Structural and functional characterization of CFE88: Evidence that a conserved and essential bacterial protein is a methyltransferase

KEITH L. CONSTANTINE,¹ STANLEY R. KRYSTEK,² MATTHEW D. HEALY,³ MICHAEL L. DOYLE,¹ NATHAN O. SIEMERS,² JANE THANASSI,³ NING YAN,¹ DIANLIN XIE,¹ VALENTINA GOLDFARB,¹ JOSEPH YANCHUNAS,¹ LI TAO,¹ BRIAN A. DOUGHERTY,³ AND BENNETT T. FARMER II¹

¹Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543, USA

²Bristol-Myers Squibb Pharmaceutical Research Institute, Pennington, New Jersey 08534, USA

³Bristol-Myers Squibb Pharmaceutical Research Institute, Wallingford, Connecticut 06492, USA

(RECEIVED February 1, 2005; FINAL REVISION March 11, 2005; ACCEPTED March 12, 2005)

Abstract

CFE88 is a conserved essential gene product from *Streptococcus pneumoniae*. This 227-residue protein has minimal sequence similarity to proteins of known 3D structure. Sequence alignment models and computational protein threading studies suggest that CFE88 is a methyltransferase. Characterization of the conformation and function of CFE88 has been performed by using several techniques. Backbone atom and limited side-chain atom NMR resonance assignments have been obtained. The data indicate that CFE88 has two domains: an N-terminal domain with 163 residues and a C-terminal domain with 64 residues. The C-terminal domain is primarily helical, while the N-terminal domain has a mixed helical/extended (Rossmann) fold. By aligning the experimentally observed elements of secondary structure, an initial unrefined model of CFE88 has been constructed based on the X-ray structure of ErmC' methyltransferase (Protein Data Bank entry 1QAN). NMR and biophysical studies demonstrate binding of S-adenosyl-L-homocysteine (SAH) to CFE88; these interactions have been localized by NMR to the predicted active site in the N-terminal domain. Mutants that target this predicted active site (H26W, E46R, and E46W) have been constructed and characterized. Overall, our results both indicate that CFE88 is a methyltransferase and further suggest that the methyltransferase activity is essential for bacterial survival.

Keywords: bioinformatics; biophysical methods; conserved essential bacterial gene; methyltransferase; mutagenesis; nuclear magnetic resonance; protein modeling; *Streptococcus pneumoniae*

Supplemental material: see www.proteinscience.org

Reprint requests to: Keith L. Constantine, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, NJ 08543; e-mail: keith.constantine@bms.com; fax: (609) 252-6030.

Abbreviations: CEG, conserved essential gene; *cfe*, conserved essential gene for expression; HSQC, heteronuclear single-quantum coherence; ITC, isothermal titration calorimetry; K_d, dissociation constant; HMMs, hidden Markov models; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PDB, Protein Data Bank; RID, residue information data structure; rRNA, ribosomal ribonucleic acid; SAH, S-adenosyl-L-homocysteine; SAM, S-adenosyl-L-methionine; τ_m, mixing time; T_m, melting temperature (transition midpoint); 2D, two-dimensional; 3D, three-dimensional; 4D, four-dimensional.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051389605>.

The high mutational rate of bacteria, combined with the widespread use and misuse of antibiotics, has selected for mutations that confer resistance to antibiotics; as a result, traditional antibacterial agents are becoming increasingly ineffective against a number of infections (Cohen 1992; Davies 1994; Nikaido 1994; Spratt 1994; Walsh 2003). To help address this issue, Bristol-Myers Squibb has developed a high-throughput gene disruption system in the Gram-positive pathogen *Streptococcus pneumoniae* (Thanassi et al. 2002), strain Rx-1 (Pozzi et al. 1996). The ultimate goal of this research is to facilitate the discovery of new antibiotics

by identifying genes that are both conserved among a number of important human pathogens and essential for cell viability, and then developing inhibitors to the expressed protein product of these genes. The effort has identified 113 conserved essential genes (CEGs) in bacteria, and has provided an expedient system for over-expressing and characterizing the identified gene products (Thanassi et al. 2002). An analogous system also has been developed by Bristol-Myers Squibb to identify novel antifungal agents (Wang et al. 2003).

The functional annotations of the 113 CEGs from the above gene knock-out campaign encompassed a variety of cellular roles, with the most common being “unknown function.” However, clues to the functional roles were obtained for some of the genes with “unknown function,” such as a weak local similarity of *cfe88* (CEG for Expression 88, the subject of this article) to S-adenosyl-L-methionine (SAM)-dependent methyltransferases. These enzymes are involved in methyl group transfers to a wide range of macromolecular substrates. A well-studied bacterial methyltransferase is ErmC', which confers antibiotic resistance against the macrolide-lincosamide-streptogramin type B antibiotics via methylation of an adenine residue in the bacterial 23S rRNA (Eady et al. 1990; Weisblum 1995). The genetic regulation of various *erm* genes has been investigated in a number of Gram-positive bacteria (for review, see Leclercq and Courvalin 2002), and structures of ErmC' methyltransferase have been determined (Bussiere et al. 1998; Schluckebier et al. 1999).

In this report, we provide a more detailed analysis of the *cfe88* gene product, hereafter referred to as CFE88. CFE88 is a 227-residue protein with no significant ($\geq 30\%$) amino-acid sequence similarity to any proteins with publicly available 3D structures (Berman et al. 2000). Combining bioinformatics analysis, protein threading (computational fold recognition), and experimental NMR data has afforded a structure-guided sequence alignment between CFE88 and ErmC' methyltransferase. Subsequently, an initial unrefined model of CFE88 has been built by using this alignment. CFE88 is shown to bind to S-adenosyl-L-homocysteine (SAH) by NMR and other biophysical methods. The NMR data show that critical conserved residues, located in or near the predicted active site, are perturbed by SAH binding. This result is consistent with the inferred methyltransferase function of CFE88. Genetic studies further suggest that CFE88 is an attractive target for antibiotic development.

Results

Bioinformatics analysis and protein threading

CFE88 did not demonstrate significant amino acid sequence similarity to any proteins with publicly available 3D

structures (Berman et al. 2000). Searches of the nonredundant database at National Center for Biotechnology Information (NCBI) revealed global sequence similarity of CFE88 to a family of bacterial “conserved hypothetical” proteins of unknown function, and weak local similarity in the N-terminal segment ($\sim 70\%$) of CFE88 to several genes annotated as methyltransferases (data not shown). A multiple sequence alignment of CFE88 and similar proteins is shown in Figure 1. Since only one related protein sequence was found in each of the examined bacterial genomes, all of the sequences shown in Figure 1 are likely to be orthologs of CFE88. The alignment demonstrates a few regions of conservation, most notably those around His26, Glu46, and Gly94. Based on nearly full-length sequences, the individual regions of conserved sequence in CFE88 and its orthologs match conserved segments in proteins with PFAM designations pfam04816, also known as DUF633 (domain of unknown function), as well as COG2384 (predicted SAM-dependent methyltransferase).

