

SUPPLEMENTARY INFORMATION for:

Tunable, post-translational hydroxylation of collagen domains in *Escherichia coli*

Daniel M. Pinkas^{†,§}, Sheng Ding^{†,§}, Ronald T. Raines[†], Annelise E. Barron^{†,*}

[†]Department of Bioengineering and, by courtesy, Department of Chemical Engineering, Schools of Medicine and of Engineering, Stanford University, Stanford, CA 94305, USA

[‡]Departments of Biochemistry and Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

[§]These authors contributed equally to this work.

*Correspondence should be addressed to A.E.B. (aebarron@stanford.edu)

Supplementary Methods

Construction of plasmids

A pET22b containing cDNAs encoding both the α and β subunits of human P4H was previously described by Kersteen *et al.* (1)

Genomic DNA from *Saccharomyces cerevisiae* (strain EBY100) was extracted using Gentra Puregene Yeast/Bact. Kit (Qiagen). cDNA encoding ALO was amplified from the genomic DNA using primers described by Lee *et al.* (2), which introduced a *Bam*HI site at the 3' end of the gene and an *Nde*I site at the 5' end. The PCR product resulting from this amplification was digested with *Nde*I and *Bam*HI (all restriction enzymes from New England Biolabs) and then ligated into a pET19b vector using T4 ligase, resulting in the plasmid named "pSD.ET19b.ALO1". In order to generate an ALO1 (the gene encoding ALO) insert with restriction sites appropriate for insertion into the co-expression vector pCOLADuet-1 (Novagen), we performed PCR on the plasmid "pSD.ET19b.ALO1" using primers "oSD-1" and "oSD-2".

The PCR produced a linear fragment with an *XhoI* site on the 3' side of the stop codon of ALO1 gene while retaining the *NdeI* site on its 5' side. The resulting fragment was then digested with the *NdeI* and *XhoI*, and ligated into the 2nd multiple cloning site (MCS) of pCOLADuet-1 vector, resulting in a plasmid named "pSD.COLADuet-1.0.ALO1".

An oligonucleotide encoding (Pro-Pro-Gly)₅ ("oSD-3") with a *Bam*HI restriction site at the 5' end and an *XhoI* site at the 3' end was amplified by PCR using primers "oSD-4" and "oSD-5" to obtain a specific double-stranded DNA. The PCR product was digested with *Bam*HI and *XhoI*, and ligated into vector pGEX4T-1 (*GE healthcare*) in order to create the fusion of (Pro-Pro-Gly)₅ to glutathione S-transferase (GST) with an intervening thrombin protease cleavage site ("pSD.GEX4T-1.GST-(PPG)₅"). In order to introduce appropriate restriction sites to ligate GST-(Pro-Pro-Gly)₅ into the 1st MCS of pCOLADuet-1 vector, PCR was carried out on the plasmid "pSD.GEX4T-1.GST-(PPG)₅", using primers "oSD-6" and "oSD-7" that introduced a *NcoI* site on the 5' side of the translation initiation codon of GST-(Pro-Pro-Gly)₅ and a *NotI* site after the 3' side of the stop codon. The PCR fragment was digested with *NcoI* and *NotI*, and ligated into the 1st MCS of both the empty pCOLADuet-1 vector and the plasmid "pSD.COLADuet-1.0.ALO1", which created plasmids "pSD.COLADuet-1.GST-(PPG)₅.0" and "pSD.COLADuet-1. GST-(PPG)₅.ALO1", respectively.

The DNA encoding (Pro-Pro-Gly)₁₀-foldon ("oSD-8") was synthesized (Genscript), and PCR amplified from the supplied vector using primers "oSD-9" and "oSD-10". The PCR product and the plasmid "pSD.COLADuet-1.GST-(PPG)₅.0" were both digested with *Bam*HI and *NotI*, and then ligated. This resulted in plasmid "pSD.COLADuet-1.GST-(PPG)₁₀-foldon.0". The DNA encoding GST-(Pro-Pro-Gly)₁₀-foldon was then isolated from the plasmid by *NcoI* and *NotI* digestion and gel extraction, and then ligated into the 1st MCS of plasmid "pSD.COLADuet-1.0.ALO1", resulting in plasmid "pSD.COLADuet-1. GST-(PPG)₁₀-foldon.ALO1".

