Published in final edited form as:

J Mol Biol. 2004 July 16; 340(4): 695–706. doi:10.1016/j.jmb.2004.05.019.

Structural Characterization and Comparative Phylogenetic Analysis of *Escherichia coli* HemK, a Protein (N5)-glutamine Methyltransferase

Zhe Yang¹, Lance Shipman¹, Meng Zhang², Brian P. Anton^{2,3}, Richard J. Roberts², and Xiaodong Cheng^{1,*}

¹Department of Biochemistry Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

²New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

³Bioinformatics Program Boston University, 48 Cummington Street, Boston MA 02215, USA

Abstract

Protein glutamine methylation at GGQ sites of protein chain release factors plays a pivotal role in the termination of translation. We report here the crystal structure of the *Escherichia coli* HemK protein (N5)-glutamine methyltransferase (MTase) in a binary complex with the methyl-donor product *S*-adenosyl-L-homocysteine (AdoHcy). HemK contains two domains: a putative substrate binding domain at the N terminus consisting of a five helix bundle and a seven-stranded catalytic domain at the C terminus that harbors the binding site for AdoHcy. The two domains are linked by a β -hairpin. Structure-guided sequence analysis of the HemK family revealed 11 invariant residues functioning in methyl-donor binding and catalysis of methyl transfer. The putative substrate-binding domains of HemK from *E. coli* and *Thermotoga maritima* are structurally similar, despite the fact that they share very little sequence similarity. When the two proteins are aligned structurally, the helical N-terminal domain is subject to approximately 10° of hinge movement relative to the C-terminal domain. The apparent hinge mobility of the two domains may reflect functional importance during the reaction cycle. Comparative phylogenetic analysis of the *hemK* gene and its frequent neighbor gene, *prfA*, which encodes a major substrate, provides evidence for several examples of lateral gene transfer.

Keywords

HemK; protein glutamine methylation; GGQ motif; NPPY motif

Introduction

Two recent studies showed how HemK, the product of the *hemK* gene, functions as a protein (N5)-glutamine MTase modulating the termination activity of release factors in ribosomal protein synthesis.^{1,2} Prior to these studies, the function of HemK had been misidentified twice.

© 2004 Elsevier Ltd. All rights reserved.

*E-mail address of the corresponding author: E-mail: xcheng@emory.edu.

Present address: Lance Shipman, Department of Chemistry, Morehouse College, Merrill Hall Room 218, 830 Westview Drive SW, Atlanta, GA 30314, USA.

Edited by I. Wilson

Supplementary data associated with this article can be found at doi: 10.1016/j.jmb.2004.05.019

The *hemK* gene was initially discovered in *Escherichia coli* during a genetic screen designed to reveal new types of heme biosynthesis mutants, prompting the suggestion that its gene product might function as a protoporphyrinogen oxidase.³ Despite this mutant phenotype, subsequent biochemical and genetic studies revealed that the *hemK* gene product appeared to have no direct involvement in the heme biosynthetic pathway.⁴ It is still unclear exactly what caused the defects in heme metabolism observed for *hemK* mutants.

Sequence analysis of HemK revealed the presence of an NPPY motif, thought to be restricted to members of the (adenine N6 and cytosine N4) DNA amino MTases.⁵ This led to the suggestion that HemK was itself an AdoMet-dependent DNA MTase;⁶ however, no evidence could be found that HemK was able to methylate DNA (M.Z. & R.J.R., unpublished results).

With the tremendous growth of protein sequence databases, it has been observed that *hemK* homologs are nearly omnipresent in species from all three domains of life, suggesting a function of ancient origin and broad importance,⁷ consistent with its elucidated role in protein translation. In *E. coli*, the gene is present in the *hemA-prfA-hemK* operon, where *prfA* encodes RF1, one of two class I protein chain release factors (RFs). The most highly conserved sequence feature of class I RFs from prokaryotic and eukaryotic organisms is a GGQ motif. Mass spectral analysis provided evidence that Gln252 of *E. coli* RF2, within this GGQ motif, is methylated at the N5 position. Furthermore, it was shown that this modification correlates with a large increase in the efficiency of the RF2 termination reaction in *E. coli* K12.⁸ This observation and the sequence similarities between HemK and known MTases led to the experiments that demonstrated HemK is the MTase that methylates both RF1 and RF2 *in vitro*.^{1,2} Here, we present the structure of *E. coli* HemK–AdoHcy. The structure reveals active-site residues and indicates the location of the substrate-binding surface. In addition, the effects of conserved residues and the relative hinge movement between the substrate and catalytic domains are discussed in comparison with the HemK homolog from *Thermotoga maritima*.⁹

Results

Structure of *E. coli* HemK and structural comparisons

E. coli HemK, a 277 residue protein, contains two structural domains (Figure 1A): a five helix N-terminal bundle (residues 2–73), and a C-terminal catalytic domain (residues 87–276) with a seven-stranded β -sheet, a characteristic feature of the class I AdoMet-dependent MTases.¹⁰ A β -hairpin (residues 74–86) connects the two domains. The linker hairpin interacts with the C-terminal domain *via* two hydrogen bonds formed by the main-chain atoms (N–H and C=O) of Phe83 and the side-chain atoms (O ^{δ 1} and N ^{δ 2}) of the conserved Asn152. The hairpin also interacts with the N-terminal domain *via* one salt-bridge interaction (Glu75–Arg22). The overall molecular dimensions of HemK are approximately 65 Å $\times$ 40 Å $\times$ 30 Å, with an interdomain cleft forming a large concave surface between the two domains (Figure 1B).

