: MW524054 (Nb-P4H1), MW524055 (Nb-P4H4), MW524056 (Nb-P4H9) and MW524057 (Nb-P4H10). For expression in insect cells the catalytically active domain was amplified without the membrane anchor. The baculovirus expression system was used to multiply the pVT-Bac-His-1 vector (Tessier et al., 1991) carrying the P4H sequences in Sf9 cells, finally the obtained viral stocks were used to co-transfect *Trichoplusia ni* High five cells to express the recombinant P4H enzymes. The enzymes were His-tag purified from the cell supernatant with immobilized metal affinity chromatography using an Äkta purifier and 1 ml His-Trap FF columns (GE Healthcare Life Sciences, Vienna, Austria).

P4H Activity Assay and LC-ESI-MS Measurements

The enzymatic assays to verify peptide hydroxylation were performed in a total of 50 µL 100 mM MES buffer (pH 6.3) containing 10 µL of purified enzyme, 100 µM synthetic peptide; 300 µM 2-oxoglutarate; 2 mM ascorbate; 50 µM FeSO₄ for 16 h at 30°C. The exact amounts of the recombinant enzymes used were Nb-P4H1: 7.01 µg; Nb-P4H4: 8.25 µg; Nb-P4H9: 4.20 µg; Nb-P4H10: 7.26 µg as determined with the micro BCA protein assay kit (Thermo Fisher Scientific, Vienna, Austria). All the synthetic peptides were ordered from JPT (JPT Peptide Technologies, Berlin, Germany) and denatured by heating to 97°C for 10 min, followed by rapid cooling before using in the activity assays. The reaction mixtures were first passed through a 10mg Hypersep-96 C18 cartridge (Thermo Fisher Scientific) and eluted with 85% Acetonitrile in 80 mM Ammonium-formiate buffer (pH 3) to elute the peptides. For MS measurements the peptides were either loaded onto a Biobasic capillary column (150 mm × 0.32 mm; Thermo Fisher Scientific) or onto an Acclaim PepMap nano column (250 mm × 0.075 mm; Thermo Fisher Scientific) using a Dionex Ultimate 3000 LC system (Thermo Fisher Scientific) coupled to a Bruker maXis 4G Q-TOF MS with the standard or the nano ESI source (Bruker Daltonics). To assist relative quantification, a fixed amount of another synthetic peptide (PTTTPTTTTTVTPTPTPTGTQTK) was added to each sample as internal standard. To quantify the hydroxylated products, it was assumed that the relation between the standard and the substrate is the same as the relation between the standard and the product.

RNAi-Mediated Transient Silencing of Endogenous Nb-P4H1 and Nb-P4H10 Genes

For RNAi-mediated gene silencing of endogenous Nb-P4H1 and Nb-P4H10, a sense-intron-antisense construct was generated following a previously established procedure (Strasser et al., 2008). A subcloned synthetic DNA fragment consisting of a Nb-P4H1 or Nb-P4H10 sense DNA fragment fused to the intron 2 from the *Arabidopsis thaliana* β1,2-xyllosyltransferase gene (Supplementary Data 2) was obtained from GeneArt Gene Synthesis (Thermo Fisher Scientific). The synthetic DNA was used as a template for PCR amplification of the corresponding antisense fragment using the primers listed in Supplementary Table 1. The antisense PCR fragment was *KpnI/BamHI* digested and ligated into the respective cloning vector carrying the sense-intron DNA. The sense-intron-antisense sequence was subsequently excised from the cloning vector and ligated into the *XbaI/BamHI* digested plant expression vector pPT2M (Strasser et al., 2005).

The pEAQ-IgA1 light and heavy chain expression vectors were available from a previous study (Göritzer et al., 2017). IgA1 constructs were additionally cloned into the *XbaI/BamHI* sites pPT2M (Strasser et al., 2005) after amplification of the light and heavy chain constructs from the previously described GeneArt String fragments (Göritzer et al., 2017). In the pPT2M vector the expression is under the control of a cauliflower mosaic virus 35S promoter.

For transient gene silencing, overnight cultures of agrobacteria carrying the expression vector for the IgA1 heavy chain (either pEAQ-IgA1-HC or pPT2M-IgA1-HC) were diluted to an optical