Stop codons were introduced just after (Pro-Pro-Gly)₁₀ in both plasmids “pSD.COLADuet-1.GST-(PPG)₁₀-foldon.0” and “pSD.COLADuet-1.GST-(PPG)₁₀-foldon.ALO1” by site-directed mutagenesis per the Stratagene Quickchange protocol using the primers “oSD-11” and “oSD-12”. This resulted in plasmids “pSD.COLADuet-1.GST-(PPG)₁₀.0” and “pSD.COLADuet-1.GST-(PPG)₁₀.ALO1”. Similarly, by introducing stop codons right after (Pro-Pro-Gly)₇ (noted as (Pro-Pro-Gly)₆-Pro-Pro to emphasize the difference with (Pro-Pro-Gly)₅ in the peptide sequence orientation towards the termini in Supplementary Figure 4) using primers “oSD-13” and “oSD-14”, plasmids “pSD.COLADuet-1.GST-(PPG)₇.0” and “pSD.COLADuet-1.GST-(PPG)₇.ALO1” were obtained. Plasmids encoding GST-((Pro-Pro-Gly)₅-Pro-Pro) were generated in the same way by primers “oSD-15” and “oSD-16”.

In order to create plasmids encoding GST-foldon, GST-(Pro-Pro-Gly)₅-foldon and GST-(Pro-Pro-Gly)₇-foldon, we performed deletion mutagenesis on the plasmids “pSD.COLADuet-1.GST-(PPG)₁₀-foldon.ALO1” and “pSD.COLADuet-1.GST-(PPG)₁₀-foldon.0” according to the strategy described by Liu et al. (3). Briefly, primers were designed to contain “non-overlapping” sequences (primer-plasmid complementary) at their 3' end and “primer-primer complementary” sequences at the 5' end. The melting temperature of non-overlapping sequences ($T_{m\ no}$) was 5 to 10 °C higher than the melting temperature of the primer-primer complementary sequences ($T_{m\ pp}$). Twelve cycles of PCR were performed of the following treatment: 95 °C for 1 min, $T_{m\ no} - 5$ °C for 1 min, and 72 °C for 10 min. The PCR cycles were followed by $T_{m\ pp} - 5$ °C for 1 min and 72 °C for 30 min. The PCR mixture was incubated with *DpnI*, and then transformed into NovaBlue competent cells, followed by screening the colonies to check the DNA sequence. The primers “oSD-17” and “oSD-18” were used to generate plasmid “pSD.COLADuet-1.GST-foldon.0”. Using primers “oSD-19” and “oSD-20”, plasmids “pSD.COLADuet-1.GST-(PPG)₅-foldon.ALO1” and “pSD.COLADuet-1.GST-(PPG)₅-foldon.0” were generated, and PCR with primers “oSD-21” and “oSD-22” resulted in plasmids “pSD.COLADuet-1.GST-(PPG)₇-foldon.ALO1” and “pSD.COLADuet-1.GST-(PPG)₇-foldon.0”.

