Characterization of a Second Arabidopsis thaliana Prolyl 4-Hydroxylase with Distinct Substrate Specificity*

Received for publication, September 28, 2004, and in revised form, October 22, 2004
Published, JBC Papers in Press, November 4, 2004, DOI 10.1074/jbc.M411109200

Paivi Tiainen, Johanna Myllyharju, and Peppi Koivunen‡

From the Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, FIN-90014 Oulu, Finland

4-Hydroxyproline is found in collagens, collagen-like proteins, elastin, and the hypoxia-inducible transcription factor in animals and in many hydroxyproline-rich glycoproteins in plants. We report here on the cloning and characterization of a second plant P4H (prolyl 4-hydroxylase), At-P4H-2, from Arabidopsis thaliana. It consists of 299 amino acids and shows 33% sequence identity to the first characterized isoenzyme, At-P4H-1. A characteristic feature of the At-P4H-2 polypeptide is a 49-amino-acid C-terminal toxin homology domain with 6 cysteines that is not found in At-P4H-1 but is present in a putative rice P4H homologue. At-P4H-2 differed distinctly from At-P4H-1 in its substrate specificity. Recombinant At-P4H-2 hydroxylated poly(L-proline) and tetraspan- and arabinogalactan-like peptides effectively but with much higher K_m values than At-P4H-1, suggesting different roles for the two At-P4Hs in the plant cell. Unlike At-P4H-1, At-P4H-2 hydroxylated collagen-like peptides only very inefficiently and did not hydroxylate hypoxia-inducible transcription factor α-like peptides at all. All the peptides efficiently hydroxylated by At-P4H-2 had at least 3 consecutive prolines, suggesting that these may represent a minimum requirement for efficient hydroxylation by this isoenzyme. N-terminal sequencing of an extensin-like peptide SPPPVYKSPPPPVKHYSPPPV indicated that At-P4H-2 preferentially hydroxylated the 3rd proline in the C-terminal PPP triplet. The K_m values of At-P4H-2 for the reaction cosubstrates Fe^3+, 2-oxoglutarate, and ascorbate were similar to those of At-P4H-1 with the exception that the K_m for iron was about 3-fold lower. Pyridine-2,4-dicarboxylate and pyridine-2,5-dicarboxylate, well known competitive inhibitors of the vertebrate P4Hs with respect to 2-oxoglutarate, were also competitive inhibitors of At-P4H-2 but with K_i values 5-100-fold higher than those of human type I collagen P4H. It thus seems that there are some distinct differences in the structure of the 2-oxoglutarate-binding site between At-P4H-2 and the animal collagen P4Hs.

* This work was supported by grants from the Health Science Council (Grant 200471) and the Finnish Centre of Excellence Programme 2000-2005 (Grant 44843) of the Academy of Finland, and from the S. Juselius Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Collagen Research Unit, Biocenter Oulu and Dept. of Medical Biochemistry and Molecular Biology, P. O. Box 5000, University of Oulu, FIN-90014, Finland. Tel.: 358-8-5375822; Fax 358-8-5375811; E-mail: peppi.koivunen@oulu.fi.

1 The abbreviations used are: P4H, prolyl 4-hydroxylase; C-P4H, collagen P4H; At-P4H-1, A. thaliana P4H isoform 1; At-P4H-2, A. thaliana P4H isoform 2; HIF, hypoxia-inducible transcription factor; HRGP, hydroxyproline-rich glycoprotein; PBCV-1, P. bursaria Chlorella virus-1; HPLC, high pressure liquid chromatography.
wise been found to hydroxylate poly(L-proline), other proline-rich peptides, and collagen-like peptides (16).

All P4Hs require Fe$^{2+}$, 2-oxoglutarate, O$_2$, and ascorbate, and their reaction involves stoichiometric decarboxylation of 2-oxoglutarate (1, 2). The Fe$^{2+}$ atom is bound to the catalytic site by 2 histidines and 1 aspartate, whereas the C-5 carboxyl group of the 2-oxoglutarate is bound to a lysine located in position +10 with respect to the 2nd iron-binding histidine (17). The C-P4Hs have a peptide substrate-binding domain that is distinct from the C-terminal catalytic domain (18, 19), but it is currently unknown whether the other P4Hs also have a separate peptide substrate-binding domain or whether their substrates become bound directly to the catalytic domain.

We report here on the cloning and characterization of a second A. thaliana P4H isoenzyme, At-P4H-2. This is a 33-kDa monomer that differs distinctly from At-P4H-1 in its substrate specificity. It was found to hydroxylate poly(L-proline) and extensive and arabinogalactan-like peptides but with much higher $K_m$ values than At-P4H-1, suggesting a different role in the plant cell. Unlike At-P4H-1, it hydroxylated collagen-like peptides only very inefficiently and did not hydroxylate HIFα-like peptides at all. We also determined the $K_m$ values of At-P4H-2 for the reaction cosubstrates and $K_v$ values for two 2-oxoglutarate analogue inhibitors.