**Determination of *K*ₘ and *v*ₘₐₓ Values**

To determine the *K*ₘ and *v*ₘₐₓ values the activity assays were downscaled to a final volume of 20 µL maintaining the same concentrations and temperature. The recombinant enzymes used were Nb-P4H1: 3.51 µg; Nb-P4H4: 1.65 µg; Nb-P4H9: 0.42 µg; Nb-P4H10: 1.45 µg per reaction. To calculate a *K*ₘ and a *v*ₘₐₓ value the concentration of the synthetic IgA1 peptide varied from 0.022 mM to 5.61 mM. For Nb-P4H1 and Nb-P4H4 a reaction time of 30 min was chosen, while for the other two enzymes the reaction was stopped after 20 min. The assays were immediately purified with 25 µg Hypersep-96 C18 cartridges (Thermo Fisher Scientific) using 85% Acetonitrile in 80 mM Ammonium-formiate buffer (pH 3) to elute the peptides. For MS measurements the peptides were either loaded onto a Biobasic capillary column (150 mm × 0.32 mm; Thermo Fisher Scientific) or onto an Acclaim PepMap nano column (250 mm × 0.075 mm; Thermo Fisher Scientific) using a Dionex Ultimate 3000 LC system (Thermo Fisher Scientific) coupled to a Bruker maXis 4G Q-TOF MS with the standard or the nano ESI source (Bruker Daltonics). To assist relative quantification, a fixed amount of another synthetic peptide (PTTTPTTTTTVTPTPTPTGTQTK) was added to each sample as internal standard. To quantify the hydroxylated products, it was assumed that the relation between the standard and the substrate is the same as the relation between the standard and the product.
density (OD600) of 0.15, bacteria carrying the IgA1 light chain (pEAQ-IgA1-LC or pPT2M-IgA1-LC) or the gene silencing constructs were diluted to an OD600 of 0.1. Equal volumes were mixed and the mixture was infiltrated into *N. benthamiana* wild-type plants as described previously (Strasser et al., 2008). Two (pPT2M-constructs) or 4 days (pEAQ-constructs) after infiltration, IgA1 was purified from leaves as described previously (Göritzer et al., 2017) and subjected to LC-ESI-MS analysis. To determine the gene silencing efficiency, leaves of 5-week-old *N. benthamiana* wild-type plants were infiltrated with the Nb-P4H1-RNAi or Nb-P4H10-RNAi silencing constructs. Infiltration with pPT2M-IgA1-HC was used as a mock control. RNA was isolated 2 days after infiltration from 25 mg leaves using the Monarch Total RNA Miniprep Kit (New England Biolabs). Total RNA was reverse transcribed using the LunaScript RT SuperMix Kit (New England Biolabs). qPCR was performed in triplicate using the GoTaq qPCR Master Mix (Promega) and the CFX96 Touch Real-Time PCR System (Bio-Rad, Feldkirchen, Germany). Protein phosphatase 2A (PP2A) expression was used for normalization of qPCR data. Data analysis was done with the CFX-Manager software version 3.1 (Bio-Rad). Three independent biological replicates were used to calculate the mean values and standard deviation.

Subcellular Localization

For Nb-P4H1-RFP expression, the full-length Nb-P4H1 coding region was amplified from the subcloned Nb-P4H1 pMiniT 2.0 clone with the primers listed in Supplementary Table 1, XbaI/BamHI digested and cloned into XbaI/BamHI digested expression vector p31 (Hütter et al., 2012). The Nb-P4H4-GFP expression construct was generated by PCR amplification from the NEB pMiniT 2.0 clone, XbaI/BglII digestion and ligation into XbaI/BamHI digested expression vector p47 (Hütter et al., 2014). Nb-P4H9 and Nb-P4H10 were cloned in a similar manner into p47. All constructs were transformed into *Agrobacterium tumefaciens* strain U1A43 and used for syringe-mediated agroinfiltration (Strasser et al., 2008). Leaves of 5-week-old wild type *N. benthamiana* plants were infiltrated with *Agrobacterium* suspensions carrying the protein(s) of interest with an OD600 of 0.1. Confocal images were acquired 2 days post infiltration (dpi) on a Leica SP5 confocal microscope (Leica Microsystems, Wetzlar, Germany). Samples were excited using 488- and 561-nm laser lines for GFP and RFP, respectively, and observed using a ×63/l.4 NA or 100 × /1.4 NA CS2 STED white oil immersion objective, respectively. Signals were collected simultaneously from 500 to 530 nm for GFP and 600–630 nm for RFP. Post-acquisition image processing was performed in Adobe Photoshop CS6 (Adobe Systems Incorporated, San Jose, CA, United States). For co-localization GFP/GFP-tagged P4Hs were co-expressed with the *cis*/*medial-Golgi* marker AtGnTI-RFP or AtGnTI-GFP (Schoberer et al., 2019).

Transient Nb-P4H Overexpression in *Nicotiana benthamiana* Leaves

Agrobacteria carrying one of the Nb-P4H-GFP or -RFP fusion constructs were diluted to an OD600 of 0.1 and mixed with agrobacteria carrying pEAQ-IgA1-HC (OD600 of 0.15) and agrobacteria carrying pEAQ-IgA1-LC (OD600 of 0.1). The mixture was infiltrated into leaves of 5-week old wild-type *N. benthamiana* using syringe-mediated agroinfiltration (Strasser et al., 2008). 4 days after infiltration, IgA1 was purified from leaves as described in detail previously (Göritzer et al., 2017) and subjected to LC-ESI-MS analysis.

