Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase

Annamari Vuorela, Johanna Myllyharju, Ritva Nissi, Taina Pihlajaniemi and Kari I. Kivirikko

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52 A, FIN-90220 Oulu, Finland

1Corresponding author

Prolyl 4-hydroxylase, the key enzyme of collagen synthesis, is an $\alpha_2\beta_2$ tetramer, the $\beta$ subunit of which is protein disulfide isomerase (PDI). Coexpression of the human $\alpha$ subunit and PDI in *Pichia* produced trace amounts of an active tetramer. A much higher, although still low, assembly level was obtained using a *Sacharomyces* pre-pro sequence in PDI. Coexpression with human type III procollagen unexpectedly increased the assembly level 10-fold, with no increase in the total amounts of the subunits. The recombinant enzyme was active not only in *Pichia* extracts but also inside the yeast cell, indicating that *Pichia* must have a system for transporting all the cosubstrates needed by the enzyme into the lumen of the endoplasmic reticulum. The 4-hydroxyproline-containing procollagen polypeptide chains were of full length and formed molecules with stable triple helices even though *Pichia* probably has no Hsp47-like protein. The data indicate that collagen synthesis in *Pichia*, and probably also in other cells, involves a highly unusual control mechanism, in that production of a stable prolyl 4-hydroxylase requires collagen expression while assembly of a stable collagen requires enzyme expression. This *Pichia* system seems ideal for the high-level production of various recombinant collagens for numerous scientific and medical purposes.

**Keywords:** chaperone/collagen/expression/prolyl 4-hydroxylase/yeast

Introduction

The collagens are a family of extracellular matrix proteins that contain the repeating triplet sequence -Gly-X-Y- in which the Y position amino acid is often 4-hydroxyproline, and have the potential for three polypeptide chains with such sequences to fold into triple-helical molecules. At least 19 proteins representing more than 30 gene products have now been defined as collagens, and more than 10 additional proteins have collagen-like domains. Many of the recently discovered collagens are present in tissues in such small amounts that they have so far been characterized mainly or only at the cDNA rather than the protein level (for reviews on collagens, see Kiely *et al.*, 1993; Mayne and Brewton, 1993; Pihlajaniemi and Rehn, 1995; Prockop and Kivirikko, 1995). It is thus obvious that an efficient expression system for recombinant collagens would have numerous applications. Nevertheless, most recombinant systems now available for the production of proteins cannot be used as such for the expression of recombinant collagens, as bacteria and yeast have no prolyl 4-hydroxylase activity and insect cells have insufficient levels of it (Lamberg *et al.*, 1996).

Prolyl 4-hydroxylase (EC 1.14.11.2) plays a central role in the synthesis of all collagens, as 4-hydroxyproline residues are essential for the formation of molecules with stable triple helices. The vertebrate enzyme is an $\alpha_2\beta_2$ tetramer (see Kivirikko *et al.*, 1989, 1992; Helaakoski *et al.*, 1995), the $\beta$ subunit of which is identical to another enzyme, protein disulfide isomerase (PDI, EC 5.3.4.1) (Koivu *et al.*, 1987; Pihlajaniemi *et al.*, 1987). The role of PDI in the enzyme is not related to its disulfide isomerase activity, as a double mutant PDI polypeptide in which both -Cys-Gly-His-Cys- catalytic sites (Vuori *et al.*, 1992a; Freedman *et al.*, 1994) have been inactivated to -Ser-Gly-His-Cys- forms a fully active prolyl 4-hydroxylase tetramer (Vuori *et al.*, 1992c). The main function of the PDI polypeptide in prolyl 4-hydroxylase appears to be to keep the highly insoluble $\alpha$ subunits in a catalytically active, non-aggregated conformation, a function that resembles the role of certain chaperones such as Hsp90 in other proteins (Vuori *et al.*, 1992b; John *et al.*, 1993; Freedman *et al.*, 1994). All attempts to construct an active prolyl 4-hydroxylase tetramer from its subunits in *vitro* have been unsuccessful, but an active recombinant tetramer has been obtained by coinfection with two baculoviruses in insect cells (Vuori *et al.*, 1992b) or cotransfection in COS-1 cells (John and Bulleid, 1996). Tetramer assembly probably requires molecular chaperones, one such chaperone being BiP (John and Bulleid, 1996; Veijola *et al.*, 1996b). Prolyl 4-hydroxylase requires Fe$^{2+}$, 2-oxoglutarate, O$_2$ and ascorbate, and an active system appears to exist in vertebrate cells for the transport of 2-oxoglutarate and ascorbate into the lumen of the endoplasmic reticulum (see Kivirikko *et al.*, 1989, 1992).

Yeast would seem to be an ideal system for studying the various assembly steps in collagen synthesis and for the recombinant production of collagens for various scientific and medical purposes. Yeast has no prolyl 4-hydroxylase or collagen synthesis, and thus provides a system in which endogenous prolyl 4-hydroxylase plays no role and in which the recombinant collagen is not contaminated by any non-recombinant collagen. Nevertheless, studies were required to determine whether yeast would contain the appropriate chaperones and other conditions required for the assembly of an active recombinant
prolyl 4-hydroxylase tetramer and whether a system of transport into the lumen of the yeast endoplasmic reticulum would be available for all the cosubstrates required by the enzyme, and thus whether the enzyme would be active inside the yeast cell. Studies were also required to confirm whether yeast could synthesize the long collagen polypeptide chains with the repetitive \(-\text{Gly-X-Y}\)- sequences and whether such chains would form triple-helical molecules in yeast. This aspect was of particular interest as Hsp47 is regarded as a collagen-specific chaperone (Nagata, 1996; Satoh et al., 1996) has indicated that \textit{Saccharomyces cerevisiae} has no Hsp47-like sequences. We report here on the expression and assembly of the subunits of human prolyl 4-hydroxylase and the pro\(\text{z}\)1 chains of human type III procollagen in the yeast \textit{Pichia pastoris}. In the course of this work we made the unexpected finding that the production of a stable prolyl 4-hydroxylase tetramer requires the expression of collagen polypeptide chains. This appears to constitute a novel control mechanism that may not be limited to recombinant expression in \textit{Pichia} but may also exist in a number of other cell types.

