BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Recombinant expression of hydroxylated human collagen in *Escherichia coli*

Christoph Rutschmann · Stephan Baumann ·
Jürg Cabalzar · Kelvin B. Luther · Thierry Hennet

Abstract Collagen is the most abundant protein in the human body and thereby a structural protein of considerable biotechnological interest. The complex maturation process of collagen, including essential post-translational modifications such as prolyl and lysyl hydroxylation, has precluded large-scale production of recombinant collagen featuring the biophysical properties of endogenous collagen. The characterization of new prolyl and lysyl hydroxylase genes encoded by the giant virus mimivirus reveals a method for production of hydroxylated collagen. The coexpression of a human collagen type III construct together with mimivirus prolyl and lysyl hydroxylases in *Escherichia coli* yielded up to 90 mg of hydroxylated collagen per liter culture. The respective levels of prolyl and lysyl hydroxylation reaching 25 % and 26 % were similar to the hydroxylation levels of native human collagen type III. The distribution of hydroxyproline and hydroxylysine along recombinant collagen was also similar to that of native collagen as determined by mass spectrometric analysis of tryptic peptides. The triple helix signature of recombinant hydroxylated collagen was confirmed by circular dichroism, which also showed that hydroxylation increased the thermal stability of the recombinant collagen construct. Recombinant hydroxylated collagen produced in *E. coli* supported the growth of human umbilical endothelial cells, underlining the biocompatibility of the recombinant protein as extracellular matrix.

The high yield of recombinant protein expression and the extensive level of prolyl and lysyl hydroxylation achieved indicate that recombinant hydroxylated collagen can be produced at large scale for biomaterials engineering in the context of biomedical applications.

Introduction

The structural and functional versatility of collagen in vertebrates makes it a coveted protein for tissue and biomaterials engineering. Yet, the considerable size of collagen polypeptides and the requirement for post-translational modifications have impeded the large-scale production of recombinant collagen featuring the biophysical properties of natural collagen. All types of collagen feature a triple helical conformation composed of repeats of the G-x-y motif, in which proline and lysine often occur at the x and y positions. During translation in the endoplasmic reticulum, selected proline and lysine residues are hydroxylated by dedicated hydroxylases, thereby yielding hydroxyproline (Hyp) and hydroxylysine (Hyl) (Myllyharju and Kivirikko 2004). The formation of Hyp is essential to stabilize the collagen triple helix and confer its thermal stability at body temperature (Shoulders and Raines 2009). Lysyl hydroxylation is involved in the formation of covalent intra- and inter-molecular crosslinks, contributing to condensation and fibril formation (Takaluoma et al. 2007). Hyl also serves as acceptor for the attachment of collagen-specific glycans (Schegg et al. 2009). Defects of lysyl hydroxylation lead to diseases such as Ehlers-Danlos type-VI (Hyland et al. 1992), Bruck syndrome (van der Slot et al. 2003), and skeletal dysplasia (Salo et al. 2008),

Christoph Rutschmann and Stephan Baumann contributed equally to this work.

C. Rutschmann · S. Baumann · J. Cabalzar · K. B. Luther ·
T. Hennet (✉)
Institute of Physiology, University of Zurich, Winterthurerstrasse
190, 8057 Zurich, Switzerland
e-mail: thennet@access.uzh.ch

demonstrating the biological importance of this post-translational modification.

The multimeric organization and limited stability of animal prolyl 4-hydroxylases and lysyl hydroxylases make them poor choices for the efficient production of recombinant collagen in conventional protein expression systems such as bacteria and yeasts, which lack endogenous prolyl and lysyl hydroxylases. Human collagen prolyl 4-hydroxylase has been expressed in *Escherichia coli* (Neubauer et al. 2005; Pinkas et al. 2011), although with limited activity towards short collagenous substrates. The coexpression of human prolyl 4-hydroxylase subunits and collagen constructs in the yeast *Pichia pastoris* has enabled the production of prolyl hydroxylated collagen up to 1.5 g per liter of culture (Nokelainen et al. 2001). Dual hydroxylation of proline and lysine has not yet been achieved in *P. pastoris*. Coexpression of human prolyl 4-hydroxylase subunits and the lysyl hydroxylase LH3 has been described in tobacco plants, in which recombinant human collagen type I was expressed at up to 200 mg per kilogram of fresh leaves (Stein et al. 2009). The use of animal cells, such as Sf9 insect cells (Lamberg et al. 1996; Tomita et al. 1995) and HEK293 human cells (Fichard et al. 1997) that express prolyl and lysyl hydroxylase endogenously, yields recombinant collagen in the microgram to milligram range per liter of culture, thereby precluding large-scale applications of the recovered collagen product.

The description of several aquatic giant viruses belonging to *Phycodnaviridae* (Van Etten 2003) and *Mimiviridae* (Raoult et al. 2004) has shown that collagen-like genes are not restricted to metazoans and some prokaryotes. In addition to collagen genes, these viruses harbor genes encoding prolyl 4-hydroxylase (Eriksson et al. 1999) and lysyl hydroxylase enzymes (Luther et al. 2011). These viral hydroxylases are soluble and active when expressed in *E. coli*, thus opening new possibilities for the production of recombinant hydroxylated collagen in bacterial expression systems. So far, human collagen type II has been produced at amounts exceeding 10 g per liter (Guo et al. 2010), although without post-translational modifications. To circumvent this limitation, we now exploit bacterially active prolyl and lysyl hydroxylase enzymes from the giant virus mimivirus (Luther et al. 2011) to produce recombinant hydroxylated collagen at high yield in *E. coli*.

Materials and methods

Cloning of mimivirus hydroxylase expression vectors The mimivirus lysyl hydroxylase L230 (Luther et al. 2011) and prolyl 4-hydroxylase L593 open reading frames were amplified by PCR from mimivirus genomic DNA using primers including *Xho*I and *Bam*HI sites. The primers were 5′-TGACCTCGAGATTAGTAGAACTTATGTAATT-3′ and 5′-CAGGGATCCGTCCAATAAAGTGTATCAAC-3′ for L230, and

5′-TGACCTCGAGAAAACTGTGACTATCATTACAATA-3′ and 5′-CAGGGATCCATTTTGTGTTAAAAAAATTTTAGG-3′ for L593. The resulting amplicons were ligated as *Xho*I-*Bam*HI fragments into the *Xho*I-*Bam*HI linearized expression vector pET16b, yielding the pET16b-L230 and pET16b-L593 vectors. Expression vectors lacking His-tags were prepared by first amplifying the L230 and L593 genes using the primers 5′-GTCGACGAGCTCACCATGGGCATTAGTAGAAC-3′ and 5′-GTAATGACATATGCGCAAGCCCAG-3′ for L230, and 5′-ATACCATGGTATTGTCAAAATCTTGTGTGT-3′ and 5′-CAGGGATCCATTTTGTGTTAAAAAAATTTTAGG-3′ for L593. The corresponding amplicons were introduced into pET16b linearized with *Nco*I-*Nde*I for L230 and with *Nco*I-*Bam*HI for L593, yielding pET16b-noH-L230 and pET16b-noH-L593. The bicistronic vector pET16b-L593/L230 was prepared by inserting the expression cassette of the pET16b-L593 as a *Bgl*II-*Hin*dIII fragment into the *Bam*HI-*Hin*dIII-linearized pET16b-L230 vector. The bicistronic vector pET16b-noH-L593/L230 featuring L230 and L593 without His-tag was prepared in the same way.