RESULTS

Identification of Prolyl 4-Hydroxylases in *Nicotiana benthamiana*

For the identification of P4H paralogs in *N. benthamiana*, the nucleic acid sequences of Arabidopsis thaliana P4Hs were used to perform BLAST searches against the gene models in the *Nicotiana benthamiana* database (see text footnote 1) (Supplementary Table 2). With this approach, 11 putative P4Hs could be determined. A phylogenetic analysis was carried out including protein sequences of these *N. benthamiana* P4H candidates and other P4H sequences of different plant origin. Only the putative catalytical domains were aligned to minimize algorithm errors. Some of the known plant P4Hs (eg., *P. patens*, *C. reinhardtii*, and *C. sorokiniana*) displayed very low sequence identity to *N. benthamiana* P4H candidates, therefore these sequences were omitted from the final alignment. The resulting tree depicted four clearly separating branches (Figure 1). The catalytically active domains of one paralog from each group were selected and successfully amplified from *N. benthamiana* cDNA libraries. The candidates were chosen based on the expression levels in the leaf tissue (Supplementary Table 3), these data were obtained from the Gene Expression Atlas (version 6) of the *N. benthamiana* database (see text footnote 1). The selected Nb-P4H candidate genes were the following: Nbv6.1TRP71678 (Nb-P4H1), Nbv6.1TRP32386 (Nb-P4H4), Nbv6.1TRP22823 (Nb-P4H9), and Nbv6.1TRP31841 (Nb-P4H10). Interestingly, aligning several biological replicates of amplified and cloned sequences of Nb-P4H4 and Nb-P4H10 with the sequences of the *N. benthamiana* database showed some discrepancies in sequence identities (Supplementary Table 4). This potentially represents natural differences in *N. benthamiana* accessions (Schiavinato et al., 2019). However, this did not disrupt enzyme activity probably attributed to the fact that the amino acids in key positions for binding cofactors and substrate were unaffected by these mutations. Supplementary Figure 1 shows the amino acid sequence alignment of the obtained sequences next to the *A. thaliana* P4H1 sequence. As clearly marked, the three Fe2+-binding residues (two histidines and an aspartate), and the lysine binding the C-5 carboxyl group of the 2-oxoglutarate are conserved for all Nb-P4H candidates.

To confirm the expression and monitor the subcellular localization of the Nb-P4H proteins, they were transiently expressed in *N. benthamiana* leaves with a fluorescent protein fused to the C-terminus. Co-localization experiments with the *cis*/*medial-Golgi* resident N-acetylglucosaminyltransferase I (GnTI) (Schoberer et al., 2019) showed that all Nb-P4H candidates were located in the Golgi apparatus (Figure 2).
addition to the Golgi localization, Nb-P4H4 was also present in the ER. Overall, the previous localization of the Nb-P4H candidates is consistent with a previous study that detected a tobacco P4H in the ER and Golgi in tobacco BY2 cells (Yuasa et al., 2005).

**Enzyme Activity and Substrate Specificity of Recombinant Nb-P4Hs**

The DNA sequences coding for the catalytic domains from all 4 Nb-P4Hs were cloned into a baculovirus expression vector and expressed in insect cells. Prolly 4-hydroxylase activity was assessed for each recombinant enzyme by mass spectrometry using synthetic peptides designed based on experiments conducted previously. These peptides had the following four sequences: VTVPVPSTPPTPSPSTPPTPSPS from IgA1 and AQKEAISPPDAAS from EPO, representing the critical parts of two biotechnologically relevant naturally O-glycosylated human proteins, AAGGSPSPPADGGSSPPPADG from the arabinan and arabinogalactan carrying *Artemisia vulgaris* pollen allergen Art v 1 (Leonard et al., 2005) and DGTPFNTSIITPPPPCDPPPATHR from the *Arabidopsis thaliana* protein STRUBBELIG (Vaddepalli et al., 2011).

Using the IgA1 peptide, $K_m$ and $v_{max}$ values were determined by mass spectrometry (Table 1 and Supplementary Figures 2, 3, 4). Although using a very specific substrate and utilizing a novel method for determination of these values make the comparison difficult, our results fit in well with the published results of other plant P4Hs (Eriksson et al., 1999; Hieta and Myllyharju, 2002; Tiainen et al., 2005; Keskiaho et al., 2007). At first glance, all four recombinant P4Hs from *N. benthamiana* displayed similar activity and substrate specificity (Figure 3). Each enzyme consistently oxidized up to five Pro residues of the IgA1 peptide during overnight assays, and they all acted on all other synthetic peptides with the exception of Nb-P4H4 on EPO. In the EPO peptide, there are two prolines present that are next to each other but only one of them (Pro$^{149}$) was oxidized by Nb-P4H1, Nb-P4H9, and Nb-P4H10. Nb-P4H1 was the most active enzyme on the extension-like sequence of Art v 1, while Nb-P4H9 and Nb-P4H10 preferred mainly the STRUBBELIG and IgA1 sequences. The cofactors 2-oxoglutarate, ascorbate and Fe$^{2+}$ were all essential to the activity of these enzymes, concluded from inactivity in the absence of these substances (data not shown). A closer inspection by MS/MS, however, revealed clear differences of the four P4Hs in site and substrate
preferences (Figure 4 and Supplementary Figure 5). The MS fragmentation data showed that each enzyme – despite having broad substrate specificity – shows a preference for specific sites. The preferred sequence of recombinant Nb-P4H1 as well as Nb-P4H10 is PPTPS in which the second proline gets oxidized (marked bold). Interestingly, this sequence occurs repeatedly in the IgA1 hinge-region peptide and while Nb-P4H1 mostly acts on $P_{18}$, Nb-P4H10 preferably oxidizes $P_{10}$. Furthermore,
TABLE 1 | Michaelis–Menten constant (K_\text{m}) and maximum velocity (v_\text{max}) values of recombinant \textit{N. benthamiana} P4Hs using the synthetic peptide from IgA1 as substrate (VTVPVTSTPPSTPSSTPS).

