Chemistry & Biology, Vol. 11, 1317–1324, September, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.07.011

Reactivity and Chemical Synthesis of *L*-Pyrrolysine—the 22nd Genetically Encoded Amino Acid

Bing Hao,[1,5] Gang Zhao,[2,5] Patrick T. Kang,[3] Jitesh A. Soares,[4] Tsuneo K. Ferguson,[4] Judith Gallucci,[2] Joseph A. Krzycki,[3,4] and Michael K. Chan[1,2,3,*]
[1]Department of Biochemistry
[2]Department of Chemistry
[3]Ohio State University Biochemistry Program
[4]Department of Microbiology
The Ohio State University
Columbus, Ohio 43210

Summary

***L*-pyrrolysine, the 22nd genetically encoded amino acid, was previously deduced to be (4R, 5R)-4-substituted-pyrroline-5-carboxylate attached to the ε-nitrogen of lysine based on the crystal structure of the *M. barkeri* monomethylamine methyltransferase (MtmB). To confirm L-pyrrolysine's identity, structures of MtmB have been determined following treatment with hydroxylamine, N-methylhydroxylamine, or dithionite. Analysis of these structures has provided additional support for the presence of the pyrroline ring and, together with previous mass spectroscopy data, has led us to assign the C_4-substituent to a methyl group. Based on this assignment, synthetic *L*-pyrrolysine was prepared by chemical methods. Detailed study of this chemically synthesized *L*-pyrrolysine has allowed us to characterize its physical properties, to study its chemical stability, and to elucidate the role of its C_4 substituent. Future applications of this synthetic *L*-pyrrolysine include its in vivo incorporation into recombinant proteins.**

Introduction

The first step in the growth of methanogens on methylamines (mono-, di-, and trimethylamine) is catalyzed by one of three classes of methylamine methyltransferases whose role is to position and activate the methyl group of the methylamine substrate for transfer to an associated corrinoid protein [1–7]. While the methanogen methyltransferases that act on different methylamine substrates exhibit no sequence homology, one surprising feature is the common presence of an in-frame amber (UAG) codon in each of their genes [3, 6, 8].

Recently, based on the structure of the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB), we have determined that translation of the UAG codon in these methyltransferases leads to insertion of a novel amino acid [9]. Analysis of electron density maps for two different crystal forms of the enzyme allowed us to suggest that this UAG-encoded amino acid is comprised of a 4-substituted pyrroline-5-carboxylate in amide linkage to the ε-nitrogen of lysine. It should be noted that for consistency, the ring numbering scheme used in this manuscript is for pyrrolines (Figure 1). For prolines, the ring numbering scheme is reversed. Based on its chemical composition, we named this amino acid *L*-pyrrolysine. The codiscovery of amber tRNA downstream of the *mtmB* gene that recognizes the UAG codon, and an aminoacyl-tRNA synthetase specific for *L*-pyrrolysine has allowed us to propose that *L*-pyrrolysine is the 22nd genetically encoded amino acid found in nature [9–11]—the 21st being the UGA-encoded selenocysteine [12].

Despite our assignment of *L*-pyrrolysine from the two structures of the MtmB protein, several questions remained to be addressed. One issue we wished to verify was the basic structure of *L*-pyrrolysine, particularly the presence of the pyrroline N_1-C_2 imine bond. Interest in obtaining additional support for this assignment stemmed from the fact that while the two previous structures of MtmB were determined to high resolution [1.55 Å for the NaCl crystal form; and 1.7 Å for the $(NH_4)_2SO_4$ crystal form], in each case, the electron density for the five-membered ring of *L*-pyrrolysine required fitting as a combination of two different orientations. For example, while the presence of the five-membered ring was readily apparent in the NaCl crystal form, in order to completely remove the surrounding difference density, a disordered model having 85% occupancy in one orientation (orientation 1), and 15% occupancy in the other (orientation 2), was required [9].

The electron density of the five-membered ring for the $(NH_4)_2SO_4$ crystal form could be fit to a disordered model that had the same two orientations as the NaCl crystal form, but with different occupancies: 40% in orientation 1, and 60% in orientation 2 (Figure 1). One distinct feature observed in orientation 2 of the $(NH_4)_2SO_4$ crystal form was the presence of electron density consistent with an additional atom attached to the C_2 position of the five-membered ring. This atom was assigned to an NH_2 group since the $(NH_4)_2SO_4$ precipitant was the only species not present in the NaCl crystal form. The apparent addition of amine to the C_2 carbon of five-membered ring in the $(NH_4)_2SO_4$ crystal form was significant because it provided the key evidence for the N_1-C_2 imine bond of the pyrroline ring and insight into the putative site of methylamine binding and activation. This assignment was also consistent with the observation that the side chains of Glu 229 and Glu 259 are perfectly positioned to serve as proton donors to the N_1 ring nitrogen. Protonation of the N_1 nitrogen would activate the C_2 carbon for nucleophilic attack [9].

Our recent efforts, described herein, have focused on strengthening the evidence for the identity of *L*-pyrrolysine, particularly the functionally critical imine bond, and toward synthesizing *L*-pyrrolysine to facilitate related functional studies. To address the first of these issues, we undertook the structure determination of complexes formed by reaction of MtmB with strong nucleophiles. It was predicted that the complexes formed with these substrates might exhibit more complete addition than ammonia, and thus greater uniformity, enabling a more definitive assignment of *L*-pyrrolysine's features. Alter-

*Correspondence: chan@chemistry.ohio-state.edu
[5]These authors contributed equally to this work.