### Results

**Coexpression of the human prolyl 4-hydroxylase \(\alpha\) subunit and PDI polypeptide in \textit{P pastoris} produces a small amount of an active enzyme tetramer**

The \(\text{his}4\), \(\text{arg}4\) \textit{P pastoris} host strain, which has defects in enzymes required for the synthesis of histidine and arginine, was used in these studies. In order to study whether the recombinant human prolyl 4-hydroxylase \(\alpha\) subunit and PDI polypeptide are able to form an active enzyme tetramer in yeast cells, cDNAs for the human prolyl 4-hydroxylase \(\alpha\) subunit and PDI polypeptide were cloned into the \textit{Pichia} expression vectors \textit{pARG815} (complementing for \(\text{arg}4\) in the host) and \textit{pAO815} (complementing for \(\text{his}4\) in the host), respectively (Table I). An additional expression vector, \textit{pARG815\(\alpha\)PDI}, coding for both the \(\alpha\) subunit and PDI polypeptide was also generated. The \(\text{his}4\), \(\text{arg}4\) strain was then transformed separately with \textit{pARG815\(\alpha\)} and \textit{pAO815PDI}, or cotransformed with either \textit{pARG815\(\alpha\)} and \textit{pAO815PDI} or \textit{pARG815\(\alpha\)PDI} and \textit{pAO815PDI}. \(\text{Arg}^+\), methanol utilizing (Mut\(^+\)) transformants of the \(\alpha\) subunit strain, \(\text{His}^+\), Mut\(^+\) transformants of the \(\alpha\) subunit strain and \(\text{Arg}^+\), \(\text{His}^+\), Mut\(^+\) cotransformants of the \(\alpha\)PDI and \(\alpha\)PDI/PDI strains were selected. The presence of one copy of each transformed DNA in the \(\alpha\)PDI and \(\alpha\)PDI/PDI strains was verified by dot blot analysis (data not shown). Cells were cultured in BMGY medium supplemented with yeast extract and peptone, and expression was induced in BMM medium, methanol being added every 24 h to a final concentration of 0.5%. The cells were harvested 60 h after induction, broken in a buffer containing Triton X-100 and centrifuged. Samples of the soluble and insoluble fractions of the cell lysates were then analysed by SDS–PAGE and non-denaturing PAGE followed by Western blotting with a polyclonal antibody to the human PDI polypeptide, or a monoclonal or polyclonal antibody to the human prolyl 4-hydroxylase \(\alpha\) subunit (Figure 1). A band corresponding to the PDI polypeptide was found in the soluble fraction of cell lysates from \(\alpha\)PDI and \(\alpha\)PDI/PDI strains (Figure 1A, lanes 2 and 3). The prolyl 4-hydroxylase \(\alpha\) subunit could not be detected in the soluble fractions (Figure 1B, lanes 2 and 3), whereas bands corresponding to it and its degradation products were found after treating the insoluble fractions with 1% SDS (Figure 1C, lanes 2 and 3). In non-denaturing PAGE a faint band corresponding to the enzyme tetramer was seen only in the soluble extract from the \(\alpha\)PDI/PDI strain, which has two copies of the PDI DNA (Figure 2, lanes 2 and 3).

Prolyl 4-hydroxylase activities of the Triton X-100 soluble extracts were assayed by two methods, one based on the hydroxylation-coupled decarboxylation of 2-oxo[1\(\text{\textsuperscript{14}C}\)]glutarate with a synthetic peptide substrate and the other on the formation of 4-hydroxy[1\(\text{\textsuperscript{14}C}\)]proline in a [\(\text{\textsuperscript{14}C}\)]proline-labelled procollagen substrate. The enzyme activities in the soluble extracts from the untransformed \(\text{his}4\), \(\text{arg}4\) host strain before and after methanol induction and from the strains expressing only the \(\alpha\) subunit or the PDI polypeptide never exceeded the background values. The enzyme activity level in the extract from the \(\alpha\)PDI/PDI strain was very low and could only be detected by the more sensitive method using the [\(\text{\textsuperscript{14}C}\)]proline-labelled procollagen as a substrate (Table II). Prolyl 4-hydroxylase activity was four times higher in the extract from the \(\alpha\)PDI/PDI strain than in that from the \(\alpha\)PDI strain, but was still less than 200 d.p.m./200 \(\mu\)g soluble protein when the method based on the hydroxylation-
PDI secretion both prolyl 4-hydroxylase tetramer assembly and with a yeast pre-pro sequence markedly increases et al. magnitude lower than the activity level obtained in insect cells by coinfection with two separate baculoviruses (Vuori et al., 1992b, 1994). The positions of the variably glycosylated forms of the α subunit are indicated by an arrow. (C) As (A), but insoluble fractions of the lysates were analysed with a polyclonal antibody K17 to the human α subunit. The positions of the variably glycosylated forms of the α subunit are indicated by an arrow. Bands corresponding to various degradation products of the α subunit are also seen.

Replacement of the signal peptide of human PDI with a yeast pre-pro sequence markedly increases both prolyl 4-hydroxylase tetramer assembly and PDI secretion

As the assembly of an active prolyl 4-hydroxylase tetramer from the inactive monomers was poor in *Pichia*, the effect of replacing the signal sequences of the α subunit and PDI polypeptide with yeast signal sequences was studied. The *Pichia* cotransformants (Table I) α/PDI-PHO1 (in which PHO1 is the *Ppastoris* acid phosphatase 1 signal sequence), α/PDI-αMF and α/PHO1/PDI-αMF (αMF is the *S.cerevisiae* α mating factor pre-pro sequence) were cultured, induced and harvested as above. The presence of only one copy of the α subunit and PDI DNAs was verified by dot blot analysis. Samples of the soluble and insoluble fractions of cell lysates were analysed by SDS-PAGE as above. Bands corresponding to the PDI polypeptide were again seen in the extracts from all three strains (Figure 1A, lanes 4–6). Faint bands corresponding to the α subunit were seen in the soluble fractions from the α/PDI-PHO1 and α/PHO1/PDI-αMF strains (Figure 1B, lanes 4 and 6), whereas slightly stronger α subunit bands were seen in the soluble fraction from the α/PDI-αMF strain (Figure 1B, lane 5). Nevertheless, strong bands corresponding to the α subunit and its degradation products were seen in the 1% SDS extract from the insoluble fraction in all three strains (Figure 1C, lanes 4–6). Nondenaturing PAGE analysis of the soluble extracts from all three strains gave a distinct band that could be immunostained with the polyclonal α subunit antibody and corresponded to the prolyl 4-hydroxylase tetramer (as shown for the α/PDI-αMF strain in Figure 2, lane 4).

