Co-translational Incorporation of Trans-4-Hydroxyproline into Recombinant Proteins in Bacteria*

Received for publication, September 12, 2002, and in revised form October 18, 2002
Published, JBC Papers in Press, October 23, 2002, DOI 10.1074/jbc.M209364200

Douglas D. Buechter†, David N. Paolella§, Bridget S. Leslie¶, Melissa S. Brown*, Karen A. Mehos, and Elliott A. Gruskin**

From the Life Sciences Division, United States Surgical Corporation, North Haven, Connecticut 06473

Trans-4-hydroxyproline (Hyp) in eukaryotic proteins arises from post-translational modification of proline residues. Because the modification enzyme is not present in prokaryotes, no natural means exists to incorporate Hyp into proteins synthesized in Escherichia coli. We show here that under appropriate culture conditions Hyp is incorporated co-translationally directly at proline codons in genes expressed in E. coli. The use of Hyp by E. coli protein synthesis machinery under typical culture conditions is not adequate to support protein synthesis; however, intracellular concentrations of Hyp sufficient to compensate for the poor use are achieved in media with hyperosmotic sodium chloride concentrations. Hyp incorporation was demonstrated in several recombinant proteins including human Type I collagen polypeptides. A fragment of the human collagen Type I (a1) polypeptide with global Hyp for Pro substitution forms a triple helix. Our results demonstrate a remarkable pliancy in the biosynthetic apparatus of bacteria that may be used more generally to incorporate novel amino acids into recombinant proteins.

Amino acids not specified by the genetic code are common in native proteins and can be essential to both structure and function. Proteins that contain novel non-natural amino acid analogues, furthermore, are of value in structure and function studies and may possess beneficial therapeutic properties. Noncoded amino acids in proteins arise either by enzymatic modification of a transfer RNA (tRNA) aminoacylated with one of the 20 coded amino acids (e.g. selenocysteine from serine) or by post-translational modifications. Reproduction of these reactions when recombinant proteins are being expressed in heterologous hosts is often either inefficient or not possible. Because of these difficulties, many proteins that contain noncoded amino acids are not available in quantities sufficient for detailed biophysical and biochemical studies.

Current methodology to make proteins that contain noncoded amino acids in in vitro systems is limited to the production of relatively small amounts of protein. Use of in vitro acylated suppressor tRNAs to insert amino acid analogues at termination codons in genes translated in vitro results in site-specific incorporation (1) but suffers from inefficiency and low yield. These problems occur in part because of limitations in producing acylated tRNA, constraints in achieving appropriate concentrations of other translational components, and variability in suppression efficiency. In vivo approaches that use an aminoacyl-tRNA synthetase with both altered amino acid and tRNA recognition and suppressor tRNAs are promising (2–4), but experimental hurdles inherent to this approach have not yet been fully overcome.

In a general strategy to produce proteins containing novel amino acids, the use of DNA coding triplets to insert the amino acid analogue should be more efficient than use of termination codons. If an aminoacyl-tRNA synthetase is sufficiently promiscuous, it will aminoacylate cognate wild-type tRNA with an amino acid analogue, and the misacylated-tRNA can be used as usual by the translational machinery. This phenomenon has been exploited to produce proteins in prokaryotic systems that contain, for example, analogues of phenylalanine (5) and tryptophan (6). Two requirements must be met for success in these experiments: (1) the analogue must be acylated onto a tRNA at a demonstrable rate, and (2) the analogue must accumulate in the cell at concentrations high enough to give adequate acylation. The first requirement can, in theory, be met either by exploiting the promiscuity of a wild-type synthetase or through mutagenesis to alter the substrate specificity of a wild-type synthetase. The second requirement, although it is a prerequisite for analogue incorporation, is only beginning to be addressed (4).

Escherichia coli prolyl-tRNA synthetase (ProRS)1 is particularly intriguing with respect to these considerations. ProRS will activate several proline analogues, and some of these can be incorporated into heterologously produced proteins in E. coli. These analogues include L-azetidine-2-carboxylic acid (7, 8) and 3,4-dehydroproline (9, 10). In general, proline analogues demonstrated to be incorporated in vitro into proteins are activated in vitro by ProRS at a rate approaching that of proline (11). In contrast to the proline analogues listed above, trans-4-hydroxyproline (Hyp) has been reported not to be activated by ProRS (11), and incorporation of Hyp into E. coli-produced proteins has not been demonstrated (10). In none of these cases has the intracellular concentration of the analogue necessary to support expression and incorporation into protein been determined.