Protein expression and purification

The pCOLADuet-1 vectors with different gene constructs in the 1st and 2nd MCS were transformed or co-transformed with pBK1.PDI1.P4H7 into Origami 2 (DE3) competent cells. In each case a starter culture was grown overnight in M9 medium (0.4% w/v tryptone was used as carbon source if not otherwise specified) supplemented with 30 μ ml⁻¹ Kanamycin when a pCOLADuet-1 vector construct was transformed and 200 μ g ml⁻¹ Ampicillin when pBK1.PDI1.P4H7 was transformed. The starter culture was used to inoculate flasks of 1 L of the culture medium indicated in experiments with appropriate antibiotics. The culture was incubated at 37 °C (250 rpm) until OD₆₀₀ reached 0.25-0.3, and then induced with 50 μ M IPTG and simultaneously supplemented with 1 mM Fe(II)SO₄ (Sigma), and expressed at 23 °C (250 rpm) for 14-18 h. Cells were harvested, resuspended in lysis buffer (DPBS plus 5mM EDTA), and then lysed by sonication. The lysate supernatants were collected after centrifugation (30,000 g, 45 min, 4 °C) and incubated with glutathione affinity resin (GE Healthcare) at 4 °C for 1 h. The resin was washed with DPBS, and then the GST tagged proteins was eluted with 50mM Tris-HCl, 10 mM reduced glutathione, pH 8.0. The proteins were then concentrated and buffer exchanged into DPBS using 10 kDa cut off Amicon protein concentrator (Millipore).

DNA oligomers

The DNA oligomers used in the gene construction as listed below were synthesized by PAN facilities at Stanford, except sequence oSD-8 which was synthesized as a plasmid insert by GenScript (Piscataway).

oSD-1: 5'-CCCGAAAGGAAGCTCGAGTTGGCTGCTG-3'

oSD-2: 5'-CAGCAGCCAACTCGAGCTTCCTTCGGG-3'

oSD-3: 5'-GCTAGGATCCCCGCCGGGTCCGCCAGGCCACCGGGTCCACCTGGCCCGCCTGGTTAAAGGAGAA-GCAGGTGCTGGACAGGGCGAGGCATCGTGCCTGGTTAACTCGAGCTAG-3'

oSD-4: 5'-GCTAGGATCCCCGCCGGGTC-3'

oSD-5: 5'-CTAGCTCGAGTTAACCAGGC-3'

oSD-6: 5'-CAGCTACCATGGGTTCCCCTATACTAGGTTATTGGAAAATTAAGGGCC-3'

oSD-7: 5'-CCTGACGGGCTTGTCTGCTCCC-3'

oSD-8: 5'-GGATCCCCTGGTGCCGCGTGCCAGCGGTCCGCCGGGCCCGCCGGGCCCGCCGGGTCCGCCGGGAC-
CTCCGGGTCTCCTGGCCCTCCTGGTCCGCCGGGACCCCCGGGTCCGCCGGGCAGCGGTTATATTCCGGAAGC-
ACCGCGTGATGGTCAGGCATACGTGCGTAAAGATGGCGAATGGGTTCTGCTGTCTACTTTTCTGTAAGCGGCCGC-
3'

oSD-9: 5'-GTGCCGCGTGGATCCGGTCCGCC-3'

oSD-10: 5'-ACGATGCGGCCGCTTACAGAAAGGTAGACAG-3'

oSD-11: 5'-GACCCCCGGGTCCGCCGTGAGCGGTTATATTC-3'

oSD-12: 5'-GAATATAACCGCTCACGGCGGACCCGGGGGTC-3'

oSD-13: 5'-CTGGCCCTCCTTGACGGGACCCCCGGGTCCG-3'

oSD-14: 5'-CGGACCCGGGGTCCCGTCAAGGAGGGCCAG-3'

oSD-15: 5'-CTCCGGTCTCCTTGACCTCCTGGTCCGCC-3'

oSD-16: 5'-GGCGGACCAGGAGGTCAAGGAGGACCCGGAG-3'

oSD-17: 5'-GGCAGCGGTTATATTCCGGAAGCACCG-3'

oSD-18: 5'-ATATAACCGCTGCCGGATCCACGCGGAACCAGATCC-3'

oSD-19: 5'-CGCGTGATCCGGTCTCCTGGCCCTCCTGGTC-3'

oSD-20: 5'-GGATCCACGCGGAACCAGATCCGATTTTG-3'

oSD-21: 5'-GGCAGCGGTTATATTCCGGAAGCACCG-3'

oSD-22: 5'-ATATAACCGCTGCCAGGAGGGCCAGGAGGACCC-3'

Peptide sequences

Below is the list of the collagenous peptides after thrombin cleavage, from N- to C-terminus:

(Pro-Pro-Gly)₅: GSPPGPPGPPGPPGPPG

(Pro-Pro-Gly)₅-Pro-Pro: GSGPPGPPGPPGPPGPPGPP

(Pro-Pro-Gly)₆-Pro-Pro: GSGPPGPPGPPGPPGPPGPPGPP

(Pro-Pro-Gly)₇: GSGPPGPPGPPGPPGPPGPPGPP

(Pro-Pro-Gly)₁₀: GSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP

Foldon: GSGSGYIPEAPRDGQAYVRKDG EWVLLSTFL

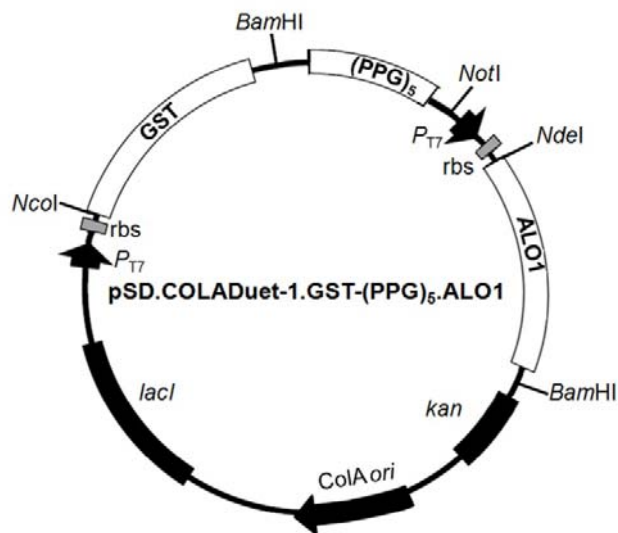
(Pro-Pro-Gly)₅-foldon: GSGPPGPPGPPGPPGPPGSGYIPEAPRDGQAYVRKDG EWVLLSTFL

(Pro-Pro-Gly)₇-foldon: GSGPPGPPGPPGPPGPPGPPGPPGSGYIPEAPRDGQAYVRKDG EWVLLSTFL

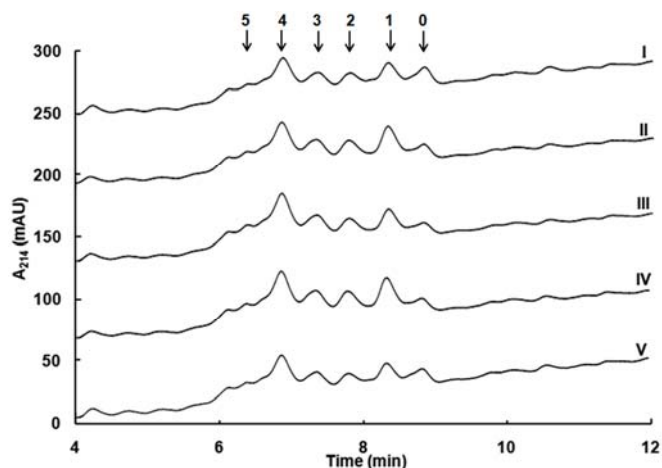
(Pro-Pro-Gly)₁₀-foldon:

GSGPPGPPGPPGPPGPPGPPGPPGPPGPPGSGYIPEAPRDGQAYVRKDG EWVLLSTFL

Supplementary Figures



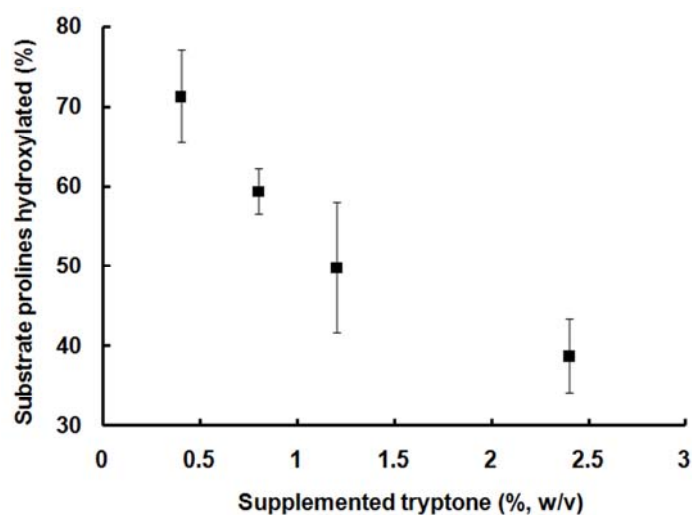
Supplementary Figure 1. Plasmid map of activator/reporter plasmid. Map of plasmid pSD.COLADuet-1.GST-(PPG)₅.ALO1, which encodes both P4H activator and activity reporter genes. The activator gene ALO1 encodes the protein D-arabinono 1,4-lactone oxidase (ALO) from *S. cerevisiae*. The P4H activity reporter encodes a fusion of the affinity tag glutathione-S-transferase (GST) to the high affinity P4H substrate (Pro-Pro-Gly)₅ ((PPG)₅) with an intervening thrombin protease cleavage site. The thrombin cleavage site coincides with one of the *Bam*HI endonuclease sites shown in the vector map.



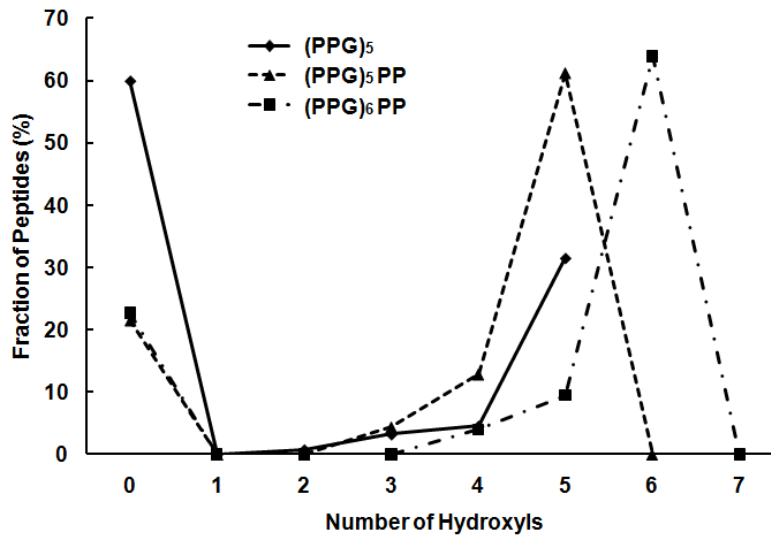
Supplementary Figure 2. Chromatograms for lactone feeding experiment. *In vivo* hydroxylation of (Pro-Pro-Gly)₅ peptides biosynthesized by cells expressing P4H and ALO and then incubated in phosphate buffer with Fe(II)SO₄ and (I) L-ascorbic acid, (II) D-arabinono-1,4-lactone, (III) L-galactono-1,4-lactone, (IV) L-gulono-1,4-lactone, or (V) nothing additional. Arrows indicate number of hydroxylated prolines in the associated peaks as determined by quadrupole mass analysis.

Supplementary Table 1. Yields and hydroxylation levels for GST-(Pro-Pro-Gly)₅ in *E. coli* when coexpressed with both P4H and ALO in different culture medium

