barkeri Fusaro and *D. hafniense pyl* gene products not shown in the figures are as follows: PylB, 63%; PylC, 49%; and PylD, 45% (see fig. S2). In each case, >90% of the two genes were involved in the alignment, and the Blosum62 matrix was used to establish similarity of residues in alignments.

23. A. J. Gale, P. Schimmel, *Biochemistry* **34**, 8896 (1995).
24. C. R. Woese, G. J. Olsen, M. Ibba, D. Söll, *Microbiol. Mol. Biol. Rev.* **64**, 202 (2000).
25. BLASTP searches at the National Center for Biotechnology Information Web site against the nonredundant databases demonstrated a strong similarity of the predicted *pylB* gene product to biotin synthase from a number of sources (alignments at 49% similarity and expect values of 3e-29), which may reflect involvement in formation of five-membered heterocylic rings. An RPS-BLAST search indicated a fainter but significant relationship to D-alanine:D-alanine ligases in pfam01820 for the *M. barkeri* (expect value 0.005) and *D. hafniense* (expect value 10e-6) PylC proteins, suggestive of a potential role in forming the amide bond in pyrrolysine.
26. M. Ibba, D. Söll, *Annu. Rev. Biochem.* **69**, 617 (2000).
27. S. Commans, A. Bock, *FEMS Microbiol. Rev.* **23**, 335 (1999).
28. T. Ferguson, T. Lienard, G. Gottschalk, J. Krzycki, unpublished.
29. U.K. Laemmli, *Nature (London)* **227**, 680 (1970).
30. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
31. S. Henikoff, J. G. Henikoff, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915 (1992).
32. The authors wish to thank C. Daniels, F. Grundy, T. Henkin, M. Ibba, and J. Reeve for helpful discussions, and the DOE Joint Genome Institute (www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html) for making available the preliminary sequence data from the genomes of *M. barkeri* Fusaro and *D. hafniense*. This work was supported by the National Science Foundation (MCB-9808914).

Supporting Online Material
www.sciencemag.org/cgi/content/full/296/5572/1459/DC1
Materials and Methods
figs. S1 and S2

4 January 2002; accepted 27 March 2002

A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase

Bing Hao,[1] Weimin Gong,[1] Tsuneo K. Ferguson,[2] Carey M. James,[2] Joseph A. Krzycki,[2]* Michael K. Chan[1]*

Genes encoding methanogenic methylamine methyltransferases all contain an in-frame amber (UAG) codon that is read through during translation. We have identified the UAG-encoded residue in a 1.55 angstrom resolution structure of the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB). This structure reveals a homohexamer comprised of individual subunits with a TIM barrel fold. The electron density for the UAG-encoded residue is distinct from any of the 21 natural amino acids. Instead it appears consistent with a lysine in amide-linkage to (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate. We suggest that this amino acid be named L-pyrrolysine.

The catabolism of methylamines by methanogens involves a conserved arrangement of proteins. A specific monomethylamine (MMA), dimethylamine (DMA), or trimethylamine (TMA) methyltransferase activates the substrate for methyl transfer to a cognate corrinoid protein (*1*, *2*). A second methyltransferase catalyzes the transfer of the methyl group from the methylated corrinoid cofactor to coenzyme M (CoM). Methyl-CoM is subsequently used to generate methane by methyl-CoM reductase (*3*).

All known methanogen methylamine (MMA, DMA, or TMA) methyltransferase genes contain a single in-frame amber (UAG) codon that does not appear to stop translation during protein synthesis (*4*, *5*). This strict conservation contrasts with the lack of sequence similarity between different MMA, DMA, and TMA methyltransferase gene families. Analysis of tryptic fragments of MMA methyltransferase (MtmB) by mass spectrometry and Edman degradation suggested that the amber codon serves as a sense codon that corresponds to lysine (*6*). The harsh conditions of peptide isolation, however, left open the possibility that the amber codon signals a modified lysine residue. The use of what is normally a stop codon to signal the incorporation of an unusual amino acid has precedent in the use of UGA to encode selenocysteine, the 21st natural amino acid found in Bacterial, Eucaryal, and Archaeal proteins (*7*, *8*).

The structure of the *Methanosarcina barkeri* MS monomethylamine methyltransferase was determined to clarify the identity of the UAG-encoded residue (Tables 1 and 2). Two forms of the enzyme were obtained from crystallization conditions that differed only in the precipitating salt used [NaCl for form 1, and $(NH_4)_2SO_4$ for form 2] and were solved to 1.55 Å and 1.7 Å resolution, respectively. The global conformations for these two forms are virtually identical, except for the region around the UAG-encoded amino acid. The overall MtmB structure consists of a homohexamer arranged into a dimer of trimers with overall D_3 symmetry (Fig. 1A). Each subunit adopts an α/β TIM barrel fold (*9*, *10*) (Fig. 1B) that is reminiscent of other corrinoid-cofactor associated proteins: tetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (*11*), diol dehydratase (*12*), and racemases/mutases (*13*, *14*), although no significant sequence similarity is found between MtmB and these proteins. The eight-stranded β barrel is formed from strands β6 to β13 (*15*). Consistent with other methyltransferases, this barrel forms a deep cavity that for the MtmB is negatively charged (*15*).