Soluble extracts from the α/PDI-PHO1, α/PDI-αMF and α-PHO1/PDI-αMF strains were analysed for prolyl 4-hydroxylase activity with the assay based on the hydroxylation-coupled decarboxylation of 2-oxo[1-14C]glutarate. In accordance with the strongest prolyl 4-hydroxylase tetramer band observed in non-denaturing PAGE, the highest prolyl 4-hydroxylase activity level, ~2600 d.p.m./200 μg, was found in the extract from the α/PDI-αMF strain, while the levels in the extracts from the α/PDI-PHO1 and α-PHO1/PDI-αMF strains were distinctly lower, ~1100 d.p.m./200 μg and 300 d.p.m./200 μg, respectively (Table II). Thus the best signal sequences for prolyl 4-hydroxylase tetramer production among those studied here were the authentic one in the case of the α subunit and the *S.cerevisiae* αMF pre-pro sequence in the case of the PDI polypeptide.
Table II. Enzyme activity levels in various *Pichia* strains expressing prolyl 4-hydroxylase

<table>
<thead>
<tr>
<th>Strain</th>
<th>Protocollagen assay (d.p.m./200 μg)a</th>
<th>2-Oxoglutarate assayb</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>his</em>, <em>arg</em> (uninduced)</td>
<td>&lt;20</td>
<td>&lt;50</td>
</tr>
<tr>
<td><em>his</em>, <em>arg</em> (induced)</td>
<td>&lt;20</td>
<td>&lt;50</td>
</tr>
<tr>
<td>αPDI</td>
<td>&lt;20</td>
<td>&lt;50</td>
</tr>
<tr>
<td>αPDI/PDI</td>
<td>160</td>
<td>~200</td>
</tr>
<tr>
<td>αPDI-PHO1</td>
<td>n.d.c</td>
<td>1060 ± 180</td>
</tr>
<tr>
<td>αPDI-αMF</td>
<td>n.d.d</td>
<td>2600 ± 350</td>
</tr>
<tr>
<td>α-PHO1/PDI-αMF</td>
<td>n.d.e</td>
<td>320 ± 70</td>
</tr>
</tbody>
</table>

aAbout 200 μg of extractable cell protein was used in each assay, and therefore the values are given as d.p.m./200 μg of protein. The values obtained with the 2-oxoglutarate assay are given as mean ± SD of three to seven independent experiments.

bThe assay based on the formation of hydroxy[1-14C]proline in a [14C]proline-labelled protocollagen substrate is more sensitive than that based on the hydroxylation-coupled decarboxylation of 2-oxo[1-14C]glutarate.

cn.d. = not determined.

Fig. 3. Secretion of PDI in *P. pastoris*. The αPDI/PDI (lane 1), αPDI-PHO1 (lanes 2 and 3), αPDI-αMF (lane 4) and αPDI-αMF-HDEL (lane 5) strains were cultured and induced as in Figure 1. The cell culture medium was harvested 60 h after induction, and unconcentrated (lanes 1, 2, 4, and 5) or concentrated (lane 3) samples were analysed by 10% SDS–PAGE followed by Western blotting using a polyclonal antibody K38 to human PDI. Strains are identified at the top, and the PDI polypeptide is indicated by an arrow.

Fig. 4. Analysis of purified recombinant human prolyl 4-hydroxylase tetramers. (A) Recombinant human prolyl 4-hydroxylases from *Sf*9 insect cells (lane 1) and the αPDI-PHO1 (lane 2) and αPDI-αMF (lane 3) *Pichia* strains were affinity-column purified, and samples analysed by 8% PAGE under non-denaturing conditions followed by Coomassie staining. The enzymes are identified at the top, and the prolyl 4-hydroxylase tetramer is indicated by the arrow αβ. (B) Purified recombinant human prolyl 4-hydroxylases from *Sf*9 insect cells (lane 1) and the αPDI-αMF (lane 2) *Pichia* strain were analysed by 8% SDS–PAGE under reducing conditions followed by Coomassie staining. The enzymes are identified at the top, and the locations of the PDI polypeptide, the PDI polypeptide containing the αMF propeptide (αMFpro) and the various glycosylated forms of the α subunit are indicated by arrows.

Secretion of prolyl 4-hydroxylase and its subunits into the culture medium was studied by analysing samples of the minimal medium by SDS–PAGE followed by Coomassie staining or Western blotting using antibodies to the α subunit and the PDI polypeptide. The α subunit could never be detected in any cell culture medium sample (data not shown). The PDI polypeptide band was seen in unconcentrated medium samples only in the case of strains expressing PDI with the αMF pre-pro sequence (Figure 3, lanes 1, 2 and 4). The amount of PDI in the culture medium, based on Coomassie staining, was ~10 mg/l (data not shown). No differences were found in the cellular PDI levels or PDI secretion levels between strains having a KDEL or HDEL endoplasmic reticulum retention signal (Pelham, 1990) in the PDI-αMF polypeptide (Figure 3, lanes 4 and 5). PDI with the PHO1 signal sequence was much less efficiently secreted and could only be detected in Western blots of medium samples concentrated 30 times (Figure 3, lane 3), while PDI with its own signal sequence was never detected even in the concentrated samples. Determination of the N-terminal amino acid sequence of PDI partially purified from the culture medium of the αPDI-αMF strain demonstrated that the αMF pre-pro sequence had been cleaved as expected, resulting in an N-terminal sequence Tyr-Val-Glu-Phe-Asp-Ala-, in which the first four amino acids originate from the pHIL-S1 plasmid and the next two correspond to the N terminus of human PDI.