Recently, a structure was determined for the HemK analog in *T. maritima*.⁹ We used the coordinates of the *T. maritima* HemK for molecular replacement (see Materials and Methods). When the two proteins are aligned structurally, the root-mean-square deviation (rmsd) between the corresponding backbone atoms is 1.6 Å for the whole protein, 1.2 Å for the C-terminal domain, and 1.4 Å for the N-terminal domain and the linker hairpin. When the two structures are superimposed on the basis of the C-terminal domain, the helical N-terminal domain is subject to approximately 10° of hinge movement, with a closer interdomain distance in *E. coli* HemK than in *T. maritima* HemK (Figure 1C). The hinge movement between the two domains may be of functional importance during the reaction cycle, induced by enzyme-cofactor and/or enzyme–substrate interactions, as observed in M.HhaI–DNA interactions.¹¹ In addition, we have found that there is one preferred protease V8 cleavage site, on the carboxyl side of Glu76, between the two domains of the *E. coli* HemK (data not shown).

A DALI¹² homology search with the coordinates of the N-terminal helical domain identified 13 weakly similar structures with Z-scores >3.0. The majority of top “hits” are proteins or domains involved in protein–protein interactions. Three of these are shown in Figure 1E. The first is the EPS15 homology domain (Z= 3.8); which is a eukaryotic signaling module that recognizes proteins containing NPF sequences.¹³ The second is domain I of the transcription factor TFIIIS (Z= 3.7); which has homologous regions in transcription elongation factor A and CRSP70 and which may be involved in functional interactions with the transcriptional machinery.¹⁴ The third is the KIX domain of CBP (Z = 3.6); which interacts directly with the nuclear transcription factor CREB.¹⁵ Collectively, the structural homology with protein-interacting domains is suggestive of a possible involvement of the HemK N-terminal domain in recognition of the protein substrate. The C-terminal catalytic domain is similar to many class-I MTases, with the highest DALI score (Z= 21.1; PDB 1dus, with 1.8 Å rmsd and 24 sequence identities over 172 amino acid residues) to the structure of *Methanococcus jannaschii* mj0882,¹⁶ which was annotated as a putative RNA MTase.

Conserved residues

A sequence alignment of 47 bacterial and mitochondrial members of the HemK protein family (HBG002674) from HOBACGEN¹⁷ revealed 11 conserved amino acids scattered throughout the C-terminal domain (Figure 2). These completely conserved residues include Pro91, Arg92 and Glu96 from the region between the linker hairpin and the C-terminal domain, the Gly-rich motif between strand β 1 and helix α A (Gly119 and Gly121), Asn152 of helix α B, the Asn183-Pro184-Pro185-Tyr186 (NPPY) motif after strand β 4, and the carboxyl end of strand β 5 (Glu237). Also included in the alignment are four related proteins that contain most of these same conserved residues. PapM catalyzes two successive N-methylation steps of 4-amino-L-phenylalanine leading to 4-dimethylamino-L-phenylalanine (DMPAPA),¹⁸ a precursor of pristinamycin I. YfcB is also a protein (N5)-glutamine MTases, methyl-ating the ribosomal protein L3 at Gln150.¹

The folded structure shows that these conserved residues are clustered on two surface patches (Figure 1B). Except Glu96, which is exposed to the concave surface of the cleft, the remaining residues surround the bound AdoHcy (Figure 3A). We suggest that the conserved surface patches have functional importance, being involved in one of three steps in the methylation reaction: AdoMet binding (Gly119 and Gly121), catalysis of methyl transfer (NPPY motif), and substrate binding *via* electrostatic interactions (Glu96). We note that the region immediately prior to the GGQ motif of prokaryotic RF1 and RF2 contains a conserved arginine residue (IDTFRSSGAGGQHVNT of *E. coli* RF1)¹⁹ that might be involved in electrostatic interactions.

In addition, Pro91, Arg92, Asn152, and Glu237 (see Figure 3A) play a role in the structural stability of the active site by connecting conserved residues together. Arg92, forming a salt-bridge with Glu202 (a residue following the NPPY motif), is flanked by two conserved proline residues that provide structural rigidity to the Pro-Arg-Pro loop. Glu237, the last residue of strand β 5, contributes two hydrogen bonds: one with the hydroxyl oxygen atom of Tyr186 of the NPPY motif and the other with the side-chain oxygen atom of Thr95 after the hairpin linker; the interaction connects two conserved segments together. Similarly, Asn152, located within the helix α B, connects the linker strand L2 and the Gly-rich motif: its side-chain carbonyl oxygen atom (O ^{δ 1}) forms a hydrogen bond to the backbone nitrogen atom of Phe83, while its side-chain amide nitrogen atom (N ^{δ 2}) forms hydrogen bonds with the backbone carbonyl groups of Phe83 and Thr120.