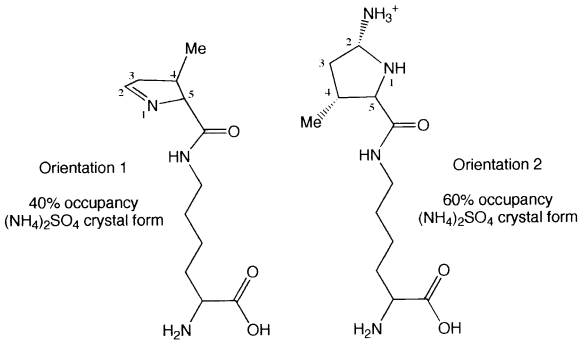

Figure 1. Simplified Stick Diagram of *L*-pyrrolysine

Left, orientation 1 in the absence of bound ligand; right, orientation 2 and bound to ammonia.

natively, it was hoped that nucleophiles containing a large atom, such as sulfur, could be used to provide additional support for the presence of two distinct orientations of the five-membered ring.

Toward this end, MtmB crystals were treated with hydroxylamine, N-methyl-hydroxylamine, or dithionite, and the structures of the resultant complexes were determined. In each case, nearly complete addition of the nucleophile to the C_2 carbon of *L*-pyrrolysine's five-membered ring was observed, providing strong support for the presence of the imine bond. Both hydroxylamine MtmB complexes were notable in that hydrogen bonding interactions between the added nucleophile and the protein side chains appeared to lock the pyrroline ring of *L*-pyrrolysine into a single conformation. Occupancy refinement of the C_4-substituent as a methyl, hydroxyl, or amino group for these hydroxylamine MtmB complexes, combined with electrospray MS measurements on the full-length MtmB protein [9], have led us to assign the C_4-substituent of *L*-pyrrolysine to a methyl group.

Having deduced the identity of the amino acid, we then prepared it by traditional chemical methods. Characterization of this synthesized *L*-pyrrolysine has provided potential insight into the role of its C_4-substituent, and has facilitated the study of its chemical stability. This synthesized *L*-pyrrolysine has also recently been shown to exhibit biological activity in that it promotes readthrough of the *mtmB1* UAG codon in *E. coli* coexpressing the genes encoding the *L*-pyrrolysine tRNA and aminoacyl-tRNA synthetase [11]—presumably due to ribosomal incorporation of the chemically synthesized *L*-pyrrolysine into the recombinant MtmB protein.

Results and Discussion

Overall Fold of MtmB Ligand Complexes

The overall fold of the three MtmB complexes is nearly identical to the native structures reported previously (Figure 2) [9]. The protein consists of a D_3 symmetric α_6-hexamer with the fold of each monomer consistent with a TIM barrel. Surface potential diagrams reveal that the barrel forms a negatively charged cavity at the bottom of which lies the *L*-pyrrolysine amino acid.

One feature worth noting is the appearance of a partially occupied disulfide bridge formed by Cys 341 and

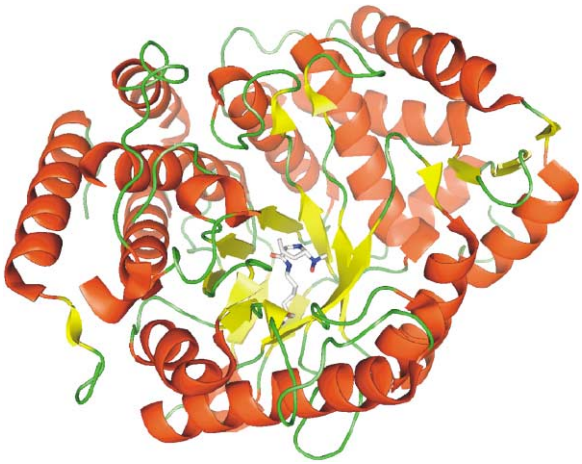

Figure 2. Overall Fold of MtmB Methyl-hydroxylamine Complex with α Helices in Red, and β Sheets in Yellow

The *L*-pyrrolysine amino acid bound to methyl-hydroxylamine is shown in stick form with carbons colored in gray, oxygens in red, and nitrogens in blue.

Cys 428 in each of the three MtmB complex structures. This disulfide bond was not present in the original 1.55 Å native MtmB structure (NaCl crystal form). Based on the observation that the ratio of dithiol/disulfide species varies in different samples, we speculate that the disulfide bond arises via slow air oxidation. Whether the close proximity of the cysteine residues and their ability to form a disulfide bond is functionally relevant, or is simply coincidental, remains to be determined.

Structure of the MtmB Sulfite Complex: Support for Two Distinct Orientations of *L*-Pyrrolysine's Pyrroline Ring

The first MtmB ligand complex was obtained by treatment of native MtmB crystals (NaCl crystal form) with dithionite [9]. Structural analysis of the resulting complex at 1.8 Å resolution revealed the presence of a sulfite group attached via its sulfur atom to the C_2 carbon of the pyrroline ring (Figure 3). The site and face of sulfite

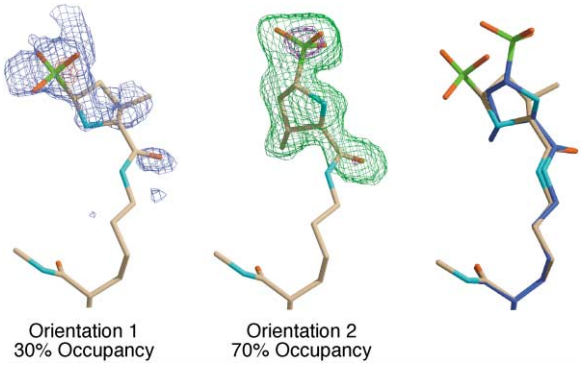

Figure 3. MtmB Sulfite Complex Active Site

Left, $F_o - F_c$ omit map (blue, 2σ; red, 6σ) with fit to orientation 2, and (middle) $F_o - F_c$ omit map (green, 4σ; purple 12σ) with fit to orientation 2. Carbons are colored in tan, oxygens in red, and nitrogens in cyan. Right, overlap of orientation 1 (tan) and orientation 2 (blue).