Cloning of collagen expression vectors A fragment of human collagen type III *COL3A1* cDNA encompassing 1,206 bp and lacking propeptide-encoding regions was custom synthesized (GenScript, Piscataway, NJ, USA) using codons optimized for bacterial expression and including *Nco*I and *Bam*HI sites at 5′- and 3′-ends (Fig. 1). The pET28a expression vector was first digested with *Nco*I-*Bam*HI, which eliminates the His-tag at the N-terminal site. The resulting hCOL3 segment was inserted as a *Nco*I-*Bam*HI fragment into pET28a, yielding pET28a-hCOL3-His.

Protein expression in E. coli The pET16b- and pET28a-based vectors were transformed into chemically competent *E. coli* BL21 (DE3) cells, which were plated on LB-agar plates containing 50 μg/ml kanamycin (Sigma-Aldrich) and 100 μg/ml ampicillin (Sigma-Aldrich) and incubated overnight at 37 °C. Protein expression followed standard protocols (Tolia and Joshua-Tor 2006). Briefly, bacteria were grown in liquid cultures at 37 °C under agitation at 220 rpm until reaching an OD_{600} value of 0.6. Isopropyl β-D-1-thiogalacto-pyranoside (Sigma-Aldrich) was added to 1 mM to induce expression, and the cultures were incubated for a further 3 h at 34 °C under agitation at 220 rpm.

Protein purification Cells were pelleted at 4,000×*g* at 4 °C for 30 min, resuspended in 3 ml of 20 mM sodium phosphate, pH 7.4, 100 mM NaCl per gram of *E. coli* wet weight, and lysed with 250 μg/ml lysozyme, 4 mg/ml deoxycholic acid under rotation at 4 °C for 20 min. DNAse I (Fluka, Buchs, Switzerland) was added to 20 μg/ml and incubation proceeded at room temperature for 30 min. Cell lysates were clarified by centrifugation at 12,000×*g* for 30 min at 4 °C and

Fig. 1 DNA and protein sequence of synthetic human hCOL3 construct. The *top panel* shows the codon-optimized DNA sequence of the truncated human hCOL3 cDNA construct flanked by a 5′ *Nco*I site and 3′ *Bam*HI site (underlined). The ATG and TGA stop codon are *bold* and *shaded*. The sequence encoding the His-tag preceding the stop codon is *dash-underlined*. The *bottom panel* shows the amino acid sequence of the truncated human hCOL3 protein. The Gly-x-y collagen domain is *shaded*

```
   1  CCATGGATGT ATGATTCGTA TGATGTCAAG TCGGGTGTGG CAGTGGGTGG TCTGGCAGGC
  61  TATCCGGGTC CGGCAGGTCC GCCGGGTCCG CCGGGTCCGC CGGGTACCTC TGGTCATCCG
 121  GGTAGCCCGG GCTCTCCGGG TTATCAGGGT CCGCCGGGTG AACCGGGCCA AGCGGGTCCG
 181  AGCGGTCCGC CGGGTCCGCC GGGCGCTATT GGTCCGAGTG GCCCGGCGGG TAAAGATGGC
 241  GAATCCGGTC GTCCGGGTCG TCCGGGTGAA CGCGGCCTGC CGGGTCCGCC GGGTATTAAA
 301  GGTCCGGCAG GCATCCCGGG TTTTCCGGGT ATGAAGGGTC ACCGCGGCTT CGACGGTCGT
 361  AACGGCGAAA AAGGTGAAAC CGGTGCCCCG GGTCTGAAGG GTGAAAACGG TCTGCCGGGT
 421  GAAAATGGTG CTCCGGGTCC GATGGGTCCG CGTGGCGCGC CGGGTGAACG TGGTCGTCCG
 481  GGTCTGCCGG GTGCCGCAGG TGCCCGCGGC AACGATGGTG CACGTGGCAG TGACGGTCAG
 541  CCGGGTCCGC CGGGTCCGCC GGGGACCGCT GGTTTTCCGG GCTCCCCGGG TGCAAAAGGC
 601  GAAGTGGGTC CGGCAGGCAG TCCGGGTTCC AATGGTGCAC CGGGTCAGCG CGGCGAACCG
 661  GGTCCGCAAG GCCATGCCGG TCCGCCGGGC CCGGTTGGTC CGGCAGGCAA GAGCGGTGAT
 721  CGTGGCGAAT CTGGTCCGGC CGGTCCGGCT GGTGCCCCGG GTCCGGCCGG TAGTCGCGGC
 781  GCACCGGGTC CGCAAGGCCC GCGTGGTGAC AAAAGGCGAAA CCGGTGAACG CGGCGCAGCT
 841  GGTATTAAGG GCCACCGTGG TTTCCCGGGC AATCCGGGTG CACCGGGCAG CCCGGGTCCG
 901  GCTGGCCAGC AGGGTGCCAT TGGCTCTCCG GGCCCGGCCG TCCGCGTGG TCCGGTTGGT
 961  CCGTCAGGTC CGCCGGGTAA AGATGGCACG TCGGGTCATC CGGGTCCGAT TGGTCCGCCG
1021  GGTCCGCGTG GTAATCGCGG TGAACGTGGC TCAGAAGGTT CGCCGGGTCA CCCGGGGCCAA
1081  CCTGGTCCGC CGGGTCCGCC GGGTGCTCCG GGTCCGTGCT GTGGCGGTGT TGGCGCGGCC
1141  GCAATCGCGG GCATCGGCGG CGAAAAGGCG GGCGGCTTTG CTCCGTATTA TCATCATCAC
1201  CATCACCATT GAGGATCC
```

```
   1  MYDSYDVKSG VAVGGLAGYP GPAGPPGPPG PPGTSGHPGS PGSPGYQGPP GEPGQAGPSG
  61  PPGPPGAIGP SGPAGKDGES GRPGRPGERG LPGPPGIKGP AGIPGFPGMK GHRGFDGRNG
 121  EKGETGAPGL KGENGLPGEN GAPGPMGPRG APGERGRPGL PGAAGARGND GARGSDGQPG
 181  PPGPPGTAGF PGSPGAKGEV GAPGSPGSNG APGQRGEPGP QGHAGPPGPV GPAGKSGDRG
 241  ESGPAGPAGA PGPAGSRGAP GPQGPRGDKG ETGERGAAGI KGHRGFPGNP GAPGSPGPAG
 301  QQGAIGSPGP AGPRGPVGPS GPPGKDGTSG HPGPIGPPGP RGNRGERGSE GSPGHPGQPG
 361  PPGPPGAPGP CCGGVGAAAI AGIGGEKAGG FAPYYHHHHH H
```

filtered through 0.22-μm membrane filters (Millipore). Imidazole was added to a concentration of 20 mM, and His-tagged proteins were purified by affinity chromatography on a 1 ml HisTrap FF Ni Sepharose 6 column (GE Healthcare) using an Äkta FPLC system (GE Healthcare). Elution of His-tagged proteins was performed with 500 mM imidazole, 20 mM sodium phosphate, pH 7.4, 100 mM NaCl. His-tagged proteins were detected after transfer to nitrocellulose (Highbound ECL, GE Healthcare) using the anti-polyHis HIS-1 monoclonal antibody (Sigma-Aldrich).