Here we demonstrate that wild-type ProRS can be exploited to incorporate co-translationally trans-4-hydroxyproline efficiently in vivo into recombinant proteins. We reasoned that even a minimal rate of misactivation and misacylation of Hyp

---

1 The abbreviations used are: ProRS, prolyl-tRNA synthetase; Hyp, trans-4-hydroxyproline; Amp, ampicillin; TGF-β1, transforming growth factor; GST, glutathione S-transferase; HPLC, high pressure liquid chromatography; IPTG, isopropyl-1-thio-β-D-galactopyranoside.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Present address: Achillion Pharmaceuticals, New Haven, CT 06511.
‡ Present address: Pfizer Corporation, Groton, CT 06340.
¶ Present address: Alexion Pharmaceuticals, Cheshire, CT 06410.
§ Present address: Bayer Corporation, West Haven, CT 06516.

** To whom correspondence should be addressed: Acorda Therapeutics, Inc., 15 Skyline Drive, Hawthorne, NY 10532. Tel.: 914-347-4300, ext. 139; E-mail: egruskin@acorda.com.

This paper is available on line at http://www.jbc.org

645
by ProRS would be sufficient for incorporation if high intracellular concentrations of Hyp could be achieved. We focused our attention, therefore, on manipulation of E. coli proline transport systems to effect the intracellular accumulation of Hyp. We show that a simple change in E. coli culture conditions results in the intracellular accumulation of Hyp to levels that support synthesis of Hyp-containing recombinant proteins. The initial effort was to produce, in E. coli, Hyp-containing recombinant Type I collagen and collagen-like proteins. In addition to the collagens, our results should make it possible to insert Hyp into any protein synthesized in E. coli and furthermore may have applicability to other amino acid analogues.

EXPERIMENTAL PROCEDURES

Cloning—The gene for prolyl-tRNA synthetase was cloned from E. coli strain XL-1 Blue (Stratagene, La Jolla, CA) by polymerase chain reaction with primers designed from the published gene sequence (GenBank™ accession no. X55518). The 5′ primer added a flanking NcoI recognition site and the 3′ primer a flanking HindIII recognition site. The PCR product was digested with both NcoI and HindIII and ligated into NcoI/HindIII-digested pTcR99+ expression vector (Amersham Biosciences). After expression, ProRS was purified to homogeneity according to published procedures (12). The gene for the Type I collagen polypeptide was cloned by polymerase chain reaction of the gene from mRNA isolated from human foreskin cells (HS27/ATCC 1634) with primers designed from the published gene sequence (GenBank™ accession no. Z74615). The 5′ primer added a flanking EcoRI recognition site and the 3′ primer a flanking HindIII recognition site. The gene was cloned into the EcoRI/HindIII site of plasmid pBR322 (Stratagene), four mutations corrected using the ExSite kit (Stratagene, La Jolla, CA), the sequence confirmed by dideoxy sequencing, and finally the EcoRI/XhoI fragment was subcloned into plasmid pGEX-4T-1 (Amer- sham Biosciences). The genes for the collagen-binding domain of human fibronectin (13) and the mature transforming growth factor β1 (TGF-β1) (14) and mature bone morphogeneic protein (15) polypeptides were expressed from vector pETHis (Invitrogen). The plasmids were a generous gift from Dr. Jane Brokaw. Mannose-binding protein was expressed from plasmid pMAL-c2 (New England Biolabs, Beverly, MA). GST was expressed from plasmid pGEX-4T-1.

Amino Acid Activation Assays—ATP-PPi exchange assays were performed at 2 mM ATP as described (16). The second order rate constant (kcat/Km) for activation of Hyp by E. coli ProRS was calculated from the slopes of plots of the initial activation rates at Hyp concentrations between 5 and 50 mM and an enzyme concentration of 5 nM. The reported value is an average of three determinations. Because radioactively labeled hydroxyproline of sufficient specific activity is not available, we would not determine the kinetics of the aminoclayation of tRNApro with Hyp.