addition are similar to that observed for ammonia addition in the $(NH_4)_2SO_4$ crystal form, as is the presence of the two distinct orientations of the pyrroline ring. One major difference, however, is that while ammonia addition was only observed in orientation 2 of the $(NH_4)_2SO_4$ crystal form, in the structure of the MtmB dithionite product, both conformations contain bound sulfite. This observation is supported by the presence of two peaks in the $2F_o - F_c$ electron density map corresponding to the positions of the sulfur atoms in the two pyrroline ring orientations. Given that ammonia addition appeared to stabilize the 2nd conformation, one might predict that the 2nd conformation would also be preferred in the MtmB sulfite complex. This is indeed the case (orientation 1:orientation 2 = 30%:70%), with the MtmB sulfite complex being the form with the highest occupancy of orientation 2 characterized to date. As in the previous MtmB structures, the dual conformations of the pyrroline ring are reflected in the disorder of the surrounding protein residues in a fashion that can be modeled [9].

The significance of the MtmB sulfite complex structure is that it provides additional evidence for a reactive site at the C_2 position of the five-membered ring, and further supports the inherent ability of *L*-pyrrolysine to adopt different conformations in both ligated and unligated states. These results also provide a rationale for excluding dithionite from the purification of proteins containing the *L*-pyrrolysine amino acid.

Structures of Two MtmB Hydroxylamine Complexes: Probing the Identity of the C4 Substituent

As mentioned previously, another major goal has been to obtain structures of MtmB with the pyrroline ring of *L*-pyrrolysine locked into a single orientation. In light of the apparent ability of ammonia to alter the orientation of this ring from orientation 1 to orientation 2, one of the first approaches explored was the use of hydroxylamines to promote complete conversion to the second form. Since hydroxylamines are known to be better nucleophiles than amines due to the α effect [13, 14], it was expected that they would add rapidly with the activated imine bond of *L*-pyrrolysine upon soaking. Moreover, it was thought that hydroxylamines would exhibit faster "on" rates than amines, resulting in a shift of the equilibrium to the bound state potentially leading to its complete addition.

The structure of the hydroxylamine MtmB complex at 2.0 Å resolution (Figure 4) reveals that, like ammonia, hydroxylamine can add to the imine carbon. As predicted, the hydroxylamine-bound *L*-pyrrolysine adopts a much more ordered conformation that provides clear electron density for both the added hydroxylamine substrate and also the C_4 substituent. There are unexpected differences in the mode of addition, however, with hydroxylamine adding more to the opposite face of the pyrroline ring than ammonia (Figure 4). This feature results in a different mode of interaction with the protein pocket. While in the previous structure of MtmB determined in the presence of $(NH_4)_2SO_4$, ammonia addition appeared to shift the equilibrium toward orientation 2; in this case, hydroxylamine addition shifts the equilibrium to a state that is more like orientation 1. The origin of this preference likely stems from the fact that in this orientation, the hydroxylamine adduct is stabilized by hydrogen bonding interactions between the oxygen of hydroxylamine and the amide NH that links the pyrroline ring to the ϵ-nitrogen of lysine of *L*-pyrrolysine, and ionic interactions between the added hydroxylamine and carboxylate side chain of Glu 205 (Figure 4).

The adduct formed from hydroxylamine addition to *L*-pyrrolysine was modeled with the hydroxylamine nitrogen atom bound to the imine carbon of the pyrroline ring. This ligation was predicted based on the greater nucleophilicity of the amine nitrogen over the hydroxyl oxygen. Further support for this orientation could be obtained by comparing this model to that with the hydroxyl group attached to the pyrroline ring. While the electron density maps with hydroxylamine bound through its nitrogen appeared reasonable, $F_o - F_c$ maps for the model with the hydroxylamine bound through its oxygen exhibited negative 3σ density.

To obtain a definitive assignment for the mode of hydroxylamine addition, however, the structure of the N-methyl-hydroxylamine MtmB complex was determined (Figure 4). The binding of N-methyl-hydroxylamine in this 2.2 Å resolution structure is similar to that of the hydroxylamine adduct except for the presence of an additional methyl group. The location of the extra electron density on the hydroxylamine atom attached to the pyrroline ring provides additional support that nucleophilic addition of hydroxylamines occurs via their amine nitrogen.

Perhaps the most important feature of the two hydroxylamine MtmB complexes is the fact that hydroxylamine addition stabilizes the *L*-pyrrolysine five-membered ring in a single conformation. This assertion is based on the absence of difference electron density around the ring, and from results obtained by occupancy refinement of the pyrroline ring both as a group and as individual atoms (Supplemental Table S1, available online at http://www.chembiol.com/cgi/content/full/11/9/1317/DC1). The uniformity in these hydroxylamine structures is significant given the inherent difficulties that the dual conformation creates in characterizing the amino acid.

In particular, while the magnitude of the electron density for the C_4-substituent of the pyrroline ring in the previous native and ammonia bound structures was consistent with a methyl, hydroxyl, or amino group, due to the disorder, we were unable to make a more definitive assignment [9]. Taking advantage of the single pyrroline conformation for both the hydroxylamine and N-hydroxylamine MtmB complexes, occupancy refinement for the atoms of their L-pyrrolysine pyrroline rings was performed with the C_4 substituent as a methyl, hydroxyl, or amino group [9]. In both cases, a methyl group provided the best fit for the identity of the C_4-substituent.