Prolyl and lysyl hydroxylase activity assays Prolyl and lysyl hydroxylase activities were measured as described previously (Luther et al. 2011). Briefly, 5 μg His-tag purified L230 lysyl hydroxylase or L593 prolyl hydroxylase was added to acceptor peptides at 0.5 mg/ml in 50 mM Tris–HCl, pH 7.4, 100 μM $FeSO_4$, 1 mM ascorbate, 100 μM DTT, 60 μM 2-oxoglutarate, and 100 nCi of 2-oxo[^{14}C]glutarate (PerkinElmer Life Sciences) in a total volume of 100 μl and incubated at 37 °C for 45 min. Released [^{14}C]O_2 was captured in a filter paper soaked in NCS II Tissue Solubilizer (GE Healthcare) suspended above the assay in a sealed 30 ml vial (VWR, Dietikon, Switzerland). Assays were stopped by addition of 100 μl ice-cold 1 M KH_2PO_4, and the filter papers were transferred to scintillation vials filled with 10 ml of IRGA-Safe Plus scintillation fluid (PerkinElmer Life

Sciences). Radioactivity was measured in a Tri-Carb 2900TR scintillation counter (PerkinElmer Life Sciences).

Amino acid analysis Purified collagens (10 μg) were hydrolyzed in 500 μl of 6 M HCl for 12 h at 105 °C. Hydrolysates were dried down under nitrogen, then washed twice with 500 μl of H_2O, and dried down again. Samples were resuspended in 100 μl of H_2O and derivatized using 9-fluorenylmethoxycarbonyl chloride (FMOC) following the procedure of Bank et al. (Bank et al. 1996). Derivatized amino acid samples were analyzed by reverse phase HPLC as described in Schegg et al. (2009).

Mass spectrometry Purified collagens (2 μg) were alkylated with iodoacetamide and digested with trypsin (Shevchenko et al. 2006). Briefly, after diluting the sample in 100 mM ammonium bicarbonate, 0.1 % (*w/v*) RapiGest (Waters, Saint-Quentin, France) and 5 mM dithiothreitol, the sample was heated for 30 min at 60 °C, cooled, and alkylated in 15 mM iodoacetamide for 30 min in the dark. Proteins were digested with trypsin overnight at 37 °C and acidified with trifluoroacetic acid to a final concentration of 0.5 % prior to desalting using a C18 ZipTip (Millipore). Tryptic digests were subjected to reverse phase LC-MS/MS analysis using a custom packed 150×0.075 mm Magic C18- AQ, 3 μm, 200 Å, column (Bischoff GmbH, Leonberg, Germany) and an

Orbitrap Velos mass spectrometer (Thermo Scientific). Peptides were separated with an 80 min gradient of 3 to 97 % of a buffer containing 99.8 % acetonitrile and 0.2 % formic acid. Spectra were recorded in the higher energy collisional dissociation mode acquiring 10 MS/MS spectra per MS scan with a minimal signal threshold of 2,000 counts. Peptides were identified and assigned using Matrix Science Mascot version 2.4.1 and verified with the Scaffold version 4 software (Proteome Software, Inc.) using the X! Tandem search engine. Variable modifications included 16 Da on methionine, proline, and lysine.

Circular dichroism Proteins were purified by gel filtration using a Superdex 200 10/300 GL Column (GE Healthcare). Protein fractions were concentrated in a 10 kDa Spin-X^R UF 500 centrifugal concentrator (Corning) in PBS, pH 7.4, and kept at 4 °C at a concentration of 0.1 mg/ml prior to analysis. Human collagen type III was purchased from Sigma-Aldrich. Measurements were performed with a wavelength between 200 and 250 nm in a spectropolarimeter (J-810, Jasco) with a thermostated quartz cell of 1 mm length. Thermal stability was analyzed at 221.5 nm under heating at a rate of 0.5 °C/min from 4 to 70 °C.

Trypsin digestion of collagen Recombinant hCOL3 (15 μg) in PBS pH 7.4 was digested with 15 ng trypsin (Roche) for 2 h at temperatures ranging from 10 to 35 °C. Digestions were stopped by addition of 2X Laemmli sample buffer, and proteins were separated in 10 % SDS-PAGE under reducing conditions.

Endothelial cell culture Human umbilical vein endothelial cells (HUVEC) were cultured on 0.1 % gelatin (Sigma-Aldrich), 0.1 % recombinant hydroxylated hCOL3, 0.1 % recombinant hCOL3, or 0.25 % poly-D-lysine in ECM endothelial cell medium (ScienCell, Carlsbad, CA) at 37 °C in 5 % CO_2. For immunofluorescence, cells were seeded on glass cover slips at 1,000 cells/cm^2, 13.3 μg coating matrix/cm^2 and cultured for 60 h. After washing twice with PBS, pH 7.4, cells were fixed with 2 % paraformaldehyde for 10 min at room temperature, washed twice with 20 mM glycine in PBS, and permeabilized with 1 mg/ml saponin. Cells were incubated with mouse anti-β-tubulin SAP.4G5 monoclonal antibody (Sigma-Aldrich) diluted 1:200 and labeled with rabbit anti-mouse IgG Alexa-488 (Life Technology) diluted 1:500 for 30 min. Nuclei were stained with DAPI (Biotium, Hayward, CA). Viability was assayed by methylthiazolyldiphenyl tetrazolium reduction using standard protocols (Mosmann 1983).

Sequence data The L230 and L593 nucleotide sequences reported in this paper have the GenBank accession number NC_014649.1. The L230 and L593 protein sequences have the UniProtKB/Swiss-Prot accession numbers Q5UQC3 and Q5UP57, respectively. The nucleotide sequence of the human hCOL3 construct has the EMBL/EBI accession number HG779440.

Results

The genome of the giant virus mimivirus contains seven collagen-like genes and open reading frames annotated as putative lysyl and prolyl hydroxylases (Raoult et al. 2004). We have previously demonstrated that the open reading frame L230 encodes a bifunctional collagen lysyl hydroxylase and glucosyltransferase enzyme (Luther et al. 2011). To confirm the activity of the putative prolyl 4-hydroxylase encoded by the open reading frame L593, we expressed a His-tagged version of the protein in *E. coli*. The 669-bp open reading frame L593 yielded a 26-kDa protein, which could be enriched on Ni^{2+} beads (Fig. 2a). The prolyl hydroxylase activity of the purified L593 protein was assayed using acceptor peptides featuring proline in sequences derived from human collagen type I, type II, adiponectin, and mannose-binding lectin. The L593 protein was active as a prolyl hydroxylase on the artificial peptide sequence $(GPP)_7$ and on the peptides GDRGETGPAGPPGAPGAPGAP and GLRGLQGPPGKLGPPGNPGPS derived, respectively, from collagen type I and mannose-binding lectin, each featuring the GPP motif (Fig. 2b). By contrast, prolyl hydroxylase activity was minimal on the peptides GPMGPSGPAGAR GIQGPQGPR and GIPGHPGHNGAPGRDGRDGTP derived, respectively, from collagen type II and adiponectin, which lack the GPP motif (Fig. 2b). The L593 prolyl 4-hydroxylase was also active on the non-collagenous peptide $(SPAP)_5$ derived from proline-rich mimivirus proteins, thus indicating that L593 was not strictly specific towards G-x-y repeats (Fig. 2b).