Characterization of human prolyl 4-hydroxylase tetramers expressed in *Pichia*

The αPDI-PHO1 and αPDI-αMF strains were cultured, induced and broken as above, and prolyl 4-hydroxylase was purified using a polyl-L-proline affinity column and gel filtration. Non-denaturing PAGE analysed by Coomassie staining showed no differences in migration between the recombinant human prolyl 4-hydroxylases purified from the αPDI-PHO1 *Pichia* strain and from the *Spodoptera frugiperda* (*Sf*9) insect cells (Figure 4A, lanes 1 and 2). N-terminal sequencing of the PDI polypeptide of the purified enzyme tetramer showed that the PHO1 signal peptide had been correctly cleaved, resulting in an N-terminal sequence Arg-Glu-Phe-Asp-Ala-, in which the first three amino acids originate from the pPIC9 plasmid, in which the first three amino acids originate from the pPIC9 plasmid.

PDI-4 hydroxylase purified from the αPDI-αMF strain migrated in non-denaturing PAGE as a broad band with a slightly slower mobility than the enzyme from the αPDI-PHO1 strain (Figure 4A, lane 3). In Coomassie stained SDS–PAGE several bands corresponding to polypeptides of ~70 kDa were seen in addition to the α subunit and PDI polypeptide bands (Figure 4B, lane 2). These were stained by the PDI antibody in Western blotting in addition to the 60 kDa PDI polypeptide band (Figure 5,
had expressed the highest level of prolyl 4-hydroxylase activity. Cells were cultured, induced, harvested and broken as above. To make sure that the cells still expressed active prolyl 4-hydroxylase, the enzyme activity was measured in the soluble fraction of the cell lysates using the assay based on the hydroxylation-coupled decarboxylation of 2-oxo[1-14C]glutarate. Other aliquots of this fraction and samples of the cell culture medium were analysed with a radioimmunoassay for the N-propeptide of human type III procollagen.

The prolyl 4-hydroxylase activity level in all the transformants but one was ~10 times that in the original α/PDI-αMF strain (Table III). Subsequent radioimmunoassays for the N-propeptide of type III procollagen indicated that all the transformants but one expressed essentially identical amounts of type III procollagen (Table III). The transformant that expressed no type III procollagen was the same one that showed no increase in prolyl 4-hydroxylase activity level. This highly surprising finding has now been confirmed in four subsequent experiments.

The presence of only one copy each of the prolyl 4-hydroxylase subunit, PDI and proc1(III) DNAs in all the α/PDI-αMF/proc1(III) transformants was verified by dot blot analysis (data not shown). No differences were found in the mRNA levels for the α subunit or the PDI polypeptide between the original α/PDI-αMF strain and any of the α/PDI-αMF/proc1(III) transformants when studied by Northern blotting (data not shown).

The soluble fraction and medium samples from the α/PDI-αMF/proc1(III) strain were compared with those of the α/PDI-αMF strain in SDS–PAGE followed by Western blotting with the polyclonal PDI antibody or the monoclonal α subunit antibody. Essentially identical amounts of the PDI polypeptide were seen in the untreated sample (Figure 5, lane 4), indicating that the multiple bands were of ~70 kDa remained after the pre-propeptide had been correctly cleaved while the propeptide was still attached to PDI. A single additional band of ~70 kDa remained after N-glycosidase F treatment (Figure 5, lane 2), indicating that the multiple bands were due to differentially glycosylated forms of the αMF propeptide. The human PDI polypeptide itself has no N-glycosylation sites (Pihlajaniemi et al., 1987), and no heterogeneity was seen in the PDI polypeptide of the enzyme purified from the α/PDI-PHO1 strain (data not shown).

Western blotting with the polyclonal α subunit antibody indicated that three forms of the α subunit are present in prolyl 4-hydroxylases purified from the α/PDI-PHO1 (data not shown) and α/PDI-αMF strains (Figure 5, lane 3). N-glycosidase F treatment produced a single polypeptide with a mobility similar to that of the lowest of the three bands present in the untreated sample (Figure 5, lane 4), indicating the presence of two differentially glycosylated forms of the α subunit. This finding agrees with the presence of two potential N-glycosylation sites in the human α subunit (Helaakoski et al., 1989). N-terminal sequencing of the α subunit of the enzyme purified from the α/PDI-αMF strain showed that the signal peptide had been correctly cleaved.

The specific activities of the recombinant enzymes purified from the α/PDI-PHO1 and α/PDI-αMF strains and from insect cells were essentially identical. The $K_m$ values for Fe$^{2+}$, 2-oxoglutarate, ascorbate and the peptide substrate were likewise identical (data not shown).

**Coexpression of collagen polypeptide chains markedly increases the amount of poly 4-hydroxylase tetramer**

In order to study whether it is possible to produce recombinant human type III procollagen with a stable triple helix in *Pichia*, a construct termed pPICZ Bproc1(III) was transformed into the α/PDI-αMF strain (Table I), which

**Table III. Prolyl 4-hydroxylase activity and expression of type III procollagen in various α/PDI-αMF/proc1(III) *Pichia* transformants**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Prolyl 4-hydroxylase activity (d.p.m./200 μg)</th>
<th>Expression of type III procollagen (ng/200 μg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>α/PDI-αMF</td>
<td>2 300</td>
<td>0</td>
</tr>
<tr>
<td>α/PDI-αMF/proc1(III) 1</td>
<td>20 920</td>
<td>380</td>
</tr>
<tr>
<td>α/PDI-αMF/proc1(III) 2</td>
<td>19 360</td>
<td>430</td>
</tr>
<tr>
<td>α/PDI-αMF/proc1(III) 3</td>
<td>2 690</td>
<td>0</td>
</tr>
<tr>
<td>α/PDI-αMF/proc1(III) 4</td>
<td>22 100</td>
<td>340</td>
</tr>
<tr>
<td>α/PDI-αMF/proc1(III) 5</td>
<td>22 410</td>
<td>410</td>
</tr>
</tbody>
</table>

*aExpression of type III procollagen was measured by a radioimmunoassay for the N-propeptide of human type III procollagen.