For the hydroxylamine MtmB complex, for example, group occupancy refinement of the five-membered ring as a 4-methyl-pyrroline yielded a value of 1.01, consistent with full occupancy of the ring. Subsequent occupancy refinement as individual atoms gave an occupancy of 0.87 for the C_4-methyl substituent (Supplemental Table S7). Conversely, occupancy refinement as a 4-hydroxypyrroline resulted in a much lower occupancy (occu-

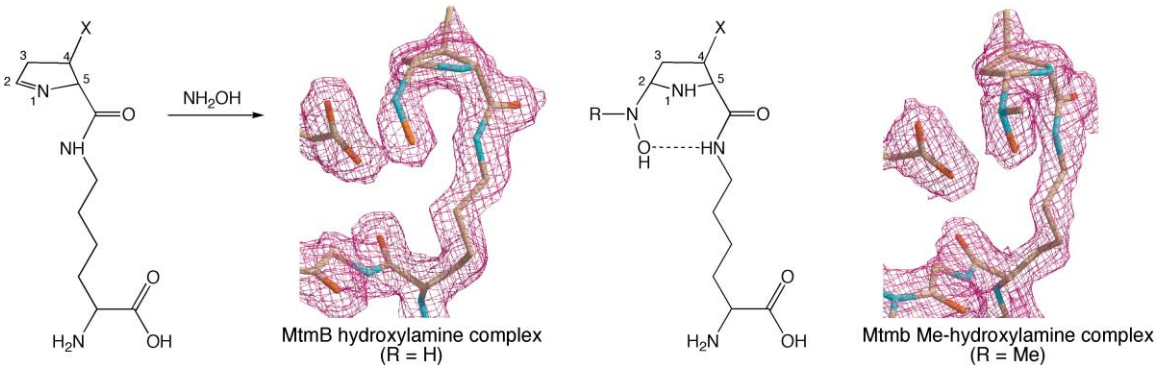

Figure 4. MtmB Active Site Following Reaction with Hydroxylamines

Left product, $2F_o - F_c$ electron density map (1.5σ) of MtmB hydroxylamine complex. Middle, Stick diagram of *L*-pyrrolysine bound to hydroxylamine. Right, $2F_o - F_c$ electron density map (1.5σ) of MtmB methyl-hydroxylamine complex. Carbons are colored in tan, oxygens in red, and nitrogens in cyan.

pancy = 0.58) for the oxygen atom of the hydroxyl group - indicating that the fitted atom was too large for the experimental electron density. Similar refinements with an amine as the C_4-substituent were slightly better, but the occupancy of the amine nitrogen (occupancy = 0.69) was still significantly lower than the other atoms of the pyrroline ring. Nearly the same results are obtained from analysis of the MtmB methyl-hydroxylamine complex. Together, these occupancy refinement studies support a methyl group as the identity of the C_4-substitutent, with only a slight possibility of it being an amino group.

In order to explore the possibility of an amino group more critically, we examined the potential hydrogen bonding interactions of the C_4-substituent with the protein. If the C_4-substituent was an amine, it should be protonated and should prefer to interact with carboxylate side chains that lie near the active site. Analysis of the various MtmB structures revealed no evidence for a salt bridge in any the conformations of *L*-pyrrolysine. While we initially considered the possibility of a potential hydrogen-bonding interaction of the pyrroline C_4-substituent in orientation **1** with the hydroxyl of Tyr 335, further examination revealed these interactions to be long, ~3.1 Å, and the position of Tyr 335 to be generally unaffected by the different pyrroline ring orientations. Thus we were unable to find evidence for the C_4-substituent being a protonated amino group. These data combined with the electrospray MS data of the full-length protein, 50,105 ± 2 Da (methyl group [50,107 Da], amino group [50,108 Da], hydroxyl group [50,109 Da]) [9], have led us to assign to a 4-methyl-pyrroline-5-carboxylate attached to the ε-nitrogen of lysine.

Chemical Synthesis of *L*-Pyrrolysine

With L-pyrrolysine chemically identified, we wished to isolate it in order to characterize its properties and to enable studies directed toward elucidating its mechanism of incorporation into proteins. Since *L*-pyrrolysine had not been previously obtained by either biochemical or synthetic methods, it was necessary to develop a strategy by which to prepare it. We report here our progress toward this objective, with the first chemical synthesis of *L*-pyrrolysine.

The main challenge in the chemical synthesis of *L*-pyrrolysine was the preparation of the (*4R*, *5R*)-4-methyl-pyrroline-5-carboxylic acid. While proline derivatives would appear to be ideal precursors for preparing the pyrroline ring, we were unable to identify a method by which to specifically oxidize it to give a pyrroline with the desired N_1-C_2 double bond (Figure 1). One route that might initially be considered is the Shono oxidation [15]. The Shono oxidation can be used to oxidize N-acetylated prolines to give an eneamide that upon deprotection yields an **enamine** with a C_2-C_3 pyrroline double bond. In contrast, the pyrroline in *L*-pyrrolysine contains an **imine** with an N_1-C_2 double bond (Figure 1). This tendency to form a C_2-C_3 double bond is a general problem for most of the nonenzymatic methods for converting prolines to pyrrolines [16]. For those transformations of proline that do form an imine bond, the problem is that these methods generally lead to an imine bond between the N_1-C_5 bond, not the N_1-C_2 bond [17, 18]. Thus these approaches to generate the desired pyrroline ring from proline are nontractable, since they lead to the wrong regioisomers.

A viable strategy, however, appeared to be via synthesis of the 4-Me-substituted glutamate γ-semialdehyde, which upon cyclization, would give the desired pyrroline ring. Following the methodologies developed by Belokon et al. (Figure 5) [19–22], asymmetric Michael addition of (*E*)-2-butenal to the glycine moiety of the nickel complex **4** gave the nickel-associated glutamate γ-semialdehyde **5** with the desired stereochemistry as confirmed by single crystal X-ray analysis. Degradation of **5** in acidic methanol yielded the 4-methyl-pyrroline methyl ester **6**. Hydrolysis and subsequent amide coupling of the generated carboxylate with amine **2** gave the trifluoroacetate-protected *L*-pyrrolysine **7**. Deprotection of **7** following treatement with LiOH gave *L*-pyrrolysine **8** as the lithium salt. The detailed experimental procedures are provided as supplemental data (see http://www.chembiol.com/cgi/content/full/11/9/1317/DC1).