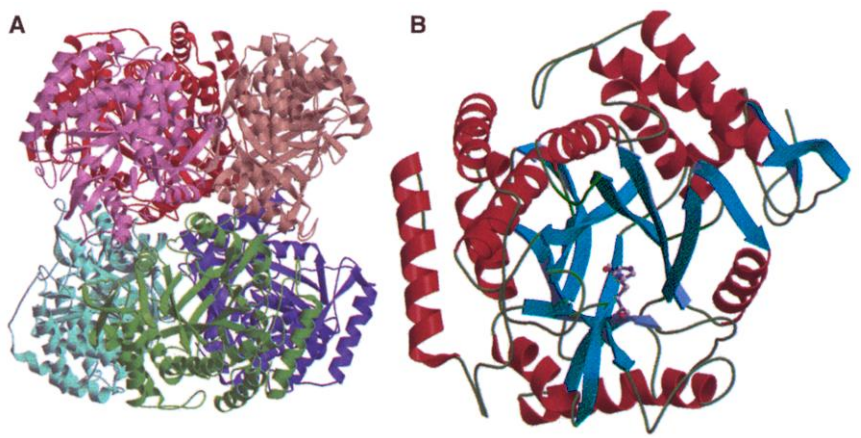

Fig. 1. (**A**) Ribbon diagram of the MtmB hexamer. The subunits forming the two trimers are shaded in red and blue hues, respectively. (**B**) Ribbon diagram of one MtmB subunit colored by secondary structure element (α helices, red; β sheets, cyan; random coil, green). The atoms of the UAG-encoded residue are shown as ball-and-stick models and are colored by their elements, with carbon as gray, nitrogen as blue, and oxygen as red. These figures were prepared with the programs MOLSCRIPT, and Raster3D (*22*, *23*).

[1]Departments of Biochemistry and Chemistry, and [2]Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, USA.

*To whom correspondence should be addressed: E-mail: chan@chemistry.ohio-state.edu and krzycki.1@osu.edu

This may facilitate binding of the methylammonium cation. The location of the amber-encoded residue at the bottom of this cavity suggests its possible role in catalysis.

The structure of MtmB from both NaCl and $(NH_4)_2SO_4$ crystal forms suggests that the amber-encoded residue is distinct from the 21 known natural amino acids. Both structures support its biochemical assignment as a lysine core modified by a group attached to its epsilon nitrogen (Fig. 2). The identity of the modifying group appears to be (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate, with the carboxylate of the modifying group in amide linkage to the epsilon nitrogen of lysine (Fig. 2B). In both crystal forms, the modifying group is disordered between two distinct orientations of the pyrroline ring (*15*). Although this disorder is present in both crystal forms, the relative occupancy of each orientation in each crystal form differs.

The initial assignment of the UAG-encoded amino acid was based on the 1.55 Å resolution structure of the NaCl form. This form gave the clearest density because one orientation of the ring dominates (orientation 1). The initial 5σ $F_O - F_C$ density could be fit to a residue resembling β-methyl-D-proline in amide linkage with the epsilon nitrogen of lysine (Figs. 2A and 3A). Position 1 in the five-membered ring was assigned to a nitrogen atom because it forms hydrogen bonds with the carboxylate side chains of Glu^{259} and Glu^{229}, and because its refinement as nitrogen gave a better fit to the density than either carbon or oxygen. The identity of the atom attached to the C-4 ring carbon is currently unclear. Based on the current fit to the electron density, it could be a methyl, ammonium, or hydroxyl group. It is within hydrogen-bonding distance of Tyr^{335}, but this distance is long, 3.16 Å, and two waters form a hydrogen bond to Tyr^{335} with a better geometry.