The active site and HemK–AdoHcy interactions

The AdoHcy molecule is clearly observed in the $F_o - F_c$ electron density map (Figure 3B). It is located in a deep acidic pocket on the surface of the C-terminal domain (Figure 1D), with approximately 2% of its surface accessible to solvent as calculated by the program AREA-MOL.²⁰ The sulfur group of AdoHcy is visible through a narrow and negatively charged channel (Figure 1D), where the target glutamine residue is likely to be inserted. The loop after strand β_4 contributes most of the residues (NPPY motif) in the active-site pocket. The invariant NPPY of the *E. coli* HemK–AdoHcy binary complex is superimposable onto the corresponding motif in the ternary complex of *T. maritima* HemK–AdoMet–Gln (free Gln) (Figure 3C), as well as the ternary complex of DNA MTase M.TaqI with DNA and a non-reactive AdoMet analog (Figure 3D), suggesting that the conformation of NPPY is quite stable and the interactions with the target amino group will be highly conserved among DNA adenine N6 and cytosine N4 MTases and protein (N5) glutamine MTases.^{9,21,22} The NPPY motif is not nucleotide-specific, but is selective for nitrogen atoms conjugated to a planar system such as an amide moiety in proteins (Gln, Asn) or a nucleotide base in DNA or RNA (adenine, cytosine).^{9,10}

HemK–AdoHcy interactions can be grouped according to the three components of AdoHcy (Figure 3B). (i) The adenine ring of the bound AdoHcy is sandwiched through van der Waals contacts between the aromatic side-chain of Trp168 after strand β_3 and the side-chain methylene portion of Arg141 after strand β_2 . An Arg residue in this position is present in the closest members of the *E. coli* HemK family of homologs, such as *Salmonella typhimurium* (71% sequence identity)²³ and *Pasteurella multocida* (46% identity).²⁴ It is often replaced by Ile/Leu/Val in more distantly related HemK homologs, such as *T. maritima* (23% identity). The exocyclic N6 amino group of the adenine ring forms a hydrogen bond with the backbone carbonyl oxygen atom of Asp167 after strand β_3 . (ii) The hydroxyl groups of the ribose ring form bifurcated hydrogen bonds with an acidic residue (Asp140) of strand β_2 ; this is nearly universal to class I MTases.²⁵ (iii) The homocysteine moiety of AdoHcy is surrounded by three segments of conserved amino acids: residues 88–92 after the hairpin linker, residues 117–123 (the Gly-rich motif I), and residues 183–186 (the NPPY motif). The oxygen atoms from the carboxyl and amino groups of the homocysteine form hydrogen bonds either with side-chain atoms pointing towards the binding cleft or with the surrounding backbone atoms of the protein.

Sequence analysis and phylogeny of HemK and PF1

In *E. coli*, *prfA* and *hemK* form the second and third genes in what appears to be a tightly linked operon with *hemA*, which encodes a glutamyl-tRNA reductase required for an early step in heme biosynthesis. The initiation codon for *hemK* overlaps with the termination codon of *prfA*. We have examined a total of 126 bacterial HemK protein sequences to explore both the syntenic relationship between the two genes and their phylogenetic properties. A summary of the syntenic analysis is shown in Table 2, where it can be seen that in 79 of the genomes, the two genes lie adjacent to one another with *prfA* upstream, as in *E. coli*. (More detailed results of the analysis are available as Supplementary Material.)

Bayesian inference methods were used to construct phylogenetic trees for HemK, RF1 and RF2. Given the close functional relationship between the *prfA* and *hemK* genes, and their synteny in so many genomes, it was anticipated that the phylogenies of their products would be comparable. Whereas the RF1 and RF2 phylogenies were similar to the 16 S rRNA phylogeny for the organisms examined, the HemK tree was incongruent in a number of specific instances. In at least two notable cases, the phylogeny of the HemK protein was strikingly

Supplementary Material consisting of one Table is available on Science Direct

different from that of RF1 and RF2 (Figure 4). The RF proteins of *Helicobacter* and its close relatives group with the other proteobacteria, but the HemK proteins are more closely related to those of the *Clostridia* and *Mycoplasmas*. Similarly, RF1 and RF2 of the *Rickettsiae* group with the proteobacteria, but the HemK proteins are much more closely related to those of the Bacilli and Bacteroides. In all eight species comprising these two cases, many open reading frames separate the *prfA* and *hemK* genes, a rare condition among the proteobacteria (15 of 59 species examined). These anomalies are highly suggestive of lateral gene transfer.

Discussion

Comparison of the HemK and substrate RF1 and RF2 phylogenies gave the surprising finding that while the RF trees are largely congruent with a 16 S rRNA tree for the same organisms (not shown), the HemK tree is not. Congruence of the RF1, RF2 and 16 S rRNA trees suggests that frequent lateral transfer of *prfA* and *hemK* together does not explain the widespread synteny observed for these two genes. Improved fitness from tightly regulated expression of these two genes provides a more likely explanation of their close physical linkage.

Aberrancies in the HemK phylogeny suggest that lateral transfer of the *hemK* gene alone may have occurred in at least two instances (Figure 4). Although loss of *hemK* is not lethal,² the ubiquity of the gene in all domains of life suggests some important advantage is conferred by its presence. In a lineage where mutations in the *hemK* gene had affected, perhaps deleteriously, the ability of HemK to methylate its substrates, there would be an opportunity for a horizontally transferred gene to establish itself. Since there are only two known substrates, RF1 and RF2, this would make it much easier for a foreign *hemK* gene to become established than if multiple substrates were involved or if many protein–protein interactions were required for viability. Oddly, the precise advantage conferred by *hemK* remains unclear. Loss or reduction of RF2 methylation by *hemK*, whether by *hemK* knockout or RF2 overexpression, strongly inhibits growth of *E. coli* K12 strains.^{1,2,26} However, the RF2 proteins of K12 strains contain an Ala246Thr mutation relative to other *E. coli* strains, rendering them partially defective and particularly sensitive to loss of HemK methylation. The effect of *hemK* loss in strains with fully functional RF2 proteins has yet to be examined.