| Recombinant enzyme | K_\text{m} (\mu\text{M}) | v_\text{max} (\mu\text{M/min}) |
|--------------------|--------------------------|-------------------------------|
| Nb-P4H1            | 8                        | 0.03                          |
| Nb-P4H4            | 1900                     | 0.1                           |
| Nb-P4H9            | 72                       | 0.4                           |
| Nb-P4H10           | 66                       | 0.02                          |

\(K_\text{m}\) and \(v_\text{max}\) values were calculated based on Hanes-Woolf plots Supplementary Figure 4, panel B1-B4.

Nb-P4H4 displays activity on the same site as Nb-P4H1 but mostly prefers the sequence SPSTP. Nb-P4H9 on the other hand was most active at the site TPSPS. Surprisingly, the prolines at the sites VPVPS, VPSTP and the C terminal SPS harboring \(P_6\), \(P_6\), and \(P_{22}\) residues of the IgA1 sequence were seemingly untouched by any of the recombinant enzymes tested. A total of 7 Hyp residues could be identified as target substrates of the IgA1 peptide, indicating that only 70% of prolines are oxidized by these enzymes. Lastly, Nb-P4H activity was probed with HEK293-cell produced IgA1 substrates as well, however, no conversion of proline to hydroxyproline could be detected (data not shown) possibly caused by different folding or the occupancy of the nearby O-glycosylation sites.

Overexpression of Nb-P4H1, Nb-P4H4, Nb-P4H9, and Nb-P4H10 Fusion Proteins and RNAi-Mediated Transient Silencing of Nb-P4H1 and Nb-P4H10 Genes

Monomeric IgA1 was co-expressed with different fluorescent protein-tagged Nb-P4Hs to see whether their overexpression results in increased Hyp formation in the IgA1 hinge region. Recombinant IgA1 was purified from crude leaf extract, digested with trypsin and applied to ESI-LC-MS analysis to investigate the changes in proline hydroxylation on the IgA1 hinge region. Relative quantification by peak integration revealed highest degrees of hydroxylation through co-expression with Nb-P4H1-RFP and Nb-P4H10-GFP (Table 2A and Supplementary Figure 6). We therefore generated Nb-P4H1 and Nb-P4H10 RNAi silencing constructs for co-infiltration with IgA1 expression vectors in \textit{N. benthamiana}. The used RNAi construct is based on a design that we successfully used previously in stable or transient silencing approaches and gives good expression 2 to 3 days after infiltration (Strasser et al., 2008; Shin et al., 2017). In addition to the pEAQ-IgA1 expression vector that shows highest expression 4 days after infiltration (Göritzer et al., 2017), IgA1 was expressed using the same binary expression vector (pPT2M) that was used for expression of the gene silencing constructs and gives maximal expression levels 2 to 3 days after infiltration. By co-infiltration of pPT2M-IgA1 with \(\text{Nb-P4H1}/\text{Nb-P4H10}-\text{RNAi}\) the highest percentage of non-modified IgA1 hinge-region could be reached (Figure 5). Even though the proline modification was reduced, keeping over 50% of the peptides unmodified, the Nb-P4H activity could not be completely abolished by only silencing two of the \(\text{Nb-P4Hs}\) (Table 2B). This is in agreement with the observed site-specificity of different recombinant Nb-P4H on synthetic substrates suggesting that several enzyme paralogs are active on a given peptide substrate in \textit{N. benthamiana}. On the other hand, a surprising feature of Nb-P4H1 and Nb-P4H10 silencing was the greatly reduced attachment of pentoses to the remaining Hyp residues (Figure 5).

DISCUSSION

In addition to already well-established mammalian cell systems, the plant-based production of recombinant immunoglobulins is gaining more and more attention as plants allow the production of customized homogeneous N-glycans with few engineering steps (Strasser et al., 2014; Loos et al., 2014). The next step to tailor immunoglobulin’s posttranslational glycosylation is to direct our attention toward O-glycans and the elimination of plant-specific modifications. These start with a proline 4-hydroxylase oxidizing a proline residue in the substrate polypeptide chain. This is then followed by glycosyltransferases building up arabinins or arabinogalactans (Leonard et al., 2005; Leonard et al., 2010). Arabinose residues attached to O-glycosylation sites have been described for maize-produced IgA and for a mucin-tandem repeat containing peptide expressed in \textit{N. benthamiana} (Karnoup et al., 2005; Pinkhasov et al., 2011). These plant-specific glycans could cause immunogenic or allergic reactions in humans (Leonard et al., 2005). Engineering mucin-type O-glycans onto MUC1 was successful in \textit{N. benthamiana}, but the unmodified hydroxyprolines were still a concern in the work of Yang et al. (2012).