As the carbanion intermediate generated by removal of the C_5 proton would be stabilized by conjugation with

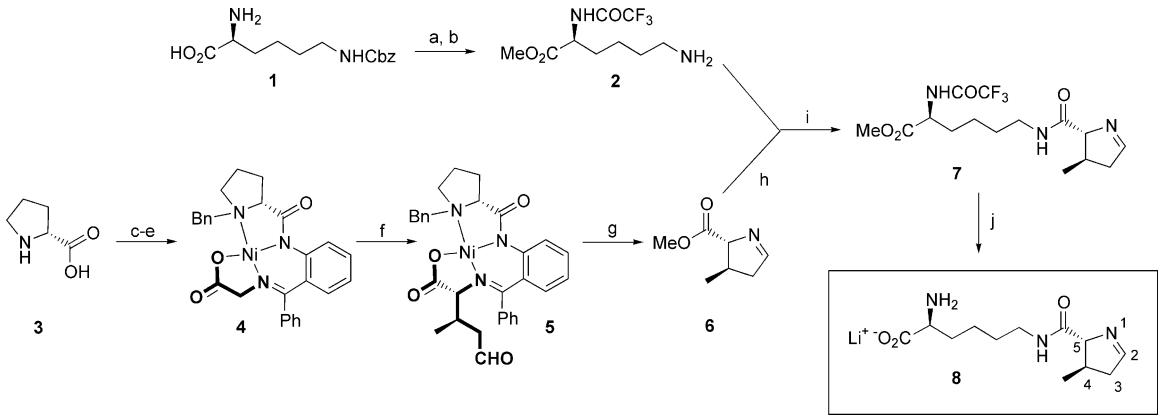

Figure 5. Synthetic Scheme Used to Synthesize *L*-Pyrrolysine

Reagents and conditions: (a) TFAA, Et_3N, CH_2Cl_2 rt, 85%; (b) H_2, Pd/C, MeOH, rt, 99%; (c) ^{i}PrOH, KOH, BnCl, rt, 83%; (d) $SOCl_2$, CH_2Cl_2, then *o*-aminobenzophenone, rt, 89%; (e) glycine, $Ni(NO_3)_2$, KOH, MeOH, reflux, 96%; (f) DBU, CH_2Cl_2, crotonaldehyde, rt, 97%; (g) HCl (concd.), MeOH, reflux, then TMSCl, MeOH, rt, 43%; (h) LiOH, THF:H_2O (3:1), rt, ~100%; (i) **2**, DPPA, Et_3N, DMF, rt, 31%; (j) LiOH, THF-MeOH-H_2O (2:2:1), rt, 98%.

the imine and amide pi-bonds, we were concerned that the likely acidity of the C_5 proton could lead to deprotonation and subsequent epimerization during the base deprotection step (Scheme 1, step j). Molecular mechanics calculations using MOPAC (Cambridge Soft: Chem3D PRO), however, predict that the form of *L*-pyrrolysine with a *trans* geometry for the C_4-methyl and C_5-amide substituents has a 5 kcal/mol lower heat of formation than the corresponding *cis* isomer. Thus one might expect that the greater thermostability of the *trans* geometry could hinder epimerization—a result that would be significant in that it would provide a potential role and explanation for the presence of the C_4-substituent in *L*-pyrrolysine.

The expected lability of the C_5 proton was verified by dissolving compound **7** in CD_3OD, and monitoring the loss of the C_5 proton resonance in the presence of catalytic NaOD, and its reappearance following a 1-day MeOH incubation (Figure 6). To test whether the obligate carbanion intermediate leads to epimerization, COSY and NOESY experiments were carried out on the purified *L*-pyrrolysine lithium salt **8**. These studies revealed that while the C_5 proton did undergo exchange, the *trans* geometry of the C_4-methyl substituent and the C_5-carboxylate was maintained—thus no epimerization took place.

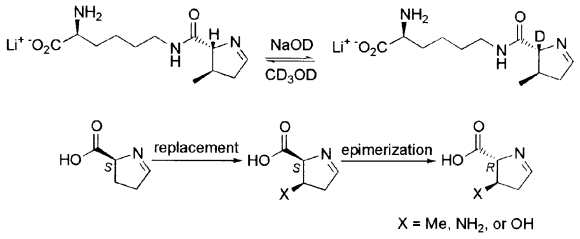

Figure 6. Potential Insight into the Role of the C_4-Substituent

Top, treatment of *L*-pyrrolysine with NaOD in MeOH leads to deprotonation, but no epimerization. Bottom, one possible pathway for *L*-pyrrolysine biosynthesis based on the chiral stability provided by the C_4-substituent.

These findings have important implications for the role and presence of the C_4-substituent in *L*-pyrrolysine. In addition to the likely role of the C_4-substituent in maintaining the stereointegrity at the otherwise chirally labile C_5 site, the greater stability of the *trans* geometry suggests one plausible route for the biosynthesis of the 4-substituted (4*R*, 5*R*)-pyrroline-5-carboxylate component from natural (*S*)-configured amino acids. For example, enantioselective replacement of a methyl group onto C_4 of *L*-pyrroline-5-carboxylate, an intermediate in the interconversion between proline and glutamate, followed by epimerization at C_5 would yield the 4-substituted (4*R*, 5*R*)-pyrroline-5-carboxylate component in two steps, as would epimerization followed by replacement (Figure 6). While the brevity of this route would be consistent with the limited number of proteins encoded in the gene cluster associated with *L*-pyrrolysine biosynthesis [10], it should be emphasized that there are other precursors and pathways that nature could use to synthesize *L*-pyrrolysine.