The tentative identification of CFE88 as a methyltransferase is supported by computational protein threading of the CFE88 sequence against all known protein folds from the Protein Data Bank (PDB) (Berman et al. 2000). The top scoring protein was L-isoadipate O-methyltransferase (PDB code 1JG1), with a threading index of 41.5 (Table 1A). Of the top 10 threading hits, four were from the same 3D fold: The SCOP superfamily is SAM-dependent methyltransferases, and in the CATH Protein Structure Classification Database the superfamily is numbered 3.40.50.150 (transferase/methyltransferase). The protein fold filter was then applied, and the top 20 fold family members are shown in Table 1A. The best scoring protein has 19% identity over 150 aligned amino acids. The fold family contains a majority of methyltransferase proteins. To better understand the putative biochemical function of CFE88, the members of these fold superfamilies were examined by using the enzyme nomenclature (E.C.) for methyltransferases, E.C. 2.1.1. (transferases, transferring one-carbon groups, methyltransferases). Table 1B shows the 10 families of methyltransferases that are members of the fold superfamily. All 10 family members use SAM to provide the methyl group for transfer. Conformational similarities between methyltransferase domains of superfamily members can be seen in Figure 2, which shows the top-ranked methyltransferase L-isoadipate O-methyltransferase (green) overlaid with two other highly ranked family members, catechol O-methyltransferase (blue) and ErmC' methyltransferase (red). This fold superfamily has the Rossmann fold, which contains a core β -sheet of seven strands surrounded by helices forming a three-layer ($\alpha\beta\alpha$) sandwich.

NMR assignments

To further characterize CFE88, heteronuclear multidimensional NMR studies have been undertaken.

```

          *                               *
Spn_CFE88 -----MISKRELVASFVSQGAII--LLDVGSDHAYLPIELVERGQIKSAIAGEVV
SPy0931 -----MDSQLSNRLAQVAAYVPKGVK--LLDVGSDHAYLPFLVETNQISAAIAGEVV
SAG1202 -----MDLQLSKRLQKVANYVPKGAR--LLDVGSDHAYLPFLLMQGYCDFAIAGEVV
SMU.1463c --MPDKKREVSLSHRLQEVASFVPKGAR--LLDVGSDHAYLPYLLQGLIDFAVAGEI
Lla_ykiC -----MNEIKLSKRLKAVANYVDNGAR--LADIGSDHAYLPYLMQKLVIDFAVAGEVV
EF1381 -----MNEQLSKRLATVGDLPGRSR--LADIGSDHAYLPVALMLENKLFAVAGEVV
PP3495 -----MNEQTLSRRLERVAHVPPQGAR--LADIGSDHGYLPVALMLRGVLEAAVAGEVA
LMOF2365_1471 MKVGNQMNEEQLSKRLKVASIYTKNER--IADIGSDHAYLPCFAVNQATASFAIAGEVV
SE1246 -----MINLNQRLSIVCSFIKRG--LADIGSDHAYLPYAIQNDLCTKVIAGEVI
SA1389 -----MISLNNRLTTSRFLKQGT--IADIGSDHAYLPYAIQNLCECGIAGEVI
BA4512 -----MNEVLSKRLKEEVVREIPVGST--VADIGSDHAYLPCYTIINNIATKAVAGEVV
Bsu_yqfN -----MADIGSDHAYLPCYAVLNHKASGAIAGEIT
CPE2004 -----MELSKRLKRIAEHVDKCES--VADIGTDHGYIPIYLVKEGICKKAIASDIN
MG249 -----MKKRISTIANLVQSFNPKLVYDIGCDHSYLTSLIKTNQNLTIIVNSDIS
MPN351 -----MKKRISTIANLVQKPKAFYDIGCDHGYLARELVQNPPLVGVNSDIS

Spn_CFE88 EGPYQSAVKNEAHLKEKIQVRLANGLAAFEETDQVSVITIACGMGGRLIARILEEG--L
SPy0931 RGPYESALKNVTSQGLAEHIQVRLANGLAAFEEDVTAITICCGMGGRLIADILEAG--K
SAG1202 NGPYQSALKNVSEHGLTSKIDVRLANGLSAFEEDNIDTITICCGMGGRLIADILNND--I
SMU.1463c KGPYESAVANVSESELGQIAVRLADGLAALNDNDIDITICCGMGGRLIADILAAG--S
Lla_ykiC KGPFEIAKNHVSEADLSRIEVRANGLGAIENTDKIDTIVIACGMGGILISEILEAG--K
EF1381 EGPYQSAKTQVSKSNLTDKIIIVRLANGLDAIEPEDQIDVISICCGMGGTLIRIDILEAGRK
PP3495 QTPFASAQNRNRRNGLQEQVTVRLADGLTAVEPQDRISVVSICCGMGGDTMCETLEAG--K
LMOF2365_1471 DGPFFQSAQKQVRSSGLTEQIDVRKGNGLAVIEKKDAIDTIVIACGMGGTLIRITILEEG--A
SE1246 QGPYKAAKNRIANYELNQVDVRLGDGLSVINSEDDQIDNITVCCGMGGPLIAKILNDG--K
SA1389 QGPFQAAVKNAANHLVERIDVRLGDGLSVIHPEVDIDNITICCGMGGPLIAKILKDG--Q
BA4512 DGPFFSAQATVAESGLQDKVDVRKGNGLAVITPGEV--DVITVAGCGMGGALIRDILESG--K
Bsu_yqfN DGPFFLSAKRQVEKSGLSNHSISVRQGDGLEVIKKGEA--DAITVAGCGMGGALIAHILEAG--K
CPE2004 KGPIEKAKVNVAFEGVSNKVKCLLGPGLNPLKVGEV--NGVILAGCGMGGNLTRDILLAD--M
MG249 KNALLSNYQKFKNNNNIHFVSDGFNNLPELNINPK--IGVIAGLGGLKIINIISQK---
MPN351 ANALGAAKGLLKDHTNMHFIITSDGFDLLPNLNLDP--VGVIAGLGGGLKIMQILAQH---

Spn_CFE88 GKLANVERLILQPNNREDDLRWLQDHGFQIVAESILEEAGKFYEILVVEAGQM-----K
SPy0931 EKLQGIERLVLQPNNREDDLRWLVSNAFKIVAETIMAENDKYEIIVAEHGEK-----A
SAG1202 DKLQHVKTLLVLQPNNREDDLRKWLAAANDFEIVAEDILTENDKRYEILVVKHGHM-----N
SMU.1463c DKLNSVKQLILQPNNCEDDLRSWLAVNDFMIKAEKMKDRHKYIEILVVEAGKI-----T
Lla_ykiC EKLGHVKRLILQPNNHEESLRQWLNVHQFVIKKEIILEAGKFYEIIVAEPLSKLMT--EK
EF1381 NRLTGKERLVLQPNIGEPTRLRRWLMANDYSIIDETIVEENRKYEIIVAEKTEQSV---S
PP3495 QRLGGVTRLVLQPNNGGERELRQWLAGNGYQIVSEELLRENRFDEYIIVAEPRG-----V
LMOF2365_1471 AKLAGVTKLILQPNIAAWQLREWSEQNWLITSEAILREDNKVYEIMVLAPEKPT--WT
SE1246 DKLVNHPRLILQSNITQALRQTLNKLSEIVDERIEEKGHIYEIVVAEFNNN-----L
SA1389 DKLSQHPRILILQSNITQENLRQTLQQLNYEIIDETIIMEEKGHIYEIVVAEHSTQ-----L
BA4512 EKLEGVTRLILQPNIAAHHIREWFIENGWELIHEKIVKEDGKIYEILVGERGNIEVP--YS
Bsu_yqfN DKLTGKERLILQPNIHAVHIREWLYKERYALIDEVILEEDGKCYEVLVAEAGDRDAA--YD
CPE2004 DKVKKYDFIILQPAQNPEVLRREFLYKNDYEIIDEDLIKDEGRFYELFKVKNENSEK--L
MG249 E--NFINRFVILQPSNLIELRSLSLNSWDIVNETLVQDREFIYPILVIEKLKPKFK--
MPN351 N--NFLKHFLVLPQPSNIIELRRFLSANQWTTTETLVEERGFIYLVLAVDKTKAQTYQY--

Spn_CFE88 L--SASDVRFPGPFLSKEVSPVVFQKQKEAEKLEFALGQIPEKNL--EERQ----VLVDK
SPy0931 L--SATELRFGPYLSQEKSVVFKQKQREMDKLAYALSCIEPEKT--QERQ----LLLT
SAG1202 L--TAKELRFGPFLSNNTTVFKEKWQNELNKLTFALNSIPNSKM--EERA----ILEDK
SMU.1463c L--SDKDLRFGPFLRQERSSIFKERWRKELAKLELALTKVPAKK--ADNM----FLSTK
Lla_ykiC L--SVNDLTFGPFLSKEKSTVFQKQKELNLTNKLIIARLPEEQ--EKRQ----EVLNQ
EF1381 Y--TDQELLFGPVLIIKKQGPVFTKKWQRELKQKRTVLAQLAKASG--EHIE--KQARLQD
PP3495 Y--SAEQLYFGPVLMEKSEAFVVKWRMLRQKQQTLANFQRARD--AVPQAKIDDFNQ
LMOF2365_1471 K--Q--EIFFGPGCLLKEQSAIFKSKWRHEANTWQNIQITISNNQPVSTENQAKIRELEHK
SE1246 VKLNILQEKFGPFLRECNNIFQKKWQRELEALRDIKSQNLSTSH--HERL--KEIEDE
SA1389 IELSSDELKFGPKLLNNKNEYFIKKWQRELEALYHIKSKLNTQEH--HQRL--AQINDE
BA4512 G--NKQTELFMGPFLLIKESEAFVEKWEKELKNFQNLKQLERAAD--SEETKAKRAEVEAK
Bsu_yqfN GISLSAGMLVGPFLAKEKNAVFLKKWTQELQHTQSIYEQISQAAD--TEQNQKQKLADLR
CPE2004 VFEDELYEVSPFLREKGHPLFEFIEEKINRCETILSFIKEDTEAAKKRK---SDLEEK