**MATERIALS AND METHODS**

**Cloning and Expression of Recombinant A. thaliana P4H-2 in Insect Cells**—PCR primers 5'-GGCGGCCGCAGATTTAGTCTAGGTCGTTCTCGCGGCGG-3' and 5'-GGCTCTCAAGAGGAGGTTTAAAGCTTTAACAAGGTTTACG-3' were synthesized based on the sequence of the gene encoding the polypeptide NP566279.1 (GenBank™, named here At-P4H-2) and used to obtain a 947-bp PCR product coding for full-length At-P4H-2 cDNA from an A. thaliana Uni-ZAP XIT cDNA library (Stratagene). The PCR product obtained had added NotI and Xbal restriction sites at its 5' and 3' ends, respectively (underlined in the primer sequences), and it was cloned into a NotI-Xbal-digested baculovirus vector pVL1392 (Pharmingen). The sequence was verified on an ABI Prism 377 automated DNA sequencer (Applied Biosystems).

The recombinant vector was cotransfected into Spinotauga frugiperda (Sf9) insect cells with a modified Autographa californica nuclear polyhedrosis virus DNA (BaculoGold, Pharmingen) by calcium phosphate transfection, and the recombinant virus was amplified (20). Polyhedrosis virus DNA (BaculoGold, Pharmingen) by calcium phosphate transfection, and the recombinant virus was amplified (20). The presence of At-P4H-2 in the 0.1% Triton X-100-soluble fraction was verified by N-terminal sequencing. Triton X-100 and SDS-soluble samples were harvested 72 h after infection, washed with a solution of 0.15M NaCl, 0.02 M phosphate, pH 7.4, 2 mM CaCl$_2$, 10 mM dithiothreitol, 0.1% Triton X-100, and 0.01 M Tris buffer, pH 7.8, and centrifuged at 10,000 × g for 20 min. The insoluble pellet was further solubilized in two consecutive steps, first in the same buffer supplemented with 1% Triton X-100 and then in a 50% glycerol, 0.6 x NaCl, 1% Nonidet P-40, 0.1 x glycine, 10 mM dithiothreitol, and, 0.01% Triton buffer, pH 7.8. The remaining pellet was solubilized in 1% SDS, and aliquots of all fractions were analyzed by 10% SDS-PAGE under reducing conditions followed by Coomassie Blue staining. The presence of At-P4H-2 in the 0.1% Triton X-100-soluble fraction was verified by N-terminal sequencing. Triton X-100 and SDS-soluble samples were transferred from SDS-PAGE to a ProBlot$^\text{TM}$ membrane (Applied Biosystems) by electroblotting and stained with Coomassie Blue. The At-P4H-2 polypeptide visible in the SDS-soluble sample and a slice of the corresponding position of the Triton X-100-soluble sample were excised from the blot and analyzed in a PreciseTM 492 sequencing system (Applied Biosystems).

**P4H Activity Assay—P4H activity was analyzed by a method based on the hydroxylated-coupled decarboxylation of 2-oxo-1,4C5-glutarate (22). The reaction was performed in a final volume of 0.5 ml, which contained 10–50 μl of the 0.1% Triton X-100-soluble fractions of the insect cells expressing recombinant At-P4H-2 or Erp57, which was used as a negative control, as sources of the enzyme, differing amounts of various peptide substrates (Sigma, Invogen), 0.025 μmol of FeSO$_4$, 0.15 μmol of 2-oxo-[1,4C]$^5$glutarate (150 000 dpm), 1 μmol of ascorbate, 30 μg of catalase (Sigma), 0.05 μmol of dithiothreitol, 1 μg of bovine serum albumin (Sigma), and 25 μmol of Tris-HCl buffer, adjusted to pH 7.8 at 25 °C. The enzyme reaction was carried out at 37 °C for 30 min. The amounts of total cellular protein in the Triton X-100 extracts were assayed by Roti-Quant (Roth) and adjusted to be equal in the activity assay reactions. The $K_m$ and $V_{max}$ values of At-P4H-2 were determined as described previously (17, 22) with poly(L-proline) as a substrate. The amount of 4-hydroxypoline formed in a poly(L-proline) substrate was determined by a colorimetric method in samples hydrolyzed with 6 M HCl at 120 °C overnight (23).

**Other Assays—**The hydroxylase site specificity of At-P4H-2 in the assay was assayed by Roti-Quant (Roth) and adjusted to be equal in the activity assay reactions. The $K_m$ and $V_{max}$ values of At-P4H-2 were determined as described previously (17, 22) with poly(L-proline) as a substrate. The amount of 4-hydroxyproline formed in a poly(L-proline) substrate was determined by a colorimetric method in samples hydrolyzed with 6 M HCl at 120 °C overnight (23).