We determined the crystal structure of HemK from *E. coli*. We found that the highly conserved residues are clustered in the catalytic domain, essential for cofactor binding and methyl transfer. Structure-based sequence alignment among 47 members of HemK family revealed the N-terminal helical domain, putatively involved with substrate recognition, displays little sequence conservation despite structural similarity of the domain between the two family members from *E. coli* and *T. maritima*. In the N-terminal domain, eight conserved residues among the five HemK members shown in Figure 2 (green colored) are associated with the hydrophobic core or intra-molecular interaction, while the surface residues, potentially important for substrate recognition, are not conserved. To date, among the suspected substrates, only the structures of *E. coli* RF2^{27,28} and the eukaryotic class I release factor eRF1²⁹ have been solved. Nonetheless, RF2 and eRF1 have little structural similarity, aside from the GGQ motif positioned within a surface loop, suggesting broad latitude in enzyme–substrate interactions. Taken further, this work provides an example of a conserved sequence motif (NPPY) that carries out a common function (amino methylation) on a range of substrates that are distinct chemically (guanine *versus* cytosine *versus* glutamine) as well as structurally (DNA *versus* RNA *versus* protein). With this information, our structure provides useful starting points for more detailed studies of this interesting enzyme.

Materials and Methods

E. coli HemK expression and purification

The full-length *E. coli* *hemK* gene was amplified by polymerase chain reaction (PCR) using the forward primer containing an NdeI restriction site (GGGA ATTCCATATGGAATATCAACACTGG) and a reverse primer containing a BamHI restriction site (CGCGGATCCTTGTCATTGATAATAGCGGCCGAG). The amplified insert was cloned into the expression vector pAII17³⁰ and transformed into *E. coli* ER2566. The cells were grown in LB medium overnight, diluted tenfold with LB medium and grown for one hour before induction. HemK expression was induced with 0.5 mM isopropyl-1-thio- β -galactoside (IPTG) for three hours. The cells were then harvested by centrifugation, re-suspended in buffer A (20 mM Hepes (pH 7.5), 50 mM NaCl, 0.5 mM EDTA), and lysed using a French press. After centrifugation, ammonium sulfate was added to the clarified supernatant to 25% saturation; the supernatant was again collected, and more ammonium sulfate was added to reach 45% saturation. The pellet of 45% saturation was collected and dissolved in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 0.5 mM EDTA. The solution was then loaded onto a hydroxy-apatite column equilibrated with the same buffer, and the bound proteins were eluted with a linear gradient from 20 mM to 500 mM sodium phosphate. Fractions containing the HemK protein were pooled and dialyzed against buffer A. For further purification, the partially purified HemK was loaded onto a Q Sepharose column (Pharmacia) equilibrated with buffer A, and eluted with a linear gradient from 50 mM to 500 mM NaCl. Fractions containing HemK were then combined and loaded onto a Superdex 75 column (Pharmacia) equilibrated with 20 mM Hepes (pH 7.0), 50 mM NaCl, and 0.5 mM EDTA. Finally, the purified HemK protein was concentrated to 10 mg/ml prior to crystallization.

Crystallography

Native crystals of HemK were grown at 16 °C by the hanging-drop method. A 1 μ l volume of 10 mg/ml of protein in 0.5 M NaCl, 20 mM Hepes (pH 7.0) was mixed with 1 μ l of a well solution containing 1.3–1.6 M ammonium sulfate, 50 mM Tris (pH 7.5). The mixture was then allowed to equilibrate over the well solution. Both native and SeMet-incorporated crystals were small needles, diffracted X-rays to ~ 3.6 Å resolution or lower, and suffered severe radiation damage. However, a native crystal soaked briefly in mercury solution diffracted better, improved resolution to 3.2 Å and lasted longer in the synchrotron X-ray beam, producing a derivative suitable for multiwavelength anomalous dispersion (MAD) phase determination. This is reminiscent of a yeast proliferating cell nuclear antigen (PCNA) crystal, whose mercury derivative diffracted better than the native crystal.³¹

Diffraction data were collected at Brookhaven National Laboratory using beamline X25 of the National Synchrotron Light Source. The mercury MAD data sets were collected at three wavelengths; only the inflection and the edge wavelengths were used (Table 1), and the data from the remote wavelength were discarded due to intensities decay and non-isomorphism. The data were subsequently reduced and scaled using the DENZO/SCALEPACK package.³² The crystal belongs to the $P6_3$ space group, with one molecule per asymmetric unit.

