Co-expression of recombinant human collagen α1(III) chain with viral prolyl 4-Hydroxylase in *P. pastoris* GS115

Jiayuan Fang[a,1], Ze Ma[a,1], Dongyue Liu[b,1], Zhaoguo Wang[a], Shuo Zheng[a], Hongyan Wu[a], Peijun Xia[a], Xi Chen[a], Rui Yang[a], Linlin Hao[a,*]

[a] College of Animal Science, Jilin University, 5333 Xi'an Road, Changchun, Jilin 130062, China

[b] Guoda New Materials Co., LTD, 3999 Air Street, Changchun, Jilin 130062, China

* Corresponding author.

E-mail address: haolinlin@jlu.edu.cn (L. Hao).

[1] Jiayuan Fang, Ze Ma and Dongyue Liu contributed equally to this study.

Abstract

Prolyl 4-hydroxylase (P4H) is essential to maintain the stable triple-helix structure and function of human collagen α1(Ш) chain (COL3A1). To obtain hydroxylated human COL3A1, the human *COL3A1* and the viral *P4H* A085R were co-expressed in *P. pastoris* GS115. The sequence of human *COL3A1* without N-terminal and C-terminal was selected for expression. Colony PCR analysis and sequencing after transfection showed that the target gene had inserted successfully. Real-time quantitative PCR (RT-qPCR) indicated that human *COL3A1* and *P4H* were expressed at the mRNA levels. SDS-PAGE and Western blotting analysis of supernatant from the recombinant methylotrophic yest culture showed that recombinant human COL3A1 (rhCOL3A1) was secreted into the culture medium with an apparent molecular mass of approximately 130 kDa. It was noted that the rhCOL3A1 expession quantity was higest at 120 h of induction. Furthermore, mass spectrometry analysis demonstrated that the rhCOL3A1 was expressed successfully. His-tagged rhCOL3A1 protein was purified by Ni-affinity column.

23   Keywords: hydroxylation, purification, recombinant human collagen α1(Ⅲ) chain, virus prolyl

24   4-hydroxylase

25   **1. Introduction**

26   Collagens are the main structural protein of vertebrates, widely distributed in connective

27   tissues of animals, and are also one of the most abundant protein families in the human body.

28   Collagens play the indispensable role in diverse biological functions, such as maintaining normal

29   physiological functions, repairing injuries, and effecting cell attachment and proliferation (Gelse

30   et al. 2003). There are many types of collagens, all of which composed of triple-helical structure,

31   in which COL3A1 is the key to supporting the epidermis (Vesentini et al. 2005). The application

32   of COL3A1 is wide , mainly in sponge dressings (Rao 1995), membrane dressings, gel dressings

33   and other medical fields. Nowadays, COL3A1 used in these fields is mainly from connective

34   tissue of animal, such as porcine, bovine, asinine and piscine skin. However, it needs to be

35   considered that animal collagen extracted from bone and buffalo skin by traditional methods has

36   certain safety risks. For this reason, expressing recombinant collagen by genetic engineering with

37   low cost and high efficiency has received increasing attention (Olsen et al. 2003).

38   During the synthesis of collagen, hydroxyproline plays an important role in maintaining the

39   triple-helix structure and biological properties of collagen (Shoulders and Raines 2009). And

40   many organisms, such as mammalian cells with P4H, can directly form hydroxylated collagen by

41   post-translational modification (Geddis and Prockop 1993; Fichard et al. 1997). However, due to

42   the lack of prolyl hydroxylase, some prokaryotic expression systems, like bacteria, cannot

43   hydroxylate proline residues in collagen (Lamberg et al. 1996). In general, human collagen is

44   coexpressed with P4H to achieve hydroxylation modification (Pinkas et al. 2011; Xu et al. 2011;

45   Nokelainen et al. 2001). Human P4H is an α2β2 heterotetramer, and its limited stability and

46   complex structure may limit efficient production of recombinant hydroxylated collagen in

47   bacterial and yeast populations lacking endogenous hydroxylase (Neubauer et al. 2007; Winter

48   and Page 2000).

49   Recently, some non-human P4Hs were of concern. Several researchers have demonstrated

50   that the co-expression of P4H α and PDI/β genes in the *Escherichia coli* (*E.coli*) could produce

51   active 4-hydroxylase (Neubauer et al. 2005). And some plant-derived P4H showed substrate

52   preference (Kaska et al. 1987). Recombinant hydroxylated collagen was obtained by co-

53   expression with proline hydroxylase L593 from giant mimicry virus in *E.coli* (Rutschmann et al.

54   2014). Moreover, it has been confirmed that recombinant hydroxylated collagen can be co-

55   expressed with collagen in mammalian cells, plant cell cultures and yeast (Xu et al. 2011). Besides,

56   a strain of P4H virus isolated from *Paramecium bursaria Chlorella* virus-1 (PBCV-1) with 242-

57   residue polypeptide was similar to the α subunit sequence of human P4H. The substrates of this

58   P4H were not restricted (Eriksson et al. 1999). In addition, the methylotrophic yeast *P. pastoris*

59   has proved to be an effective production system for the production of heterologous proteins (Baez

60   et al. 2005).

61   At present, rhCOL3A1 is mostly expressed in small fragments, because its full-length

62   expression requires the hydroxylation of proline hydroxylase. However, the commonly used

63   bacterial expression system and yeast expression system do not contain proline hydroxylase, and

64   the stability of the product obtained by the recombinant expression is very low (Lamberg et al.