Bacteriology and Protein Expression—E. coli strain JM109 was cured of its proline auxotrophy-complementing episprome by treatment with acridine orange (17). Loss of the episprome was confirmed by the inability of JM109 (F-) to grow on minimal medium that lacked proline. In expression experiments, cultures of JM109 (F−) that harbored the expression plasmid in Luria broth (LB) media containing 100 μg/ml ampicillin (Amp) were grown overnight. Cultures were centrifuged and the cell pellets washed twice with M9 minimal media that contained 100 μg/ml Amp and supplemented with 0.5% glucose and 100 μg/ml of all amino acids except glycine and alanine, which were at 200 μM and contained no proline. The cells were then resuspended in the above medium at 37 °C for 30 min. For isolation of hydroxyproline, osmolyte, or IPTG was added when appropriate. After 3 h at 37 °C, the cells were finally resuspended in the above media. Each culture was spun down, washed with 10 ml of 0.1 absorbance units. Cells were collected by centrifugation and washed with 10 ml of M9 media. Each cell pellet was resuspended in 20 ml of M9/Amp media supplemented with 0.5% glucose and 100 μg/ml of all of the amino acids except proline. Cultures were grown at 37 °C for 30 min to deplete endogenous proline. After growth, NaCl was added to the indicated concentration, Hyp was added to 40 mM and IPTG to 1.5 mM. After 3 h at 37 °C, cells from three 5-ml aliquots of each culture were collected separately on polycarbonate filters and washed twice with 5 ml of M9 media that contained 0.5% glucose and the appropriate concentration of NaCl. Cells were extracted with 1 ml of 70% ethanol by vortexing for 30 min at room temperature. Extract supernatants were taken to dryness, resuspended in 100 μl of 2.5% NaOH, and assayed for Hyp by the method of Neuman and Logan (18). Total protein was determined with the BCA kit (Pierce) after cell lysis by three sonication/freeze-thaw cycles. The data are the means ± S.E. of three separate experiments.

Protein Chemistry and Analysis—For purification of protein D4, crude GST-D4 was dissolved in 0.1% HCl in a round-bottom flask with stirring. After the addition of a 2-10-fold molar excess of BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 h, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid and purified on a Vydac C4 reverse phase-HPLC column (10 × 250 mm, 5, 300 Å). D4 eluted as a single peak at 28% acetonitrile, 0.1% trifluoroacetic acid during a gradient of 15–40% acetonitrile, 0.1% trifluoroacetic acid during a 45-min period. Cleavage with BrCN in 70% formic acid resulted in extensive formylation of D4, presumably at the hydroxyl groups of the Hyp residues. Formylation of BrCN/formic acid-cleaved chains has been noted previously (19). Amino acid analysis was carried out at the W. M. Keck Foundation Biotechnology Resource Laboratory (New Haven, CT) on a Beckman ion exchange instrument with post-column derivatization. N-terminal sequencing was performed at this facility on an Applied Biosystems sequencer equipped with an on-line HPLC system. Electrospray mass spectra were obtained with a VG Biotech BIO-Q quadrupole analyzer by M-Scan, Inc. (West Chester, PA). Circular dichroism (CD) spectra were obtained on an Aviv model 62DS spectropolarimeter (Yale University, Molecular Biophysics and Biochemistry Department). A 1-mm path-length quartz suprasil fluorimeter cell was used. After a 10-min incubation period at 4 °C, standard wavelength spectra were recorded from 260 to 190 nm using 10-s acquisition times and 0.5-nm scan steps. For thermal melts, the temperature was raised in 0.5 °C increments from 4 to 85 °C with a 4-min equilibration between steps. Data were recorded at 221.5 nm. The thermal transition was calculated using the program ThermostDynaCD version 0.9.8, (written by P. Predki).