**RESULTS**

**Cloning of At-P4H-2 and Comparison of Its Amino Acid Sequence with Those of Other P4Hs—**A sequence homology search of the A. thaliana genome indicated that it encodes at least six polypeptides of 280–392 residues that show 21–27% identity to the catalytically important C-terminal regions of the α subunits of the human type I and type II C-P4Hs (15). One of these, a 283-residue polypeptide with a GenBank™ accession number AAC64297, was cloned and characterized recently and named At-P4H-1 (15), being the first plant P4H to be cloned from any source. We have now cloned a second 299-amino-acid A. thaliana P4H based on the GenBank™ sequence NP566279.1 and have named it At-P4H-2, its gene being located in chromosome 3. This polypeptide corresponds to AAP08583 as identified in the first sequence homology search (15), the first 178 amino acids encoded by NP566279.1 being identical to those of AAP08583, whereas the subsequent AAP08583 sequence has a deletion corresponding to 21 amino acids. A third GenBank sequence, AAM67123.1, is identical to NP566279.1 except that it lacks the first two amino acids of the encoded polypeptide.

We used PCR to amplify a 947-nucleotide product from an A. thaliana cDNA library encoding 11 nucleotides of the 5'-untranslated sequence, the entire coding region, and 16 nucleotides of the 3'-untranslated sequence. This cDNA encodes a 299-amino-acid polypeptide identical to NP566279.1. The calculated molecular mass of At-P4H-2 is 33,019 Da, its isoelectric point is 5.93, and the polypeptide contains three potential N-glycosylation sites. The 3 amino acids, 2 histidines, and 1 aspartate that bind the iron atom and the lysine that binds the C-5 carboxyl group of the 2-oxoglutarate (17) are conserved in its sequence (Fig. 1).

At-P4H-2 shows a 33% amino acid sequence identity to At-P4H-1 (Fig. 1), the identity of the C-terminal catalytic regions being distinctly higher. The identity of the 126-amino-acid stretches starting at residue 118 of At-P4H-2 and at residue 159 of At-P4H-1 is 51%, for example (Fig. 1). A characteristic feature of the At-P4H-2 polypeptide is a 49-amino-acid C-terminal extension with 6 cysteines that is not found in At-P4H-1 (Fig. 1). Interestingly, a closer homologue of At-P4H-2 is a 310-amino-acid polypeptide found in rice (AAO72377.1), which is 53% identical to At-P4H-2. These two polypeptides also share the C-terminal cysteine-rich extension (Fig. 1). The P. bursaria Chlorella virus-1 (PBCV-1) P4H (16) shows a 25% overall sequence identity to At-P4H-2, the C-terminal catalytic regions being more alike, the identity of a 126-amino-acid stretch starting at residue 118 of At-P4H-2 to that starting at residue 131 of the PBCV-1 P4H being 34% (Fig. 1). Interestingly, a 24-amino-acid stretch in the N-terminal region of At-P4H-2, starting at residue 52, is 67 and 54% identical to the corresponding regions
in the rice and PBCV-1 P4Hs, whereas its identity to the corresponding region in At-P4H-1 is only 42% (Fig. 1). The sequence identity between At-P4H-2 and the 517-residue α subunit of the human type I C-P4H is 23% (data not shown), being thus very similar to the identity of 25% between At-P4H-1 and this α subunit (15).

Expression of Recombinant At-P4H-2 in Insect Cells—A baculovirus expressing At-P4H-2 was generated and used to infect Sf9 and H5 insect cells. The cells were harvested 72 h after infection, homogenized in a buffer containing 0.1% Triton X-100, and centrifuged. The pellet obtained was further solubilized consecutively in a buffer containing 1% Triton X-100 and in a buffer containing 50% glycerol, 0.6 M NaCl, and 1% Nonidet P-40, and 1% SDS. Analysis of the glycerol and SDS extracts by SDS-PAGE under reducing conditions indicated that both insect cell types expressed the recombinant At-P4H-2 as a polypeptide of about 35 kDa, as shown for Sf9 cells in Fig. 2 (lanes 3 and 4), whereas its solubility in the two Triton X-100-containing buffers was poor (Fig. 2, lanes 1 and 2). The 0.1% Triton X-100 and SDS-soluble samples were electrophoblated, and the membrane was stained by Coomassie Blue. N-terminal sequencing of the band representing the SDS-soluble At-P4H-2 polypeptide gave the sequence MSMSRGLLILF, corresponding to the N terminus of At-P4H-2.

When a slice of the blot cut from the corresponding migration position of the 0.1% Triton X-100-soluble fraction was analyzed, the same sequence was identified, thus verifying the presence of At-P4H-2 in this fraction. High levels of At-P4H-2 activity were seen in the 0.1% Triton X-100-soluble fraction, the activity generated in Sf9 cells being twice that obtained in H5 cells (data not shown), and therefore the former cell line was chosen for further investigation. The At-P4H-2 activity in the 50% glycerol-soluble fraction was only 16% of that in the 0.1% Triton X-100-soluble fraction when expressed per 100 μg of protein in the fraction (data not shown), suggesting that a considerable percentage of the glycerol-soluble polypeptide was incorrectly folded. The 0.1% Triton X-100-soluble fraction was therefore used as the At-P4H-2 source in all subsequent experiments.