65   1996). In order to obtain effective and stable rhCOL3A1, the gene encoding for human *COL3A1*

66   without N-terminus and C-terminus was constructed into pPIC9K expression vector, and the viral

3

67   *P4H* A085R derived from PBCV-1 was constructed into pPICZA expression vector, then the two

68   vectors were co-expressed in *P.pastoris* GS115 to obtain hydroxylated collagen. The expression

69   vectors of *COL3A1* and *P4H* were successfully constructed and secreted rhCOL3A1 was obtained.

70   **2. Materials and methods**

71   2.1. Strains, Plasmids

72   *E.coli* strain DH5 (ANGYU, Shanghai, China) was used to amplify plasmids. *P. pastoris*

73   GS115 strain (ANGYU) was used as the host to express the heterologous protein. The expression

74   vectors pPIC9K and pPICZA were procured from Invitrogen.

75   2.2. Construction of rhCOL3A1 expression plasmid

76   The expression vector of the human COL3A1 was constructed as follows: First, the cDNA

77   sequence of the core functional region of human *COL3A1* (GenBank: BC028178.1) was amplified

78   by PCR with primers designed. Next, the product was digested with *SnaBI-NotI* and cloned into

79   expression vector pPIC9K at the same sites to obtain the pPIC9K-COL3A1 plasmid. Besides, the

80   sequence of human *COL3A1* was without N-terminus and C-terminus but with the α-factor signal

81   sequence. The expression vector of human *P4H* was constructed as follows: the cDNA sequence

82   of sub A085R protein gene sequence(GenBank:NP-048433.1) was selected as proline

83   hydroxylase gene sequence (Lu et al. 1995). Next, the gene encoding P4H was amplified by PCR,

84   and the product was digested with *EcoRI-KpnI* and cloned into expression vector pPICZA at the

85   same sites resulting in the pPICZA-P4H plasmid.

86   2.3. Transformation

87   Plasmid pPIC9K-COL3A1 and plasmid pPICZA-P4H were linearized at the site with sacI

88   before electrotransformation. After mixing 20 μl linearized recombinant plasmids with 80 μl

4

89 competent *P. pastoris* GS115 cells, the cells were immediately transferred to a prechilled

90 electroporation cuvette and incubated on ice for 15 min. The charging voltage was 1.5 kV,

91 capacitance was 25 mF, and resistance was 200Ω. The transformed plasmids were coated on yeast

92 extract peptone dextrose medium (YPD) containing bleomycin resistance, and the culture plates

93 were incubated at 30℃ for 2~3 days to grow.

94 2.4. PCR analysis of integrants

95   To further confirm the integration of pPIC9K-COL3A1 and pPICZA-P4H plasmids into the

96 genome of *P.pastoris* GS115 growing on YPD containing bleomycin resistance plates, genomic

97 PCR was performed. Briefly, the clones growing on YPD plates were expanded in the liquid YPD

98 medium, and their genomic DNA was isolated using TIANamp Yeast DNA Kit (TIANGEN,

99 Beijing, China). Two pairs of primers for PCR were 5'/3'*AOX1* and α-factor/3' *AOX1*, and the

100 sequence of 5'*AOX1* was 5'-GACTGGTTCCAATTGACAAGC-3', 3'*AOX1* was 5'-

101 GCAAATGGCATTCTGACATCC-3' and *α-factor* was 5'-TACTATTGCCAGCATTGCTGC-3'.

102 The integration of pPIC9K-COL3A1 gene was verified by PCR using 300 ng genomic DNA as

103 template. For amplification controls, 100 ng of recombinant plasmids pPIC9K-COL3A1 and

104 pPICZA-P4H were used as positive controls; 100 ng of empty plasmids pPIC9K were used as

105 negative control; 100 ng of plasmids pPIC9K-COL3A1 were used as single plasmid control.

106 2.5. Shake-Flask cultivation

107   The successfully transfected colonies were selected for protein induction expression analysis.

108 Single colony was cultured in buffered glycerol medium [BMGY(0.2% biotin, 1% yeast nitrogen

109 base, 1% glycerol, 1% yeast extract, 2% peptone, 100 mM potassium phosphate)] at 30 ℃ with

110 vigorous shaking (200 rpm) for 24 h. The colonies were harvested by centrifugation at 5,000 *g*

5

111   for 20 min at 4 ℃, and then the cell pellets were resuspended in buffered methanol medium

112   [BMMY(0.2% biotin, 1% methanol, 1% yeast nitrogen base, 1% yesat extract, 2% peptone, 100

113   mM potassium phosphate)] for rhCOL3A protein induction. Methanol was supplemented every

114   24 h to maintain 1% methanol concentration, and the culture lasted for 144 h. After induction, the

115   culture supernatant was collected by centrifugation at 5,000 g for 20 min at 4 ℃.

116   2.6. RNA extraction and RT-qPCR

117   Total RNA was isolated using TRIzol Reagent (Takara, USA) and total RNA was harvested

118   at different times after induction. The RNA was immediately used by the Thermo Scientific Revert

119   Aid First Strand cDNA Synthesis Kit(Thermo, Waltham, MA, USA) for reverse transcription

120   reactions. The qPCR reactions were carried out in a 96-well plate in CFX96 Real-time PCR

121   Detection System (Bio-Rad). The primers were listed in Table 1. All samples were performed in

122   triplicate. Each mRNA quantitative data represented the average of at least three measurements.