```

Figure 1. (Continued on next page)

Three different isotopic labeling strategies were used to obtain NMR assignments for wild-type CFE88. Uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled and uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled samples were produced. In addition, samples with $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ leucine, isoleucine, and valine residues

incorporated into an otherwise uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled protein were prepared (Metzler et al. 1996). Henceforth, these latter samples will simply be referred to as ILV-labeled samples. Figure 3 shows a 2D ^1H - ^{15}N heteronuclear single quantum correlation (HSQC)

MG249	--LTKELVILG PKL INFKDKHCLMKHYQCLLRVYQPKQKPSLMDL-----KI
MPN351	--LTEEQMLIG PHL VNFTDKQMLVKYYNGLLNFTARLNPSQFDS-----QV
Spn_CFE88	IQAIKEVLHVSK--
SPy0931	IQQIKEVIVHES--
SAG1202	IQDIKEVLDES---
SMU.1463c	IEQIKEVLYES---
Lla_ykiC	IARIEEVLRL----
EF1381	QQLIEEVLANGCER
PP3495	VGWITQVLA-----
LMOF2365_1471	IALVEDVLK-----
SE1246	INLIQEVLINEN--
SA1389	IAVIERVL-----
BA4512	MKMIGEVLIS-----
Bsu_yqfN	MELLKEVIDHG---
CPE2004	INKLKGML-----
MG249	IETLNKIITSYESS
MPN351	IHVLNKKII-----

Figure 1. Multiple sequence alignment of CFE88 and orthologous proteins. The conserved His26 and Glu46 residues mutated in this report are denoted by asterisks. Strictly conserved residues are shown in bold. Genes in the alignment are from the following organisms: Spn_CFE88 (*Streptococcus pneumoniae*), SPy0931 (*Streptococcus pyogenes*), SAG1202 (*Streptococcus agalactiae*), SMU.1463c (*Streptococcus mutans*), Lla_ykiC (*Lactococcus lactis*), EF1381 (*Enterococcus faecalis*), PP3495 (*Pseudomonas putida*), LMOF2365_1471 (*Listeria monocytogenes*), SE1246 (*Staphylococcus epidermidis*), SA1389 (*Staphylococcus aureus*), BA4512 (*Bacillus anthracis*), Bsu_yqfN (*Bacillus subtilis*), CPE2004 (*Clostridium perfringens*), MG249 (*Mycoplasma genitalium*), and MPN351 (*Mycoplasma pneumoniae*).

spectrum of an ILV-labeled CFE88 sample. The peaks in the HSQC spectrum are well dispersed, indicative of a folded protein.