After refinement of the initial model built from the 5σ electron density, weak but broad 3σ difference density was observed (*15*). This residual density suggested that the five-membered ring also adopts a second orientation (orientation 2), though at much lower occupancy. In this second orientation, this ring is rotated by approximately 90° relative to its position in orientation 1 and has different set of hydrogen bonding interactions with the protein. Refinement of the occupancies and thermal parameters of the modifying group in both orientations suggests that the relative occupancies of orientations 1 and 2 in

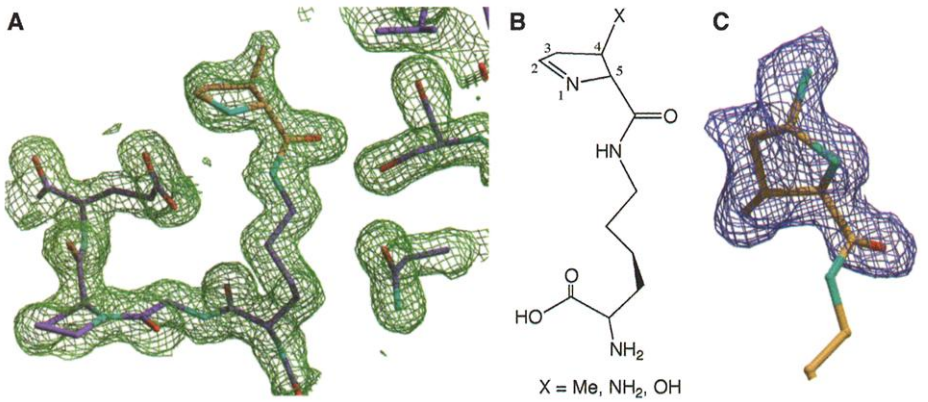

Fig. 2. (**A**) Fit of (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate to the $2F_O - F_C$ 3σ density of the NaCl crystal form (orientation 1). The substituent attached to the C-4 carbon is shown as a methyl, but it could also be an ammonium or hydroxyl group. (**B**) Stick-diagram of proposed L-pyrrolysine amino acid. (**C**) Residual $F_O - F_C$ difference map of NH_4SO_4 crystal form after incorporation of a 40% occupancy model consisting of L-pyrrolysine in orientation 1 and an exogenous ammonium ion. This remaining omit density suggests that L-pyrrolysine adopts a different orientation (orientation 2) at 60% occupancy in NH_4SO_4 with an amine added to the C-2 carbon of the pyrroline ring. These figures were prepared with the programs XtalView, MOLSCRIPT, and Raster3D (*22–24*).

Table 1. Statistics for data collection, and phase determination. The MtmB crystals contain one MtmB subunit per asymmetric unit and belong to space group $P6_322$ with unit cell dimensions $a = b = 158.8$ Å, $c = 136.5$ Å. The details regarding the crystallization, data collection, and structure determination are provided as supplementary material (*15*). All the calculations were done with the program PHASES (*20*). For completeness and R_{sym}, numbers in parentheses represent the statistics for the shell comprising the outer 10% (theoretical) of the data. Phasing power is the mean value of the heavy atom structure factor divided by the lack of closure. Figure of merit is the mean value of the cosine of the error in phase angles. The combined figure of merit for all datasets was 0.63.

Parameter	Native 1 NaCl	Native 2 $(NH_4)_2SO_4$	NaI	I_2	EMTS	Me_3PbOAc	K_2OsO_4	CsCl
				Data statistics				
Heavy atom								
Concentration			2 M	saturated	1 mM	5 mM	2 mM	4.5 M
Soaking time			20 min	5 min	cocrystal	cocrystal	24 hours	cocrystal
Number of sites			27	9	6	4	3	12
Wavelength (Å)	1.0000	1.0000	1.3869	1.3869	0.9924	0.9488	1.1402	1.5497
Resolution (Å)	1.55	1.7	2.3	3.4	2.7	2.9	2.9	3.2
Observations	549727	513202	187164	66528	152898	157617	126772	66423
Unique	140499	105886	43796	14232	26712	22397	22544	17003
Completeness (%)	97.3 (93.7)	96.7 (94.8)	97.4 (80.3)	98.6 (91.5)	98.6 (91.5)	97.3 (97.2)	98.6 (90.7)	99.6 (99.8)
Redundancy	3.6	4.6	4.2	4.6	5.6	7.2	5.5	4.0
Mosaicity (°)	0.42	0.33	0.42	0.31	0.48	0.46	0.17	0.54
R_{sym}* (%)	7.1 (51.0)	4.9 (28.1)	9.2 (25.0)	15.0 (29.0)	9.8 (26.9)	11.5 (28.1)	11.5 (23.6)	14.6 (33.6)
				Phasing statistics				
Resolution (Å)			20–2.6	20–3.4	20–2.7	20–2.9	20–2.9	20–3.2
Phasing power (iso/ano)			1.81/1.30	0.84/0.46	0.92/0.87	0.99/0.91	0.49/0.89	0.89/0.64
R_{cullis}† (centric)			0.53	0.67	0.65	0.65	0.75	0.67
R_{kraut}‡ (iso/ano)			0.11/0.12	0.20/0.33	0.19/0.23	0.19/0.23	0.20/0.28	0.19/0.23
Figure of merit			0.34	0.17	0.19	0.17	0.11	0.20

*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. †$R_{cullis} = \Sigma(||F_{PH} - F_P| - |F_{H(calc)}||)/\Sigma|F_{PH} - F_P|$, where F_{PH} is the structure factor of the derivative and F_P is that of the native data. The summation is valid only for centric reflections. ‡$R_{kraut} = \Sigma\varepsilon_{iso}/\Sigma\Delta_{iso}$ for isomorphous differences and $\Sigma\varepsilon_{ano}/\Sigma\Delta_{Bijvoet}$ for anomalous differences, where ε_{iso} and ε_{ano} are the isomorphous and anomalous lack of closure respectively, and Δ_{iso} is the isomorphous difference and $\Delta_{Bijvoet}$ is the Bijvoet difference. The summation is taken over all acentric reflections.

the NaCl crystal form are 85% and 15%, respectively.