123   The Ct value of GAPDH gene was normalized to the Ct value of target gene. Data analysis was

124   performaed using the following formula: $2^{-\triangle\triangle Ct}$($\triangle$Ct=Ct$_{COL3A1/P4H}$-Ct$_{GAPDH}$)

125   Table 1 Primer sequence of *COL3A1* , *P4H* and *GAPDH*.

Gene	Primers	Length (bp)
COL3A1	(F) 5'-GAAAGGGGTCGTCCAGGTTT-3'	173
	(R) 5'-GCACCGTTAGATCCAGGTGA-3'	
P4H	(F) 5'- CCCGGCGAACACGAAGTTAT-3'	168
	(R) 5'- TGCATCGTCGCAGTCATCTC-3'	
GAPDH	(F) 5'- AATGGGCAGCCGTTAGGAAA-3'	168
	(R) 5'- GCGCCCAATACGACCAAATC-3'	

126   2.7. SDS-PAGE and Western Blotting

127   The expression of rhCOL3A1 was evaluated by SDS-PAGE analysis of microbial supernatant

6

128   with 12% Tris-glycine gels by the method of Laemmli(Laemmli 1970). The protein sizes on the

129   gel was measured by approximate comparison with the color-mixed Protein Marker (Solarbio，

130   Beijing, China). Then, the 8% Tris-glycine gels containing proteins were transferred to

131   polyvinylidene fluoride (PVDF) membranes. After blocking with western blocking buffer at 25 ℃,

132   the first memberane incubated with anti-COL3A1 mouse monoclonal antibody (Santa Cruz, USA)

133   at 1:100 dilution overnight at 4 ℃. The second memberane incubated with His-Tag Monoclonal

134   antibody (Proteintech, USA) at 1:3000 dilution overnight at 4 ℃. Then the two membranes were

135   washed with TBST three times (5 min each), and the first memberane was incubated with goat

136   anti-mouse IgG labeled with horseradish peroxidase (TIANGEN) at 1:1000 dilution at 25 ℃ for

137   2 h. BeyoECL Plus (Beyotime, Shanghai, China) assessed immune reactivity.

138   2.8. Mass spectrometry analysis

139   After the recombinant protein was isolated by SDS-PAGE, the 130 kDa protein band of

140   rhCOL3A1 was cut from gels and sent it to the Sangon Biotech for Liquid chromatography

141   coupled with tandem mass spectrometry (LC-MS/MS) analysis.

142   2.9. Purification

143   The yeast fermentation medium was harvested after 120 h culture. Medium supernatant was

144   first clarified by centrifugation at 5,000 g for 20 min at 4 ℃. Filtering the centrifuged supernatant

145   with a 0.4 μm filter, then were directly loaded onto a Ni-NTA 6 FF Pre-Packed Gravity column

146   (Sangon Biotech, Shanghai, China) preequilibrated with 10 × Binging Buffer. The column was

147   washed with Elution Buffer and collected 4 mL protein. The purified rhCOL3A1 was analyzed

148   by SDS-PAGE with 8% Tris–glycine gels, and protein ladder was used as reference. The

149   concentration of rhCOL3A1 was determined by the BCA protein.

150   2.10. Statistical Analysis

151   All experiments were performed in triplicate. All statistical analyses were performed using

152   GraphPad Prism 9.0 (GraphPad Software, San Diego, CA). Differences were considered

153   significant at $p < 0.05$.

154   **3. Results and Discussion**

155   3.1. Design of COL3A1 sequence and Construction of vector

156   Many studies have demonstrated that recombinant collagen can be obtained by genetic

157   engineering (Olsen et al. 2005; Luther et al. 2011), and rhCOL3A1 had the same properties as

158   human COL3A1. In this study, we cloned and co-expressed the *P4H* and human *COL3A1* in

159   *P.pastoris* GS115. The cDNA sequence of COL3A1 core functional region (GenBank:

160   BC028178.1) was not contain N-terminus and C-terminus, which were replaced by α-MF signal

161   peptide sequence and six histidine coding sequences, respectively (Li et al. 2015). The *P4H* gene

162   sequence was obtained from viral A085R protein gene sequence(LU. et al. 1995). According to

163   the codon preference of *P. pastoris* cells, *COL3A1* and *P4H* gene sequences were optimized for

164   expression in *P. pastoris* GS115.

165   High copy number *P. pastoris* expression vectors pPIC9K and pPICZA were selected. Vector

166   pPIC9K was digestion with *SnaBI-NotI* and vector pPICZA was digestion with *EcoRI-KpnI*. Fig.

167   1A and B showed the construction of pPIC9K-COL3A1 and pPICZA-P4H plasmids. DNA

168   sequencing results suggested that the target genes were successfully constructed into the

169   expression vector (data not shown). Then, the constructed plasmid would be transfected into *P.*

170   *pastoris* GS115 expression host.

171   3.2. Colony identification and RT-qPCR

8

172   To screen the GS115 colonies integrated rh*COL3A1* gene, Colonies PCR were applied to

173   select recombinant strains(Xu et al. 2015), and two primer pairs were designed, respectively.

174   The results showed that the recombinant plasmids were successfully transferred into *P.pastoris*

175   GS115 (Fig. 2A). The result of CP1 and CP2 colonies PCR using 5'/3'*AOX1* primers had two

176   bands at 2250 bp and 741 bp, respectively. The results of CP1 and CP2 colonis PCR using α -F

177   /3' primers had a band at 3994 bp respectively. The results were in accord with expectations.Then

178   the single colony Cp1 and Cp2 on the [1.0 mg/ml] zeocin plate were picked for induction culture.