Even sialylated structures could be engineered into \textit{N. benthamiana} without causing phenotype changes, losing biomass or any adverse effect on expression yield (Castilho et al., 2012). Still, hydroxyproline remained integrated in the product, which asks for further engineering steps. In this work four different homology groups of potential P4Hs from \textit{N. benthamiana} were identified. Members of each of these groups, that were assumed as potentially relevant for heterologous protein production were expressed and characterized. Consistent with previous data on BY2 cells (Yuasa et al., 2005) or transient expression of Arabidopsis P4Hs in \textit{N. benthamiana} (Velazquez et al., 2015), all selected Nb-P4H candidates were localized in the Golgi apparatus, with Nb-P4H4 showing presence in the ER as well.

Our selected method to evaluate P4H activity was LC-ESI-MS. Prolyl 4-hydroxylation is clearly detectable by searching the MS spectra for additional peaks with the mass increment of 16 Da. Four synthetically produced peptides, two from human proteins of biomedical interest and two from plant proteins were used as substrate for the recombinant enzymes. Differences in substrate preferences were observed, but in general all 4 enzymes were active on all four substrates with the exception of Nb-P4H4 on human erythropoietin peptide. Moreover, LC-ESI-MS/MS also revealed remarkable differences in the order of hydroxylation of different proline residues within a substrate. Eventually, although in different progression, all recombinant Nb-P4Hs oxidized the same target sequences on our substrates, showing...
no differences in specificity. Human IgA1 peptide was used as substrate for developing a new method to obtain the Michaelis-Menten constants and values of the reactions’ maximum velocity. These values didn’t show greater differences between the different *N. benthamiana* enzymes than those reported in *A. thaliana* P4H orthologs (Tiainen et al., 2005). Based on our results, Nb-P4H1 (K_m = 8 µM) and Nb-P4H10 (K_m = 66 µM) have the highest affinity to bind to the IgA1 hinge region peptide, which is in agreement with the overexpression results.

Substrate specificity of plant P4Hs is still not well defined and poly-L-proline was accepted as a general substrate (Tanaka et al., 1980). Initially, this suggested the recognition of the characteristic
three-residue-per-turn extended helix, a conformation rather than a sequence. However, Hyp and glycosylated Hyp occur on sequences lacking this polyproline II conformation (Kieliszewski and Lamport, 1994), which implies that plant P4Hs, just like animal P4Hs, are indeed essentially directed by amino acid sequence. In green algae, contiguous Pro residues, Pro residues alternating with some other amino acid, mostly Ser, and PPSPX repeats have been identified in hydroxyproline-rich glycoproteins (where X means any amino acid), with many of the Pro residues being subsequently 4-hydroxylated (Adair and Snell, 1990; Sumper and Hallmann, 1998; Ferris et al., 2001). The Chlamydomonas reinhardtii P4H1 also used SPKPP as substrate amongst other motifs already mentioned above, albeit with lesser efficiency (Keskiaho et al., 2007). This is a rare occurrence, because Pro followed by a Lys residue is not hydroxylated in higher plants but is a good substrate of a viral P4H (Kieliszewski and Lamport, 1994; Eriksson et al., 1999). In BY-2 tobacco cells the 5 amino acid sequence required for proline 4-hydroxylation was described as [AVSTG]–P–[AVSTG]–[GAVPSTC]–[APSDE] (modification site marked bold) (Shimizu et al., 2005). Our findings showed that all investigated Nb-P4Hs used the following 5 sequence motif in the IgA1 peptide: [VSTP]–P–[STP]–[TP]–[SP]. This complements the sequence requirements of Shimizu et al. (2005) with Pro residues in both +1 and −1 position to the modification site. In a tomato AGP, all prolines of the peptide sequence TGQTPAAAXVGAKGTTPPAAP were converted to Hyp, creating similar motifs with prolines in both +1 and −1 positions (Zhao et al., 2002). With the help of the EPO peptide, another motif (PPDAA) was revealed to be substrates of some Nb-P4Hs. Nb-P4H1, Nb-P4H9 and Nb-P4H10 utilized this motif where Asp is in +1 position of the hydroxylation site. This motif was also discovered as substrate for moss P4H1 (Parsons et al., 2013). We proved that Nb-P4Hs are acting on various different substrates and new hydroxylation sites were discovered. It is clear that the sequence requirements are far from specific. This might cause issues with not only IgA and EPO, but at the production of other glycoproteins as well. There is a clear need for further understanding and successful elimination of P4H activity in order to utilize N. benthamiana as a safe host for the expression of recombinant glycoproteins. By overexpression of the selected candidates in N. benthamiana leaves, some candidates seemed to act on the IgA1 proline-rich peptide more, therefore silencing constructs were developed for transient co-expression with glycoproteins. Using this approach, the quantity of Hyp residues and the