The HemK structure was first solved by molecular replacement with AMoRe.³³ The starting model was a HemK homolog from *T. maritima* (PDB code 1NV8). *E. coli* and *T. maritima* HemK proteins share 23% sequence identity (Table 2). However, only after extensive modification of the *T. maritima* HemK model by deleting the non-conserved side-chains was the correct solution obtained in the inflection dataset with a correlation coefficient of 0.19 and an *R* factor of 0.48 over the resolution range of 10–4.0 Å. The molecular replacement phases were used to locate the mercury sites *via* an anomalous difference Fourier map in the MAD data. The experimental phases were calculated using the program SHARP.³⁴ A combination

of the molecular replacement and experimental phases, implemented in DMMULTI,³⁵ improved the quality of the electron density map greatly (Figure 3E) and made it suitable for interpretation of the structure. The initial chain tracing and all subsequent model buildings were performed using O.³⁶

Crystallographic refinement in the resolution range of 25–3.19 Å employed the programs CNS³⁷ and Refmac38 alternately. The two programs gave options of strong (CNS) and loose (Refmac) restraints on the protein geometry during each individual refinement cycle, followed by manual refitting. The final model includes all the residues of *E. coli* HemK, with an *R*-value of 0.289 and a free *R*-value of 0.315. Because of the limited resolution (3.19 Å), we took a conservative approach: discontinuous densities do exist and assignment of solvent molecules to these densities would further reduce the value of the *R*-factor but had little effect on *R*_{free}. Thus, we did not add any solvent molecules to the final model. Furthermore, we compared our structure to other structures deposited in the PDB with a similar resolution (3.2 Å): our *R*_{free} value of 0.315 is comparable, for example, to 0.305 of PDB 1A9B,³⁹ and 0.336 of 1ADV.⁴⁰ However, the *R*-value varies from 0.206 in 1ADV (13% difference between *R* and *R*_{free}), 0.251 in 1A9B (5.4% difference), to 0.289 in our structure (2.6% difference). It is interesting to note the non-conserved residues in the surface mediate the protein–protein contacts in the crystal lattice, and thus contribute to the diffraction limit:⁴¹ while *T. maritima* HemK diffracted X-rays to 2.2 Å resolution,⁹ *E. coli* HemK with 23% sequence identity diffracted to 3.2 Å; similarly, human monoamine oxidase (MAO) B diffracted X-rays up to 1.7 Å resolution,⁴² rat MAO A with 70% sequence identity diffracted X-ray to 3.2 Å.⁴³

The Figures were drawn with the programs XtalView,⁴⁴ MOLSCRIPTS,⁴⁵ GRASP⁴⁶ and Raster3D.⁴⁷

Phylogenetic analysis

The *hemK* and *prfA* orthologs were identified using BLAST.⁴⁸ Corresponding protein sequences were conceptual translations obtained from GenBank, and were aligned with CLUSTAL X v1.83⁴⁹ using the default parameters, followed by minor manual refinement. Phylogenies were calculated using MrBayes v3.0b4,^{50,51} using the WAG rate matrix as the amino acid substitution model.⁵²

Abbreviations

AdoHcy, S-adenosyl-L-homocysteine; RF, release factor; MAD, multiwavelength anomalous dispersion..

Acknowledgements

We are very grateful to Heidi L. Schubert (University of Utah) for the coordinates of the *T. maritima* HemK before publication, and to Annie Heroux and Michael Becker (Brookhaven National Laboratory) for help with X-ray data collection at beamlines X25 and X26C in the facilities of the National Synchrotron Light Source. The study was supported, in part, by US Public Health Service grants (GM49245 and GM61355) to X.C., NIH/NIGMS IRACDA Fellowships in Research and Science Teaching grant (GM00680) to L.S., and New England Biolabs (to M.Z., B.A. and R.J.R.).

References

1. Heurgue-Hamard V, Champ S, Engstrom A, Ehrenberg M, Buckingham RH. The hemK gene in *Escherichia coli* encodes the N(5)-glutamine methyltransferase that modifies peptide release factors. EMBO J 2002;21:769–778. [PubMed: 11847124]
2. Nakahigashi K, Kubo N, Narita S, Shimaoka T, Goto S, Oshima T, et al. HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release