179   The induced strains were collected every 24 h and RNA was extracted for RT-qPCR (Xu et al.

180   2015). Fig. 2B and C showed that RT-qPCR results of time-course of *P4H*. The results suggested

181   that the expression of *P4H* gene in CP1 and CP2 strains was the highest at 96h. Fig. 2D and E

182   showed that RT-qPCR results of time-course of *COL3A1*. The mRNA relative expression levels

183   of *COL3A1* gene was highest at 120 h. The results might confirm that the expression of P4H can

184   stabilize the structure of COL3A1, and this result needed to be confirmed by further experiments .

185   Besides, the study of Vuorela et al. had shown that the co-expression of *P4H* and *COL3A1* not

186   only maintains the stability of COL3A1 triple helix structure, but also markedly increased the

187   amount of active prolyl 4-hydroxylase tetramer (Vuorela et al. 1997).

188   3.3. Protein identification and Purification

189   To confirm the rhCOL3A1 protein, SDS-PAGE and western blotting were applied. The Fig.3A

190   was the SDS-PAGE analysis of extracelluar proteins. SDS-PAGE showed a thin protein at 130

191   kDa (marked in Fig.3A), and western blotting showed that the apparent molecular weight of

192   protein was approximately 130 kDa (marked in Fig.3B), and this result was similar to the studies

193   results of Li et al. (Li et al. 2015). Therefore, we can confirm that the *COL3A1* was expressed in

194   the *P.pastoris* GS115. Besides, Fig.3B showed the rhCOL3A1 protein levels of two time-courses,

195   and the results demonstrated that the expression levels of protein was highest for 120 h. Fig.3C

196   suggested the P4H protein expression levels for 96 and 120 h of NC and Cp2, and the 96 h was

197   higher, which were consistent with the qRT-PCR results (Fig.2C). Studies have shown that

198   increasing the copy number of expressing plasmid could increase the production of recombinant

199   protein in *P. pastoris* (Clare et al. 1991; Sreekrishna et al. 1997). Then, the rhCOL3A1 proteins

200   were purified by Ni-affinity column and gel filtration chromatography (Shi et al. 2017). The

201   results suggested that the rhCOL3A1 was concentrated in the first 2 ml of the eluent (Fig.3D),

202   and whether the first 2 ml contains rhCOL3A1 fragments was something we needed to further

203   verify in the futureand. Thus, rhCOL3A1 could be successfully purified with his-tagged.

204   3.4. LC-MS/MS results

205   LC-MS/MS was becoming an increasingly important analytical technology in the

206   identification of peptides and proteins (Brun et al. 2009; Kirkpatrick et al. 2005). Besides, Lund

207   et al. identified lysine methylation sites in globle proteins by mass spectrometry (Lund et al. 2019).

208   At the same time, Li et al. directly detected mercaptan modification of various proteins in

209   pancreatic β cells by mass spectrometry (Li et al. 2021). To confirm the extent of the proline

210   hydroxylation in the rhCOL3A1 protein, the gel pieces containing the 130 kDa protein were cut

211   and identified by LC-MS/MS. Fig.4A-F were the secondary spectrum of peptides identified by

212   mass spectrometry. The abscissa was m/z and the ordinate was relative intensity. The peaks in the

213   figure were signal peaks of fragment ions, and the peaks marked in color were matched fragment

214   ions. It was generally believed that the more fragments ions, the more sufficient the identification

215   evidence. The results showed that the sequence coverage was 19%, because the concentration of

216   protein samples was low. Then, compared with UniProt proteins (rhCOL3A1 and P02461), the

217   protein was identified as human collagen α1(III) chain (Fig.4G). Among the amino acids detected

218   in the peptide sequence was hydroxyprolines, as shown in Fig.4H that mark in gray represented

219   the hydroxylated amino acids. The results suggested that the proline was hydroxylated and its

220   structure was maintained by the formation of intramolecular hydrogen bonds (Berisio et al. 2004).

221   In summary, P4H was the key enzyme in collagen synthesis (Myllyharju 2003). The results

222   confirmed that virus-derived P4H could play the same function as human P4H in Pichia pastoris

223   cells.

224   **4. Conclusions**

225   In this study, we successfully constructed the eukaryotic expression system of rhCOL3A1,

226   and also have shown that the heterologous protein rhCOL3A1 with molecular weight up to 130

227   kDa could be secret by *P. pastoris* GS115. The results indicate that *P. pastoris* GS115 was a good

228   host for rhCOL3A1 protein expression. The results of RT-qPCR suggested that high expression

229   of *P4H* can maintain the stability of COL3A1. Besides, SDS-PAGE and western blotting results

230   showed that the protein expression level of strain was the highest after 120 h of induction,

231   consistent with the RT-qPCR results. At the same time, we confirmed that rhCOL3A1 was

232   modified by P4H, and the presence of hydroxyproline residues was verified by LC-MS/MS

233   analysis. In summary, we confirmed the co-expression of rhCOL3A1 with viral P4H in *P. pastoris*

234   GS115. The strategy provided an alternative for production and purification of the rhCOL3A1

235   protein.