With successful synthesis of *L*-pyrrolysine confirmed, the next step was to characterize its physical properties. While most amino acids have no absorption at 280 nm, Trp and Phe absorb in this range due to π-π^* transitions of their aromatic rings, and thus one might expect *L*-pyrrolysine to have similar absorptions due to the Schiff base moiety. Indeed, in methanol, **8** exhibited a shoulder at 310 nm (ϵ, 35.4 $M^{-1}cm^{-1}$). In water, however, no major absorptions above 220 nm were observed.

Another issue of interest was the chemical stability of *L*-pyrrolysine. Given the presence of the imine and amide bonds in *L*-pyrrolysine, it might be expected to be susceptible to extremes of pH. Additionally, previous attempts to isolate MtmB peptide fragments containing the UAG-encoded amino acid found only lysine [8]. Thus, insight into chemical stability of *L*-pyrrolysine was thought to be of importance to guide attempts to isolate it from biological sources.

One observation made during the synthesis of *L*-pyrrolysine was the inherent instability of the pyrroline ring. Purified compound **6** decomposed within one week at 4°C to give a complex mixture. The pyrroline ring was

Table 1. Data Processing and Refinement Statistics

Data set	Dithionite	$NH_2(OH)$	Me-NH(OH)
Space group	$P6_322$	$P6_322$	$P6_322$
Cell parameters (Å)	a = b = 158.82, c = 136.33	a = b = 158.17, c = 136.21	a = b = 157.99, c = 136.91
Wavelength (Å)	0.8270	0.9920	1.0800
Resolution range (Å)	20.0–1.8	20.0–2.0	20.0–2.2
Observations/unique	515,977/92,028	481,711/67,189	313,011/51,381
Completeness (%)[a]	98.3 (95.9)	98.9 (99.6)	99.9 (100.0)
R_{sym} (%)[b]	6.5 (34.7)	17.7 (59.3)	10.6 (33.5)
No. of protein (additive) atoms	3491 (12)	3491 (10)	3491 (11)
No. of solvent molecules	471	462	461
$R_{crystal}$[c] (R_{free})[d] (%)	17.4 (18.6)	16.5 (18.2)	14.5 (16.7)
B factor of protein (Å^2)	19.5	21.7	25.4
B factor of pyrrolysine (Å^2)	21.4 (24.2)[e]	24.1	26.4
Rmsd of bond distance (Å)	0.005	0.005	0.005
Rmsd of bond angles (°)	1.187	1.172	1.171

[a] The numbers in parentheses are for the highest resolution shell.
[b] R_{sym} (I) = $\Sigma_h\Sigma_i|I_i - I|/\Sigma_h\Sigma_i$ I, where I is the mean intensity of the *i* observations of reflection *h*.
[c] $R_{crystal} = 100 \times \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.
[d] R_{free} is the same as $R_{crystal}$, and was calculated using 10% of the data excluded from refinement.
[e] The number in parentheses is for the 2nd conformation of pyrrolysine.

stabilized, however, upon amide linkage to lysine. Protected **7** and deprotected forms of *L* -pyrrolysine **8** were much more stable at 4°C and neutral pH, although over time, decomposition was still observed.

As amide bonds are base sensitive, it appeared likely that *L*-pyrrolysine would degrade at high pH. To examine these issues, the synthetic *L*-pyrrolysine was treated with either NaOH or LiOH and TLC was used to monitor its stability. While the stability of *L*-pyrrolysine was fairly stable under weakly basic conditions, at high pHs partial degradation of the amino acid could be slowly observed over a period of days. In the presence of 1 M NaOH, a new band formed over time with an R_f corresponding to free lysine. Treatment of **8** with 1 M LiOH also led to degradation, although here two bands were observed on the TLC plate. One band was determined to be a mixture of products by NMR, while the other band was confirmed to be lysine by its R_f, MS, and NMR. Although free pyrroline-5-carboxylate could not be detected, the appearance of lysine demonstrates that amide cleavage of *L*-pyrrolysine occurs at high pH. Based on the slow rate of this degradation, however, *L*-pyrrolysine appears to be able to tolerate short-term exposure to even strongly basic conditions.

L-Pyrrolysine was found to be extremely sensitive to acid. This instability was initially suggested during attempts to prepare the amino acid using Boc-protected lysine (Supplemental Figure S1, available at http://www.chembiol.com/cgi/content/full/11/9/1317/DC1). While amide coupling of the pyrroline carboxylate to the Boc-protected lysine could be achieved, Boc deprotection under standard acidic conditions (TFA in CH_2Cl_2) failed, leading to a mixture of decomposition products.

This observation was significant, since previous attempts to isolate fragments of the MtmB to identify the UAG amino acid used a TFA/CH_3CN gradient as the eluent [8]. Importantly, following purification of the target peptide fragments, the TFA/CH_3CN eluent was removed by evaporation, meaning that these fragments were exposed to fairly high concentrations of TFA. Given that characterization of the peptide fragment containing the UAG encoded amino acid gave MS and Edman degradation data consistent with lysine, one fundamental question was whether loss of the pyrroline component was due to this TFA exposure [8].

To evaluate apparent instability of *L*-pyrrolysine to TFA, the chemically synthesized *L*-pyrrolysine **8** was placed in a 10% v/v TFA/MeOH solution and monitored by TLC. Under these conditions, *L*-pyrrolysine was found to undergo rapid conversion to give a new, more polar band. Purification and characterization of this band by MS and NMR suggested that, while the major product contained both the lysine and pyrroline components, its pyrroline ring was modified—possibly due to tautomerization of the N_1-C_2 imine to either the C_2-C_3 enamine or N_1-C_5 imine. Thus while TFA treatment does not appear to lead to separation of *L*-pyrrolysine's lysine and pyrroline components, it does result in the decomposition of *L*-pyrrolysine to a new species.