The Recombinant At-P4H-2 Effectively Hydroxylates Poly(l-proline)—P4H activity generated in the 0.1% Triton X-100-soluble fraction of Sf9 cells expressing either At-P4H-2 or human ERp57, an endoplasmic reticulum glycoprotein-specific thiol-disulfide oxidoreductase (24, 25) that was used here as a negative control, was analyzed by a method based on the hydroxylation-coupled decarboxylation of 2-oxo-[1-14C]glutarate (22). A high amount of P4H activity was found in the soluble fraction of cells expressing At-P4H-2 when 200 μM poly(l-proline) was used as the peptide substrate, the amount of activity increasing with the molecular weight of the poly(l-proline) (Table I). The activity level found in the corresponding fraction of cells expressing ERp57 was less than 0.1% of that with At-P4H-2 (Table I).

The C-P4Hs are known to catalyze a very low rate of uncoupled decarboxylation of 2-oxoglutarate, i.e. decarboxylation without subsequent hydroxylation of a peptide substrate (1, 17). To verify that the 2-oxoglutarate decarboxylation found with At-P4H-2 indeed represented hydroxylation-coupled decarboxylation, the amount of 4-hydroxyproline formed by the insect cell extract expressing At-P4H-2 with poly(l-proline) as a substrate was measured by a colorimetric method (23). Sf9 cells have small amounts of endogenous C-P4H activity and produce small amounts of endogenous collagens (26). Therefore small amounts of 4-hydroxyproline were found in the sample from cells expressing ERp57, evidently being derived from the 4-hydroxyproline present in the cell extract used as the enzyme source (Table II). The amount of 4-hydroxyproline generated by
At P4H-2 was about 12 times of that obtained with this negative control, however (Table II).

**At-P4H-2 Effectively Hydroxylates Peptides with Sequences Corresponding to Those in Arabinogalactan Proteins and Extensins**—A synthetic peptide with a sequence (ATPPPV)3, corresponding to that in arabinogalactan proteins, and two peptides with sequences SPPPVSPPPVSPPPVSPPPPV and SPPPVYKSPPPVPKHYSPPPVSPPPPV, corresponding to those in extensins (15), were studied as substrates using the assay based on the peptides with sequences SPPPVSPPPVSPPPVSPPPPV and SPPPVYKSPPPVPKHYSPPPVSPPPPV—corresponding to that in arabinogalactan proteins, and two poly(L-proline) peptides with different M, values, however, particularly as the activities shown for the latter were obtained with lower substrate concentrations (Table I). The three plant peptides also served as substrates for At-P4H-1, but in this case, the longer extensin-like peptide was the least satisfactory substrate (15). A surprising finding was that the two extensin-like peptides were also hydroxylated to a minor extent by the endogenous P4H present in insect cells, whereas no such hydroxylation was seen with the arabinogalactan protein-like peptide (Table I). No reports are available on the substrate specificity of insect cell P4H, but the present data indicate that it appears to differ distinctly from those of the vertebrate C-P4Hs (1, 2).

**At-P4H-2 Hydrolyses Very Ineffectively Collagen and HIFα-like Peptides and Some Other Peptides Reported to Contain 4-Hydroxyproline**—(PPG)5, a collagen-like peptide that is used as a standard substrate in the assays of C-P4H activity (1, 2) and has also been reported to act as a good substrate for At-P4H-1 (15), was a very ineffective substrate for At-P4H-2, the level of activity obtained even with a 1 mM concentration being less than 3% of that obtained with a 200 μM concentration of poly(L-proline), M, 32,000 (Table I). The level of activity obtained with (PPG)10 was nevertheless much higher than that obtained with the ERp57-expressing negative control, suggesting that the endogenous insect cell enzyme (Table I).

Bradykinin, a peptide involved in the regulation of blood pressure, contains a single -PPG- triplet that can be hydroxylated by the C-P4Hs in vitro (1). Correspondingly, human plasma, urine, and ascitic fluid contain small amounts of hydroxypoline-lysyl-bradykinin with a single 4-hydroxyproline residue (1). We therefore tested whether bradykinin would serve as a substrate for At-P4H-2 but found no significant hydroxylation (Table I).