RESULTS AND DISCUSSION

We reasoned that two requirements were necessary to exploit the promiscuity of E. coli ProRS to misacylate tRNApro so that insertion of Hyp at proline codons would occur during translation. The first requirement was the ability of ProRS to activate Hyp and subsequently charge tRNApro with Hyp. We therefore examined the activation of Hyp by ProRS and found that hydroxyproline is a substrate for activation by E. coli ProRS (Fig. 1a). The Ks for Hyp is estimated to be at least 500 mM. It was not practical to assay for activation at hydroxyproline concentrations sufficiently above this concentration for a more accurate determination of the Ks. The specificity constants (kcat/Km) for activation of hydroxyproline and proline by ProRS are 0.007 s−1·μM−1 and 450 s−1·μM−1 (11), respectively. The ratio of these two constants, 1.5 × 105, is a measure of the fitness of hydroxyproline as a substrate compared with proline (20). Thus, Hyp is activated in vitro by E. coli ProRS with S orders of magnitude less efficiency than proline. At 100 mM amino acid concentration, the rate of activation of alanine by ProRS is comparable with that of Hyp (Fig. 1a). The selectivity of E. coli ProRS for proline versus Hyp and alanine in the activation step is comparable with that exhibited by other E. coli aminoclay-tRNA synthetases that activate noncognate amino acids (21). Because of experimental limitations, we were not able to measure directly the transfer of activated Hyp to tRNApro. In several cases of misacylation, synthetase-mediated editing removes the incorrectly activated or tRNA-esterified amino acid before insertion into a protein. E. coli ProRS can edit alanine misacylated tRNApro (22) but does not edit misacylated cysteine-tRNApro (23). Editing has not been demonstrated for Hyp, and because E. coli ProRS is generally promiscuous and Hyp is not naturally present in E. coli, we proceeded on the assumption that E. coli ProRS would aminoclayate tRNApro with Hyp and would not edit either Hyp adenylate or Hyp-tRNApro.

A second requirement must be met by in vitro approaches to amino acid analogue substitution; intracellular concentrations...
of the analogue sufficient to drive misacylation of the tRNA
must be achieved. Because Hyp is activated poorly in vitro by
ProRS compared with proline, we anticipated that high in vivo
concentrations of hydroxyproline would be necessary to result
in enough Hyp-charged tRNAPro to support protein synthesis.
To achieve this goal, we took advantage of the phenomenon
that proline and other "compatible" solutes are actively accu-
minated by JM109 (F\(^{r-}\)) as a function of NaCl concentration in the
/\(\text{H}11002\)
culated.
enzyme concentration was 20 nM. Amino acid concentrations are indi-
in vitro

cination assays were performed at 2 mM ATP as described (16). The
intracellular hydroxyproline concentration is proportional to the
external NaCl concentration as high as \(-600\) mM. At concentra-
trations higher than 600 mM, accumulation plateaus, possibly
because of saturation of the proline porters. Using a typical
value for the volume of an \(E.\) coli cell (\(1 \times 10^{-15}\) liters), the
intracellular concentration of hydroxyproline at 450 mM NaCl
is \(-150\) mM. This concentration represents an \(\sim 4\)-fold increase
over the extracellular concentration of 40 mM and is compara-
bale with the increases typically found for proline under osmotic
shock conditions (29). Significantly, this concentration ap-
proaches the estimated lower limit for the \(K_m\) of hydroxyproline
activation by ProRS.