FIGURE 4  | Summarized results of the site-specific analysis of P4H activity of recombinant P4H candidates from N. benthamiana. Base Peak Chromatograms of the LC-ESI-MS runs are depicted and location of oxidation sites are marked as detected by MS/MS analysis of the synthetic IgA1 sequence VTVP4VP6STP9P10TP12; SP14STP17P18TP20; SP22S. Individual MS2 spectra for each peak can be found in Supplementary Figure 5.
attached pentoses on the plant produced recombinant IgA1 were perceivably reduced. Pentoses, presumably arabinoses (Shpak et al., 2001; Xu et al., 2008), were only observed where three or more Hyp residues were present. The exact arrangement of these substrates may be relevant for arabinosyl transferases, which would explain the stronger reduction of glycosylation compared to that of hydroxylation. Glycosylation was previously shown to have more rigid sequence requirements where three or more Hyp residues were present. The average number of Hyp residues per peptide is listed below (av. #Hyp) together with percent increase or decrease (%change) compared to the wild IgA1 expression at different time points) for recombinant production of IgA1 are evaluated separately. Values are presented in percent (%) using the summed peak areas of all observed charge states. The first four isotopic peaks were considered for each of the charge states; in case of ammonium adduct formation only the monoisotopic peak was considered. The average number of Hyp residues per peptide is listed below (av. #Hyp) together with percent increase or decrease (%change) compared to the wild IgA1 construct (Nb-IgA1). The corresponding mass spectra can be found in Supplementary Figure 6 and in Figure 5. Obviously, any strategy toward elimination of proline 4-hydroxylation in *N. benthamiana* will have to consider the possibility of phenotypic changes. The effect of P4H action on plant phenotype depends on the particular P4H and its substrate. Elimination of AtP4H2, AtP4H5 and AtP4H13 in *Arabidopsis* resulted in a short root hair phenotype (Velasquez et al., 2011; Velasquez et al., 2015). In *Arabidopsis* a correct O-glycosylation on extensins seemed essential for cell-wall assembly and, hence, root hair elongation. In the moss *P. patens* and, hence, root hair elongation. In the moss *P. patens* (Parsons et al., 2013). A profound effect on general plant growth was revealed by gene silencing of putative P4Hs in tomato (Fragkostefanakis et al., 2014). Interestingly, silencing of three different P4Hs in tomato resulted in an increased leaf area due to enhanced cell division and cell expansion. While the complete elimination could potentially have adverse side effects, an increased leaf area would also be favorable in terms of biomass for recombinant protein production in *N. benthamiana*. Strategies to overcome potential adverse effects of Nb-P4H elimination include the use of powerful transient gene silencing or tissue or even cell-type specific genome editing approaches (Huang et al., 2009; Wang et al., 2020). Moreover, a better understanding of the substrate specificities of different P4Hs may allow the generation of engineered P4H variants that act specifically on endogenous plant proteins without hydroxylation (Shpak et al., 2001; Xu et al., 2008), were only observed.
FIGURE 5 | LC-ESI-MS spectra of the hinge region peptide from recombinant IgA1 produced in the presence of Nb-P4H silencing constructs (Nb-P4H1 RNAi and Nb-P4H10 RNAi). Recombinant IgA1 was expressed by infiltration of pPT2M-IgA1. (A) Unmodified peptide and additional peaks showing P4H activity. (B) A wider mass range shows the same samples carrying additional O-glycans in the form of arabinosylation. Asterisks denote impurities unrelated to the target molecules.
custom-made O-glycans lacking Hyp and associated plant-specific glycosylation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Nicotiana benthamiana database (https://benthgenome.qut.edu.au/).

AUTHOR CONTRIBUTIONS

RM was involved in most experiments, analysis, and interpretation of the data as well as drafting the manuscript. KG planned and executed plant expression and silencing experiments. DS did the enzyme kinetic experiments. DM provided insights and help with mass spectrometry and data quantification. FA and RS conceived and supervised the work. All authors have revised and approved the final version.

REFERENCES

Adair, W. S., and Snell, W. J. (1990). “The Chlamydomonas reinhardtii cell wall: structure, biochemistry, and molecular biology,” in Organization and Assembly of Plant and Animal Extracellular Matrix, eds W. S. Adair and R. P. Mecham (San Diego, CA: Academic Press), 15–84.

Castillo, A., Neumann, L., Daskalova, S., Mason, H. S., Steinkellner, H., Altmann, F., et al. (2012). Engineering of sialylated mucin-type O-glycosylation in plants. J. Biol. Chem. 287, 36518–36526. doi: 10.1074/jbc.M112.402685

Clarke, A. E., Anderson, R. L., and Stone, B. A. (1979). Form and function of arabinoxylans and arabinoxylan-proteins. Phytochemistry 18, 521–540. doi: 10.1016/S0031-9422(00)84255-7

Cox, K. M., Sterling, J. D., Regan, J. T., Gasdaska, J. R., Frantz, K. K., Peele, C. G., et al. (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant Lemna minor. Nature Biotechnol. 24, 1591–1597. doi: 10.1038/nb1260

Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O’Rourke, J., Mole, D. R., et al. (2001). C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell 107, 43–54. doi: 10.1016/S0092-8674(01)00507-4

Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999). The O-linked glycosylation pathway: protein transport, structure, biochemistry, and molecular biology, “ in Glycobiology, eds W. S. Adair and R. P. Mecham (San Diego, CA: Academic Press), 15–84.