236   **Data Availability Statement**

237   The original contributions presented in the study are includedin the article/supplementary

238   material, further inquiries can bedirected to the corresponding author

239   **Acknowledgments**

246   **Declaration of competing interest**

247   The authors declare that they have no known competing financial interests or personal

248   relationships that could have appeared to influence the work reported in this paper.

249   **References**

250   Baez, J., D. Olsen, and J. W. Polarek. 2005. Recombinant microbial systems for the production of human
251   collagen and gelatin. Appl Microbiol Biotechnol. 69, 245-52.
252   Berisio, R., V. Granata, L. Vitagliano, and A. Zagari. 2004. Imino acids and collagen triple helix stability:
253   characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats. J Am
254   Chem Soc. 126, 11402-3.
255   Brun, V., C. Masselon, J. Garin, and A. Dupuis. 2009. Isotope dilution strategies for absolute quantitative
256   proteomics. J Proteomics. 72, 740-9.
257   Clare, J. J., M. A. Romanos, F. B. Rayment, J. E. Rowedder, M. A. Smith, M. M. Payne, K. Sreekrishna,
258   and C. A. Henwood. 1991. Production of mouse epidermal growth factor in yeast: high-level
259   secretion using Pichia pastoris strains containing multiple gene copies. Gene. 105, 205-12.
260   Eriksson, M., J. Myllyharju, H. Tu, M. Hellman, and K. I. Kivirikko. 1999. Evidence for 4-
261   hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide
262   substrates. Journal of Biological Chemistry. 274, 22131-4.
263   Fichard, A., E. Tillet, F. Delacoux, R. Garrone, and F. Ruggiero. 1997. Human recombinant alpha1(V)
264   collagen chain. Homotrimeric assembly and subsequent processing. Journal of Biological
265   Chemistry. 272, 30083-7.
266   Geddis, A. E., and D. J. Prockop. 1993. Expression of human COL1A1 gene in stably transfected HT1080
267   cells: the production of a thermostable homotrimer of type I collagen in a recombinant system.

268   Matrix. 13, 399-405.

269   Gelse, K., E. Poschl, and T. Aigner. 2003. Collagens--structure, function, and biosynthesis. Adv Drug
270   Deliv Rev. 55, 1531-46.

271   Kaska, D. D., V. Günzler, K. I. Kivirikko, and R. Myllylä. 1987. Characterization of a low-relative-
272   molecular-mass prolyl 4-hydroxylase from the green alga Chlamydomonas reinhardii. Biochem
273   J. 241, 483-90.

274   Kirkpatrick, D. S., S. A. Gerber, and S. P. Gygi. 2005. The absolute quantification strategy: a general
275   procedure for the quantification of proteins and post-translational modifications. Methods. 35,
276   265-73.

277   Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage
278   T4. Nature. 227, 680-5.

279   Lamberg, A., T. Helaakoski, J. Myllyharju, S. Peltonen, H. Notbohm, T. Pihlajaniemi, and K. I. Kivirikko.
280   1996. Characterization of human type III collagen expressed in a baculovirus system. Production
281   of a protein with a stable triple helix requires coexpression with the two types of recombinant
282   prolyl 4-hydroxylase subunit. Journal of Biological Chemistry. 271, 11988-95.

283   Li, L., D. Fan, X. Ma, J. Deng, and J. He. 2015. High-level secretory expression and purification of
284   unhydroxylated human collagen alpha1(III) chain in Pichia pastoris GS115. Biotechnol Appl
285   Biochem. 62, 467-75.

286   Li, X., N. J. Day, S. Feng, M. J. Gaffrey, T. D. Lin, V. L. Paurus, M. E. Monroe, R. J. Moore, B. Yang, M.
287   Xian, and W. J. Qian. 2021. Mass spectrometry-based direct detection of multiple types of
288   protein thiol modifications in pancreatic beta cells under endoplasmic reticulum stress. Redox
289   Biol. 46, 102111.

290   Lu, Z., Y. Li, Y. Zhang, G. F. Kutish, D. L. Rock, and J. L. Van Etten. 1995. Analysis of 45 kb of DNA
291   located at the left end of the chlorella virus PBCV-1 genome. Virology. 206, 339-52.

292   LU., Zhiqian, Yu LI., Yanping ZHANG., F GERALD, KUTISH. , and L. ROCK. JAMES DANIEL,
293   L,VAN ETTEN. 1995. Analysis of 45 kb of DNA located at the left end of the chlorella virus
294   PBCV-1 genome. Virology. 206., 339-52.

295   Lund, P. J., S. M. Lehman, and B. A. Garcia. 2019. Quantitative analysis of global protein lysine
296   methylation by mass spectrometry. Methods Enzymol. 626, 475-98.

297   Luther, K. B., A. J. Hülsmeier, B. Schegg, S. A. Deuber, D. Raoult, and T. Hennet. 2011. Mimivirus
298   collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. Journal
299   of Biological Chemistry. 286, 43701-09.

300   Myllyharju, J. 2003. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol. 22,
301   15-24.

302   Neubauer, A., P. Neubauer, and J. Myllyharju. 2005. High-level production of human collagen prolyl 4-
303   hydroxylase in Escherichia coli. Matrix Biol. 24, 59-68.

304   Neubauer, A., J. Soini, M. Bollok, M. Zenker, J. Sandqvist, J. Myllyharju, and P. Neubauer. 2007.
305   Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in Escherichia coli:
306   improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer
307   anhydrotetracycline. J Biotechnol. 128, 308-21.