The $2F_O - F_C$ density for UAG-encoded residue of the $(NH_4)_2SO_4$ crystal form (1.7 Å resolution) differs significantly from that observed in the NaCl crystal form (*15*). It can be initially fit, however, to a model composed of the same two orientations found in the NaCl crystal form, although with different relative occupancies (*15*). Refinement suggests that in the $(NH_4)_2SO_4$ crystal form, orientation 2 now becomes the dominant conformation, with the relative occupancies of orientations 1 and 2 having values of 40% and 60%, respectively.

In addition to the difference in the relative occupancies, electron density maps of the $(NH_4)_2SO_4$ crystal form also reveal the presence of an atom bound to the C-2 ring carbon for orientation 2 that was not observed in the NaCl crystal form (Figs. 2C and 3C). We assign this atom to nitrogen, because $(NH_4)_2SO_4$ is the only salt not present in the solution used to grow the NaCl crystals. Consistent with this assignment, the side chains of Glu^{259} and Gln^{333} are at hydrogen-bonding distances to the proposed amine substituent. These interactions likely stabilize the L-pyrrolysine in orientation 2, thereby accounting for its higher occupancy in the $(NH_4)_2SO_4$ crystal form.

In order to account for some remaining difference density in the $(NH_4)_2SO_4$ crystal form, a free ammonium ion was required at 40% occupancy (*15*). This likely represents enzyme-bound ammonium ion that is associated with the protein when the ring is in orientation 1 (Fig. 3B). This ion does not bind or interact with the UAG-encoded residue, but instead forms a hydrogen bond with Met^{261} and Tyr^{335}. The ammonium ion occupies a position near the amine that is added to the C-2 carbon of the ring in orientation 2 of the $(NH_4)_2SO_4$ crystal form. This structure may reflect an intermediate state before amine addition.

In order for an amine to add to the ring, the C-2 carbon must be sp^2 hybridized. This hybridization is likely achieved by having an imine bond between the N-1 and C-2 atoms of the ring. This implies that the identity of the UAG-encoded amino acid before amine addition is (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate in amide linkage to the epsilon nitrogen of lysine, a species that we proposed be named L-pyrrolysine (Fig. 2B). We note that although the modifying group has analogy to Δ^1-pyrroline-5-carboxylate, the direct precursor for proline synthesis, the D chirality of the pyrroline-5-carboxylate and the additional substituent at the C^4 carbon make it distinct.

MtmB in complex with its associated corrinoid protein, MtmC, was subjected to electrospray mass spectrometry to confirm the pyrrolysine assignment (*16*). The measured masses of MtmB and MtmC were 50,105 ± 2 daltons and 23,066 ± 1 daltons, respectively. These values can be compared with the deduced mass from the predicted protein sequences from the encoding genes (*4*) using the program SHERPA (*17*). Without the corrinoid prosthetic group, the theoretical average mass of MtmC is calculated as 23,067, which is consistent with the experimentally derived molecular mass. For MtmB, the theoretical average mass is calculated as 49,998, by assuming a lysine residue at the UAG-encoded position. This calculated value is 107 atomic mass units (amu) less than the experimental mass of MtmB (50,105 daltons). With the incorporation of the 4-substituted-pyrroline-5-carboxylate group as