To assess the ability of mimivirus L230 lysyl hydroxylase and L593 prolyl 4-hydroxylase to modify collagen fragments produced in *E. coli*, we coexpressed the two mimivirus hydroxylases together with a 38-kDa fragment of human COL3A1 collagen type III. To this end, the mimivirus L230 and L593 open reading frames were expressed bicistronically under kanamycin selection and the human hCOL3 fragment on a separate plasmid under ampicillin selection. The hCOL3 protein included 119 G-x-y repeats flanked by the N- and C-telopeptide sequences but lacking the N- and C-propeptide sequences (Fig. 1). The co-transformation of *E. coli* with the hydroxylase-containing plasmid and the human hCOL3 construct yielded expression of the three His-tagged target proteins at the expected molecular masses of 101, 38, and 26 kDa corresponding to L230, hCOL3, and L593, respectively (Fig. 3a). As a next step, the L230 and L593 hydroxylases were expressed without His-tags to enable the single
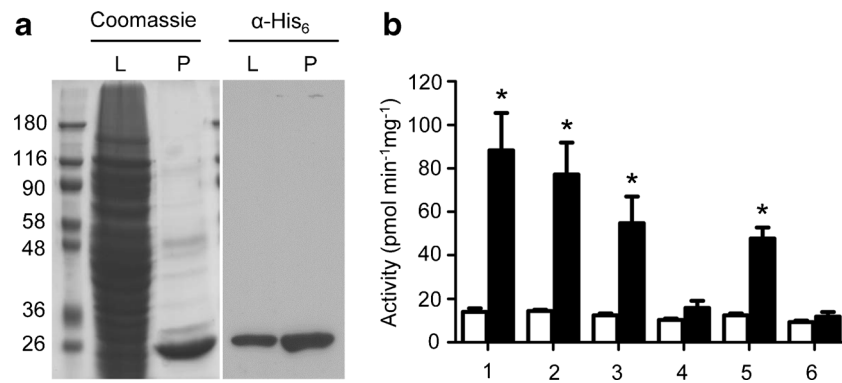
Fig. 2 Bacterial expression and characterization of mimivirus L593. **a** SDS-PAGE of mimivirus L593 expressed in *E. coli* shown as cell lysate (*L*) and after Ni^{2+}-affinity purification (*P*), either after staining with Coomassie blue R-250 or after Western blotting with anti-His$_6$ antibody. **b** Prolyl hydroxylase activity of purified mimivirus L593 assayed on the peptide acceptors [SPAP]$_5$ (*1*), [GPP]$_7$ (*2*), GDRGETGPAGPPGAPGAPGAP from human collagen type I (*3*), GPMGPSGPAGARGIQGPQGPR from human collagen type II (*4*), GLRGLQGPPGKLGPPGNPGPS from human mannose-binding lectin (*5*), GIPGHPGHNGAPGRDGRDGTP from human adiponectin (*6*). *Open bars* show prolyl hydroxylase activity measured without peptide acceptor and *black bars* with peptide acceptors. *Stars* above bars indicate statistically significant activity using two-tailed paired *t* test ($p < 0.01$)

enrichment of the hCOL3 protein from *E. coli* cell lysates. The expression of hCOL3 alone or together with L230 and L593 hydroxylases showed that the collagen fragment reached similar expression levels. After purification by Ni^{2+} affinity chromatography, 90 mg of collagen protein per liter of bacterial culture was routinely obtained. Of note, the coexpression with L230 and L593 hydroxylases produced a smear above the expected hCOL3 band, suggestive of a larger protein size or a decreased migration in polyacrylamide gels (Fig. 3b). This smear may also reflect heterogeneity at the level of prolyl and lysyl hydroxylation of the recombinant protein.

The level of prolyl and lysyl hydroxylation of the hCOL3 protein achieved by coexpression with L230 and L593 hydroxylases was determined by amino acid analysis. Native human collagen type III and the recombinant His-tagged hCOL3 protein expressed with or without L230 and L593 hydroxylases were hydrolyzed under acidic conditions and

derivatized with FMOC. The separation of FMOC-labeled amino acids by HPLC analysis showed that 54 % of proline residues and 47 % of lysine residues were hydroxylated in native human collagen type III (Fig. 4, A). By comparison, the levels of prolyl and lysyl hydroxylation reached, respectively, 25 % and 26 % in the human hCOL3 protein coexpressed with the L593 and L230 hydroxylases (Fig. 4, B). In the absence of L593 and L230 hydroxylases, no prolyl and no lysyl hydroxylation were observed in the recombinant hCOL3 protein (Fig. 4, C). The efficient hydroxylation of recombinant hCOL3 indicates that substrates and co-factors required by the L593 and L230 hydroxylases are present in sufficient amounts in *E. coli* cultured in standard LB medium.

The distribution of Hyp and Hyl residues across the recombinant hCOL3 protein was investigated by mass spectrometry. The analysis of tryptic digested hCOL3 covered 92 % of the sequence including 84 of 87 proline residues and all 12 lysine residues of the hCOL3 protein. The analysis of three different
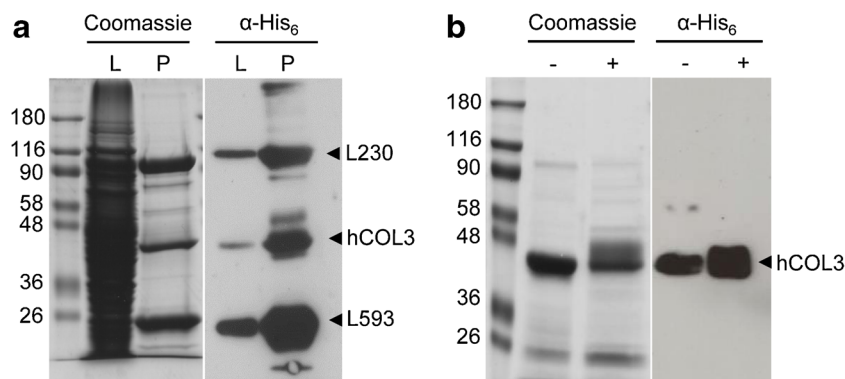

Fig. 3 Coexpression of His-tagged mimivirus hydroxylases L593 and L230 with His-tagged human hCOL3 fragment. **a** SDS-PAGE of mimivirus L593, mimivirus L230, and human hCOL3 construct expressed in *E. coli* shown as cell lysate (*L*) and after Ni^{2+}-affinity purification (*P*), either after staining with Coomassie blue R-250 or after Western blotting with anti-His$_6$ antibody. **b** SDS-PAGE of His-tagged human hCOL3 construct expressed alone (−) or with L593 and L230 hydroxylases (+), shown after staining with Coomassie blue R-250 or after Western blotting with anti-His$_6$ antibody