308   Nokelainen, M., H. Tu, A. Vuorela, H. Notbohm, K. I. Kivirikko, and J. Myllyharju. 2001. High-level
309   production of human type I collagen in the yeast Pichia pastoris. Yeast. 18, 797-806.

310   Olsen, D., J. Jiang, R. Chang, R. Duffy, M. Sakaguchi, S. Leigh, R. Lundgard, J. Ju, F. Buschman, V.
311   Truong-Le, B. Pham, and J. W. Polarek. 2005. Expression and characterization of a low

13

312   molecular weight recombinant human gelatin: development of a substitute for animal-derived
313   gelatin with superior features. Protein Expr Purif. 40, 346-57.

314 Olsen, D., C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perala, E. R. Hamalainen,
315   M. Jarvinen, and J. Polarek. 2003. Recombinant collagen and gelatin for drug delivery. Adv
316   Drug Deliv Rev. 55, 1547-67.

317 Pinkas, D. M., S. Ding, R. T. Raines, and A. E. Barron. 2011. Tunable, post-translational hydroxylation of
318   collagen Domains in Escherichia coli. ACS Chem Biol. 6, 320-4.

319 Rao, K. P. 1995. Recent developments of collagen-based materials for medical applications and drug
320   delivery systems. J Biomater Sci Polym Ed. 7, 623-45.

321 Rutschmann, C., S. Baumann, J. Cabalzar, K. B. Luther, and T. Hennet. 2014. Recombinant expression of
322   hydroxylated human collagen in Escherichia coli. Appl Microbiol Biotechnol. 98, 4445-55.

323 Shi, J., X. Ma, Y. Gao, D. Fan, C. Zhu, Y. Mi, and W. Xue. 2017. Hydroxylation of Human Type III
324   Collagen Alpha Chain by Recombinant Coexpression with a Viral Prolyl 4-Hydroxylase in
325   Escherichia coli. Protein J. 36, 322-31.

326 Shoulders, M. D., and R. T. Raines. 2009. Collagen structure and stability. Annu Rev Biochem. 78, 929-
327   58.

328 Sreekrishna, K., R. G. Brankamp, K. E. Kropp, D. T. Blankenship, J. T. Tsay, P. L. Smith, J. D.
329   Wierschke, A. Subramaniam, and L. A. Birkenberger. 1997. Strategies for optimal synthesis and
330   secretion of heterologous proteins in the methylotrophic yeast Pichia pastoris. Gene. 190, 55-62.

331 Vesentini, S., A. Redaelli, and F. M. Montevecchi. 2005. Estimation of the binding force of the collagen
332   molecule-decorin core protein complex in collagen fibril. J Biomech. 38, 433-43.

333 Vuorela, A., J. Myllyharju, R. Nissi, T. Pihlajaniemi, and K. I. Kivirikko. 1997. Assembly of human
334   prolyl 4-hydroxylase and type III collagen in the yeast pichia pastoris: formation of a stable
335   enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires
336   coexpression with prolyl 4-hydroxylase. EMBO J. 16, 6702-12.

337 Winter, A. D., and A. P. Page. 2000. Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme
338   required for exoskeleton formation and the maintenance of body shape in the nematode
339   Caenorhabditis elegans. Mol Cell Biol. 20, 4084-93.

340 Xu, J., L. N. Wang, C. H. Zhu, D. D. Fan, X. X. Ma, Y. Mi, and J. Y. Xing. 2015. Co-expression of
341   recombinant human prolyl with human collagen α1 (III) chains in two yeast systems. Lett Appl
342   Microbiol. 61, 259-66.

343 Xu, X., Q. Gan, R. C. Clough, K. M. Pappu, J. A. Howard, J. A. Baez, and K. Wang. 2011. Hydroxylation
344   of recombinant human collagen type I alpha 1 in transgenic maize co-expressed with a
345   recombinant human prolyl 4-hydroxylase. BMC Biotechnol. 11, 69.

346

347

348   **Figure captions**

349   Fig.1 Physcial maps of the vectors. The red areas were target genes. (A) Schematic map of the

350   recombinant expression vector pPIC9K-COL3A1. (B) Schematic map of the recombinant

351   expression vector pPICZA-P4H.

352

353   Fig.2 (A) Colony PCR results. Lane M, DNA ladder; Lane PC, positive control; Lane 1-3, Cp1,

354   single plasmid (pPIC9K-COL3A1) control and Cp2; Lane NC, negative control. (B) and (C) The

355   relative of P4H expression levels on mRNA. (D) and (E) The relative of COL3A1 expression

356   levels on mRNA.

357

358   Fig.3 SDS-PAGE and Western blotting results for rhCOL3A1 expression. (A) SDS-PAGE

359   results for rhCOL3A1. (B) Western blotting results for rhCOL3A1. Lane M was the standard

360   molecular weights. Lanes 1 and lane 2 were the secretory proteins of the Cp1 strain at different

361   induction times: 96 and 120 h, respectively, after induction. Lane 3, lane 4 and lane 5 were the

362   secretory proteins of the Cp2 strain at different induction times: 96, 120 and 144 h, respectively,

363   after induction. (C) Western blotting results for P4H. Lane M was the standard molecular

364   weights. Lane 1 and lane 2 were the endocrine proteins of the negative control strain at different

365   induction times: 96 and 120 h. Lane 3 and lane 4 were the endocrine proteins of the Cp2 strain

366   at different induction times: 96 and 120 h. (D) Western blotting results for purified rhCOL3A1.