Initially, a sample of uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CFE88 was used to obtain backbone ^1H N, ^{15}N , ^{13}CO , $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ assignments. Peak lists derived from 2D HSQC and 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB spectra were supplied as input to the AUTOASSIGN program (Zimmerman et al. 1997), an expert system for automatically assigning protein backbone atoms based on triple-resonance NMR experiments. AUTOASSIGN assigned 210 out of 220 backbone amide resonances by using intraresidue and sequential correlations involving the backbone ^{13}CO , $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ resonances. In order to confirm and extend the backbone assignments, data were collected by using a sample of uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CFE88. 2D HSQC and 3D HNCA, HN(CA)HA, CBCA(CO)NH, and HBHA(CO)NH spectra were acquired. These data were used to establish residue information data structures (RIDs) (Friedrichs et al. 1994; Constantine et al. 1997). The ^{13}C assignments for uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CFE88 were initially estimated by using the approach of Farmer and coworkers (Venters et al. 1996). Intraresidue and sequential $^{13}\text{C}\alpha$ and $^1\text{H}\alpha$ correlations were used to establish RID–RID links. All of the assignments obtained by AUTOASSIGN were confirmed, and additional backbone amide assignments were obtained. In summary, backbone amide assignments were obtained for all nonproline residues except Ile2, Glu81, and Ala146. $^1\text{H}\alpha$ assignments were obtained for

218 out of 227 residues, and $^1\text{H}\beta$ assignments were obtained for 141 out of 210 nonglycine residues.

A partial set of side-chain assignments was obtained from data collected on uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled and ILV-labeled samples of CFE88. For the former samples, 3D H(C)(CC)(CO)NH, 3D HCCH-TOCSY, and 4D HCNH-NOESY ($\tau_m = 80$ msec) spectra were collected. For the latter samples, 3D H(C)(CC)(CO)NH, 3D (H)C(CC)(CO)NH, 4D HNNH-NOESY ($\tau_m = 200$ msec), 4D HCNH-NOESY ($\tau_m = 90$ and 300 msec), and 4D CHCH-NOESY ($\tau_m = 90$ and 220 msec) spectra were recorded. All of these spectra were analyzed interactively. The methyl ^1H and ^{13}C resonances for all alanine, threonine, leucine, isoleucine, and valine residues have been assigned, and limited side-chain ^1H , ^{13}C , and ^{15}N assignments have been obtained for other residue types. Tables of assignments for protonated CFE88 are supplied as Supplemental Material.

Secondary structure, alignment to ErmC', and modeling

$^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ NMR assignments immediately provide information on local protein backbone conformation (Spera and Bax 1991; Wishart and Sykes 1991). Differences are computed between the observed $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts and random-coil values to produce $\Delta^{13}\text{C}\alpha$ and $\Delta^{13}\text{C}\beta$ values, respectively. $\Delta\text{C}\alpha\text{C}\beta$ defined as $\Delta^{13}\text{C}\alpha - \Delta^{13}\text{C}\beta$ (Metzler et al. 1993), is positive for residues in helical conformations and is negative for residues in extended (β -strand) conformations. By taking the difference, biases introduced by any

Table 1. A. Results from the protein threading using ProHit summarized by the top hit fold family

ThdIdx ^a	% identity	Template	Name
41.5	19.33	1jgl.A	PROTEIN-L-ISOASPARTATE O-METHYLTRANSFERASE
38.72	20.67	1dus.A	MJ0882
32.6	20.44	2dpml.A	ADENINE-SPECIFIC METHYLTRANSFERASE DPNII 1
31.09	13.87	1vid.-	CATECHOL O-METHYLTRANSFERASE
29.82	17.62	1inl2.C	SPERMIDINE SYNTHASE
24.79	14.09	1kph.B	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE 1
22.17	14.55	1lle.A	MYCOLIC ACID SYNTHASE
21.4	11.59	1yubl.-	RRNA METHYLTRANSFERASE
19.9	17.16	1kpi.A	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE 2
19.8	16.9	1f311.A	PROTEIN ARGININE METHYLTRANSFERASE PRMT 3
19.32	13.94	1qaml.A	ERMC' METHYLTRANSFERASE
18.42	14.59	1im8.B	YECCO
18.17	14.79	1j4f.A	RNA METHYLTRANSFERASE
17.72	13.23	1dl5.A	PROTEIN-L-ISOASPARTATE O-METHYLTRANSFERASE
17.58	17.39	1g6ql.3	HNRNP ARGININE N-METHYLTRANSFERASE
14.8	14.66	1khh.B	GUANIDINOACETATE METHYLTRANSFERASE
14.65	12.2	3mag.-	VP39
13.56	13.51	1ej64.A	LAMBDA2
13.04	14.18	1g551.A	DNA CYTOSINE METHYLTRANSFERASE DNMT2
12.44	9.94	1af72.-	CHEMOTAXIS RECEPTOR METHYLTRANSFERASE CHER

B. Enzyme nomenclature and name for the methyltransferases 2.1.1. that are members of the SCOP superfamily

E.C. number ^c	Enzyme name
EC 2.1.1.2	guanidinoacetate <i>N</i> -methyltransferase
EC 2.1.1.6	catechol <i>O</i> -methyltransferase
EC 2.1.1.16	methylene-fatty-acyl-phospholipid synthase
EC 2.1.1.20	glycine <i>N</i> -methyltransferase
EC 2.1.1.28	phenylethanolamine <i>N</i> -methyltransferase
EC 2.1.1.37	DNA (cytosine-5-)-methyltransferase
EC 2.1.1.48	rRNA (adenine - <i>N</i> ⁶ -)-methyltransferase
EC 2.3.1.72	indoleacetylglucose-inositol <i>O</i> -acyltransferase
EC 2.1.1.77	protein-L-isoaspartate (D-aspartate) <i>O</i> -methyltransferase
EC 2.1.1.80	protein-glutamate <i>O</i> -methyltransferase

^aThdInx = ProHit Threading Index.^bThis protein was used to model CFE88.^cSAM-dependent methyltransferases and/or the CATH Protein Structure Classification Database the superfamily is numbered 3.40.50.150, transferase/methyltransferase.

possible chemical-shift referencing errors are largely eliminated (Constantine et al. 1997). Figure 4 shows the $\Delta C\alpha C\beta$ values determined for CFE88. These data demonstrate that CFE88 has a mixed helical/extended fold for residues 1–163 and a primarily helical fold for residues 164–227. Most of the helices are verified by medium-range NOEs (see Supplemental Material). As discussed in more detail below, the NMR data on CFE88 are consistent with a 163-residue N-terminal domain that is similar to that observed in the Erm methyltransferases, and with a mainly helical 64-residue C-terminal domain that is similar in length to the C-terminal domains of the Erm methyltransferases.