*bValues are given as d.p.m./200 μg of extractable cell protein."
Assembley of prolyl 4-hydroxylase and collagen in Pichia

Fig. 6. Coexpression of polypeptide chains of type III procollagen increases the amount of enzyme tetramer. The α/PDI-αMF (lanes 1, 3, 5, 7 and 9) and α/PDI-αMF/proα1(III) (lanes 2, 4, 6, 8 and 10) strains were cultured, induced and harvested as in Figure 1. Samples of the soluble (lanes 1–4) and insoluble (lanes 5 and 6) fractions of the cell lysates and the cell culture medium (lanes 7 and 8) were analysed by 10% SDS–PAGE followed by Western blotting with a polyclonal PDI antibody K38 (lanes 1, 2, 7 and 8) or a monoclonal α subunit antibody M95L (lanes 3–6). Other aliquots of the soluble fractions were analysed by 8% non-denaturing PAGE followed by Western blotting with a polyclonal α subunit antibody K17 (lanes 9 and 10). Locations of the PDI-αMFpro and the PDI polypeptides, the α subunit and the prolyl 4-hydroxylase tetramer (αβ₂) are indicated by arrows.

products in the insoluble fraction from the α/PDI-αMF/proα1(III) strain was smaller than that in the insoluble fraction from the α/PDI-αMF strain (Figure 6, lanes 5 and 6).

Collagen polypeptide chains coexpressed in Pichia with prolyl 4-hydroxylase form molecules with stable triple helices

α/PDI-αMF/proα1(III) and Proα1(III) strains (Table I) were cultured and induced as above, ascorbate being added to the culture medium every 12 h during the induction. Cells were harvested 60 h after induction and broken in a 5% glycerol, 1 mM EDTA and 50 mM sodium phosphate buffer, pH 7.4. Samples of the soluble fraction of the cell lysates and the cell culture medium were analysed by means of a radioimmunoassay for the N-propeptide of type III procollagen. In both strains the majority of the recombinant human type III procollagen was present in the soluble fraction of the cell lysates, only 10–20% being found in the culture medium (data not shown).

Samples of the soluble fraction of the cell lysates were analysed by SDS–PAGE under reducing and non-reducing conditions followed by Coomassie staining or Western blotting with an antibody to the N-propeptide of type III procollagen, and aliquots of the samples were also studied after digestion with pepsin for 2 h at 22°C. The triple helix of collagen is resistant to pepsin while non-triple-helical proα1(III) chains and the propeptides of triple-helical procollagen molecules are digested. Only one band corresponding to full-length proα1(III) chains was seen in immunoblots of non-panepsinized samples from both strains when analysed by SDS–PAGE under reducing conditions (as shown for the Proα1(III) strain in Figure 7, lane 1). Essentially all the proα1(III) chains were found as disulfide-bonded trimers when analysed by SDS–PAGE under non-reducing conditions (as shown for the Proα1(III) strain in Figure 7, lane 2). Pepsin-resistant α chains of type III collagen were seen in the digested samples from both strains when analysed by SDS–PAGE under reducing conditions (Figure 7, lanes 3 and 5). However, a substantial fraction of the α1(III) chains from the Proα1(III) strain was digested by pepsin to smaller products and these α1(III) chains could only be detected by Western blotting with a monoclonal antibody that recognizes the collagenous regions of various collagen chains (Figure 7, lane 3), whereas essentially all the α1(III) chains from the α/PDI-αMF/proα1(III) strain were of full length and could readily be seen by Coomassie staining (Figure 7, lane 5). Moreover, even the full-length α1(III) chains from the Proα1(III) strain were not disulfide-bonded (Figure 7, lane 4), whereas all the α1(III) chains from the α/PDI-αMF/proα1(III) strain were present as disulfide-bonded trimers when analysed by SDS–PAGE under non-reducing conditions (Figure 7, lane 6).

The thermal stability of the pepsin-resistant type III collagens was studied using digestion with a mixture of trypsin and chymotrypsin after heating to various temperatures (Bruckner and Prockop, 1981). The recombinant type III collagen produced in the Proα1(III) strain was completely digested at 27°C (Figure 7, lane 7). In agreement with this result, no 4-hydroxyproline was detected in amino acid analysis of the recombinant collagen purified from this strain (data not shown). The thermal stability of the recombinant type III collagen produced in the α/PDI-αMF/proα1(III) strain was much higher, the Tm being ~39°C (Figure 8). The amino acid composition of the recombinant collagen purified from this strain was essentially identical to that of the non-recombinant human type III collagen (data not shown), except that the degree of 4-hydroxylation of the incorporated proline residues (i.e. the percentage of 4-Hyp of the sum of 4-Hyp + Pro) was 44.2%, while the corresponding value for the recombinant human type III collagen produced in insect cells, which had a Tm identical to that of the non-recombinant human type III collagen, was 50.1% (Lambert et al., 1996) and that for the non-recombinant human type III collagen determined here (Chemicon) was 51.6%. The only other significant difference between the recombinant collagen and the corresponding non-recombinant protein was that the Pichia-derived collagen contained no hydroxylsine while the latter has five hydroxlysine residues per 1000 amino acids (Chung and Miller, 1974).

The level of expression of the recombinant type III procollagen in the α/PDI-αMF/proα1(III) strain, as estimated from the amount of purified type III collagen obtained
A.Vuorela et al.