The prion protein involved in the pathogenesis of transmissible spongiform encephalopathies contains a 4-hydroxyproline residue in its N-terminal domain that may have a poly(L-proline) type II helix (27). A 17-amino-acid peptide corresponding to this site was found to serve as a substrate for At-P4H-2, and a low level of activity was also found with the ERp57-expressing negative control, suggesting that the endogenous insect cell P4H was able to hydroxylate this peptide at a low

---

**Table I**

| Substrate | Peptide concentration | Activity |
|-----------|-----------------------|----------|
| Poly(L-proline), M, 5,000 | 200 | 9840 |<| 2 |
| Poly(L-proline), M, 8,000 | 200 | 24,690 |
| Poly(L-proline), M, 32,000 | 200 | 33,120 | 30 |
| (ATPPPV)3 | 500 | 5370 | 70 |
| SPPPVSPPPVSPPPVSPPPPV | 500 | 9690 | 1100 |
| SPPPVYKSPPPVPKHYSPPPVSPPPPV | 500 | 13,790 | 1050 |
| (PPG)10 | 250 | 320 |
| (PPG)10 | 1000 | 950 | <20 |
| (PAG)5 | 500 | 270 | <20 |
| (APG)5 | 500 | 140 |
| RPPGSPFR | 500 | 200 | 210 |
| RYPGGQGSPGNRKYPQQ | 150 | 190 |
| RYPGGQGSPGNRKYPQQ | 750 | 1230 | 205 |
| DLQEFYIAPPDDFDQQRSDL | 500 | 40 |
| DALTLLAPAGDTSLSLPQ | 500 | <20 |
| ELDLETAPYMDGEDFQ | 200 | <20 |
| EPEELQAPTPGDAISL | 500 | <20 |
| ENIMQLQAPAGTCFDDLDAEKCKYGMEIPTNL | 500 | <20 |

| Enzyme | 4-Hydroxyproline formed |
|--------|-------------------------|
| At-P4H-2 | 3.82 |
| ERp57a | 0.32 |

a Soluble extracts of insect cells expressing ERp57 were used as a negative control.

b Sequences present in arabinogalactan protein (15).

c Sequences present in extensin (15).

d Bradykinin (1).

e Sequence representing the hydroxylation site of the prion protein (27).

f Sequence representing the C-terminal hydroxylation site of HIF-1α (31, 32).

g Sequence representing the N-terminal hydroxylation site of HIF-1α (31, 32).

h Sequence representing the C-terminal hydroxylation site of HIF-2α (31, 32).

i Sequence representing the N-terminal hydroxylation site of HIF-2α (31, 32).

j Sequence representing the potential hydroxylation site of the large subunit of RNA polymerase II (28).

| Enzyme | 4-Hydroxyproline formed |
|--------|-------------------------|
| At-P4H-2 | 3.82 |
| ERp57a | 0.32 |
Second A. thaliana Prolyl 4-Hydroxylase

Table III

| Substrate                             | $K_m$ (µM) At-P4H-2 | $K_m$ (µM) At-P4H-1 | $K_m$ (µM) PBCV-1 |
|---------------------------------------|---------------------|---------------------|-------------------|
| Poly(t-proline), $M_r = 5,000–10,000$ | 30                  | 0.2                 | 500               |
| Poly(t-proline), $M_r = 10,000–20,000$ | ND                  |                     | 100               |
| Poly(t-proline), $M_r = 30,000–40,000$ | 13                  |                     |                   |
| Poly(L-proline), $M_r = 750$          | 1260                | 20                  |                   |
| Poly(L-proline), $M_r = 2800$         | 680                 | 10                  |                   |
| Poly(L-proline), $M_r = 2900$         | 380                 | 40                  |                   |
| Poly(L-proline), $M_r = 3000$         | 2800                | 60                  | 2900              |
| Poly(L-proline), $M_r = 5,000–10,000$ | 30                  |                     |                   |
| Poly(L-proline), $M_r = 10,000–20,000$| ND                  |                     |                   |
| Poly(L-proline), $M_r = 30,000–40,000$| 13                  |                     |                   |

*a Ref. 15.  
*b Ref. 16.  
*c ND, not determined.  
*d Sequences present in arabinoxylan protein (15).  
*e Sequences present in extensin (15).  
*f Sequence representing the hydroxylation site of the prion protein (27).  

The level of activity generated by At-P4H-2 with a 750 µM peptide concentration was nevertheless only about 4–12% of those generated with 200 µM concentrations of the three poly(t-proline) peptides (Table I).

At-P4H-1 has been found to hydroxylate synthetic peptides corresponding to the two hydroxylation sites in human HIF-1α (15), which was a highly surprising finding as the sequences of these peptides are very different from those of the proline-rich plant proteins and as the HIF system has not been characterized in plants. At-P4H-2, however, showed no activity with any of the four peptides representing the C-terminal and N-terminal hydroxylation sites in HIF-1α and HIF-2α (Table I).

It has been suggested that human RNA polymerase II may become bound to the von Hippel Lindau protein by a similar 4-hydroxyproline-mediated mechanism as HIF-α, which would suggest that a LXCLAP sequence present in this polymerase may be hydroxylated (28). We therefore tested whether At-P4H-2 would act on peptides of different lengths corresponding to this sequence, but no hydroxylation was found (as shown in Table I for the longest peptide studied).