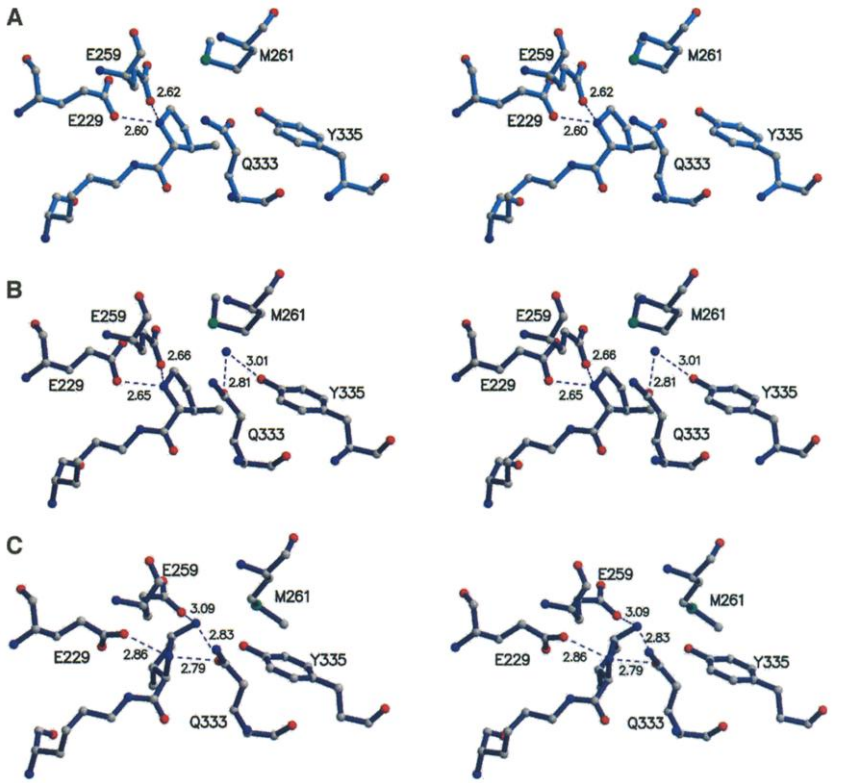

Fig. 3. Stereoview of primary forms of the active site around the amber-encoded ligand: (**A**) NaCl crystals; (**B**) $(NH_4)_2SO_4$ crystals, 40% occupancy orientation that is similar to NaCl crystals; (**C**) $(NH_4)_2SO_4$ crystals, 60% occupancy orientation with amine added to ring. These figures were prepared with the programs XtalView, MOLSCRIPT, and Raster3D. (*22–24*).

Table 2. Refinement statistics. The starting R and Free-R factors after the first rigid body refinement were 0.45 and 0.44. The model was refined by simulated annealing, followed by several cycles of minimization using a maximum likelihood target based on the amplitudes and incorporation of an flat bulk solvent correction and an overall anisotropic B factor. The quality of the final model of the MtmB was assessed using the program PROCHECK and was found to be acceptable. The statistics for each of the stereochemical parameters was inside or better than the expected values, and 90% of the residues were found to occupy the most favored regions of the Ramachandran plot (*21*). The refined model consists of all the residues from 2 to 458, except the NH_2-terminal methionine. This is consistent with its posttranslational cleavage as determined by NH_2-terminal sequencing of the isolated protein (*4*). There are a total of three residues having the *cis* conformation, Pro^{204}, Glu^{235}, and Val^{367}. rms, root mean square.

Data set	Native 1, NaCl	Native 2, $(NH_4)_2SO_4$
Resolution range (Å)	20.0–1.55	20.0–1.70
Total number of nonhydrogen atoms	3490	3490
Number of water molecules	464	533
$R_{crystal}$* (R_{free})† (%)	17.6 (18.8)	16.0 (17.4)
Average B factor (Å^2)		
Protein atoms	21.4	15.1
Water molecules	36.5	34.2
rms bond length deviation (Å)	0.005	0.005
rms bond angle deviation (°)	1.198	1.176

*$R_{crystal} = 100 \times \Sigma|F_o - F_c|/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. †R_{free} is the same as $R_{crystal}$ and was calculated by using 10% of the data excluded from refinement.

found in the crystal structure, however, the theoretical molecular mass for MtmB increases by 109 amu to 50,107 daltons assuming a methyl group is attached to the C-4 ring carbon. If the substituent attached to the ring is either an amine or hydroxyl group, then the theoretical mass would be higher by 110 amu (50,108 daltons) and 111 amu (50,109 daltons), respectively.

MtmB assists in the transfer of the methyl group of monomethylamine to the corrinoid cofactor of MtmC. A mechanism for how pyrrolysine could play a role in activation of methylamine substrates is suggested on the basis of the two forms of the enzyme (Fig. 4). In this model, the role of pyrrolysine is to position and display the methyl group of methylamine for attack by the corrinoid cofactor. A similar mechanism could also be envisioned for the TMA and DMA methyltransferases, which also have UAG-encoded residues.

The two conformations of pyrrolysine and the surrounding side chains are consistent with induced fit mechanisms observed in other systems. Here, the two orientations of the ring may facilitate different interactions with the carboxylate side chain of Glu^{259}. We propose a mechanism in which the carboxylate of Glu^{259} is protonated before methylamine addition and positioned so that the carboxyl hydrogen bonds to the pyrroline ring nitrogen of pyrrolysine. Protonation of the imine nitrogen would make it more electron withdrawing and help to activate the C-2 carbon for nucleophilic addition. After methylamine addition to the C-2 carbon, the new deprotonated carboxylate of Glu^{259} shifts to be a hydrogen bond acceptor with the bound methyl ammonium group. The amide side chain of Gln^{333} provides a second hydrogen bond acceptor to the bound methyl ammonium group. These two hydrogen bonds and the covalent bond to the pyrroline ring serve to position the methylammonium group so that its methyl group is directed toward the surface of the binding cleft. Here, it presumably is positioned to interact with the corrinoid-cofactor of MtmC upon formation of the MtmBC complex. Preliminary docking models of MtmB with the B_{12}-binding domain of MetH (*18*), a homolog of MtmC (*4*) have indicated the feasibility of this concept. Additional work will be required to confirm this mechanism and our proposed structure for L-pyrrolysine; however, these models provide fertile grounds for future experimentation.