The question remained whether the intracellular hydro-
xyproline level in hyperosmotically shocked JM109 (F\(^-\))
resulted in an adequate amount of hydroxyproline mischarged
tRNA\(^{Pro}\) to support synthesis of a recombinant protein. To
address this question, we tested expression of a fragment of the
human Type I collagen chain (Fig. 2, D4) fused to the C
terminus of GST in JM109 (F\(^-\)) under conditions of hyperos-
motic shock. The collagen fragment comprises the C-terminal
193 amino acids of the triple helical region and the 26-amino
acid C-terminal nonhelical telopeptide. To preclude the possi-
bility that expression of the highly repetitive D4 gene would be
limited by differences in codon usage in \(E.\) coli compared with
humans, we synthesized the D4 gene from synthetic oligonu-
cleotides designed to reflect optimal \(E.\) coli codon usage (30,
31). Protein GST-D4 is efficiently expressed in JM109 (F\(^-\)) in
minimal media lacking proline but supplemented with Hyp and
NaCl (Fig. 3). At a fixed NaCl concentration of 500 mM, expres-
sion is minimal at Hyp concentrations less than \(-10\) mM,
whereas the expression level plateaus at Hyp concentrations
greater than \(20\) mM (Fig. 3a). Likewise, at a fixed Hyp concen-
tration of \(40\) mM, NaCl concentrations less than \(300\) mM result
in little protein accumulation, and expression decreases at
concentrations greater than \(700\)–\(800\) mM NaCl (Fig. 3b). The
sodium chloride concentrations that allow for the greatest ac-
cumulation of GST-D4 roughly correspond to those that cause
the intracellular concentration of hydroxyproline to increase,
which suggests that expression is limited in this range by the
intracellular accumulation of hydroxyproline. Either sucrose or
KCl can be substituted for NaCl as the osmolyte (Fig. 3b). Thus,
the osmotic shock-mediated intracellular accumulation of Hyp
is the critical determinant of expression rather than the precise
chemical identity of the osmolyte. Despite the large number of
prolines (14 in GST and 52 in D4) in GST-D4, its size (46 kDa),
and nonoptimal growth conditions, it is expressed at \(-10\)% of
the total cellular protein. Expressed proteins of less than full-
length that were indicative of aborted transcription or transla-
tion or mRNA instability were not detected.

In the experiments shown in Fig. 3, we expected that Hyp
would be inserted at each of the 52 proline codons of protein D4. To confirm this, GST-D4 was cleaved with BrCN at methionines within GST and at the unique methionine at the N-terminal end of D4, and D4 was purified by reverse phase HPLC. Electrospray mass spectroscopy of this protein gave a single molecular ion corresponding to a mass of 20,807 Da. This mass is within 0.05% of that expected for D4 if it contains 100% Hyp in lieu of proline. Proline was not detected in amino acid analysis of purified D4, a finding again consistent with complete substitution of Hyp for proline. To confirm further that Hyp substitution had only occurred at proline codons, we sequenced the N-terminal 13 amino acids of D4. The first 13 codons of D4 specify the protein sequence H2N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly. The sequence found was H2N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly. The sequence found was H2N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly. The sequence found was H2N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly. Taken together these results indicate that Hyp was inserted only at proline codons and that the fidelity of the E. coli translational machinery was not otherwise altered by either the high intracellular concentration of Hyp or hyperosmotic culture conditions.

One of our goals was to develop the methodology to produce Hyp-containing collagen polypeptides in E. coli. The defining feature of the collagens is the Gly-X-Y repeating tripeptide. In vertebrate fibrillar collagens, Hyp in the Y position is critical for the formation of a stable triple helical structure and subsequent fibrils (32). Trans-4-hydroxyproline is not typically found in the X-Y region of the bovine Type I α1 collagen chain comparable in length to D4 forms homotrimeric helices with a melting temperature of 29 °C (36). Thus, despite global Hyp for proline substitution in both the X and Y positions, D4 forms triple helices of stability similar to comparably sized fragments of bovine collagen containing Hyp at the normal percentage and only in the Y position. These results support the notion that Hyp in the X position does not significantly destabilize the triple helix a priori but may, in specific situations, contribute to triple helix stability. Because Ac-(Gly-Hyp-Thr)10-NH2 forms a triple helix but (Hyp-Pro-Gly)10 does not, Pro in the Y position may counteract the otherwise stabilizing effect of Hyp in the X position (34). D4 contains Gly-X-Y triplets that place Hyp in multiple contexts with respect to neighboring amino acids; however, because of global Hyp for Pro substitution no triplets occur with contiguous Hyp and Pro. This fact, along with the overall increase in Hyp content in D4, may also contribute to its triple helix stability.