In order to characterize the secondary structure and folding topology of CFE88 in greater detail, a 4D HNNH NOESY spectrum ($\tau_m = 200$ msec) was

recorded on a uniformly $^2H/^{13}C/^{15}N$ -labeled sample, after replacement of 2H with 1H at the exchangeable sites. This spectrum contains 635 manually verified peaks. The 1HN - 1HN NOEs that have been assigned with high confidence are reported in the Supplemental Material. Long-range NOE data identify a five-stranded parallel β -sheet, as shown in Figure 5. This comprises the major portion of the seven-stranded β -sheet expected for the fold superfamily.

Secondary structures and folding topologies were examined for methyltransferase X-ray structures in the PDB. This analysis took into consideration the CFE88 chemical shift and NOE information, as well as the bioinformatics, threading, and ligand-binding results (see below) that point to the identification of CFE88 as a methyltransferase. Considering the domain sizes, domain organizations, and

Figure 2. Superposition of three methyltransferase fold family members that were highly ranked by protein threading with the CFE88 sequence. The top-ranked protein L-isoaspartate O-methyltransferase (PDB code 1JG1, green), catechol O-methyltransferase (PDB code 1VID, blue), and ErmC' methyltransferase (PDB code 1QAN, colored red) are overlaid, with the protein backbones rendered in the ribbon representations.

secondary structures, the Erm family of methyltransferases appears to be most similar overall to CFE88 and its orthologs. The Erm methyltransferases are two-domain enzymes containing an N-terminal catalytic (methyltransferase superfamily) domain and a (primarily helical) C-terminal domain that has been implicated in the recognition of the macromolecular rRNA substrates (Yu et al. 1997). ErmC' methyltransferase (PDB entries 2ERC, 1QAM, 1QAN, 1QAO, and 1QAQ), a member of the Erm family of methyltransferases, was chosen as a representative for the SAM-dependent methyltransferase superfamily.

Most of the helical and extended regions of CFE88 can be approximately matched to α -helices and β -strands in ErmC' (Fig. 6). The secondary structure correspondences were used to manually refine a sequence alignment between ErmC' and CFE88 (Fig. 6). This refined alignment served as input to the LOOK program (Levitt 1992) for generating a model of CFE88 (Fig. 7). In this alignment, 39 out of 227 CFE88 residues (17.2%) are aligned to identical residues in ErmC'. Based on secondary structure and protein threading results, CFE88 residues 163–164 appear to be located at a domain boundary. It should be noted that the initial CFE88 model has not been explicitly refined by using NMR-derived restraints, and that many of the details regarding local backbone conformations, starting and ending residue positions of the elements of secondary structure, etc., are not expected to be highly accurate. This approximate model of the overall fold of CFE88 is intended to facilitate the interpretation of additional data.

Justification for using ErmC' methyltransferase for modeling CFE88 is provided above. The observation (see below) of the binding of SAH (chart 1 below) to CFE88 has yielded additional evidence for the biochemical

Chart 1

function of CFE88. SAH is an endogenous inhibitor of methyltransferases that utilize SAM as a methyl-donating substrate (cofactor). Enzymes that utilize SAM also include hydrolases and oxidoreductases; however, our results are consistent with a methyltransferase function.

Ligand binding to wild-type CFE88

An X-ray crystal structure of SAH bound to ErmC' has been determined (Schluckebier et al. 1999); this structure is shown in Figure 8A. SAH was titrated into an ^{15}N -labeled sample of CFE88 (see Materials and Methods), whereupon numerous peaks in the 2D ^1H - ^{15}N HSQC spectrum have been observed to display extensive chemical-shift changes and/or line broadening effects. These perturbations, mapped onto the CFE88 model in Figure 8B, are consistent with site-specific binding of SAH. All of the perturbed residues are in the N-terminal domain of CFE88 (see Supplemental Material). Comparing panels A and B of Figure 8

Figure 3. ^1H - ^{15}N HSQC spectrum of wild-type CFE88. This spectrum was recorded on a 200- μM sample of “ILV-labeled” CFE88 (see Materials and Methods).

Figure 4. Plot of raw (nonsmoothed) $\Delta^{13}\text{C}\alpha\text{-}\Delta^{13}\text{C}\beta$ values (δCACB ; see text) versus the CFE88 residues. $^{13}\text{C}\alpha/\beta$ chemical shifts derived from fully protonated $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CFE88 were used to compute $\Delta^{13}\text{C}\alpha\text{-}\Delta^{13}\text{C}\beta$ values.

clearly reveals that the SAH-induced perturbations of CFE88 residues are centered on the SAH binding site predicted by the ErmC'-based model. This result provides strong support both for the overall structural alignment between CFE88 and ErmC' and for identical cofactor specificities. Therefore, we conclude that CFE88 is a two-domain enzyme, with an N-terminal catalytic (methyltransferase) domain that utilizes SAM as a cofactor, and with a C-terminal domain that is probably involved in macromolecular recognition.

The binding of SAH to CFE88 was also studied by isothermal titration calorimetry (ITC). Figure 9 shows the results of titrating a concentrated solution of SAH (700 μM) into a dilute solution of CFE88 (initial concentration 28.8 μM). Fitting the experimental data to a single-site binding equation yielded the following parameters: K_d (dissociation constant) = 7.3 ± 0.2 μM , N (number of sites) = 0.985 ± 0.006 , and ΔH (observed enthalpy change) = -14.9 ± 0.1 kcal/mol. These results (enthalpy-driven single-site binding) are consistent with the site-specific binding observed by NMR.

NMR and biophysical studies of CFE88 mutants

Site-directed mutagenesis is a powerful tool for investigating protein function. CFE88 residues His26 and Glu46 were chosen for mutagenesis based on the following criteria: (1) both residues are highly conserved in the orthologs of CFE88 (Fig. 1); (2) based on the initial CFE88 model, these residues are expected to be in or near the SAH binding groove (Fig. 7); (3) both residues are hydrophilic and likely to be solvent and ligand accessible; and (4) histidines, in particular, are often involved in the catalytic mechanisms of enzymes. His26 was changed to a tryptophan, producing the H26W mutant. Glu46 was

changed to an arginine and to a tryptophan, producing the E46R and E46W mutants, respectively.