In conclusion, the in-frame amber codon in the gene encoding MtmB encodes a novel amino acid with properties favoring catalysis of methyltransferase reactions involving amines. As described in the accompanying paper (*19*), an unusual aminoacyl-tRNA synthetase and $tRNA_{CUA}$ are encoded near the *mtmB* genes in *M. barkeri*. Although it formally remains possible that modification of lysine occurs after UAG directed insertion of lysine into the protein, the use of a canonical stop codon and dedicated tRNA is most consistent with the direct translational encoding of L-pyrrolysine. Thus current evidence supports the case for L-pyrrolysine representing the 22nd naturally occurring amino acid to be identified.

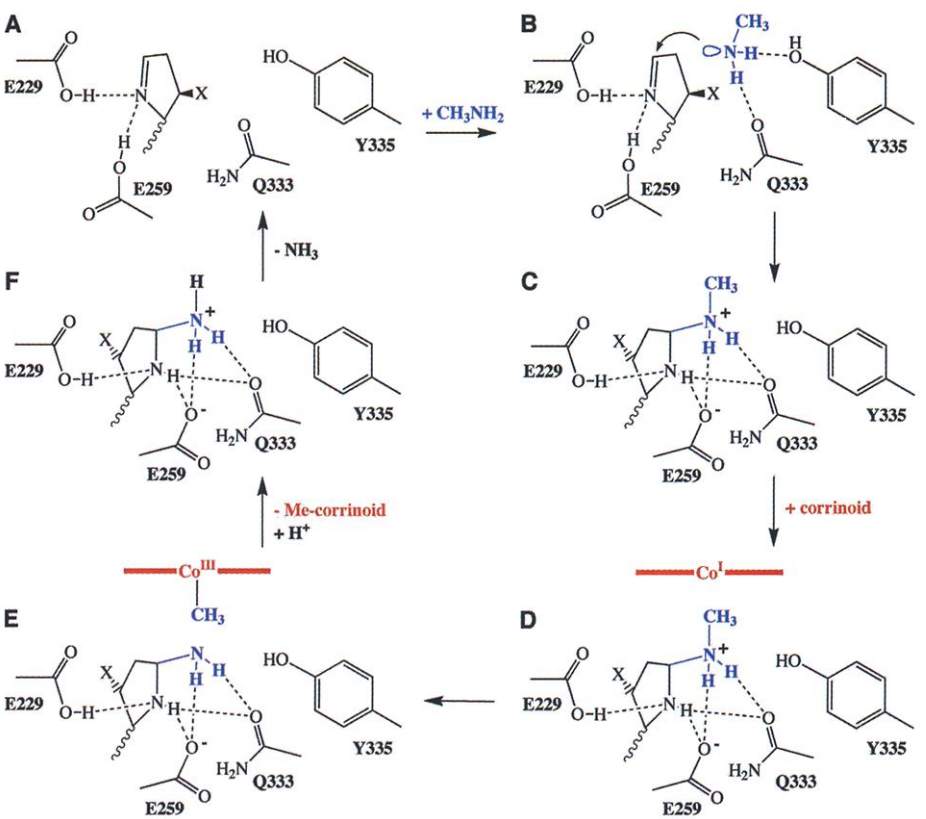

Fig. 4. Hypothetical model for the role of the amber-encoded residue in catalysis. The proposed intermediates for (**A**), (**B**), (**C**), and (**F**) are based on the structures of L-pyrrolysine (X = Me, NH_2, or OH) in its 85% occupancy orientation in NaCl crystal form and on both orientations of L-pyrrolysine in the $(NH_4)_2SO_4$ crystal form. Intermediates (**D**) and (**E**) are based on a preliminary docking model of MtmB with its cognate corrinoid protein, MtmC.