The $K_m$ Values of At-P4H-2 for Peptide Substrates Are Markedly Higher Than Those of At-P4H-1—$K_m$ values were determined for poly(t-proline) peptides of two molecular weights and five other synthetic peptides that were found to serve as substrates for At-P4H-2. In agreement with the data in Table I, the lowest $K_m$ values were found for the $M_r = 30,000–40,000$ and $5000–10,000$ poly(t-proline) peptides, being 13 and 30 µM, respectively (Table III). These values are distinctly higher than those for At-P4H-1 as its $K_m$ for poly(t-proline), $M_r = 10,000–20,000$, was as low as 0.2 µM (Table III). The two extensin-like peptides, SPPPPVKSPPPVPKHYSPPPV and SPPPPVSPPPVP, had the lowest $K_m$ values among the other synthetic peptides studied, 380 and 680 µM, respectively, whereas the arabinoxylan-like peptide had a $K_m$ of more than 1 mM (Table III). All these values are about 10–60 times higher than those of At-P4H-1 (Table III). The $K_m$ values of At-P4H-2 for poly(t-proline) were much lower than those of the P. bursaria Chlorella virus-1 P4H, whereas the $K_m$ values of these two P4Hs for (PPG)$_{10}$ were essentially identical (Table III).

At-P4H-2 Preferentially Hydroxylates in the Extensin-like Peptide SPPPPVKSPPPVPKHYSPPPV the Third Proline in the C-terminal PPP Triplet—To study the hydroxylation pattern of the most efficient synthetic peptide substrate, the longer extensin-like peptide SPPPPVKSPPPVPKHYSPPPV, this peptide was partially hydroxylated with the recombinant At-P4H-2. The peptide was then purified from the enzyme reaction mixture, and the molecular mass of the peptide fraction was analyzed by mass spectrometry. The majority of the purified peptide had a molecular mass identical to that of the non-hydroxylated peptide, whereas about 20% had a higher molecular mass. About 85% of this fraction had a molecular mass 16 Da higher and the rest had a molecular mass that was 32 Da higher than that of the non-hydroxylated peptide, suggesting the presence of one and two 4-hydroxyproline residues, respectively (data not shown). N-terminal sequencing of the peptide fraction showed that 4-hydroxyproline residues were present in the 9th and 20th positions of the peptide, i.e. the first residue of the central PPP sequence and the last residue of the C-terminal PPP triplet (Fig. 3). 8.8% of the former and 41.3% of the latter being hydroxylated, whereas no 4-hydroxyproline was found in the N-terminal PPP triplet. No data are available on the hydroxylation pattern of At-P4H-1 with these peptides or any other plant peptide, but it hydroxylated in (PPG)$_{10}$ both consecutive prolines, although it preferentially acted on the prolines preceding the glycine, this pattern being similar to that of the vertebrate C-P4Hs (1–3, 15).

The $K_m$ Values of At-P4H-2 for Reaction Cosubstrates Are Similar to Those of At-P4H-1—The $K_m$ values of At-P4H-2 were determined for Fe$^{2+}$/2-oxoglutarate, and ascorbate (Table IV). The $K_m$ for iron was 5 µM, about one-third of that of At-P4H-1 but more than twice that of the type I C-P4H, whereas the viral P4H has a much lower value (Table IV). The $K_m$ for 2-oxoglutarate was about 100 µM (Table IV), which would suggest that the $K_m$ for Fe$^{2+}$ is about 500 µM.
Second A. thaliana Prolyl 4-Hydroxylase

| Cosubstrate or inhibitor | Constant | At-P4H-2 | At-P4H-1<sup>a</sup> | PBCV-1<sup>b</sup> | C-P4H-F<sup>c</sup> |
|--------------------------|----------|----------|-----------------|----------------|----------------|
| Fe<sup>2+</sup>          | $K_m$    | 5        | 16              | 0.4            | 2              |
| 2-Oxoglutarate           | $K_m$    | 170      | 130             | 20             | 20             |
| Ascorbate                | $K_m$    | 300      | 300             | 300            | 300            |
| Pyridine-2,4-dicarboxylate| $K_i$    | 10       | 2               | 2              | 0.8<sup>d</sup>|
| Pyridine-2,5-dicarboxylate| $K_i$    | 80       |                 |                |                |

<sup>a</sup> Ref. 15.  
<sup>b</sup> Ref. 16.  
<sup>c</sup> Ref. 17.  
<sup>d</sup> Ref. 1.

The amino acid sequence identity between At-P4H-2 and At-P4H-1 is 33%, the most striking difference being a 49-amino-acid C-terminal extension with 6 cysteines present in At-P4H-2. A putative rice homologue of At-P4H-2 also possesses this C-terminal extension, the overall sequence identity between these two polypeptides being 53%. This extension, which may represent a separate domain, is also found in several proteins of *C. elegans*, *Caenorhabditis briggsae*, *Hydra vulgaris*, and *Podocoryne carnea* with different functions, such as a matrix metalloproteinase, thyroid peroxidase, and acetylcholine receptor, and it is known as a toxin homology domain. Interestingly, the N-terminal region of At-P4H-2 contains a stretch of 24 amino acids in which the sequence identity to the rice and viral P4Hs is much higher than to At-P4H-1. If the plant P4Hs have a separate peptide-binding region, this sequence could explain some of the major differences found in the hydroxylation patterns of At-P4H-2 and the viral P4H relative to At-P4H-1.