These results have important implications on the study of *L*-pyrrolysine. For example, electrospray MS data of peptides are collected in the presence of a small amount of a weak organic acid to generate the positively charged ion. In light of the extreme acid sensitivity of *L*-pyrrolysine, these results suggest that while acid treatment may not necessarily alter the measured mass, the species being measured following such acid exposure may not in fact be *L*-pyrrolysine. While the TFA concentrations used to study the stability of *L*-pyrrolysine are higher than typical MS experiments, they nevertheless highlight the importance of maintaining neutral pH conditions when attempting to study this amino acid.

Perhaps the most exciting finding that resulted from the chemical synthesis of *L*-pyrrolysine, however, was the demonstration that its presence could promote readthrough of the *mtmB* UAG codon in *E. coli*. coexpressing the genes encoding the *L*-pyrrolysine tRNA and pyrrolysyl-tRNA synthetase [11]. Together with the previously reported mass spectroscopic measurements

[9] and the crystallographic data presented within this manuscript, these findings build a strong case for *L*-pyrrolysine's correct identification.

Significance

The 22nd genetically encoded amino acid, *L*-pyrrolysine was first identified based on the electron density from two different crystal forms of a methanogen monomethylamine methyltransferase (MtmB). While *L*-pyrrolysine was disordered in both crystal forms, the electron density could be modeled as two distinct conformations of a 4-substituted-(4R, 5R)-pyrroline-5-carboxylate attached to the ϵ-nitrogen of lysine. The key evidence for the presence of the pyrroline imine bond was based on the observed partial (60%) addition of ammonia to the pyrroline C_2 carbon in one of the crystal forms. The work detailed in this manuscript provides additional support for its current assignment as determined from structures of MtmB following reaction with hydroxylamine, N-methyl-hydroxylamine, or dithionite. In each case, nearly complete addition of the substrate to the C_2 position pyrroline ring is observed. Importantly, in the structures of both hydroxylamine MtmB complexes, the *L*-pyrrolysine's pyrroline ring adopts a single conformation. Occupancy refinement of the C_4-substituent for these complexes as either a methyl, amino, or hydroxyl group, together with previous electrospray MS data on the MtmB protein have led us to assign the C_4-substituent as a methyl group. Based on this assignment, the 4-methyl substituted form of *L*-pyrrolysine has been prepared. Studies of this synthetic *L*-pyrrolysine have facilitated an analysis of its chemical stability, and have yielded insight into the role of its C_4-substituent—namely to provide chiral stability at the otherwise chirally labile C_5 site. This chemically synthesized *L*-pyrrolysine has been shown to promote in vivo readthrough of the *mtmB1* UAG codon in *E. coli* coexpressing the genes encoding *L*-pyrrolysine's tRNA and pyrrolysyl-tRNA synthetase [11].

Experimental Procedures

Crystallization and Derivative Preparation

The MtmB crystals were grown using a reservoir solution containing 4.3 M NaCl, and 0.1 M HEPES (pH 7.5) as previously described [9]. The additive-bound crystals were prepared by soaking the MtmB crystals under various conditions. Hydroxylamine and N-methyl-hydroxylamine solutions were neutralized before being added to the synthetic mother liquor. The hydroxylamine-bound crystals were obtained by soaking the MtmB crystals with solutions of synthetic mother liquor containing 1.0 M hydroxylamine for 7 hr. The N-methyl-hydroxylamine-bound crystals were prepared by treating the MtmB crystals with 0.5 M N-methyl-hydroxylamine for approximately 5 hr. The dithionite-bound crystals were obtained by transferring the MtmB crystals to degassed synthetic mother liquor containing saturated sodium dithionite for a few minutes. Following chemical treatment, each of the above complex crystals were transferred stepwise through synthetic mother liquor solutions containing increasing amounts of glycerol up to a maximal concentration of 30% (vol/vol) prior to being flash-cooled in liquid nitrogen.

Data Collection and Structure Refinement

The diffraction data were collected at beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL), and the BIOCARS 14D and SBC 19BM beamlines at the Advance Photon Source (APS) at Argonne National Laboratory. Data processing and reduction were performed by either using the program HKL2000, or the related programs DENZO and SCALEPACK (Table 1) [23]. The structure of each additive-bound complex was determined using the native MtmB structure (PDB entry: 1NTH [9]) as the starting model. Model building and refinement were performed using the programs O [24] and CNS [25], respectively (Table 1). The additives were modeled based on the resulting $2F_o - F_c$ and $F_o - F_c$ electron density maps. The figures depicting the protein molecule and its electron density were prepared with the programs PyMOL [26], Xtalview [27], and Raster 3D [28].

Occupancy refinement of the pyrroline ring both as a group and as individual atoms was performed using the program CNS [25]. The occupancies of the C_4-substituent as a methyl, amino, and hydroxyl group were determined by two methods. The first method involved refining the coordinates and thermal values with three different substituents followed by occupancy refinement. While this method was effective, it was found that some of the occupancy differences were masked by increased thermal parameters. Thus for the second method, the MtmB coordinates with a methyl group C_4-substituent was used as the starting model. Following alteration of the substituent to either an amino or hydroxyl group, the coordinates were refined, but no refinement of the thermal factors was performed prior to the individual occupancy analysis. These data, which are reported in Supplemental Table S1, clearly support a methyl group as the identity of the C_4-substituent.

Supplemental Data

Supplemental Data for this article is available online at http://www.chembiol.com/cgi/content/full/11/9/1317/DC1.

Acknowledgments

We thank Dr. Tomasz Fekner and Dr. Sunney I. Chan for their detailed comments and insights. This work was supported by grants from the National Institutes of Health (GM 061796) to M.K.C. and J.A.K., and the National Science Foundation (MCB-9808914) and the Department of Energy (DE-FG02-92ER20042) to J.A.K. Work was done partially at the Stanford Synchrotron Radiation Laboratory (SSRL), operated by the Department of Energy, Office of Basic Energy Sciences. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources.