Fig. 7. Expression of recombinant human type III procollagen in *P. pastoris*. The transformant strains Proα1(III) and α/PDI-αMF/proα1(III) were cultured and induced as in Figure 1, ascorbate being added every 12 h during the induction. Cells were harvested 60 h after induction and broken in a glycerol-phosphate buffer. Aliquots of the soluble fraction of the cell lysates were digested first with pepsin and then with a mixture of trypsin and chymotrypsin. Samples were electrophoresed on 5% SDS–PAGE under reducing (lanes 1, 3, 5 and 7) or non-reducing (lanes 2, 4 and 6) conditions. Undigested samples were analysed by Western blotting with an antibody to the N-propeptide of type III procollagen (lanes 1 and 2). The pepsin-digested samples from the Proα1(III) strain were analysed by Western blotting with a monoclonal collagen antibody 95D1A (lanes 3 and 4). The pepsin-digested samples from the α/PDI-αMF/proα1(III) strain (lanes 5 and 6) and the sample from the Proα1(III) strain digested with a mixture of trypsin and chymotrypsin at 27°C (lane 7) were analysed by Coomassie staining. Strains are identified at the top, and the positions of the trimeric [proα1(III)]₃ and [α1(III)]₃ and the monomeric proα1(III) and α1(III) chains are shown by arrows. The mobilities of these chains were identical to those of the corresponding chains expressed in insect cells.

and the recovery of the purification, was ~15 mg/l (data not shown).

**Discussion**

The present data indicate that coexpression of the α subunit of human prolyl 4-hydroxylase and the human PDI polypeptide in *P. pastoris* produces a fully active recombinant enzyme tetramer. It is thus obvious that *Pichia* contains appropriate chaperones and other conditions required for the assembly of an active prolyl 4-hydroxylase. The recombinant enzyme was active not only in *Pichia* extracts but also inside the yeast cell, indicating that *Pichia* must have an active or passive system for transporting all the cosubstrates needed by the enzyme into the lumen of the endoplasmic reticulum. No prolyl 4-hydroxylase activity was produced when the α subunit was expressed alone, indicating that the human α subunit does not form an active tetramer with the *Pichia* PDI. This result differs from the situation in insect cells, in which the α subunit, when expressed alone, produces significant amounts of enzyme activity due to tetramer formation with the endogenous insect cell PDI (Vuori et al., 1992b; Veijola et al., 1994). *Pichia* would thus seem to be more suitable than insect cells for detailed studies of the mechanisms involved in prolyl 4-hydroxylase assembly, as tetramer formation with the endogenous PDI polypeptide complicates interpretation of some of the data obtained in insect cells. The *Pichia* PDI thus appears to differ distinctly from the insect cell and *Caenorhabditis elegans* (Veijola et al., 1996a) PDI polypeptides, which do form an active prolyl 4-hydroxylase tetramer with the human α subunit.

Fig. 8. Analysis of the thermal stability of the recombinant human type III collagen expressed in the α/PDI-αMF/proα1(III) *Pichia* strain. Pepsin-digested samples were treated with a mixture of trypsin and chymotrypsin at the temperatures indicated at the bottom. Samples were electrophoresed on 8% SDS–PAGE under reducing conditions and analysed by Coomassie staining. The arrow shows the position of the α1(III) chain.

The signal sequence of PDI was found to play a major role in prolyl 4-hydroxylase assembly in *Pichia*. The authentic human signal sequence appeared to be particularly ineffective for the transport of PDI into the lumen of the endoplasmic reticulum, as only trace amounts of tetramer were produced even when the α subunit DNA was used together with two copies of the PDI DNA. A much higher tetramer assembly level was obtained by using the *P. pastoris* acid phosphatase 1 signal sequence, but even this level was only ~40% of that obtained with the *S. cerevisiae* α mating factor pre-pro sequence. The latter gave the highest amount of tetramer among the various constructs studied here even though this sequence also markedly increased the secretion of PDI into the culture medium. The tetramer produced with this construct contained both molecules in which the propeptide was retained and others in which the propeptide had been
cleaved at the correct site. The presence of the latter molecules suggests that either the PDI polypeptide which subsequently associated with the α subunit or the tetramer had been retrieved from the secretory pathway after the point at which the propeptide had been cleaved. The presence in the tetramer of a PDI polypeptide containing the propeptide agrees with recent data indicating that the N terminus of PDI is not involved in prolyl 4-hydroxylase assembly, as a PDI polypeptide containing a poly-histidine tag in its N terminus forms an enzyme tetramer as efficiently as the wild-type polypeptide (Annumen et al., 1997).

A highly surprising finding to emerge here was that coexpression of collagen polypeptide chains markedly increased the amount of active prolyl 4-hydroxylase tetramer. Northern analyses indicated that this increase was not due to increased transcription of the respective genes or to increased stability of the mRNAs, as the levels of the α subunit and PDI mRNAs were not increased. The amount of PDI polypeptide was likewise not increased. The amount of α subunit in the soluble fraction, i.e. in the soluble enzyme tetramer, was markedly increased, whereas the amount in the insoluble fraction was decreased. The latter decrease cannot be quantified exactly, as more than 95% of the total α subunit was present in the insoluble fraction in the absence of collagen synthesis and more than 50% was insoluble even in its presence. Nevertheless, the data appear to exclude any increase in the total amount of α subunit by more than ~20–30%, a value that is insignificant compared with the ~1000% increase in the amount of enzyme tetramer. It thus seems that the main reason, and probably the only reason, for the ~10-fold increase in the amount of active enzyme tetramer was an ~10-fold increase in the level of association of enzyme subunits.

It has been demonstrated previously that an α subunit present without the PDI polypeptide forms insoluble, inactive aggregates (Kivirikko et al., 1989, 1992; Vuori et al., 1992b). The α subunit may form soluble complexes with BiP (John and Bulleid, 1996; Veijola et al., 1996b), but they have no prolyl 4-hydroxylase activity (Veijola et al., 1996b). It thus seems unlikely that an unassembled α subunit would recognize collagen expression or that collagen expression would influence subunit assembly. A much more likely mechanism is that the α subunit and PDI polypeptide may effectively assemble to form the tetramer, but in the absence of collagen synthesis this tetramer, which has no covalent interchain bonds (Kivirikko et al., 1992), will rapidly dissociate back to its subunits and the α subunit will form insoluble aggregates. The enzyme tetramer is quite stable in the presence of conditions suboptimal for protein production in bioreactors (Romanos et al., 1992; Cregg et al., 1993). As the Km of O2 in the prolyl 4-hydroxylase reaction is as high as ~40 μM (Tuderman et al., 1977), the O2 concentration within the lumen of the endoplasmic reticulum is also likely to be rate-limiting for hydroxylation under shaker-flask conditions. Thus the minor differences
in 4-hydroxyproline content and $T_m$ between the recombinant and non-recombinant collagens are likely to disappear when the protein is produced in a bioreactor. It has further been demonstrated previously that the levels of expression of various proteins in *Pichia* increase markedly with the number of DNA copies at least up to 30–50 copies (Buckholz and Gleeson, 1991; Romanos et al., 1992; Scorer et al., 1994). It would thus seem to be easy to optimize the present system under bioreactor conditions and by using high-copy integrants for the production of very large amounts of various collagens.