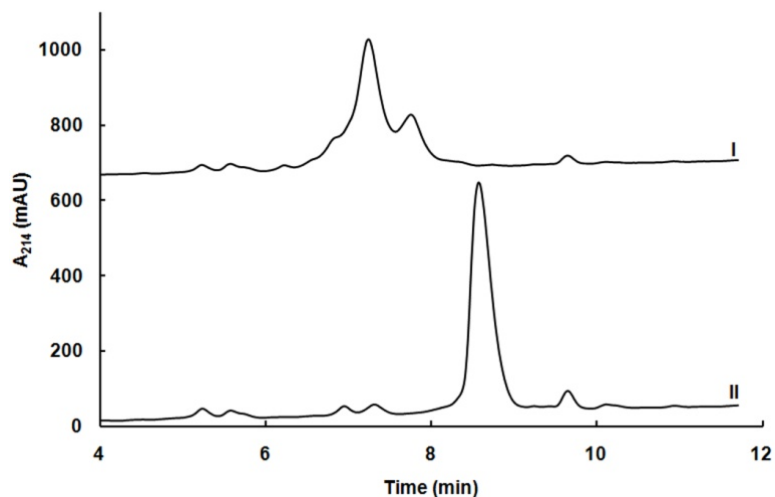
Medium	M9 minimal medium + 0.4% w/v tryptone	M9 minimal medium + 0.4% w/v tryptone + 0.4% v/v glycerol	Terrific Broth
Yield (mg L ⁻¹)	0.2 ± 0.1	0.6 ± 0.1	0.8 ± 0.1
Hydroxylation level (%)	71 ± 6	48 ± 4	37 ± 3



Supplementary Figure 3. Relationship between hydroxylation level and the amount of tryptone in the culture medium. Hydroxylation levels of (Pro-Pro-Gly)₅ peptides expressed in *E. coli* system. The culture medium was M9 minimal medium with different amounts of tryptone as a carbon source (0.4%, 0.8%, 1.2%, or 2.4%, w/v).



Supplementary Figure 4. Relationship between fraction of peptides and the number of hydroxylation events per protein. These data demonstrate both the recurrence and the generality of a preferential “all-or-none” hydroxylation pattern, and clearly indicate processivity of P4H activity in the absence of ascorbate supplementation. Hydroxylation pattern observed for (Pro-Pro-Gly)₅, (Pro-Pro-Gly)₅-Pro-Pro and (Pro-Pro-Gly)₆-Pro-Pro, when each was co-expressed with P4H only (*i.e.*, no ascorbate) in M9 minimal media plus 0.4% w/v tryptone and 0.4% v/v glycerol. Quantitative determination of peptide fraction was accomplished by integrating peak areas from ion chromatograms obtained by LC-MS. For instance, for the **(PPG)₆PP** substrate, 23% of peptides bore no hydroxylation whatsoever; none of the peptides were found to have just one, two or three hydroxyls; 4% of the peptides underwent four hydroxylations; 10% of the peptides bore five hydroxyls; 60% of the peptides had six prolines hydroxylated (hence, this was the most common outcome) and finally, 0% were found to have seven of their prolines hydroxylated.



Supplementary Figure 5. *In vitro* P4H activity assay. UV absorbance chromatograms of (Pro-Pro-Gly)₅ peptides after different treatments. 0.2 mg of purified, unhydroxylated GST-(Pro-Pro-Gly)₅ fusion protein was incubated in 50 mM Tris-HCl buffer, pH 7.8 containing bovine serum albumin (1 mg/mL), catalase (100 µg/mL), dithiothreitol (100 µM), FeSO₄ (50 µM), α-ketoglutarate (500 µM), and P4H (1.5 µM), when additionally: (I) 2 mM ascorbate, or (II) no ascorbate, was added to the final mixture. The reactions took place for 15 h at 37 °C. The samples were then incubated with thrombin. After boiling, the recovered peptides in the supernatant were analyzed by LC-MS.

Supporting References

1. Kersteen, E. A., Higgin, J. J., and Raines, R. T. (2004) Production of human prolyl 4-hydroxylase in *Escherichia coli*, *Protein Expr Purif* 38, 279-291.
2. Lee, B. H., Huh, W. K., Kim, S. T., Lee, J. S., and Kang, S. O. (1999) Bacterial production of D-erythroascorbic acid and L-ascorbic acid through functional expression of *Saccharomyces cerevisiae* D-arabinono-1,4-lactone oxidase in *Escherichia coli*, *Appl Environ Microbiol* 65, 4685-4687.
3. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol, *BMC Biotechnol* 8, 91.