References and Notes

1. S. A. Burke, J. A. Krzycki, *J. Biol. Chem.* **272**, 16570 (1997).
2. D. J. Ferguson Jr., N. Gorlatova, D. A. Grahame, J. A. Krzycki, *J. Biol. Chem.* **275**, 29053 (2000).
3. R. K. Thauer, *Microbiology* **144**, 2377 (1998).
4. S. A. Burke, S. L. Lo, J. A. Krzycki, *J. Bacteriol.* **180**, 3432 (1998).
5. L. Paul, D. J. Ferguson, J. A. Krzycki, *J. Bacteriol.* **182**, 2520 (2000).
6. C. M. James, T. K. Ferguson, J. F. Leykam, J. A. Krzycki, *J. Biol. Chem.* **276**, 34252 (2001).
7. J. F. Atkins, A. Böck, S. Matsufuji, R. F. Gesteland, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), pp. 637–673.
8. A. Böck *et al.*, *Mol. Microbiol.* **5**, 515 (1991).
9. D. W. Banner *et al.*, *Nature* **255**, 609 (1975).
10. C. A. Orengo *et al.*, *Structure* 5, 1093 (1997).
11. T. Doukov, J. Seravalli, J. J. Stezowski, S. W. Ragsdale, *Structure* **8**, 817 (2000).
12. J. Masuda, N. Shibata, Y. Morimoto, T. Toraya, N. Yasuoka, *Structure* **8**, 775 (2000).
13. F. Mancia *et al.*, *Structure* **4**, 339 (1996).
14. R. Reitzer *et al.*, *Structure* **7**, 891 (1999).
15. Supplementary figures and details of experimental procedures are available on *Science* Online.
16. Electrospray ionization (ESI) experiments were performed on a Micromass Q-TOF II (Micromass, UK) mass spectrometer equipped with an orthogonal electrospray source (Z-spray) operated in positive ion mode. We prepared proteins in a solution containing 50% acetonitrile/50% water, and 0.1% formic acid. Optimal ESI conditions were capillary voltage 3000 V, source temperature 110°C and a cone voltage of 60 V. The ESI gas was nitrogen. We deconvoluted ESI data by using MaxEnt I provided by Micromass. Standard deviations were calculated from the mass values obtained from individual peaks in the ion series of each protein. The calculated masses allowed for the known processing of the NH_2-terminal methionine from MtmB and MtmC.
17. J. A. Taylor, K. A. Walsh, R. S. Johnson, *Rapid Commun. Mass Spectrom.* **10**, 679 (1996).
18. C. L. Drennan, S. Huang, J. T. Drummond, R. G. Matthews, M. L. Ludwig, *Science* **266**, 1669 (1994).
19. G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **296**, 1459 (2002).
20. W. Furey, S. Swaminathan, *Methods Enzymol.* **277**, 590 (1997).
21. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
22. P. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
23. E. Merritt, M. Murphy, *Acta Crystallogr.* **D50**, 869 (1994).
24. D. E. McRee, *J. Struct. Biol.* **125**, 156 (1999).

25. This work was supported by grants from the National Institutes of Health (GM43268 to M.K.C and J.A.K), the National Science Foundation (MCB-9808914 to J.A.K.), the Department of Energy (DE-FG0202-91ER200042 to J.A.K.), and a fellowship

from the Alfred P. Sloan Foundation (to M.K.C.). Work was done partially at Stanford Synchrotron Radiation Laboratory (SSRL), which is operated by the Department of Energy, Office of Basic Energy Sciences. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707. The beamline X4A at the National Synchrotron Light Source, a Department of Energy facility, is supported by the Howard Hughes Medical Institute. The authors thank K. Greenchurch at the Ohio State University Chemical Instrumentation Center for her assistance with mass spectrometric analysis. The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1L2Q (NaCl form) and 1L2R [$(NH_4)_2SO_4$ form].

Supporting Online Material
www.sciencemag.org/cgi/content/full/296/5572/1462/DC1
Materials and Methods
Figs. S1 to S3

4 January 2002; accepted 27 March 2002

Combinatorial Synthesis of Genetic Networks

Călin C. Guet,[1,3] Michael B. Elowitz,[3] Weihong Hsing,[1] Stanislas Leibler[1,2,3*]

A central problem in biology is determining how genes interact as parts of functional networks. Creation and analysis of synthetic networks, composed of well-characterized genetic elements, provide a framework for theoretical modeling. Here, with the use of a combinatorial method, a library of networks with varying connectivity was generated in *Escherichia coli*. These networks were composed of genes encoding the transcriptional regulators LacI, TetR, and lambda CI, as well as the corresponding promoters. They displayed phenotypic behaviors resembling binary logical circuits, with two chemical "inputs" and a fluorescent protein "output." Within this simple system, diverse computational functions arose through changes in network connectivity. Combinatorial synthesis provides an alternative approach for studying biological networks, as well as an efficient method for producing diverse phenotypes in vivo.