Fig. 4 Amino acid analysis of native and recombinant human hCOL3. Purified hCOL3 proteins were acid hydrolyzed and the resulting amino acids labeled with FMOC and separated by HPLC. The positions of amino acids are indicated by the *single-letter code*. The positions of hydroxyproline (*Hyp*) and hydroxylysine (*Hyl*) are marked by *arrows. A* native human COL3A1, *B* recombinant hCOL3 construct coexpressed with mimivirus L593 and L230 hydroxylases, *C* recombinant hCOL3 construct expressed alone
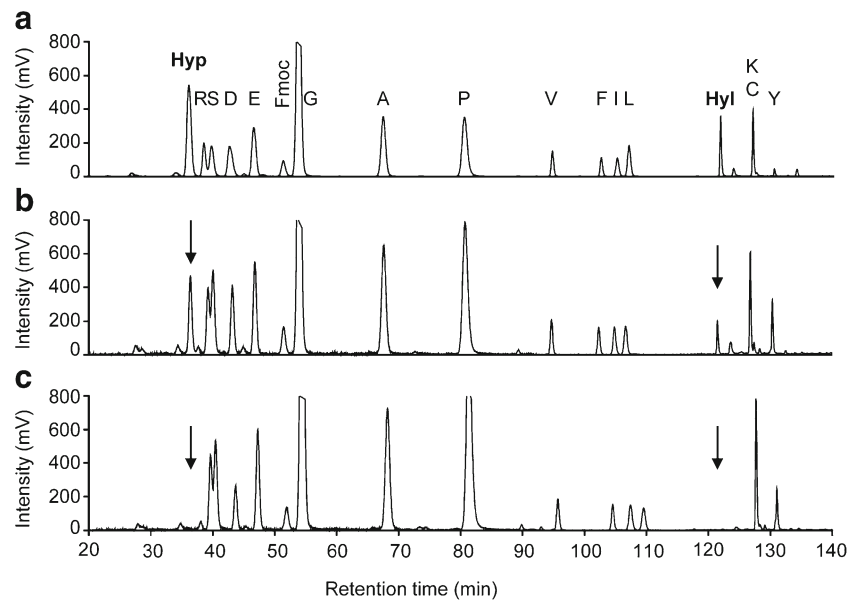
batches of recombinant hCOL3 protein revealed that between 66 and 83 % of covered proline residues were hydroxylated (Table 1). For lysine residues, between 55 and 80 % were detected as hydroxylated (Table 1). The assembly of tryptic peptides showed that hydroxylation was evenly distributed across the hCOL3 protein (Fig. 5a, Table S1). The mimivirus prolyl 4-hydroxylase did not appear to prefer proline residues at either the x or y position of the G-x-y motif. Several G-P-P motifs even included Hyp residues at both x and y positions. The recombinant hCOL3 protein included 12 lysine residues, of which 6 to 8 were hydroxylated (Fig. 5a). As observed for Hyp, the positions of Hyl residues within the G-x-y motif indicated that the mimivirus lysyl hydroxylase enzyme efficiently hydroxylated residues at both x and y positions. We compared the pattern of proline and lysine hydroxylation between native human collagen type III and the recombinant hCOL3 protein expressed in *E. coli*. The analysis revealed a similar distribution of hydroxylated amino acids across both polypeptide sequences (Figs. 5b and S1). Overall, more Hyp

residues were identified in the recombinant hCOL3 protein than in native collagen type III, although differences were minimal across the sequence regions surveyed. These sequences included only three lysine residues, only one of which was found to be hydroxylated in native collagen type III. By contrast, these three lysine residues were hydroxylated in the recombinant hCOL3 protein (Fig. 5b). Hyl residues on recombinant hCOL3 were not further modified, for instance, by glycosylation. We recently showed that the L230 protein is a bifunctional enzyme including both lysyl hydroxylase and Hyl glucosyltransferase domains (Luther et al. 2011). Whereas L230 efficiently converted lysine to Hyl, the enzyme failed to glycosylate the resulting Hyl residues on recombinant collagen, suggesting that the substrate UDP-Glc was not accessible in amounts sufficient to enable the L230-mediated glycosylation of recombinant collagen in *E. coli*.

The triple helical conformation and the thermal stability of the recombinant hCOL3 protein were investigated by circular dichroism. The ellipticity spectra obtained for non-

Table 1 Hydroxylation efficiency of recombinant hCOL3 protein

	AA[a]	Coverage [%]	Pro	Hyp	Hyp/Pro [%]	Lys	Hyl	Hyl/Lys [%]
hCOL3	401		87			12		
Batch 1	346	86	80	56	70	11	6	55
Batch 2	343	86	80	53	66	11	6	55
Batch 3	291	73	63	52	83	10	8	80
Combination[b]	368	92	84	59	70	12	9	75

Tryptic digests were analyzed for hydroxylation of proline (Pro) and lysine (Lys) by mass spectrometry

[a] Covered amino acid length

[b] Combined assembly of tryptic peptides from batches 1 to 3

Fig. 5 Distribution of Hyp and Hyl on recombinant human hCOL3. **a** The occurrence of hydroxylated residues was determined by mass spectrometric analysis of tryptic digests from recombinant human hCOL3 protein. *Grayed* sequences represent portions of the sequences not covered in the analysis. Proline (*P*) and lysine (*K*) residues identified as hydroxylated are shaded. **b** Comparison of Hyp and Hyl distribution on stretches of native human COL3A1 (*nat*) and recombinant human hCOL3 (*rec*) produced in *E. coli*. Proline (*P*) and lysine (*K*) residues identified as hydroxylated are *shaded*

```
a      1  MYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPP
      51  GEPGQAGPSGPPGPPGAIGPSGPAGKDGESGRPGRPGERGLPGPPGIKGP
     101  AGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGLPGENGAPGPMGPRG
     151  APGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEV
     201  GPAGSPGSNGAPGQRGEPGPQGHAGPPGPVGPAGKSGDRGESGPAGPAGA
     251  PGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAG
     301  QQGAIGSPGPAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSE
     351  GSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAGIGGEKAGGFAPYYHHHHH
     401  H
```

```
b
nat   308  GRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGS
rec   156  GRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGS

nat        PGSNGAPGQRGEPGPQGHA  376
rec        PGSNGAPGQRGEPGPQGHA  224


nat  1039  GPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPR 1080
rec   225  GPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPR  266


nat  1109  GFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGKDGTSGHPGP
rec   285  GFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGKDGTSGHPGP

nat        IGPPGPR 1165
rec        IGPPGPR  341
```

hydroxylated and hydroxylated hCOL3 proteins showed the typical shape for triple helical collagen with a maximum peak around 221 nm and a negative peak below 200 nm (Fig. 6a). The changes in ellipticity at 221.5 nm during heating were monitored for non-hydroxylated and hydroxylated hCOL3 proteins to assess the thermal stability of both constructs. The triple helical conformation of the non-hydroxylated hCOL3 protein was unstable and showed an approximate 50 % loss of ellipticity by 19.5 °C (Fig. 6b). Because non-hydroxylated hCOL3 did not yield constant ellipticity values at low temperatures, 19.5 °C cannot be defined as a true T_m value. By contrast, the hydroxylated hCOL3 protein showed a 50 % loss of ellipticity by 24.3 °C, indicating that hydroxylation increased the thermal stability of the construct (Fig. 6b). We also compared the degree of triple helical conformation in non-hydroxylated and hydroxylated hCOL3 by digestion with trypsin, which cleaves denatured collagen but not triple helical collagen (Bruckner and Prockop 1981). Hydroxylated hCOL3 was resistant to trypsin up to 30 °C, whereas non-hydroxylated hCOL3 was mostly degraded by 30 °C (Fig. 6c). Both forms of hCOL3 were completely degraded by 35 °C, which confirmed their low thermal stability below 37 °C.