P4Hs have been identified in species ranging from mammals to insects, nematodes, plants, and viruses and have been found to have multiple isoenzymes in almost all organisms studied (1–3, 14, 16). The vertebrate C-P4Hs hydroxylate only prolines in collagen-like XPG sequences, poly(L-proline) being a competitive inhibitor (1–3, 6, 7), whereas the three HIF-P4Hs act on LXXLAP sequences in HIFα (8, 9). The functions of 4-hydroxylproline in collagens and HIFα are distinctly different as the 4-hydroxylproline residues are required for the stability of collagen molecules, whereas in HIFα, the 4-hydroxyproline acts as signal for protein degradation (1–3, 31, 32). Plants and green algae have many different HRGPs that play important structural roles in their cell walls but are also involved in various aspects of plant growth, development, and differentiation and responses to stress (12, 13). 4-Hydroxyproline residues serve as sites for O-glycosylation in plant HRGPs (12), whereas no reports are available on the glycosylation of 4-hydroxyproline residues in animal proteins. According to a contiguity hypothesis, the extent of aromaticosylation of 4-hydroxyproline residues by small 4–6-residue glycosylated modules increases with 4-hydroxyproline contingency, clustered non-contiguous 4-hydroxyproline residues being sites for the addition of larger arabino-nogalactan polysaccharides (33).

The first characterized *A. thaliana* isoenzyme, At-P4H-1, is a highly unusual enzyme among the various P4Hs as it hydroxylates substrates varying from poly(L-proline) and plant HRGP-like peptides to the collagen model peptide (PGP)β<sub>10</sub> and HIFα-like peptides (15). Its substrate specificity also contradicted the previous suggestions that a requirement for the hydroxylation by the plant P4Hs may be a substrate in the poly(L-proline) type II helix conformation (34). Others have suggested that the hydroxylation of proline residues in HRGPs may depend on multiple, sequence-specific P4Hs rather than a single poly(L-proline) type II conformation-dependent enzyme (35). At-P4H-2 was found to differ distinctly from At-P4H-1 in its substrate specificity. All the peptides studied here were much better substrates, defined by a lower $K_m$, for At-P4H-1 than for At-P4H-2, and At-P4H-2 did not hydroxylate the HIFα-like peptides at all. At-P4H-2 resembled the viral PBCV-1 P4H (16) in that they both hydroxylated (PPG)β<sub>10</sub> only very inefficiently, but the viral P4H had $K_m$ values that were even much higher for poly(L-proline) than those for At-P4H-2. Unlike the C-P4Hs, At-P4H-2 failed to hydroxylate bradykinin at any significant rate, but it did hydroxylate a 17-residue peptide corresponding to the sequence around the hydroxylation site in the prion protein, although ineffectively and with a high $K_m$ (27). The prion peptide RYPQQGSPGGRNPQQG studied here has altogether 4 proline residues, 2 of them in XPG sequences and 2 consecutive ones in a PPQ sequence. The 4-hydroxyproline
identified in this peptide is located in the SPG triplet (27), but due to the low hydroxylation level, it was not possible to analyze which one(s) of the prolines became hydroxylated by At-P4H-2. At-P4H-1 preferentially hydroxylated the prolines preceding glycine in the (PPG)10 peptide, but the other prolines were also hydroxylated to a small extent (15).

The extensin and arabinoxylan-like peptides that acted as good substrates for At-P4H-2 all have at least 3 consecutive prolines, suggesting that this isoenzyme may require a stretch of at least 3 successive prolines for effective hydroxylation. (PPG)10, which has stretches of 2 consecutive prolines, was hydroxylated at a low rate, and it is possible that hydroxylation of the prion peptide also occurred in its PPQ triplet with 2 consecutive prolines, whereas (PAG)10, (APG)10, and the HIF-like and RNA polymerase II-like peptides, with single proline residues, were not hydroxylated at all. Bradykinin, in which the two consecutive prolines are located in the sequence RP-PGFSPFR very close to the N terminus of the peptide, was likewise not hydroxylated at all. N-terminal sequencing of the partially hydroxylated peptide SPPVVYSPPPVKHYSPPPV indicated that At-P4H-2 acted preferentially on the 3rd consecutive proline in the C-terminal PPP triplet, but a small amount of 4-hydroxylation was also found in the 1st proline of the central PPPP sequence. As this peptide was hydroxylated to a low extent also by the endogenous prolyl 4-hydroxylase(s) present in the insect cell extract (Table I), the possibility cannot be excluded that hydroxylation of the latter position was carried out by the insect cell enzyme(s) rather than At-P4H-2. Nevertheless, it is clear that the enzyme can also act on more than 3 consecutive prolines as poly(L-proline) was an effective substrate. The lack of hydroxylation of any proline in the N-terminal PPP triplet and the last proline in the central PPPP triplet may be due to a preferential hydroxylation in short peptide prolines in the C-terminal region. The cell types and developmental stages expressing At-P4H-1 and At-P4H-2 in vivo are currently unknown. It is therefore likewise unknown whether At-P4H-2 or At-P4H-1 has any major role in hydroxylating poly(L-proline)-like sequences in HRGPs in vivo in cases in which both At-P4H isoenzymes are expressed at the same time and in the same location as At-P4H-1 possesses much lower Km values for such peptides. No substrates with lower Km values for At-P4H-2 than for At-P4H-1 have yet been identified.