D4 was heated to 85 °C for 5 min before the CD spectrum was obtained, the magnitude of the absorbance at 198 nm was decreased and the absorbance at 221 nm was abolished (Fig. 4). This behavior is typical of the triple helical structure of collagen (35). A thermal melt profile of D4 in phosphate buffer showed a melting temperature of 29 °C. A fragment of the C-terminal region of the bovine Type I α1 collagen chain comparable in length to D4 forms homotrimeric helices with a melting temperature of 29 °C (36). Thus, despite global Hyp for proline substitution in both the X and Y positions, D4 forms triple helices of stability similar to comparably sized fragments of bovine collagen containing Hyp at the normal percentage and only in the Y position. These results support the notion that Hyp in the X position does not significantly destabilize the triple helix a priori but may, in specific situations, contribute to triple helix stability. Because Ac-(Gly-Hyp-Thr)10-NH2 forms a triple helix but (Hyp-Pro-Gly)10 does not, Pro in the Y position may counteract the otherwise stabilizing effect of Hyp in the X position (34). D4 contains Gly-X-Y triplets that place Hyp in multiple contexts with respect to neighboring amino acids; however, because of global Hyp for Pro substitution no triplets occur with contiguous Hyp and Pro. This fact, along with the overall increase in Hyp content in D4, may also contribute to its triple helix stability.
The full-length human Type I α1 and α2 collagen polypeptides (ColECol(α1) and ColECol(α2), Fig. 2), although more than four times the size of D4, also express as N-terminal fusions with GST in JM109 (F−) in HYP/NaCl media (Fig. 5a). Like D4, the genes for the collagen portions of these proteins were constructed from synthetic oligonucleotides designed to mimic codon usage in highly expressed E. coli genes. In contrast, expression from a GST-human Type I α1 gene fusion (GST-HCol), identical to GST-ColECol(α1) in coded amino acid sequence but containing the human codon distribution, could not be detected in Coomassie Blue-stained SDS-PAGE of total cell lysates of induced JM109 (F−)/pHCol cultures (Fig. 5a). Thus, sequence or structural differences between the genes for ColECol(α1) and HCol are critical determinants of expression efficiency in E. coli. This fact is likely the result of codon distribution in these genes and ultimately of differences in tRNA isoacceptor levels in E. coli compared with humans. However, additional effects on other transcription or translation steps cannot be excluded definitively.

Interestingly, neither GST-ColECol(α1), GST-ColECol(α2), GST-D4, nor GST-HCol accumulates in either hyperosmotic shock media containing only proline or in rich media (data not shown). GST-ColECol(α1), GST-ColECol(α2), and GST-D4 are insoluble when expressed in HYP-containing media at either 37 or 30 °C, and insolubility may segregate them from degradation pathways. Alternatively, the Hyp-containing proteins may escape degradation because they adopt a protease-resistant structure, whereas the proline-containing proteins do not. The large number of codons nonoptimal for E. coli found in the human gene and the instability of proline-containing collagen polypeptides in E. coli may, in part, explain why expression of human collagen in E. coli has not been reported previously.

We sought to generalize our observations further by determining the expression of other proteins in HYP/NaCl-containing media (Fig. 5b). In addition to GST, mannose-binding protein, the mature TGF-1 polypeptide, a 70-kDa fragment of human fibronectin, a chimera of fibronectin and the mature TGF-β1 polypeptide, and a chimera of fibronectin and the mature bone morphogeneic protein 2 polypeptide are expressed in JM109 (F−) in HYP/NaCl. In each case, expression is dependent on NaCl and Hyp. The number of prolines in these proteins varies between 9 (TGF-β1) and 32 (FN-TGF-β1). These results, along with successful expression of GST-ColECol(α1) containing 255 prolines, suggest that any number of prolines can be substituted with hydroxyproline, provided the protein accumulates in E. coli. Incorporation does not depend on the post-translational enzymatic oxidation of proline to hydroxyproline, and as expected for a mechanism operating at a translational rather than post-translational level, successful substitution does not depend on the precise sequence, structure, or activity of the recombinant protein.