Living cells respond to information from their environment on the basis of the interactions of a large yet limited number of molecular species that are arranged in complex cellular networks (*1*, *2*). A classic example of such biochemical computation is the chemotaxis behavior of *Escherichia coli*, which is mediated by a well-characterized signal transduction network (*3*). However, despite growing knowledge about the molecular components of the cell, the dynamics of even simple cellular networks are not well understood. For instance, a quantitative explanation of the high sensitivity and exact adaptation observed in bacterial chemotaxis is still lacking (*3*). Similarly, many other cellular networks, such as the ones responsible for signal transduction, regulation of gene expression, or metabolism, are poorly understood from a quantitative point of view. Thus, simple and modular experimental systems are needed to study how the genetic structure and connectivity of cellular networks are related to their function. To this end, we devised an in vivo synthetic system that enables the generation of combinatorial libraries of genetic networks.

We have generated a combinatorial library composed of a small set of transcriptional regulatory genes and their corresponding promoters with varying connectivity (Fig. 1). We chose genes of three well-characterized prokaryotic transcriptional regulators: LacI, TetR, and lambda cI (*4*). The binding state of LacI and TetR can be changed with the small molecule inducers, isopropyl β-D-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc), respectively. We also chose five promoters regulated by these proteins, which cover a broad range of regulatory characteristics such as repression, activation, leakiness, and strength. Two of the promoters are

[1]Howard Hughes Medical Institute, Department of Molecular Biology, [2]Department of Physics, Princeton University, Princeton, NJ 08544, USA. [3]The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. Laboratory for Living Matter, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

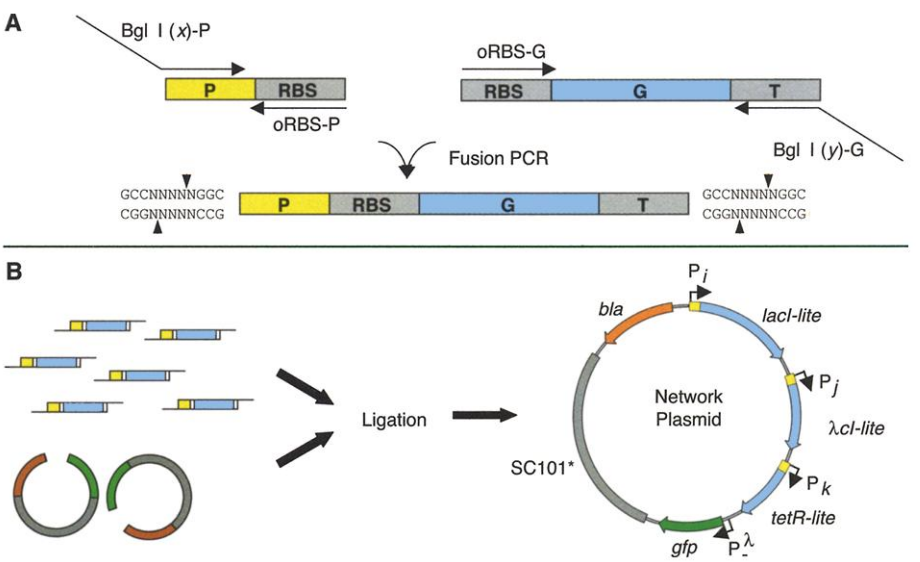

Fig. 1. The modular genetic cloning strategy used to generate combinatorial libraries of logical circuits. Construction of the library proceeded in two steps. In the first step (**A**), we built all 15 possible promoter-gene units. Individual promoters and genes were first amplified by PCR. The genes [denoted "*–lite*" in (B)] have an *ssrA* tag that reduces the half-life of the proteins encoded by the modified gene (*21*). The five promoters used were P^L_1 and P^L_2 (repressed by LacI), P^T (repressed by TetR), and P^λ_- and P^λ_+ (repressed and activated, respectively, by λ cI) (*5*). The transcriptional terminator T1 was present at the end of each gene. Identical RBS were used as internal primers for the subsequent fusion PCR step to form promoter-gene units (*27*). In order to control the number of promoter-gene units and the position of a given gene in the network, Bgl I sites were incorporated in PCR primers, as shown. The special recognition and restriction properties of Bgl I (*28*) allow various sticky ends to be produced by Bgl I cleavage. Here we designed the Bgl I sites such that specific cohesive ends *x* and *y* were associated with each regulatory gene (for *lacI*, x_{lac} = GCC, y_{lac} = TTC; for λ *cI*, x_{cI} = AAG, y_{cI} = GTG; and for *tetR*, x_{tet} = CAC, y_{tet} = TCG). Note that y_{lac} is compatible with x_{cI}, y_{cI} is compatible with x_{tetR}, and so on. Thus, in step (**B**) when all 15 possible fusion PCR products were mixed together and ligated, the resulting products contained exactly three promoter-gene units in one particular order (*lacI*, λ *cI*, *tetR*). These products were cloned into a low copy number plasmid (three to four copies per cell) (*23*), carrying the reporter gene *gfpmut3* under the control of P^λ_- (*29*).