The biocompatibility of hydroxylated and non-hydroxylated recombinant hCOL3 produced in *E. coli* was assessed by using the protein as a matrix supporting the growth of HUVEC. These cells prefer to grow on extracellular matrix proteins such as fibronectin and collagen (Smeets et al. 1992). The growth of HUVEC was compared between poly-D-lysine, bovine gelatin, recombinant non-hydroxylated hCOL3, and recombinant hydroxylated hCOL3 used as support. Cell morphology was examined by immunofluorescent staining of microtubules. When cultured on recombinant hydroxylated and non-hydroxylated hCOL3, cell viability after 60-h culture reached, respectively, 64 and 49 % of the viability observed when cells grew on 0.1 % gelatin. As expected, viability was lowest when cells were cultured on poly-D-lysine (Fig. 7a). Cell morphology assessed by staining of microtubules showed that cells were evenly spread and tightly attached to the gelatin and hydroxylated hCOL3 matrices as indicated by the large number of processes (Fig. 7b). By contrast, the amount of rounded cells was elevated when non-hydroxylated hCOL3 was applied as matrix and few cells were visible when cultured on poly-D-lysine (Fig. 7b). The compatibility of hydroxylated hCOL3 as support for the growth of HUVEC demonstrated that the recombinant protein was suitable for biological applications such as matrix-assisted cell proliferation and adhesion.

Discussion

The production of recombinant collagen requires post-translational modifications which are lacking in bacterial and yeast expression systems. In the present study, we show that the prolyl hydroxylase L593 and the lysyl hydroxylase L230

Fig. 6 Circular dichroism of recombinant human hCOL3. **a** Samples of purified hydroxylated (*left panel*) and non-hydroxylated recombinant hCOL3 protein (*right panel*) protein at 0.1 mg/ml were scanned between 200 and 250 nm in a spectropolarimeter. **b** Thermal transitions of hydroxylated (*left panel*) and non-hydroxylated recombinant hCOL3 protein (*right panel*) in PBS, pH 7.4 measured at 221.5 nm under a heating rate of 0.5 °C/min from 4 to 70 °C. The T_m values were determined at the midpoints of the sigmoid curves. **c** Trypsin digestion of hydroxylated (*left panel*) and non-hydroxylated recombinant hCOL3 protein (*right panel*); the *arrowhead* at the right shows the position of the hCOL3 protein
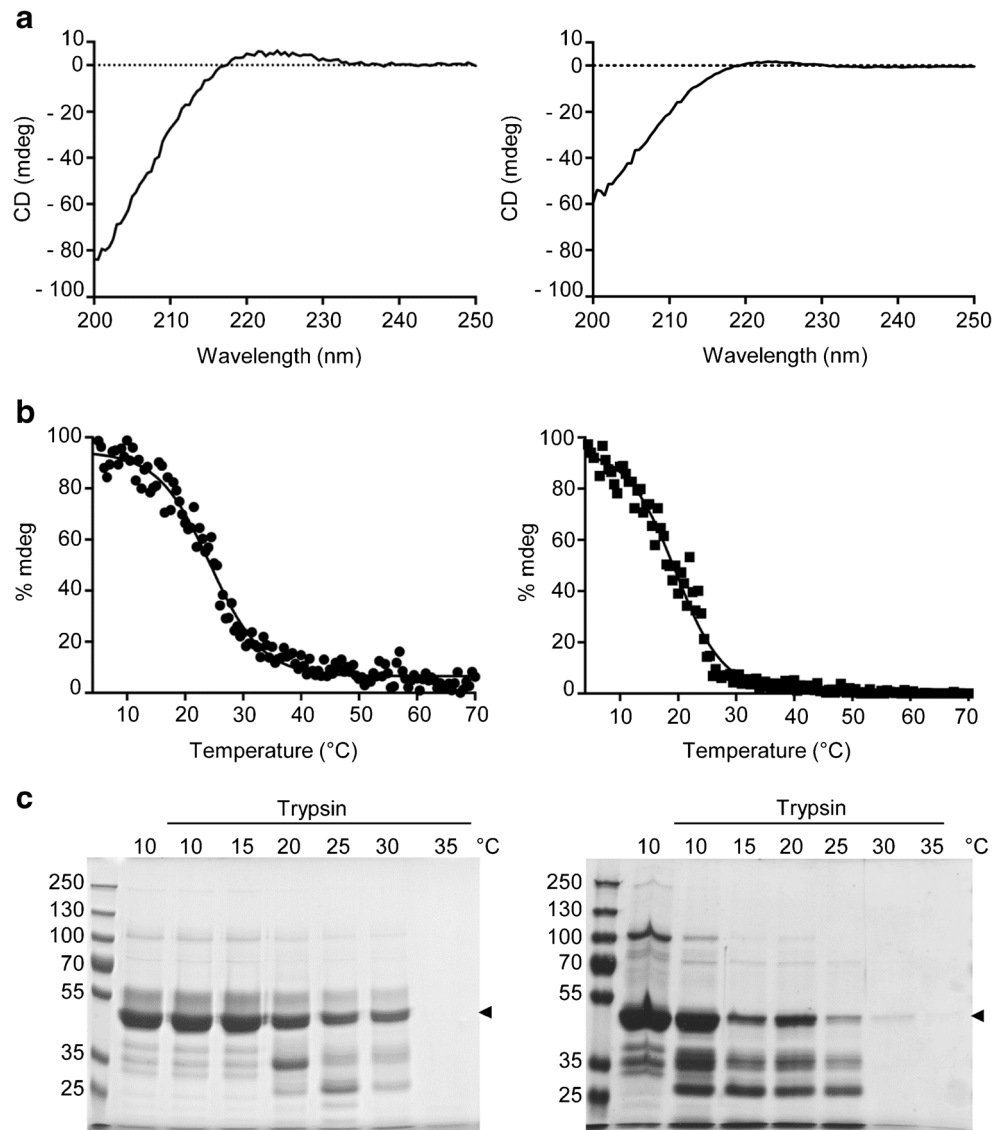

from the giant virus mimivirus can be expressed as active enzymes in *E. coli* without any toxicity for the host cells. The coexpression of these two mimivirus hydroxylases with human collagen constructs enabled the efficient hydroxylation of proline and lysine residues across collagen requiring neither supplementation of co-factors nor increased oxygen partial pressure.

To date, typical cost-effective and high-yield expression systems like yeasts and bacteria have not allowed the production of both prolyl and lysyl hydroxylated collagen because of the low activity of animal hydroxylases introduced in these hosts. The best results were obtained in the yeast *P. pastoris* expressing human prolyl 4-hydroxylase, which enabled 44 % hydroxylation of proline residues on recombinant human collagen type III (Vuorela et al. 1997). However, efficient lysyl hydroxylation of collagen has not been achieved in *P. pastoris* so far. The degree of collagen hydroxylation is also a limiting factor for expression systems based on animal cells. Accordingly, the endogenous prolyl 4-hydroxylase and lysyl hydroxylase activities of insect cells did not yield efficient modification of recombinantly expressed collagen without co-transfection with human prolyl 4-hydroxylase subunits (Lamberg et al. 1996).