Hiem, M., Jahn-Schmid, B., Dedic, A., Kelemen, P., Wopfner, N., Altmann, F., et al. (2003). Art v 1, the major allergen of mugwort pollen, is a modular glycoprotein with a defensin-like and a hydroxyproline-rich domain. FASEB J. 17, 106–108. doi: 10.1096/fj.02-0472fje

Huang, C., Xie, Y., and Zhou, X. (2009). Efficient virus-induced gene silencing in plants using a modified geminivirus DNA1 component. Plant Biotechnol. J. 7, 254–265. doi: 10.1111/j.1467-7653.2008.00395.x

Hüttner, S., Veit, C., Schoberer, J., Grass, J., and Strasser, R. (2012). Unraveling the function of Arabidopsis thaliana OS9 in the endoplasmic reticulum-associated degradation of glycoproteins. Plant Mol. Biol. 79, 21–33. doi: 10.1007/s11103-012-9891-4

Hüttner, S., Veit, C., Vavra, U., Schoberer, J., Liebminger, E., Maresch, D., et al. (2014). Arabidopsis class I α-mannosidases MNS4 and MNS5 are involved in endoplasmic reticulum-associated degradation of misfolded glycoproteins. Plant Cell 26, 1712–1728. doi: 10.1105/tpc.114.123216

Jansing, J., Sack, M., Augustine, S. M., Fischer, R., and Bortesi, L. (2019). CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in Nicotiana benthamiana for the production of recombinant proteins lacking β-1, 2-xyllose and core α-1, 3-fucose. Plant Biotechnol. J. 17, 350–361. doi: 10.1111/pbi.12981

Josè, M., and Puigdomènech, P. (1993). Structure and expression of genes coding for structural proteins of the plant cell wall. N. PhytoL 125, 259–282. doi: 10.1111/j.1469-8137.1993.tb03881.x

Kalaitzis, P., Fragkostefanakis, S., Dandachi, F., and Siomos, A. S. (2009). “The role of Prolyl 4-hydroxylases in tomato fruit ripening,” in Proceedings of the VI International Postharvest Symposium, Vol. 877, Antalya, 999–1004.

Karnoup, A. S., Turkelson, V., and Anderson, W. K. (2005). O-linked glycosylation in maize-expressed human IgA1. Glycobiology 15, 965–981. doi: 10.1093/glycob/cw077

Kato, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi: 10.1093/molbev/mst010

Keskiöho, K., Hietta, R., Sormunen, R., and Myllyharju, J. (2007). Chlamydomonas reinhardtii has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. Plant Cell 19, 256–269. doi: 10.1105/tpc.104.042739

Kieliszewski, M. J., and Lamport, D. T. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant J. 5, 157–172. doi: 10.1046/j.1365-313X.1994.05020157.x

Kivirikko, K. I., Myllylä, R., and Pihlajaniemi, T. (1989). Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four co substrates and a effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor α-like peptides. J. Biol. Chem. 277, 23965–23971. doi: 10.1074/jbc.M201865200

FUNDING

The authors acknowledge the support from the Austrian Science Fund (FWF) (P29830) and [Doctoral Program BioToP–Biomolecular Technology of Proteins (W1224)].

ACKNOWLEDGMENTS

We would like to thank Christiane Veit, Ulrike Vavra, Yun-Ji-Shin, Julia König, Karin Polacek, and Thomas Dalik for their excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.636597/full#supplementary-material
Mócsai et al. Prolyl Hydroxylase Paralogs in Nicotiana benthamiana line AXT/FT. BMC Genom. 20:594. doi: 10.1186/s12864-019-5960-2

Schoberer, J., Liebninger, E., Vavra, U., Veit, C., Grünwald-Gruber, C., Altmann, F., et al. (2019). The golgi localization of Grn1 requires a polaron acid residue within its transmembrane domain. Plant Physiol. 180, 859–873. doi: 10.1104/pp.19.00310

Shimizu, M., Igasaki, Y., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., et al. (2005). Differential regulation of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. Plant J. 42, 877–889. doi: 10.1111/j.1365-313X.2005.02419.X

Shin, Y. J., Castillo, A., Dicker, M., Sádio, F., Vavra, U., Grünwald-Gruber, C., et al. (2017). Reduced paucimannosidic N-glycan formation by suppression of a specific β-hexosaminidase from Nicotiana benthamiana. Plant Biotechnol. J. 15, 197–206. doi: 10.1111/pbi.12602

Showalter, A. M., and Varner, J. E. (1989). “Plant hydroxyproline-rich glycoproteins,” in The Biochemistry of Plants: A Comprehensive Treatise, Vol. 15, ed. A. Marcus (New York, NY: Academic Press), 485–520.