Received: May 11, 2004
Revised: July 12. 2004
Accepted: July 29, 2004
Published: September 17, 2004

References

1. Burke, S.A., and Krzycki, J.A. (1995). Involvement of the "A" isozyme of methyltransferase II and the 29-kilodalton corrinoid protein in methanogenesis from monomethylamine. J. Bacteriol. *177*, 4410–4416.
2. Ferguson, D.J., Jr., Krzycki, J.A., and Grahame, D.A. (1996). Specific roles of methylcobamide:coenzyme M methyltransferase isoenzymes in metabolism of methanol and methylamines in *Methanosarcina barkeri*. J. Biol. Chem. *271*, 5189–5194.
3. Paul, L., and Krzycki, J.A. (1996). Sequence and transcript analysis of a novel *Methanosarcina barkeri* methyltransferase II homolog and its associated corrinoid protein homologous to methionine synthase. J. Bacteriol. *178*, 6599–6607.
4. Burke, S.A., and Krzycki, J.A. (1997). Reconstitution of monomethylamine:coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. J. Biol. Chem. *272*, 16570–16577.
5. Ferguson, D.J., Jr., and Krzycki, J.A. (1997). Reconstitution of trimethylamine-dependent coenzyme M methylation with the

trimethylamine corrinoid protein and the isoenzymes of methyltransferase II from *Methanosarcina barkeri*. J. Bacteriol. *179*, 846–852.

6. Burke, S.A., Lo, S.L., and Krzycki, J.A. (1998). Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. J. Bacteriol. *180*, 3432–3440.
7. Ferguson, D.J., Jr., Gorlatova, N., Grahame, D.A., and Krzycki, J.A. (2000). Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. J. Biol. Chem. *275*, 29053–29060.
8. James, C.M., Ferguson, T.K., Leykam, J.F., and Krzycki, J.A. (2001). The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina barkeri* is translated as a sense codon. J. Biol. Chem. *276*, 34252–34258.
9. Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science *296*, 1462–1466.
10. Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459–1462.
11. Blight, S., Larue, R., Mahapatra, A., Longstaff, D., Chang, E., Zhao, G., Kang, P., Green-Church, K.B., Chan, M.K., and Krzycki, J.A. (2004). Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*.
12. Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991). Selenocysteine: the 21st amino acid. Mol. Microbiol. *5*, 515–520.
13. Grekov, A.P., and Veselov, V.Y. (1978). α-Effect in the chemistry of organic compounds. Russ. Chem. Rev. *47*, 631–648.
14. Fina, N.J., and Edwards, J.D. (1973). Alpha effect. Review. Int. J. Chem. Kinet. *5*, 1–26.
15. Shono, T., Matsumura, Y., Tsubata, K., Sugihara, Y., Yamane, S., Kanazawa, T., and Aoki, T. (1982). Electroorganic chemistry. 60. Electroorganic synthesis of enamides and enecarbamates and their utilization in organic synthesis. J. Am. Chem. Soc. *104*, 6697–6703.
16. Baldwin, J.E., Bamford, S.J., Fryer, A.M., Rudolph, M.P.W., and Wood, M.E. (1997). Towards a versatile synthesis of kainoids. I: Introduction of the C-3 and C-4 substituents. Tetrahedron *53*, 5233–5254.
17. Haeusler, J. (1987). A convenient synthesis of (R)-γ-amino-β-hydroxybutanoic acid (GABOB) from natural (2S,4R)-4-hydroxyproline. Monatsh. Chem. *118*, 865–869.
18. Ezquerra, J., Escribano, A., Rubio, A., Remuinan, M.J., and Vaquero, J.J. (1995). New enantioselective approach to α-allokainoids by Michael addition to chiral 4-substituted 2,3-didehydroprolinate. Tetrahedron Lett. *36*, 6149–6152.
19. Belokon, Y.N., Sagyan, A.S., Dzhamgaryan, S.M., Bakhmutov, V.I., and Belikov, V.M. (1988). Asymmetric synthesis of β-substituted α-amino acids via a chiral nickel(II) complex of dehydroalanine. Tetrahedron *44*, 5507–5514.
20. Belokon, Y.N., Bulychev, A.G., Ryzhov, M.G., Vitt, S.V., Batsanov, A.S., Struchkov, Y.T., Bakhmutov, V.I., and Belikov, V.M. (1986). Synthesis of enantio- and diastereo-isomerically pure β- and γ-substituted glutamic acids via glycine condensation with activated olefins. J. Chem. Soc. Perkin Trans. *1*, 1865–1872.
21. Belokon, Y.N., Tararov, V.I., Maleev, V.I., Savel'eva, T.F., and Ryzhov, M.G. (1998). Improved procedures for the synthesis of (S)-2-[N-(N′-benzyl-prolyl)amino]benzophenone (BPB) and Ni(II) complexes of Schiff's bases derived from BPB and amino acids. Tetrahedron Asymmetry *9*, 4249–4252.
22. Belokon, Y.N., Popkov, A.N., Chernoglazova, N.I., Saporovskaya, M.B., Bakhmutov, V.I., and Belikov, V.M. (1988). Synthesis of a chiral nickel(II) complex of an electrophilic glycinate, and its use for asymmetric preparation of α-amino acids. J. Chem. Soc. Chem. Commun., 1336–1338.
23. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
24. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
25. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
26. DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).
27. McRee, D.E. (1999). XtalView Xfit—A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. *125*, 156–165.
28. Merritt, E., and Murphy, M. (1994). Raster3D version 2.0—a program for photorealistic molecular graphics. Acta Crystallogr. D *50*, 869–873.

Accession Numbers

The coordinates for each of the three structures have been deposited in the Protein Data Bank. PDB IDs: 1TV4 (sulfite MtmB complex), 1TV2 (hydroxylamine MtmB complex), and 1TV3 (N-methyl-hydroxylamine MtmB complex).