367   NC stands for negative control. Lane M was the standard molecular weights. Lanes 1, lane 2

368   and lane 3 were the collected proteins of the Cp1 strain at different stage: 1st milliliter, 2nd

15

369   milliliter and 3rd milliliter, respectively. Lanes 4, lane 5 and lane 6 were the collected proteins

370   of the Cp2 strain at different stage: 1st milliliter, 2nd milliliter and 3rd milliliter.

371

372   Fig.4 The identification of rhCOL3A1 by LC-MS/MS. (A-F) Secondary mass spectrometry of

373   peptides. The abscissa of the secondary mass spectrometry was m/z, and the ordinate was the

374   relative intensity. The peaks were the signal peak of the fragment ions, and the peaks marked in

375   color were the matched fragment ions. (G) Mascot search results showed that the sequence

376   coverage was 19%, and matched peptides was shown in bold red. (H) The hydroxylated amino

377   acids marked in gray.
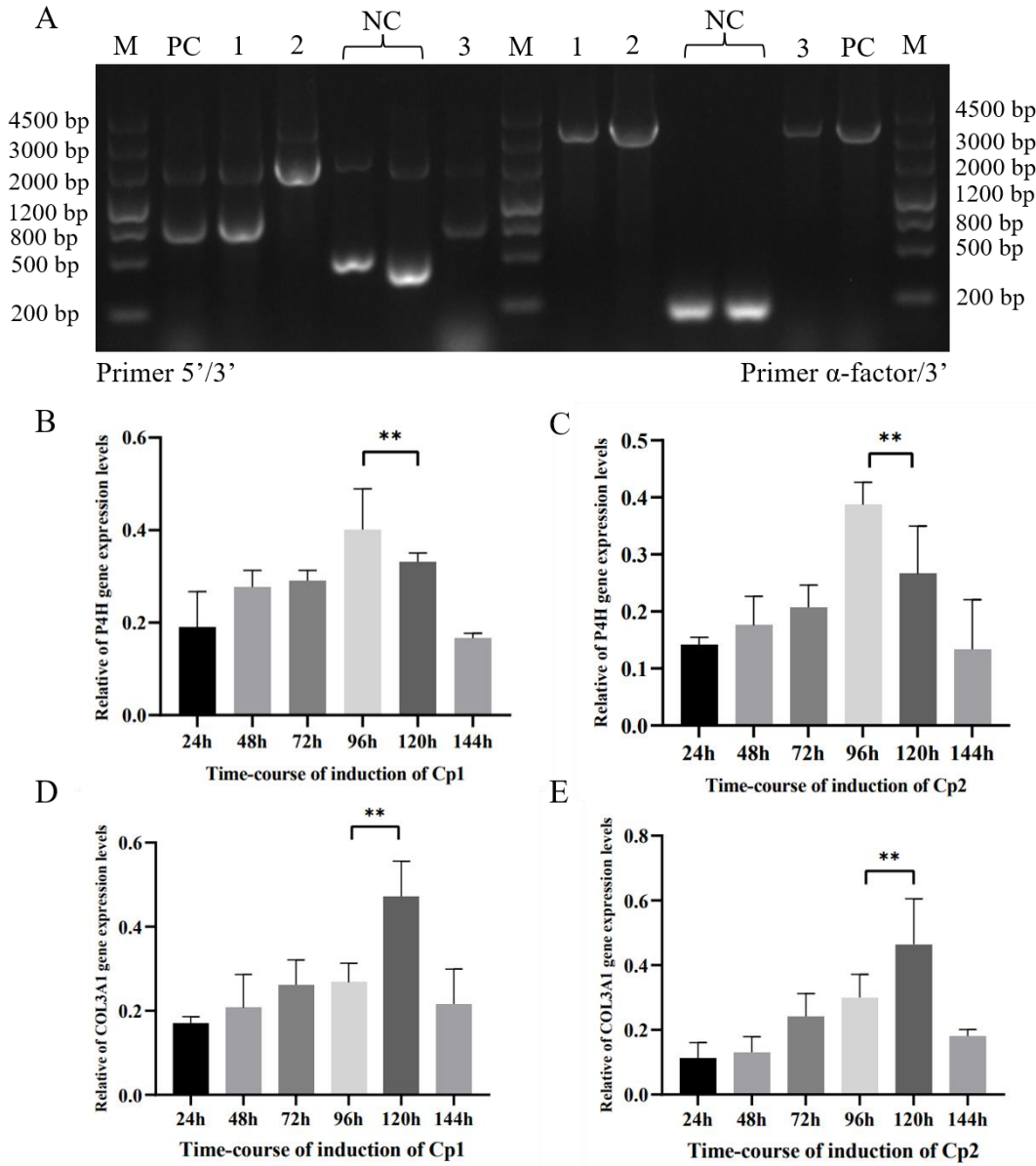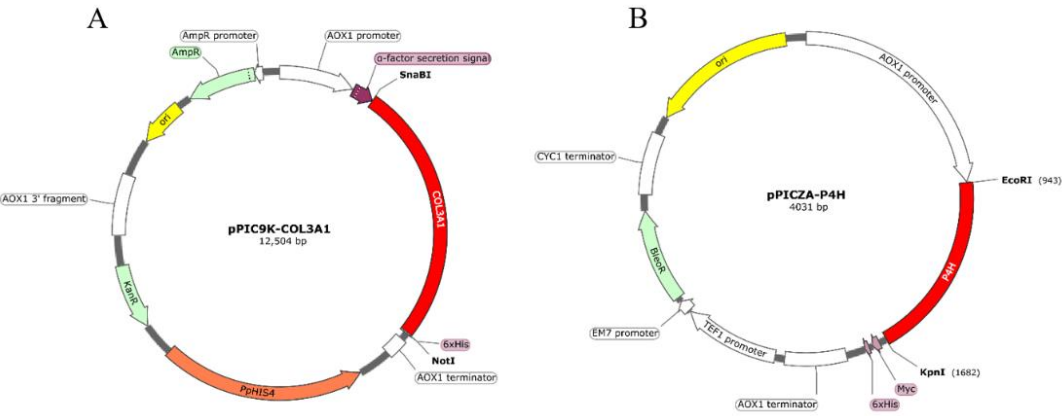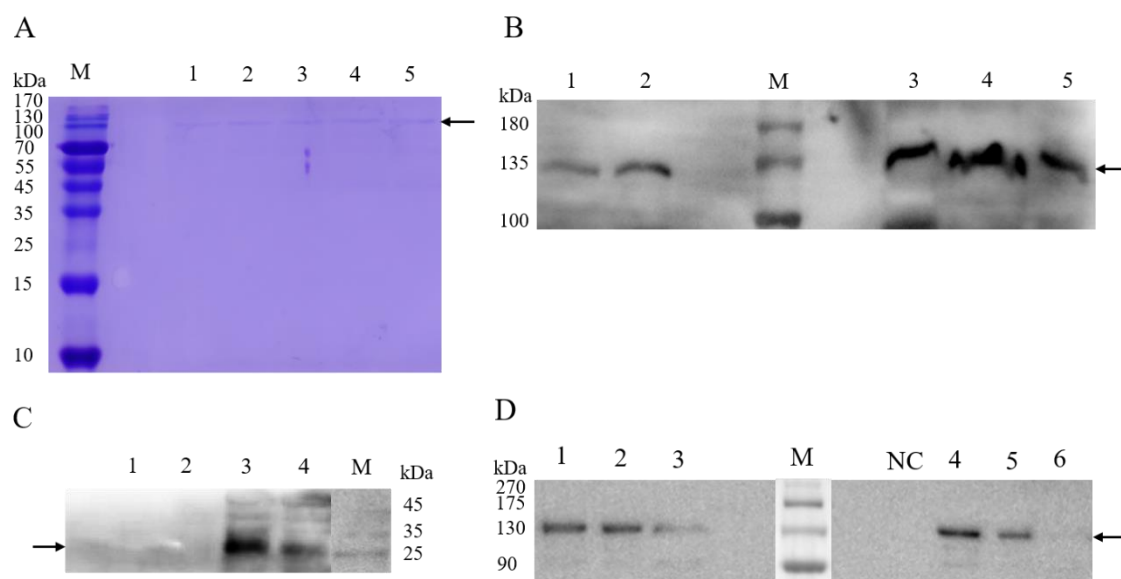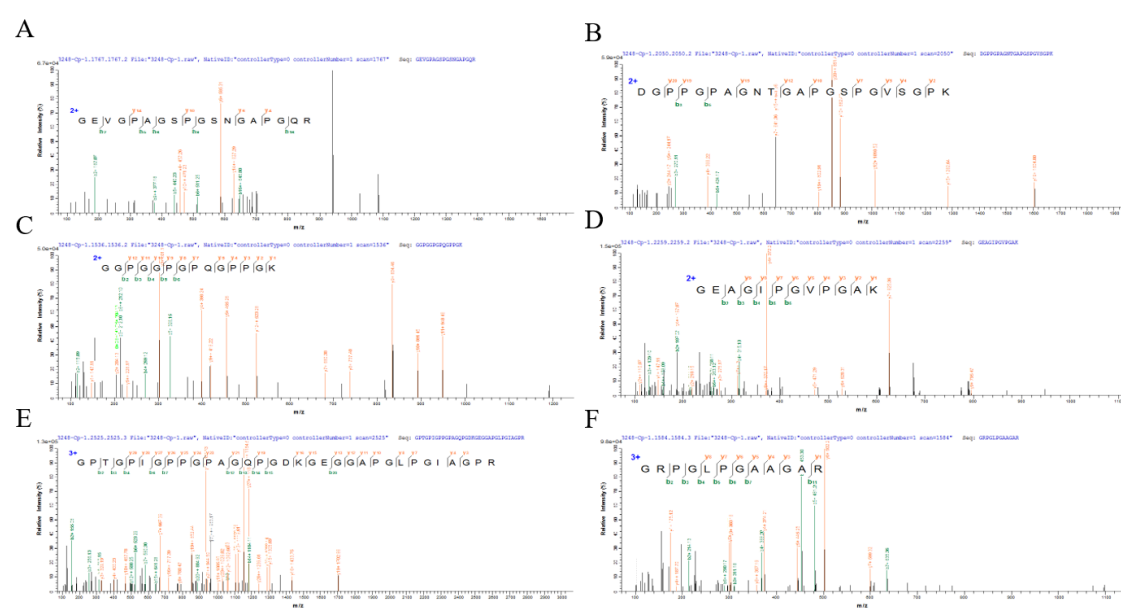
378

379

380　　Figure 1.

Primer 5'/3'　　　　　　　　　　　Primer α-factor/3'

381

382　　Figure 2.

Figure 3.

Figure 4.