Shpak, E., Barbar, E., Leykam, J. F., and Kielisiewski, M. J. (2001). Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in Nicotiana tabacum. J. of Biol. Chem. 276, 11272–11278. doi: 10.1074/jbc.M011330200

Sourrouille, C., Marquet-Blouin, E., D’Aoust, M. A., Kiefer-Meyer, M. C., Sevono, M., Pagny-Sahelabadi, S., et al. (2008). Down-regulated expression of plant-specific glycoepitopes in alfalfa. Plant Biotechnol. J. 6, 702–721. doi: 10.1111/j.1467-7652.2008.00353.x

Strasser, R. (2012). Challenges in O-glycan engineering of plants. Front. Plant Sci. 3:218. doi: 10.3389/fpls.2012.00218

Strasser, R., Altmann, F., Mach, L., Glossl, J., and Steinkellner, H. (2004). Generation of Arabidopsis thaliana plants with complex N-glycans lacking B1, 2-linked xylose and core α1, 3-linked fucose. FEBS Lett. 561, 132–136. doi: 10.1016/S0014-5793(04)00150-4

Strasser, R., Altmann, F., and Steinkellner, H. (2014). Controlled glycosylation of plant-produced recombinant proteins. Curr. Opin. Biotechnol. 30, 95–100. doi: 10.1016/j.copbio.2014.06.008

Strasser, R., Stadlmayr, J., Schahs, M., Stiegler, G., Quendler, H., Mach, L., et al. (2008). Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. Plant Biotechnol. J. 6, 392–402. doi: 10.1111/j.1467-7652.2008.00330.x

Strasser, R., Stadlmayr, J., Svoboda, B., Altmann, F., Glossl, J., and Mach, L. (2005). Molecular basis of N-acetylgalactosaminyltransferase I deficiency in Arabidopsis thaliana plants lacking complex N-glycans. Biochem. J. 387, 385–391. doi: 10.1042/BJ20041666

Sumper, M., and Hallmann, A. (1998). "Biochemistry of the extracellular matrix of Volvox," in International Review of Cytology, Vol. 180, ed. K. W. Leon (Cambridge, MA: Academic Press), 51–85. doi: 10.1002/0470-6796(08)61770-2

Tanaka, M., Shibata, H., and Uchida, T. (1980). A new prolyl hydroxylase acting on poly-L-proline, from suspension cultured cells of Nicotiana tabacum. J. Biochem. 97, 121, 332, 1401–1403. doi: 10.1262/jbc.232000922

Tessier, D. C., Thomas, D. Y., Khouri, H. E., Laliberié, F., and Vernet, T. (1991). Tri and tetrasaccharides of an art v 1 from mugwort. Biochemistry 30, 27192–27200. doi: 10.1021/bi0013300

Van der Hoff, S. (2012). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. Plant J. 42, 877–889. doi: 10.1111/j.1365-313X.2005.02419.X

Velasquez, S. M., Ricardo, M. M., Doros, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., et al. (2011). O-glycosylated cell wall proteins are essential in root hair growth. Science 332, 1401–1403. doi: 10.1126/science.1206657

Velasquez, S. M., Ricardo, M. M., Poulsen, C. P., Okawa, A., Dilkopimol, A., Halim, A., et al. (2015). Complex regulation of prolyl-4-hydroxylyases impacts root hair expansion. Mol. Plant 8, 734–746. doi: 10.1016/j.molp.2014.11.017
Mócsai et al. Prolyl Hydroxylase Paralogs in Nicotiana benthamiana

Wang, X., Ye, L., Lyu, M., Ursache, R., Löytynoja, A., and Mähönen, A. P. (2020). An inducible genome editing system for plants. *Nat. Plants* 6, 766–772. doi: 10.1038/s41477-020-0695-2

Xu, J., Tan, L., Lamport, D. T., Showalter, A. M., and Kieliszewski, M. J. (2008). The O-Hyp glycosylation code in tobacco and *Arabidopsis* and a proposed role of Hyp-glycans in secretion. *Phytochemistry* 69, 1631–1640. doi: 10.1016/j.phytochem.2008.02.006

Yang, Z., Drew, D. P., Jørgensen, B., Mandel, U., Bach, S. S., Ulvskov, P., et al. (2012). Engineering mammalian mucin-type O-glycosylation in plants. *J. Biol. Chem.* 287, 11911–11923. doi: 10.1074/jbc.M111.312918

Yuasa, K., Toyooka, K., Fukuda, H., and Matsuoka, K. (2005). Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *Plant J.* 41, 81–94. doi: 10.1111/j.1365-313X.2004.02279.x

Zhao, Z. D., Tan, L., Showalter, A. M., Lamport, D. T. A., and Kieliszewski, M. J. (2002). Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. *Plant J.* 31, 431–444. doi: 10.1046/j.1365-313X.2002.01365.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mócsai, Göritzer, Stenitzer, Maresch, Strasser and Altmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.