In media that contain a mixture of proline and hydroxyproline, the choice of which of these two amino acids is inserted at a given proline codon during translation should depend on the relative efficiency of ProRS-catalyzed aminoacylation of tRNAPro with the amino acid. Thus, selection of the appropriate ratio of proline to hydroxyproline in the growth media should make it possible to express proteins that contain any desired amount of hydroxyproline. To determine whether this is the case, we expressed GST in JM109 (F−) in media containing a fixed amount of hydroxyproline (40 mM) but increasing amounts of proline (Fig. 6). In medium containing either no proline or 1 μM proline, the migration positions of GST are identical, suggesting that these proteins contain comparable
amounts of hydroxyproline. At 2.5 mM proline, GST migrates at the same position as GST expressed in media containing proline but no Hyp. At intermediate proline concentrations in the range of 1 to 30 μM, the migratory position of the protein is between these two extremes, which suggests that these proteins contain variable amounts of Hyp and proline. We estimate that 50% substitution of hydroxyproline for proline occurs at 10 μM or less proline in media containing 40 mM Hyp. The ratio of proline to hydroxyproline in the extracellular media at this concentration is $2.5 \times 10^{-4}$. This ratio is close to the ratio of the in vitro specificity constants for activation of these two amino acids by ProRS (see above), which suggests that, in the absence of editing, the incorporation selectivity reflects the in vivo kinetics of activation and/or aminosacylation. Precise control of the degree of hydroxyproline substitution, coupled with genetic approaches to insert site-specific Hyp at the Y-position of the Gly-X-Y repeat, may make it possible to produce human collagens with any desired Hyp composition.

This system is an experimentally simple and robust way to insert Hyp into any proline-containing protein that can be synthesized in E. coli. The power of the method results from the use of DNA coding triplets and the wild-type amino acid synthetase/ARN pair. In the general case, it is likely that only modest misacylation of a tRNA with an analogue is required, provided a sufficient intracellular concentration of the amino acid analogue can be achieved. Manipulation of amino acid transport systems to achieve intracellular accumulation of the analogue may be of general use and may be feasible with amino acid analogues other than Hyp. Other organisms, including Saccharomyces cerevisiae (37), alter metabolism in response to environmental stresses, and conditions may be found in these systems to promote the cellular uptake, accumulation, and incorporation of novel amino acid analogues.

Strategically placed prolines can increase protein thermostability, and isomerization of proline residues can be the rate-limiting step in protein folding (38, 39). Evidence (40, 41) suggests that inductive effects of electronegative substituents on the proline ring affect the rate and equilibrium position of cis-trans isomerization and, consequently, triple helix stability. These effects are not likely to be specific to collagen and suggest that hydroxyproline may enhance stability or affect the rate of protein folding in noncollagenous proteins. The ability to produce large quantities of Hyp-containing proteins in E. coli should make possible biophysical and biochemical approaches to these areas and may open the door to novel proteins with unique therapeutic, biomaterial, or bioengineering applications.

Acknowledgments—We thank R. S. Lloyd (University of Texas), K. Musier-Forsyth (University of Minnesota), J. O. Hollinger (Oregon Health Sciences University), P. Schimmel (Scripps Research Institute), N. Tawil, D. Gies, and T. Turecek for thoughtful discussions and comments on the manuscript. We also thank Michael R. Russo for assistance with the figures.