- factors, and hemK knockout induces defects in translational termination. *Proc. Natl Acad. Sci. USA* 2002;99:1473–1478. [PubMed: 11805295]
3. Nakayashiki T, Nishimura K, Inokuchi H. Cloning and sequencing of a previously unidentified gene that is involved in the biosynthesis of heme in *Escherichia coli*. *Gene* 1995;153:67–70. [PubMed: 7883187]
 4. Le Guen L, Santos R, Camadro JM. Functional analysis of the *hemK* gene product involvement in protoporphyrinogen oxidase activity in yeast. *FEMS Microbiol. Letters* 1999;173:175–182.
 5. Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol* 1995;253:618–632. [PubMed: 7473738]
 6. Bujnicki JM, Radlinska M. Is the HemK family of putative S-adenosylmethionine-dependent methyltransferases a “missing” zeta subfamily of adenine methyltransferases? A hypothesis. *IUBMB Life* 1999;48:247–249. [PubMed: 10690633]
 7. Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA* 1996;93:10268–10273. [PubMed: 8816789]
 8. Dincbas-Renqvist V, Engstrom A, Mora L, Heurgue-Hamard V, Buckingham R, Ehrenberg M. A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation. *EMBO J* 2000;19:6900–6907. [PubMed: 11118225]
 9. Schubert HL, Phillips JD, Hill CP. Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase. *Biochemistry* 2003;42:5592–5599. [PubMed: 12741815]
 10. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci* 2003;28:329–335. [PubMed: 12826405]
 11. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 1994;76:357–369. [PubMed: 8293469]
 12. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol* 1993;233:123–138. [PubMed: 8377180]
 13. de Beer T, Carter RE, Lobel-Rice KE, Sorkin A, Overduin M. Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science* 1998;281:1357–1360. [PubMed: 9721102]
 14. Booth V, Koth CM, Edwards AM, Arrowsmith CH. Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J. Biol. Chem* 2000;275:31266–31268. [PubMed: 10811649]
 15. Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 1997;91:741–752. [PubMed: 9413984]
 16. Huang L, Hung L, Odell M, Yokota H, Kim R, Kim SH. Structure-based experimental confirmation of biochemical function to a methyltransferase, MJ0882, from hyperthermophile *Methanococcus jannaschii*. *J. Struct. Funct. Genomics* 2002;2:121–127. [PubMed: 12836702]
 17. Perriere G, Duret L, Gouy M. HOBACGEN: database system for comparative genomics in bacteria. *Genome Res* 2000;10:379–385. [PubMed: 10720578]
 18. Blanc V, Gil P, Bamas-Jacques N, Lorenzon S, Zagorec M, Schleuniger J, et al. Identification and analysis of genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-l-phenylalanine precursor of pristinamycin I. *Mol. Microbiol* 1997;23:191–202. [PubMed: 9044253]
 19. Frolova LY, Tsivkovskii RY, Sivolobova GF, Oparina NY, Serpinsky OI, Blinov VM, et al. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* 1999;5:1014–1020. [PubMed: 10445876]
 20. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol* 1971;55:379–400. [PubMed: 5551392]
 21. Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nature Struct. Biol* 2001;8:121–125. [PubMed: 11175899]

22. Gong W, O’Gara M, Blumenthal RM, Cheng X. Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucl. Acids Res* 1997;25:2702–2715. [PubMed: 9207015]
23. Elliott T. Cloning, genetic characterization, and nucleotide sequence of the *hemA-prfA* operon of *Salmonella typhimurium*. *J. Bacteriol* 1989;171:3948–3960. [PubMed: 2544564]
24. May BJ, Zhang Q, Li LL, Paustian ML, Whittam TS, Kapur V. Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl Acad. Sci. USA* 2001;98:3460–3465. [PubMed: 11248100]
25. Fauman, EB.; Blumenthal, RM.; Cheng, X. Structure and evolution of AdoMet-dependent methyltransferases. In: Cheng, X.; Blumenthal, RM., editors. *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions*. Singapore: World Scientific Publishing; 1999. p. 1–38.
26. Uno M, Ito K, Nakamura Y. Functional specificity of amino acid at position 246 in the tRNA mimicry domain of bacterial release factor 2. *Biochimie* 1996;78:935–943. [PubMed: 9150870]
27. Vestergaard B, Van LB, Andersen GR, Nyborg J, Buckingham RH, Kjeldgaard M. Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell* 2001;8:1375–1382. [PubMed: 11779511]
28. Rawat UB, Zavialov AV, Sengupta J, Valle M, Grassucci RA, Linde J, et al. A cryoelectron microscopic study of ribosome-bound termination factor RF2. *Nature* 2003;421:87–90. [PubMed: 12511960]
29. Song H, Mugnier P, Das AK, Webb HM, Evans DR, Tuite MF, et al. The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 2000;100:311–321. [PubMed: 10676813]
30. Kong H, Kucera RB, Jack WE. Characterization of a DNA polymerase from the hyper-thermophile archaea *Thermococcus litoralis*. *Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exo-nuclease activities*. *J. Biol. Chem* 1993;268:1965–1975. [PubMed: 8420970]
31. Krishna TS, Fenyo D, Kong XP, Gary S, Chait BT, Burgers P, Kuriyan J. Crystallization of proliferating cell nuclear antigen (PCNA) from *Saccharomyces cerevisiae*. *J. Mol. Biol* 1994;241:265–268. [PubMed: 7914545]
32. Otwinowski Z, Borek D, Majewski W, Minor W. Multiparametric scaling of diffraction intensities. *Acta Crystallog. sect.A* 2003;59:228–234.
33. Navaza J. Implementation of molecular replacement in AMoRe. *Acta Crystallog. sect.D* 2001;57:1367–1372.
34. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallog. sect. D* 2003;59:2023–2030.
35. Cowtan KD, Zhang KY. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol* 1999;72:245–270. [PubMed: 10581970]
36. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A* 1991;47:110–119.
37. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. NMR crystallography system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D* 1998;54:905–921.
38. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D* 1997;53:240–255.
39. Menssen R, Orth P, Ziegler A, Saenger W. Decamer-like conformation of a nona-peptide bound to HLA-B*3501 due to non-standard positioning of the C terminus. *J. Mol. Biol* 1999;285:645–653. [PubMed: 9878435]
40. Kanellopoulos PN, Tsernoglou D, van der Vliet PC, Tucker PA. Alternative arrangements of the protein chain are possible for the adenovirus single-stranded DNA binding protein. *J. Mol. Biol* 1996;257:1–8. [PubMed: 8632448]
41. Kang YN, Adachi M, Mikami B, Utsumi S. Change in the crystal packing of soybean beta-amylase mutants substituted at a few surface amino acid residues. *Protein Eng* 2003;16:809–817. [PubMed: 14631070]