The Km values of At-P4H-2 for the reaction cova substrates iron, 2-oxoglutarate, and ascorbate were highly similar to those reported for At-P4H-1 (15). The most striking difference in these values between the vertebrate C-P4Hs and the two characterized A. thaliana enzymes is that the Km values of At-P4H-1 and At-P4H-2 for 2-oxoglutarate were 130 and 170 μM, respectively, being thus 6.5–8.5 times the value 20 μM of the type I C-P4H (17). The Km values of the three HIF-P4Hs for 2-oxoglutarate, 55–60 μM, are higher than that of C-P4H-1 but still only about one-third of the value of At-P4H-2 (11, 17). It was therefore of interest to study whether the two well characterized competitive 2-oxoglutarate analogue inhibitors of the C-P4Hs, pyridine-2,4-dicarboxylate and pyridine-2,5-dicarboxylate (1), would also inhibit At-P4H-2. It transpired that pyridine-2,4-dicarboxylate was an effective inhibitor, with a Ki of 10 μM, whereas pyridine-2,5-dicarboxylate had a Ki of 80 μM, although it is a more effective inhibitor of the C-P4Hs than the former (1). It thus seems that there are some distinct differences in the structure of the 2-oxoglutarate-binding site between At-P4H-2 and the C-P4Hs.

Acknowledgments—We thank Dr. Ulrich Bergmann, Biocenter Oulu, for the help with mass spectrometry analysis, and the expert technical assistance of Tanja Aantsinki, Raija Juntunen, Eeva Lehtimäki, Outi Mänty, Merja Nissilä, and Liisa Aijälä is gratefully acknowledged.

REFERENCES

1. Kivirikko, K. I., and Pihlajaniemi, T. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72, 325–398
2. Myllyharju, J. (2000) Matrix Biol. 22, 15–24
3. Myllyharju, J., and Kivirikko, K. I. (2004) Trends Genet. 20, 33–43
4. Helaakoski, T., Vuori, K., Myllyla, R., Kivirikko, K. I., and Pihlajaniemi, T. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4392–4396
5. Anzalone, F., Helaakoski, T., Myllyharju, J., Veijola, P., Pihlajaniemi, T., and Kivirikko, K. I. (1997) J. Biol. Chem. 272, 17342–17348
6. Kukkola, L., Hieta, R., Kivirikko, K. I., and Myllyharju, J. (2003) J. Biol. Chem. 278, 47685–47693
7. Vasudevan, C., Papay, B., Bolender, Z., Brown, A., and Pickering, J. G. (2003) Circulation 108, 508–511
8. Bruck, R. K., and McKnight, S. L. (2001) Science 294, 1337–1340
9. Epstein, A. C. R., Gleade, J. M., McNeill, L. A., Hewitson, K. S., O’ourke, J., Dole, M. R., Mukherjee, M., Metzen, E., Wilson, M. I., Dhand, A., Tian, Y.-M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2001) Cell 107, 43–54
10. Ivan, M., Haberberger, T., Gervasi, D. C., Michelson, K. S., Gunzler, V., Kondo, K., Yang, H., Sorekina, I., Conaway, R. C., Conaway, J. W., and Kaelin, W. G., Jr. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 13459–13464
11. Hisirka, M., Koivunen, P., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2003) J. Biol. Chem. 278, 30772–30780
12. Kieliszewski, M. J., and Shpak, E. (2001) CMLS Cell. Mol. Life Sci. 58, 1386–1398
13. Showalter, A. M. (2001) CMLS Cell. Mol. Life Sci. 58, 1399–1417
14. Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1992) In Post-Translation Modifications of Proteins (Harding, J. J., and Crabbe, M. J. C., eds) pp. 1–51, CRC Press, Inc., Boca Raton, FL
15. Hieta, R., and Myllyharju, J. (2002) J. Biol. Chem. 277, 23965–23971
16. Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999) J. Biol. Chem. 274, 22131–22134
17. Myllyharju, J., and Kivirikko, K. I. (1997) EMBO J. 16, 1173–1180
18. Myllyharju, J., and Kivirikko, K. I. (1999) EMBO J. 18, 306–312
19. Hieta, R., Koikka, L., Permi, P., Pirilä, P., Kivirikko, K. I., Kilpeläinen, I., and Myllyharju, J. (2003) J. Biol. Chem. 278, 34966–34974
20. Crossen, R., and Gruenwald, S. (1998) Bacculovirus Expression Vector System: Instruction Manual, Pharmingen, San Diego, CA