The expression of prolyl and lysyl hydroxylase enzymes derived from the giant virus mimivirus yielded degrees of hydroxylation for recombinantly expressed collagen close to those of native collagen type III. The roles of the prolyl hydroxylase L593 and lysyl hydroxylase L230 in mimivirus biology are unknown, but the presence of seven collagen genes in the mimivirus genome (Raoult et al. 2004) suggests that L593 and L230 are involved in the hydroxylation of mimivirus collagen. Indeed, we have previously shown that the mimivirus collagen-like protein L71 is hydroxylated and glycosylated in vitro by the L230 enzyme (Luther et al. 2011).

Fig. 7 Growth of HUVEC on recombinant human hCOL3 matrix. **a** The viability of HUVEC seeded at 1,000 cells per cm^2 was determined by reduction of methylthiazolyldiphenyl tetrazolium to formazan after 60 h incubation at 37 °C on the matrices: 0.1 % gelatin, 0.1 % hydroxylated recombinant human hCOL3 (*hCOL3-OH*), 0.1 % non-hydroxylated recombinant human hCOL3 (*hCOL3*), and 0.25 % poly-ᴅ-lysine (*PDL*). Cell viability is expressed relatively to the values obtained for the positive control, 0.1 % gelatin. Data show the mean and standard error of the mean of three experiments. All conditions tested were significantly different (*p* value<0.05) to cell viability on PDL as determined by one-way ANOVA test with Bonferroni multiple comparison. **b** Immunofluorescence microtubule (*green*) and DNA (*blue*) staining of HUVEC grown on 0.1 % gelatin, 0.1 % hydroxylated re-combinant human hCOL3 (*hCOL3-OH*), 0.1 % non-hydroxylated recombinant human hCOL3 (*hCOL3*), and 0.25 % poly-ᴅ-lysine (*PDL*)
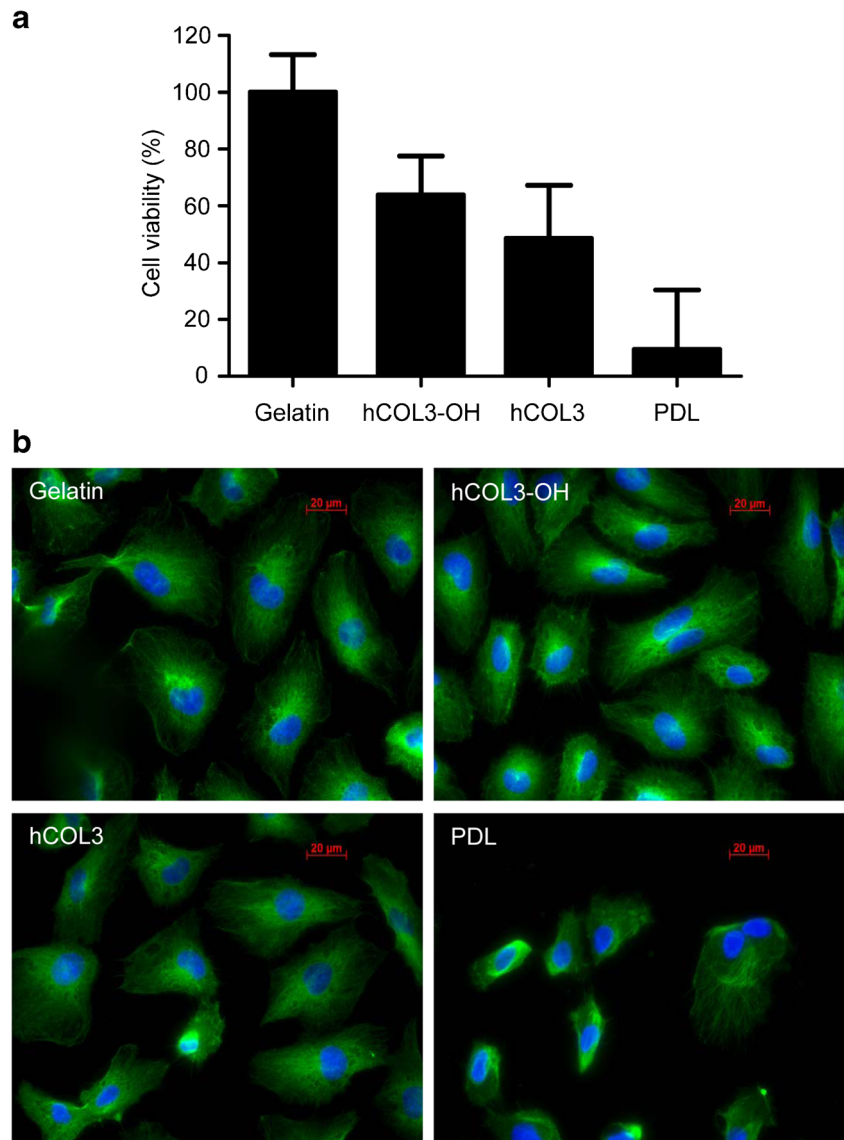
Structurally related proteins are also found in related giant viruses, such as megavirus (Arslan et al. 2011) and moumouvirus (Yoosuf et al. 2012), which also include collagen genes in their genome. Considering their stability and activity when expressed in *E. coli*, proteins from giant viruses may represent a valuable source of enzymes for biotechnological applications, as shown here for the production of hydroxylated recombinant collagen.

The distribution of Hyp and Hyl residues on recombinant hCOL3 showed that mimivirus hydroxylases were able to hydroxylate proline and lysine in various sequence contexts. The pattern of prolyl hydroxylation showed that proline at either position x or y of the motif G-x-y could be efficiently hydroxylated. Studies performed on synthetic peptides containing Hyp at positions x or y or both demonstrated that Hyp at position y strongly increases thermal stability (Jiravanichanun et al. 2006), whereas Hyp at position x

destabilizes the triple helical conformation in Gly-Hyp-y repeats (Inouye et al. 1982). The presence of Hyp at both positions x and y, by contrast, further stabilized the triple helical conformation of the peptides (Berisio et al. 2004). The detection of several Hyp residues at position x on various types of collagen makes it difficult to predict the positional impact of Hyp on the thermal stability of more complex polypeptides (Bann and Bachinger 2000; Buechter et al. 2003; Song and Mechref 2013). Although early work demonstrated that Hyp occurs exclusively at position y (Fietzek and Rauterberg 1975), recent studies showed that Hyp also occurs at position x in fibrillar collagens (Song and Mechref 2013; Weis et al. 2010).

In animal cells, the C-propeptide domain of collagen is important for the initiation of triple helix formation in the endoplasmic reticulum and contributes to the solubility of the molecules along the secretory pathway (Boudko et al.

2012). The addition of trimerization domains to short collagen constructs, such as the bacteriophage T4 foldon domain at the C terminus of a [GPP]$_{10}$ sequence, was reported to dramatically increase the thermal stability of the collagen construct (Boudko et al. 2002). Although advantageous in accelerating triple helix formation, the addition of propeptides in recombinant collagen constructs expressed in *E. coli* or *P. pastoris* later requires their removal for formation of fibrillar structures. This procedure is usually performed by pepsin digestion, which leaves the triple helical domain intact but also removes the short telopeptides sequences required for the registration of collagen molecules in order to form fibrils (Capaldi and Chapman 1982). Therefore, we chose to produce an hCOL3 construct devoid of propeptides but including the telopeptides necessary for fibrillogenesis. The absence of propeptides did not affect the solubility of the recombinant hCOL3 protein and simplified down-stream processing by avoiding protease digestion and removal from purified collagen.