In conclusion, the present data indicate that it is possible to produce a fully active recombinant human prolyl 4-hydroxyxylase tetramer and collagen molecules with stable triple helices in *P. pastoris*. *Pichia* would thus seem to provide an ideal system for detailed studies of the mechanisms involved in the assembly of the enzyme tetramer and the collagen triple helix, and of the possible role of Hsp47 in the latter. In addition, *Pichia* would seem to be an ideal system for the high-level production of various recombinant collagen types for numerous scientific and medical purposes. The data further indicate that a highly unusual control system exists in collagen synthesis in *Pichia*, and probably also in other cell types, in that production of a stable prolyl 4-hydroxylase tetramer requires expression of collagen polypeptide chains while the production of collagen molecules with stable triple helices requires expression of active prolyl 4-hydroxylase.

### Materials and methods

**Construction of expression vectors**

A cDNA for the human PDI polypeptide (Pihlajaniemi et al., 1987) extending from the translation initiation codon to the stop codon and flanked by EcoRI restriction sites was synthesized using PCR and ligated into the EcoRI site of pAO815 (Invitrogen), generating pAO815PDI. A modified *Pichia* expression vector pARG815 was constructed by replacing the HIS4 selection marker with an *S. cerevisiae* ARG4 selection marker. pYM25 (a gift from Dr James Cregg) was digested with HpaII and StuI, and the *α* subunit of the collagen tetramer was replaced with that coding for the PHO1 signal sequence by PCR and used to replace the 3′ end containing 713 bp of 3′ untranslated sequence of a full-length proα1(III) cDNA (Tromp et al., 1989) with an artificial BglII site created 6 bp upstream from the translation initiation codon (Lamberg et al., 1996). The BglII–XbaI proα1(III) cDNA was ligated into the EcoRI site of pPHL-D2 (Invitrogen) and the EcoRI–XbaI site of pPICZ B (Invitrogen), generating pPHL-D2proα1(III) and pPICZ Bproα1(III), respectively.

**Transformation, growth and induction of *P. pastoris***

The his4, arg4 *P. pastoris* host strain was a gift from Dr James Cregg. pARG815αx and pAO815PDI were separately transformed into the host strain by the electroporation method (Manual Version 3.0 of the Pichia Expression Kit, Invitrogen) to generate the αx and PDI strains, respectively. The *P. pastoris* strain expressing the prolyl 4-hydroxylase α subunit and PDI, respectively. Six *Pichia* strains expressing prolyl 4-hydroxylase: α/PDI, α/PDI/PDI, α/PDI-αMF, α/PDI-αMF–HDEL and α–PHO1/α–αMF (Table I), were generated by cotransferring the host strain with pARG815αx and pAO815PDI, pARG815αx and pHIL–S1PDI, pARG815αx and pPIC9PDI, pARG815αx and pPIC9PDI–HDEL, and pARG815αx/PDI and pPIC9PDI, respectively. In all the transformations the α subunit expression vectors and the pARG815αxPDI vector were linearized with *Sau3AII*, while the PDI expression vectors were linearized with *SstI*.

**A *Pichia* strain expressing the proα1 chain of human type III procollagen, Proα1(III), was generated by transforming the host strain with *Sau3AII*-linearized pPHL-D2proα1(III), and a strain α/PDI-αMF/proα1(III) coexpressing human prolyl 4-hydroxylase and proα1(III) chains by transforming *Pmel*-linearized pPICZ Bproα1(III) into the prolyl 4-hydroxylase-expressing strain α/PDI-αMF (Table I).**

**Analysis of recombinant proteins produced in *P. pastoris***

Cells were harvested after a 60 h methanol induction at 30°C, washed once and suspended in a cold (4°C) 0.1 M NaCl, 0.1 M glycine, 0.1% Triton X-100, 10 μM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl-fluoride (PMSF) and 10 mM Tris buffer, pH 7.8, or in a 5% glycerol, 1 mM PMSF, 10 mM EDTA, 50 mM sodium phosphate buffer, pH 7.4. Cells harvested from large-scale cultures were broken by vortexing with glass beads (0.5 mm diameter) and those from large-scale cultures by agitation with glass beads using a Minibead-Beater (Biospec Products). The lysate was centrifuged at 10,000 g for 30 min, and the insoluble fraction was further solubilized in 1% SDS. Protein concentrations were determined using the Bio-Rad Protein Assay (Bio-Rad). The cell culture medium was concentrated in Centricron-30 (Amicon). Aliquots of the soluble and insoluble fractions of the cell lysates and the cell culture medium were analysed by SDS-PAGE and non-denaturing PAGE following by staining with Coomassie Blue or by Western blotting with a polyclonal antibody K38 (Veijola et al., 1996b) to the human PDI polypeptide, a monoclonal antibody M95L or a monoclonal antibody K17 (Veijola et al., 1996b) to the human prolyl 4-hydroxylase α subunit, a polyclonal antibody to the N-propeptide of human type III procollagen (Farmos Diagnostica) or a monoclonal antibody 95D1A recognizing the collagenous regions of various collagen chains (A. Snelman and T. Pihlajaniemi, unpublished observations). Prolyl 4-hydroxylase activity was measured by a method based on the hydroxylation-coupled decarboxylation of 2-oxo-[1,2,4]C]glutarate or on the formation of 4-hydroxy-[1,2,4]C]proline using a [1,2,4]C]proline-labeled collagen substrate (Kivirikko and Myllylä, 1982). $K_m$ values were determined as described previously (Myllylä et al., 1977). Further aliquots of the soluble fraction and the cell culture medium were analysed by a radioimmunoassay for the trimeric N-propeptide of human type III
procollagen (Farms Diagnostica), and others were digested with pepsin for 2 h at 22°C. The thermal stability of the pepsin-resistant recombinant type III collagen was studied by means of a brief digestion with a mixture of trypsin and chymotrypsin at various temperatures (Bruckner and Prockop, 1981).