REFERENCES

1. Mendel, D., Cornish, V. W., and Schultz, P. G. (1995) Annu. Rev. Biophys. Biophys. Struct. 24, 435–462
2. Liu, D. R., Magliery, T. J., Pastarnak, M., and Schultz, P. G. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 10092–10097
3. Kober, C., Xie, L., Kellermeyer, S., Varshney, U., and RajBhandary, U. L. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 14310–14315
4. Liu, D. R., and Schultz, P. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4760–4765
5. Yoshikawa, E., Fournier, M. J., Mason, T. L., and Terrell, D. A. (1994) Macromolecules 27, 5471–5475
6. Hogue, C. W. V., Rasquinha, I., Szabo, A. B., and MacManus, J. P. (1992) FEBS Lett. 310, 269–277
7. Deming, T. J., Fournier, M. J., Mason, T. L., and Terrell, D. A. (1996) Macromolecules 29, 1442–1444
8. Powdne, L., and Richmond, M. H. (1963) Biochem. Biophys. Acta 71, 459–461
9. Powdne, L., Neule, S., and Trumam, H. (1963) Nature 199, 35–38
10. Deming, T. J., Fournier, M. J., Mason, T. L., and Terrell, D. A. (1994) Polym. Mater. Sci. Eng. Proc. 71, 673–674
11. Papas, T. S., and Mehler, A. H. (1970) J. Biol. Chem. 245, 1588–1595
12. Liu, H., and Musier-Forsyth, K. (1994) Biochemistry 33, 12708–12714
13. Kernsftilt, A. R., Umezawa, K., Vibe-Pedersen, K., and Baralle, F. E. (1985) EMBO J. 4, 1055–1059
14. Debynek, R., Jarrett, J. A., Chen, E. Y., Eaton, D. H., Bell, J. R., Assocen, R. K., Roberts, A. S., Scrpan, M. B., and Goeddel, D. V. (1985) Nature 316, 701–705
15. Wozney, J. M., Rosen, V., Celeste, A. J., Mitisck, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M., and Wang, E. A. (1988) Science 242, 1528–1534
16. Heacock, D., Forsyth, C. J., Shiba, K., and Musier-Forsyth, K. (1996) Bioorg. Chem. 24, 273–289
17. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
18. Neuman, R. E., and Logan, M. A. (1950) J. Biol. Chem. 185, 581–593
19. Roberts, A. B., Sporn, M. B., and Goeddel, D. V. (1985) Science 225, 547–554
20. Wada, K., and Ikemura, T. (1992) J. Biol. Chem. 267, 20129–20136
21. Wada, K. a., Wada, Y., Ishibashi, T., and Ikemura, T. (1992) J. Biol. Chem. 267, 20137–20143
22. Beuning, P. J., and Musier-Forsyth, K. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 8916–8920
23. Avel, I., Sthapathopoulos, C., Ambrogelly, A., Saurerwald, A., Teegood, H., Hartsch, T., and Sol, D. (2002) J. Biol. Chem. 277, 34743–34748
24. Wood, J. M. (1988) J. Membrane Biol. 106, 183–202
25. Glasker, E., Konings, W. N., and Poolman, B. (1996) J. Bacteriol. 178, 575–582
26. Gowrishankar, J. (1985) J. Bacteriol. 164, 443–445
27. Gowrishankar, J. (1986) J. Bacteriol. 166, 331–333
28. Rowland, J., and Trumam, H. (1975) J. Bacteriol. 123, 871–877
29. Milner, J. L., McClellan, D. J., and Wood, J. M. (1987) J. Gen. Microbiol 133, 1851–1860
30. Wada, K. a., Wada, Y., Ishibashi, T., and Ikemura, T. (1992) Nucleic Acids Res. 20, (suppl.) 2111–2118
31. Grosjean, H., and Fiers, W. (1982) Gene 18, 199–207
32. Berg, R. A., and Prockop, D. J. (1973) Biochem. Biophys. Res. Comm. 52, 113–120
33. Iino, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1982) Arch. Biochem. Biophys. 219, 198–203
34. Bann, J. G., and Bachinger, H. P. (2000) J. Biol. Chem. 275, 24466–24469
35. Bhatnagar, R. S., and Gough, C. A. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G. D., ed.) pp. 138–199, Plenum Press, New York
36. Ross, A., Dazacel, L. V., Zanaboni, G., Monzani, E., Dyne, K. M., Cetta, G., and Tenni, R. (1995) J. Mol. Biol. 250, 4780–4794
37. Horbek, K., and Blomberg, A. (1997) J. Biol. Chem. 272, 5544–5554
38. Watanabe, K., Masuda, T., Ohashi, H., Mihara, H., and Suzuki, Y. (1994) Eur. J. Biochem. 226, 277–283
39. Lang, K., and Schmid, F. X. (1990) J. Mol. Biol. 212, 185–196
40. Eberhardt, E. S., Nicholas Panasik, J., and Raines, R. T. (1996) J. Am. Chem. Soc. 118, 12251–12266
41. Holmgen, S. K., Taylor, K. M., Bretscher, L. E., and Raines, R. T. (1998) Nature 392, 666–667
Co-translational Incorporation of Trans-4-Hydroxyproline into Recombinant Proteins in Bacteria
Douglas D. Buechter, David N. Paolella, Bridget S. Leslie, Melissa S. Brown, Karen A. Mehos and Elliott A. Gruskin

J. Biol. Chem. 2003, 278:645-650.
doi: 10.1074/jbc.M209364200 originally published online October 23, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M209364200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 38 references, 15 of which can be accessed free at http://www.jbc.org/content/278/1/645.full.html#ref-list-1