The simplicity of this mimivirus hydroxylase expression system enables the efficient post-translational hydroxylation of proteins containing collagen domains. In addition to the family of true collagens, several collagenous proteins like adiponectin, mannose-binding lectin, and the surfactant proteins A and D can now be produced as hydroxylated proteins in *E. coli*. The high yield of bacterial expression combined with a high degree of prolyl and lysyl hydroxylation provides the framework for the large-scale production of recombinant collagens for human applications, in which animal collagens represent significant risks for allergic reactions and zoonotic disease transmission.

References

Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A 108(42):17486–17491

Bank RA, Jansen EJ, Beekman B, te Koppele JM (1996) Amino acid analysis by reverse-phase high-performance liquid chromatography: improved derivatization and detection conditions with 9-fluorenylmethyl chloroformate. Anal Biochem 240(2):167–176

Bann JG, Bachinger HP (2000) Glycosylation/Hydroxylation-induced stabilization of the collagen triple helix. 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. J Biol Chem 275(32):24466–24469

Berisio R, Granata V, Vitagliano L, Zagari A (2004) Imino acids and collagen triple helix stability: characterization of collagen-like poly-peptides containing Hyp-Hyp-Gly sequence repeats. J Am Chem Soc 126(37):11402–11403

Boudko S, Frank S, Kammerer RA, Stetefeld J, Schulthess T, Landwehr R, Lustig A, Bachinger HP, Engel J (2002) Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. J Mol Biol 317(3):459–470

Boudko SP, Engel J, Bachinger HP (2012) The crucial role of trimerization domains in collagen folding. Int J Biochem Cell Biol 44(1):21–32

Bruckner P, Prockop DJ (1981) Proteolytic enzymes as probes for the triple-helical conformation of procollagen. Anal Biochem 110(2):360–368

Buechter DD, Paolella DN, Leslie BS, Brown MS, Mehos KA, Gruskin EA (2003) Co-translational incorporation of trans-4-hydroxyproline into recombinant proteins in bacteria. J Biol Chem 278(1):645–650

Capaldi MJ, Chapman JA (1982) The C-terminal extrahelical peptide of type I collagen and its role in fibrillogenesis in vitro. Biopolymers 21(11):2291–2313

Eriksson M, Myllyharju J, Tu H, Hellman M, Kivirikko KI (1999) Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. J Biol Chem 274(32):22131–22134

Fichard A, Tillet E, Delacoux F, Garrone R, Ruggiero F (1997) Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing. J Biol Chem 272(48):30083–30087

Fietzek PP, Rauterberg J (1975) Cyanogen bromide peptides of type III collagen: first sequence analysis demonstrates homology with type I collagen. FEBS Lett 49(3):365–368

Guo J, Luo Y, Fan D, Yang B, Gao P, Ma X, Zhu C (2010) Medium optimization based on the metabolic-flux spectrum of recombinant *Escherichia coli* for high expression of human-like collagen II. Biotechnol Appl Biochem 57(2):55–62

Hyland J, Ala-Kokko L, Royce P, Steinmann B, Kivirikko KI, Myllyla R (1992) A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers-Danlos syndrome type VI. Nat Genet 2(3):228–231

Inouye K, Kobayashi Y, Kyogoku Y, Kishida Y, Sakakibara S, Prockop DJ (1982) Synthesis and physical properties of (hydroxyproline-proline-glycine)10: hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. Arch Biochem Biophys 219(1):198–203

Jiravanichanun N, Nishino N, Okuyama K (2006) Conformation of alloHyp in the Y position in the host-guest peptide with the pro-pro-gly sequence: implication of the destabilization of (Pro-alloHyp-Gly)10. Biopolymers 81(3):225–233

Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko KI (1996) Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. J Biol Chem 271(20):11988–11995

Luther KB, Hülsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286(51):43701–43709

Mosmann TR (1983) Rapid colorimetric assay for cellular growth and survival. Application to proliferation and cytotoxicity assays. J Immunol Meth 65:55–65

Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20(1):33–43

Neubauer A, Neubauer P, Myllyharju J (2005) High-level production of human collagen prolyl 4-hydroxylase in *Escherichia coli*. Matrix Biol 24(1):59–68

Nokelainen M, Tu H, Vuorela A, Notbohm H, Kivirikko KI, Myllyharju J (2001) High-level production of human type I collagen in the yeast *Pichia pastoris*. Yeast 18(9):797–806

Pinkas DM, Ding S, Raines RT, Barron AE (2011) Tunable, post-translational hydroxylation of collagen domains in *Escherichia coli*. ACS Chem Biol 6(4):320–324

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306(5700):1344–1350

Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, Robins SP, Myllyla R (2008) A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. Am J Hum Genet 83(4):495–503

Schegg B, Hülsmeier AJ, Rutschmann C, Maag C, Hennet T (2009) Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol 29(4):943–952

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1(6):2856–2860

Shoulders MD, Raines RT (2009) Collagen structure and stability. Annu Rev Biochem 78:929–958

Smeets EF, von Asmuth EJ, van der Linden CJ, Leeuwenberg JF, Buurman WA (1992) A comparison of substrates for human umbilical vein endothelial cell culture. Biotech Histochem 67(4):241–250

Song E, Mechref Y (2013) LC-MS/MS identification of the *O*-glycosylation and hydroxylation of amino acid residues of collagen alpha-1 (II) chain from bovine cartilage. J Proteome Res 12(8):3599–3609

Stein H, Wilensky M, Tsafrir Y, Rosenthal M, Amir R, Avraham T, Ofir K, Dgany O, Yayon A, Shoseyov O (2009) Production of bioactive, post-translationally modified, heterotrimeric, human recombinant type-I collagen in transgenic tobacco. Biomacromolecules 10(9):2640–2645

Takaluoma K, Hyry M, Lantto J, Sormunen R, Bank RA, Kivirikko KI, Myllyharju J, Soininen R (2007) Tissue-specific changes in the hydroxylysine content and cross-links of collagens and alterations in fibril morphology in lysyl hydroxylase 1 knock-out mice. J Biol Chem 282(9):6588–6596

Tolia NH, Joshua-Tor L (2006) Strategies for protein coexpression in *Escherichia coli*. Nat Methods 3(1):55–64

Tomita M, Ohkura N, Ito M, Kato T, Royce PM, Kitajima T (1995) Biosynthesis of recombinant human pro-alpha 1(III) chains in a baculovirus expression system: production of disulphide-bonded and non-disulphide-bonded species containing full-length triple helices. Biochem J 312(Pt 3):847–853

van der Slot AJ, Zuurmond AM, Bardoel AF, Wijmenga C, Pruijs HE, Sillence DO, Brinckmann J, Abraham DJ, Black CM, Verzijl N, DeGroot J, Hanemaaijer R, TeKoppele JM, Huizinga TW, Bank RA (2003) Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. J Biol Chem 278(42):40967–40972

Van Etten JL (2003) Unusual life style of giant chlorella viruses. Annu Rev Genet 37:153–195

Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997) 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. EMBO J 16(22):6702–6712

Weis MA, Hudson DM, Kim L, Scott M, Wu JJ, Eyre DR (2010) Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. J Biol Chem 285(4):2580–2590

Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG, La Scola B, Raoult D, Koonin EV (2012) Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga* moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol Evol 4(12):1324–1330