Two of the three mutants (H26W and E46R) were characterized by NMR (data not shown). Both are properly folded. For the H26W mutant, SAH induces extensive chemical-shift changes and line-broadening effects similar to those observed for the wild-type protein. In contrast, the NMR data indicate extremely weak binding of SAH to the E46R mutant. At a ligand/protein concentration ratio of $\sim 1.5:1$, very minor chemical-shift changes are observed only for Leu6 and Leu29 of the E46R mutant in the presence of SAH. In the case of the wild-type protein, the ^1H - ^{15}N HSQC peaks for these residues are broadened beyond detection by SAH binding.

Biophysical studies corroborate and extend the NMR findings. By using ThermoFluor technology, SAH was observed to bind both to the wild-type CFE88 protein and to the H26W mutant, as reflected by changes in the protein denaturation midpoint (melting) temperature upon exposure to SAH (Table 2). SAH appears to bind with higher affinity to the H26W mutant than it does to the wild-type protein, based on a larger change in melting temperature (Table 2). SAH does not induce a significant change in the melting temperature for the E46R or E46W mutants (Table 2), consistent with the NMR data indicating minimal SAH binding to the E46R mutant.

Bacterial cross-in studies of CFE88 mutants

Two mutant versions of the CFE88 gene, H26W and E46R, were crossed back in to the *S. pneumoniae* chromosome (Table 3). In both cases, the transformations

Figure 5. Schematic illustration of assigned long-range backbone ^1H - ^1H NOEs that identify a five-stranded parallel β -sheet. β -Strand numbers are shown in bold, underlined font. Residues within each strand are indicated by residue number, and dashed arrows indicate that a backbone ^1H - ^1H NOE has been assigned between a given pair of residues.

Figure 6. Manual alignment of the ErmC' and CFE88 amino acid sequences. This alignment is based on the minimal observed sequence similarity, on the secondary structure annotations given in the ErmC' X-ray structure file (PDB entry 2ERC), and on the secondary structure indicated by the CFE88 NMR data. Identical residues are shown in magenta. Below each sequence, the residue secondary structure is indicated by H (helical conformation, dark yellow) or E (extended conformation, red). Below the residue secondary structure designations, the approximate locations of the α -helices (α) and β -strands (β), as defined in Figure 1 of Schluckebier et al. 1999, are indicated (blue).

did not result in viable colonies, while control transformations behaved as expected, indicating that His26 and Glu46 both make critical contributions to the essential function of CFE88. Polarity studies of the original CFE88 knock-out (Thanassi et al. 2002) indicate that it is nonpolar; i.e., it does not appear that a downstream gene was responsible for the essential phenotype. The H26W and E46R knock-in mutants confirm and extend these results.

Discussion

An analysis of the CFE88 sequence using similarity searches and sequence alignments suggests that CFE88 is a conserved hypothetical protein found almost exclusively in Gram-positive bacteria, which has similarity to a family of genes (COG2384) annotated as SAM-binding methyltransferases. These results are further supported by computational protein threading studies, which also predicted that CFE88 is similar to methyltransferases with known 3D structures. Based on NMR assignments and secondary structure analysis, a model of CFE88 has been constructed based on conformational similarity

to the known structure of ErmC' methyltransferase. This report provides laboratory-based experimental evidence in support of the methyltransferase annotation of CFE88. As illustrated below and discussed elsewhere (Hillisch et al. 2004; Homans 2004), even an approximate model can be very useful for interpreting the results of ligand-binding studies, and for suggesting residues to target for mutagenesis. The NMR assignments and initial structural characterization of CFE88 provide a basis for additional studies of the structural, functional and dynamic features of CFE88.

SAH is an analog of SAM, a common methyl-donating substrate (cofactor) for methyltransferase reactions. SAH binds to the predicted methyltransferase active-site cleft, supporting the hypothesis that CFE88 is a methyltransferase. In addition, mutations targeting the putative active site cleft of CFE88 were designed, produced, and characterized. Mutating Glu46 to either tryptophan (E46W) or arginine (E46R) greatly reduces SAH binding affinity, indicating that this residue is important for SAH/SAM recognition. In the case of ErmC' methyltransferase, the side-chain of residue

Figure 7. Initial unrefined model of CFE88. This model was produced by using the LOOK program (Molecular Applications Group) based on the alignment with ErmC' (PDB entry 1QAN), as described in the text. The protein backbone is shown as a ribbon diagram, rendered by secondary structure: coil (orange), α -helix (pink), and β -strand (green). The side-chains of H26 and E46 are shown as magenta and red stick diagrams, respectively. These two residues were targeted for mutation studies (see text). The N-terminal domain is oriented to the bottom of the figure, and the C-terminal domain is oriented to the top of the figure.

Glu59, which aligns to CFE88 Glu46 (Fig. 6), hydrogen bonds directly to the ribose hydroxyls of SAH (chart 1) and SAM (Schluckebier et al. 1999). Our results suggest that Glu46 plays a similar role in cofactor recognition by CFE88. An acidic residue at this position is conserved across various methyltransferase families.

Mutation of the CFE88 residue His26 to a tryptophan (H26W) does not hinder SAH binding. While cofactor binding does not appear to depend on His26, the genetic cross-in studies suggest that this residue may play an important and possibly unique role in the function of CFE88 and its orthologs. While strictly conserved in the orthologs of CFE88 (Fig. 1), His26 is not a conserved residue in other methyltransferase families. The corresponding residue in ErmC' (K41) is not a conserved site in the Erm methyltransferases; a histidine at position 41 does occur in at least one member of the Erm methyltransferase family (see Maravic et al. 2003, Fig. 1). We note that, based on the X-ray structures of ErmC' with bound SAH or SAM (Schluckebier et al. 1999), a tryptophan at position 41 would not be expected to block cofactor binding in the case of the ErmC' methyltransferases. His26 may play a role in the catalytic chemistry of CFE88 and its orthologs, or it may be involved in substrate binding.