42. Binda C, Li M, Hubalek F, Restelli N, Edmondson DE, Mattevi A. Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proc. Natl Acad. Sci. USA* 2003;100:9750–9755. [PubMed: 12913124]
43. Ma J, Yoshimura M, Yamashita E, Nakagawa A, Ito A, Tsukihara T. Structure of rat monoamine oxidase A and its specific recognitions for substrates and inhibitors. *J. Mol. Biol* 2004;338:103–114. [PubMed: 15050826]
44. McRee DE. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol* 1999;125:156–165. [PubMed: 10222271]
45. Esnouf RM. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D* 1999;55:938–940.
46. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet* 1991;11:281–296. [PubMed: 1758883]
47. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. *Methods Enzymol* 1997;277:505–524. [PubMed: 18488322]
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol* 1990;215:403–410. [PubMed: 2231712]
49. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
50. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17:754–755. [PubMed: 11524383]
51. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–1574. [PubMed: 12912839]
52. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol* 2001;18:691–699. [PubMed: 11319253]
53. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol* 2000;300:1005–1016. [PubMed: 10891285]

Figure 1.

The binary structure of *E. coli* HemK-AdoHcy. A, Ribbon representation. The N-terminal domain is in green, the MTase catalytic domain is in yellow and the β -hairpin linker is in red. The AdoHcy is shown as a stick model. B, The GRASP surface representation uses the same color scheme as in A. The invariant residues (see Figure 2) are colored light blue. C, Superposition of *E. coli* HemK (red) and *T. maritima* HemK homolog (blue). D, The GRASP charge distribution. The surface is colored according to charge: positively charged groups (Arg and Lys) are blue; negatively charged groups (Glu and Asp) are red; and uncharged groups are white. E, Comparison of the HemK N-terminal domain with its structural homologs: (from left to right) *E. coli* HemK (PDB 1T43), EPS homology domain (PDB 1EH2), domain I of TFIIS

(PDB 1EO0), and KIX domain of CBP (PDB 1KDX). The green regions indicate the residues (composed of three longer helices N1, N2, and N4) used in the alignment.

Figure 2.

Structure-based sequence alignment of some selected HemK orthologs from eubacteria and mitochondria. From the original alignment of 47 sequences, five are shown here: *E. coli* (P37186), *T. maritima* (Q9WYV8), human mitochondria (Q9Y5R4), mouse mitochondria (Q921L7), and fly mitochondria (Q9VMD3). Mitochondrial localization of eukaryotic HemK homologs was predicted using the program TargetP.⁵³ The secondary structure of *E. coli* HemK is indicated above the sequence with cylinders for helices and arrows for strands; the N-terminal domain is in green, the MTase catalytic domain is in orange, and the β -hairpin linker is in red. Conserved sequence motifs characteristic of class I MTases are indicated below the sequence, labeled with roman numerals according to Malone *et al.*⁵ Residues invariant

among the 47 HemK family members are in black, and conserved residues are colored in blue. These residues are all in the catalytic domain. Additional residues conserved among the five members shown here, but not necessarily among the larger set, are colored according to the corresponding protein domain (green, red, or orange). Four other proteins exhibiting sequence similarity with *E. coli* HemK are included in the alignment: *E. coli* YfcB and two homologs, and *S. pristinaespiralis* PapM (P72542). *E. coli* YfcB is 32% identical with and 53% similar to *E. coli* HemK in the catalytic domain, from helix N5 to strand β 5. The invariant residues of the YfcB family (white against black in the *E. coli* YfcB sequence) are based on 44 BLAST hits (three are shown here: *E. coli* (P39199), HAEIN (*Haemophilus influenzae*, P45106), NEIMA (*Neisseria meningitidis* A, Q9JTA1), from the NCBI non-redundant sequence database with *E*-values below $2e^{-56}$. PapM is 29% identical with and 44% similar to *E. coli* HemK.

Figure 3.

AdoHcy binding and the active site. A, The locations of 11 invariant residues (see Figure 2). B, The ball-and-stick representation of AdoHcy is superimposed with a simulated annealing omit map ($F_o - F_c$) that contoured at 2.5σ . The hydrogen bonds are shown as broken lines. C, The superposition of the active site of *E. coli* HemK-AdoHcy (red) and *T. maritima* HemK-Gln (light blue) complexes. D, The superposition of the active site of *E. coli* HemK-AdoHcy (red) and M. TaqI-DNA (light blue) complexes. E, A representative electron density map (in stereo), contoured at 1.0σ , phased with combined molecular replacement and experimental phases, showing the central β sheet of the catalytic domain, superimposed with the refined structure.

Figure 4.