**Dot blot and Northern blot analyses**

Chromosomal DNA, 1–5 µg, was analysed by a DNA dot blot method (Ausubel et al., 1993). The DNA was isolated using the Easy-DNA Kit (Invitrogen), with *P. pastoris* PER3 (obtained from Dr James Cregg) as a single copy control. Total RNA was isolated from transformed cells 12 h after induction with methanol (0.5%) (Schmitt, 1990). RNA was analysed by Northern blotting using a 32P-labelled 1099 bp fragment of cDNA for the human prolyl 4-hydroxylase α subunit, a 828 bp fragment of cDNA for human PDI and a 862-bp fragment of cDNA for human protrI(III) chains as probes. A *P. pastoris* AOX1 fragment was used as an internal control.

**Purification of recombinant proteins**

Prolyl 4-hydroxylase was purified by breaking the cells in a 0.1 M NaCl, 0.1 M glycine, 0.1% Triton X-100, 10 mM PMSF and 10 mM Tris buffer, pH 7.8. The cell lysate was centrifuged at 10 000 g for 30 min, and the supernatant was chromatographed on a poly(L-proline) affinity column (Kivilikko and Myllylä, 1987) followed by a Superdex 200 gel filtration column (Pharmacia). The PDI polypeptide secreted into the minimal medium was partly purified in a Superdex 200 column (Pharmacia). Type III collagen was purified by breaking the cells in a 5% glycerol, 1 mM EDTA and 50 mM sodium phosphate buffer, pH 7.4. The cell lysate was centrifuged at 10 000 g for 30 min, the supernatant was digested with a final concentration of 150 µg/ml of pepsin for 2 h at 22°C, and the pepsin was inactivated by neutralization. The sample was centrifuged at 100 000 g for 40 min, concentrated in Centricon-50, chromatographed on a Superdex 200 column (Pharmacia) with a solution of 200 mM NaCl and 25 mM sodium phosphate, pH 7.4, dialysed against 0.1 M acetic acid, and lyophilized.

**Other assays**

*N*-glycosidase F treatment was performed according to the instructions provided by the manufacturer (Boehringer Mannheim). The N-terminal sequences of the subunits of the purified recombinant human prolyl 4-hydroxyases were determined on an Applied Biosystems model 477A on-line 120 liquid pulse protein sequencer (Department of Biochemistry and Biotechnology, University of Kuopio). Amino acid analysis of the purified type III collagen was performed in an Applied Biosystems 421 amino acid analyzer.

**Acknowledgements**

We thank Dr James Cregg, Oregon Graduate Institute of Science and Technology, for the highly valuable advice and suggestions, the his4, arg4 *P. pastoris* host strain, the plasmid pYM25 and the *P. pastoris* PER3 DNA; Kari Juntunen, Department of Clinical Chemistry, University of Oulu, for help with protein purification; and Minna Siurua, Anu Kauko, Pihlajaniemi for their expert technical assistance. This work was supported by grants from the Health Sciences Council of the Academy of Finland, from the European Commission (BI04-CT96-0537) and from FibroGen Inc. (San Francisco, CA).

**References**


Berg,B.R., Kay,W.W. and Kedersha,N.L. (1980) The assembly of *P. pastoris* PER3 (obtained from Dr James Cregg) as a single copy control. Total RNA was isolated from transformed cells 12 h after induction with methanol (0.5%) (Schmitt, 1990). RNA was analysed by Northern blotting using a 32P-labelled 1099 bp fragment of cDNA for the human prolyl 4-hydroxylase α subunit, a 828 bp fragment of cDNA for human PDI and a 862-bp fragment of cDNA for human protrI(III) chains as probes. A *P. pastoris* AOX1 fragment was used as an internal control.

**Purification of recombinant proteins**

Prolyl 4-hydroxylase was purified by breaking the cells in a 0.1 M NaCl, 0.1 M glycine, 0.1% Triton X-100, 10 mM PMSF and 10 mM Tris buffer, pH 7.8. The cell lysate was centrifuged at 10 000 g for 30 min, the supernatant was chromatographed on a poly(L-proline) affinity column (Kivilikko and Myllylä, 1987) followed by a Superdex 200 gel filtration column (Pharmacia). The PDI polypeptide secreted into the minimal medium was partly purified in a Superdex 200 column (Pharmacia). Type III collagen was purified by breaking the cells in a 5% glycerol, 1 mM EDTA and 50 mM sodium phosphate buffer, pH 7.4. The cell lysate was centrifuged at 10 000 g for 30 min, the supernatant was digested with a final concentration of 150 µg/ml of pepsin for 2 h at 22°C, and the pepsin was inactivated by neutralization. The sample was centrifuged at 100 000 g for 40 min, concentrated in Centricon-50, chromatographed on a Superdex 200 column (Pharmacia) with a solution of 200 mM NaCl and 25 mM sodium phosphate, pH 7.4, dialysed against 0.1 M acetic acid, and lyophilized.

**Other assays**

*N*-glycosidase F treatment was performed according to the instructions provided by the manufacturer (Boehringer Mannheim). The N-terminal sequences of the subunits of the purified recombinant human prolyl 4-hydroxyases were determined on an Applied Biosystems model 477A on-line 120 liquid pulse protein sequencer (Department of Biochemistry and Biotechnology, University of Kuopio). Amino acid analysis of the purified type III collagen was performed in an Applied Biosystems 421 amino acid analyzer.