NMR data, Thermofluor T_m values (Table 2), and thermal melting profiles (data not shown) indicate that all three mutants are properly folded. Importantly, the H26W and E46R mutations are both apparently lethal

when crossed back into bacteria. (E46W was not examined by bacterial knock-in studies.) In summary, based on the CFE88 mutant studies, it appears that the essential activity of CFE88 requires a functional methyltransferase active site. As far as we are aware, this is the first indication of an essential methyltransferase activity in bacteria. It should be noted that the bacterial enzyme SAM:tRNA (uracil-5-) methyltransferase was found to be essential for bacterial survival (Persson et al. 1992). However, evidence was present indicating that the uracil methylation activity, localized to the C-terminal methyltransferase domain, is *not* the essential function of this protein. The investigators further propose that SAM:tRNA (uracil-5-) methyltransferase contains a second RNA-modifying activity localized to the N-terminal domain and that this activity is essential.

In the cases of CFE88 and the Erm methyltransferases, the methyltransferase activity is localized to the N-terminal domains, and the C-terminal domains of these enzymes are probably involved in noncovalent macromolecular recognition processes, rather than a second covalent modification activity. The observation that the C-terminal domains of CFE88 and its orthologs are not similar to any other known sequences suggests that the putative macromolecular recognition specificities of CFE88 and its orthologs are distinct from other methyltransferases.

Figure 8. SAH bound to ErmC' and ligand-induced ^1H - ^{15}N HSQC peak perturbations mapped on to the CFE88 model. Protein structures are oriented as in Figure 7. (A) X-ray crystal structure of the ErmC'/SAH complex (PDB entry 1QAN). The protein backbone (residues 9–244) is depicted by a ribbon representation, rendered by secondary structure and colored N- to C-terminal, blue to red, respectively. SAH is shown as a stick diagram color coded according to atom type. (B) Mapping of NMR-observed SAH binding effects on to the CFE88 model. The protein backbone is depicted by a ribbon representation, rendered by secondary structure. Residues showing significant SAH-induced chemical shift changes and/or line broadening are shown in red; the remaining residues are shown in green.

Figure 9. ITC study of SAH binding to CFE88. (Top) Raw titration data. (Bottom) Binding heat per mole of SAH injected (squares) and the best-fit curve (line) for a single-site binding equation.

Recently, the effects of single-site mutations on the *in vivo* function and *in vitro* activity of ErmC' methyltransferase have been reported (Maravic et al. 2003). Eight residues were selected for mutation to alanine. Surprisingly, of the eight residues mutated, only Tyr104 was found to be essential to the catalytic activity of ErmC'. This ErmC' residue aligns to Gly94 in CFE88 (Fig. 6). Gly94 and surrounding CFE88 residues are perturbed by SAH binding (see Supplemental Material), consistent with active site localization of these residues. ErmC' residues Lys41 and Glu59, which align to His26 and Glu46 of CFE88 (Fig. 6), were not mutated in the study by Maravic et al. (2003). In the Erm methyltransferase family, the glutamate residue at position 59 (ErmC' sequence) is strictly conserved (Maravic et al. 2003).

Overall, our results strongly support the identification of CFE88 as a methyltransferase with structural similarity to the Erm family of bacterial methyltransferases. It has previously been shown that CFE88 is essential for bacterial survival (Thanassi et al. 2002), and our mutational data suggest that the essential function of CFE88 involves methyltransferase activity.

Table 2. Results of Thermofluor temperature stability studies on wild type and mutant CFE88 proteins in the absence and presence of SAH

CFE88 Form	T_m , °C (control)	T_m , °C (100 μ M SAH)	ΔT_m , °C
Wild type	52.2 $\pm$ 0.05	53.7 $\pm$ 0.05	1.5
H26W	49.5 $\pm$ 0.62	52.6 $\pm$ 0.06	3.1
E46R	55.2 $\pm$ 0.05	55.2 $\pm$ 0.06	0
E46W	50.2 $\pm$ 0.06	50.0 $\pm$ 0.19	-0.2

Small-molecule inhibitors of Erm methyltransferases have been described (Hajduk et al. 1999; Hanessian and Sgarbi 2000), indicating that this is a class of enzymes for which small-molecule inhibitors can be developed. The same is likely to be true for CFE88 and its orthologs. Therefore, CFE88 appears to be a potential target for the development of novel antibacterial compounds. Furthermore, identification of the natural substrate(s) of CFE88 should provide an understanding of the biological role played by this enzyme and may reveal a novel mechanism essential for bacterial survival.

Materials and methods

Bioinformatics analysis and threading

A DeCypher hardware-accelerated bioinformatics system (TimeLogic, Inc.) was used to perform the Smith-Waterman, FrameSearch, and HMMER searches described below. The Smith-Waterman algorithm (Smith and Waterman 1981) was used to compare the amino acid sequence of CFE88 with protein sequences from completed microbial genomes, and the FrameSearch algorithm (Edelman et al. 1995) was used to compare CFE88 with nucleotide sequences from unfinished microbial genomes, using the DeCypher system as described in Healy (2003). Selected putative orthologs of CFE88 were pasted into the NCBI public BLAST server (at www.ncbi.nih.gov) to search public protein sequence databases and the NCBI CDDDB database of conserved domains (submission of a protein sequence to the NCBI BLAST server automatically performs a CDDDB search in addition to the BLAST search). The BLAST program (Altschul et al. 1997) was also used to search internal BMS sequence databases for possible CFE88 orthologs. From the best matches found, a subset were selected to perform a multiple sequence alignment by using CLUSTALW (Higgins et al. 1996). Regions of high sequence conservation among these putative orthologs of CFE88 were identified by visual inspection, and used to generate hidden Markov models (HMMs) with which the microbial sequence databases were searched by using the HMMER program (Eddy 1996) on the DeCypher system. The HMMER program was also used to search the public PFAM database of HMMs for known protein domains (Sonnhammer et al. 1997).

Protein threading was carried out by using ProHit [P]rofessional (version 1.8.1, ProCeryon BioSciences). The sequence of CFE88 was compared against all known protein folds from the PDB (Berman et al. 2000). The results were analyzed as previously described (Flockner et al. 1997). In order to add value to the threading results, the ranked threading hits were analyzed by protein fold/domain classifications using SCOP (Structural Classification of Proteins) (Murzin et al. 1995) and CATH Protein Structure Classification Database (Orengo et al. 1997). The World Wide Web version (<http://www.chem.qmul.ac.uk/iubmb>) of enzyme nomenclature was used to characterize the fold family members based upon enzymatic function.