Summary of Bayesian phylogenetic analyses of 126 HemK proteins and comparable sets of RF1 and RF2 proteins and 16 S rRNA showing the displacement of the Rickettsiae (green) and ϵ -Proteobacteria (red) in the HemK tree relative to the other three trees. The general topology of the HemK tree is shown on the left, with posterior probabilities for certain nodes indicated. Major bacterial taxonomic groups that are largely monophyletic with high posterior probability are represented as single entries in the summary tree, with an example species listed for each. A consensus of the corresponding RF1, RF2, and 16 S rRNA trees is shown on the right. Only the topology of the proteobacterial groups is shown explicitly (continuous lines), with the relationship of the remaining groups shown as unresolved (dotted lines) for simplicity. The proteobacterial topology is that of the RF2 tree; RF1 and 16 rRNA are closely comparable, but other resolutions are possible. The original HemK tree is a majority rule consensus of

11,001 likely trees (1,500,000 generations, sampling every 100 trees, discarding the first 4000 trees as burn-in), rooted using an outgroup of putative archaeal HemK homologs. RF1 consensus: 1,000,000 generations, sampling every 100 trees, discarding the first 1250; RF2 consensus: 250,000 generations, sampling every 100 trees, discarding the first 500; 16 S rRNA consensus: 1,000,000 generations, sampling every 100 trees, discarding the first 4250. RF1 and RF2 trees were rooted using an outgroup of archaeal aRF1 sequences, and the 16 S rRNA tree was rooted using an outgroup of archaeal 16 S sequences.

Table 1

Crystallographic data and refinement statistics

Parameters	Hg peak	Hg inflection
Wavelength(Å)	1.0069	1.0090
Space group	$P6_3$	
Unit cell dimensions (Å)		
$a = b$	139.7	
c	40.7	
Resolution range (Å)	25.0 – 3.19	
Completeness (%)	99.8	
R_{merge}^a	0.092	0.075
$\langle I/\sigma \rangle$	13.8	15.3
Observed reflections	39,812	39,431
Unique reflections	7749	7675
Anomalous sites	4	
Highest resolutions shell		
Resolution range (Å)	3.30 – 3.19	
Completeness (%)	99.9	100
R_{merge}^a	0.343	0.250
$\langle I/\sigma \rangle$	4.6	6.0
<i>Refinement</i>		
Resolution range (Å)		20.0 – 3.2
Molecules/asym. unit		1
R_{factor}^b		0.289
R_{free}^c		0.315
rms deviation from ideal		
Bond lengths (Å)		0.009
Bond angles (deg.)		1.8
Dihedrals (deg.)		23.0
Improper (deg.)		1.1

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity from multiple observations.

^b $R_{\text{factor}} = \sum |F_o - F_c| / \sum |F_o|$.

^c R_{free} was calculated using a subset (5%) of the reflections not used in refinement.

Table 2
Summary of gene synteny data for bacterial *prfA* and *hemK* (see the Supplementary Material for detailed analysis)

Group ^a	No. Taxa	Gene arrangement ^b					N/D
		P + H + ^c	H + P + ^c	P + [1-2]H +	P + [>2]H + ^d	P + [>2]H - ^d	
Actinobacteria	10	9 (2)	0	0	0	0	1
Bact/Chlor group	4	0	0	0	1	2	1
Bacteroidetes	3	0	0	0	0	2	1
Chlorobi	1	0	0	0	1	0	0
Chlam/Spiro	6	5 (4)	0	0	1	0	0
Chlamydia	3	3 (3)	0	0	0	0	0
Spirochaetes	3	2 (1)	0	0	1	0	0
Chloroflexi	1	0	0	0	0	0	1
Cyanobacteria	8	0	0	0	2	4	2
Deinococcales	1	0	0	0	1	0	0
Firmicutes	31	25 (21)	4 (0)	1	0	0	1
Bacillales	19	18 (16)	0	1	0	0	0
Clostridia	6	1 (1)	4 (0)	0	0	0	1
Mollicutes	6	6 (4)	0	0	0	0	0
Fusobacteria	1	1 (1)	0	0	0	0	0
Hyperthermophiles	2	0	0	0	2	0	0
Aquificae	1	0	0	0	1	0	0
Thermotogae	1	0	0	0	1	0	0
Planctomycetes	1	1 (0)	0	0	0	0	0
Proteobacteria	59	35 (20)	0	7	4 [4]	10 [4]	3
Alpha	13	9 (8)	0	1	1 [1]	3 [2]	0
Beta	8	3 (1)	0	2	0	2	1
Gamma	27	20 (9)	0	4	0	3	0
Delta	4	3 (2)	0	0	0	0	1
Epsilon	5	0	0	0	3 [3]	2 [2]	0
Other	1	0	0	0	0	0	1
Total	124 ^e	76 (48)	4 (0)	8	11	16	9

- ^aMajor bacterial taxonomic groups are indicated in boldface. Subgroups of each group, if any, are separated beneath.
- ^bP represents *prfA*, H represents *hemK*, numerals represent the number of intervening genes, if any. Signs represent strands (++ , same strand; + - opposite strands). N/D indicates that the sequence is incomplete and the arrangement is indeterminate. No gene arrangements were found other than those listed here.
- ^cFor adjacent gene arrangements, the number in parentheses indicates the number of cases in which the genes overlap by at least one nucleotide.
- ^dThe number in brackets indicates the number of taxa from Rickettsiae or ϵ -Proteobacteria, both of which are significantly misplaced on the HemK phylogenetic tree.
- ^eAside from the outgroup, two other bacterial proteins were included in the phylogenetic analysis: namely, secondary putative HemK homologs in *D. hafniense* (Clostridia group) and *M. magnetotacticum* (α -Proteobacteria group), whose genes are both located distal to the *prfA-hemK* gene pair, for a total of 126 bacterial proteins.