Cloning, expression, and NMR sample preparation

Recombinant CFE88 proteins with a six-residue C-terminal His tag were prepared by first subcloning the cDNA sequence

Table 3. Summary of transformation constructs used for CFE88 genetic knock-in experiments

Construct	Description	Viability
No DNA	transformation control; no added DNA	—
<i>lytA</i> -KO	control non-essential gene knockout construct	+
<i>ftsZ</i> -KO	control essential gene knockout construct	—
CFE88-KO	CFE88 gene knockout construct	—
CFE88-WT	CFE88 wild-type gene knock-in construct	+
CFE88-H26W	CFE88 gene H26W knock-in construct	—
CFE88-E46R	CFE88 gene E46R knock-in construct	—

into the pET28b expression vector and then transforming *Escherichia coli* BL21(DE3) cells with that vector. The bacterial growth media contained [1,2- $^{13}\text{C}_2$, 99%]-sodium acetate (Isotec) and [^{15}N , 99%] ammonium sulfate (Isotec) as the sole carbon and nitrogen sources, respectively. To produce a uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled sample, cells that can tolerate high levels of $^2\text{H}_2\text{O}$ (>90%) were selected by published methods (Venters et al. 1995). Samples with $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ leucine, isoleucine, and valine residues incorporated into an otherwise uniformly $^2\text{H}/^{12}\text{C}/^{15}\text{N}$ -labeled protein were prepared using methods described previously (Metzler et al. 1996).

Cells were grown to an OD_{600} of ~ 1.0 and then induced with 0.4 mM isopropyl- β -D-thiogalactopyranoside for 5 h at 37°C. After harvesting, the cells were resuspended in 30 mL (per L of culture) of 50 mM Tris-HCl and 150 mM NaCl (pH 8.0; disruption buffer), sonicated, and clarified in a Sorval centrifuge at 16,000 rpm for 20 min. The supernatant was applied on a Ni^{2+} resin (3 mL of gel per L of culture). After washing, the protein was eluted with 0.5 M imidazole in the disruption buffer. Concentrated eluted protein was then applied on to a Superdex 75 26/60 column equilibrated with the disruption buffer. The protein peak was then concentrated into NMR buffer containing 50 mM d $_4$ -imidazole, 150 mM KCl, 7% $^2\text{H}_2\text{O}$ (pH 7.0). The following samples of wild-type CFE88 were prepared for NMR studies: ILV-labeled (0.2 mM and 0.6 mM), uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled (1.0 mM), and uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled (0.5 mM). In addition, uniformly ^{15}N -labeled samples (0.2 mM each) of the H26W and E46R mutants were prepared.

NMR spectroscopy

All NMR experiments were recorded at 33°C on 600-MHz Varian Inova or 600-MHz Varian Unity Plus spectrometers using 5 mm ^1H -observe, ^{13}C - ^{15}N triple resonance room temperature probes equipped with either triple- or single-axis (z) pulsed-field gradients. All spectra were processed and analyzed with a modified version of the FELIX program (Hare Research, Inc.; M.S. Friedrichs, unpubl.).

Backbone atom resonance assignments based on intra-residue and sequential ^{13}CO , $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ correlations were obtained from data collected on the 0.5 mM uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled sample. The 2D ^1H - ^{15}N spectra were acquired with the sequence of Kay et al. (1992), modified to incorporate water flip-back. Modified versions (Constantine et al. 1997) of published 3D triple-resonance pulse sequences were used to collect data. The 3D HNCO spectrum was recorded with a modified version of the basic sequence of

Kay et al. (1994). For the 3D HNCA, HN(CA)CB, HN(CO)CA, and HN(COCA)CB spectra, modified versions of the basic sequences of Shan et al. (1996) were used. The 3D HN(CA)CO spectrum was recorded with a modified version of the basic sequence of Matsuo et al. (1996).

Backbone assignment data were also recorded by using the 1.0 mM uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled sample. The 2D HSQC and 3D HNCA were recorded as described above. The 3D HN(CA)HA was recorded with a modified version of the basic sequence of Clubb et al. (1992). The CBCA(CO)NH and HBHA(CO)NH experiments were performed as described by Grzesiek and Bax (1992).

Side-chain assignment data were recorded by using the 1.0 mM uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled and 0.6 mM ILV-labeled samples. For both samples, 3D H(C)(CC)(CO)NH (Logan et al. 1992) and 4D HCNH-NOESY (Farmer and Mueller 1994) spectra were recorded by using the cited pulse sequences. For the 1.0 mM uniformly $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled sample, a gradient-enhanced 3D HCCH-TOCSY was also recorded, as described (Kay et al. 1993). For the ILV-labeled sample, 3D (H)C(CC)(CO)NH (Farmer and Venters 1995), 4D HNNH-NOESY (Farmer and Mueller 1994), 4D HCNH-NOESY (Farmer and Mueller 1994), and 4D CHCH-NOESY (Vuister and Bax 1993; B.T. Farmer II, unpubl.) were also acquired by using the cited methods.

Homology modeling

After preparing the manual alignment between CFE88 and ErmC' methyltransferase (Fig. 6), a model of CFE88 was constructed starting from the A chain of the 2.4 Å resolution X-ray structure of ErmC' methyltransferase (PDB code 1QAN) by using the LOOK program (Molecular Applications Group). The homology modeling protocol within LOOK was used to generate the model; this protocol utilizes an automated segment-matching algorithm (Levitt 1992) that is implemented in LOOK's SegMod module.

Biophysical measurements

ITC experiments were carried out at 25°C in a VP-ITC micro-calorimeter (Microcal Inc.) with CFE88 dialyzed extensively against 20 mM PIPES and 150 mM NaCl (pH 7.0). For the ITC measurements, 7 μL injections of SAH (700 μM), dissolved in the buffer above, were made into the calorimetric cell containing CFE88 (28.8 μM). Data were analyzed by using the single-site binding model in Origin 7 (Microcal) software.

The thermal stability enhancement effect of SAH on CFE88 was measured with a ThermoFluor instrument (Three-Dimensional Pharmaceuticals, Inc.) as described by Pantoliano et al. (2001). The technology measures thermal denaturation curves based on the increase in fluorescence due to preferential binding of the extrinsic, fluorescent, hydrophobic probe 1-anilino-8-naphthalene sulfonate (ANS) to the denatured protein. Reactions containing CFE88 (4 μM) in the presence and absence of SAH (100 μM) were analyzed in 20 mM PIPES (pH 7.0), 150 mM NaCl, 100 μM ANS, and 0.5% (v/v) DMSO. The 5 μL reaction wells were overlaid with 2.5 μL of silicon oil (Sigma) to prevent evaporation. Reactions were monitored under UV illumination in black 384-well plates (Abgene) by increasing temperature in 1°C increments with 60-sec equilibration at each degree from 25°C to 80°C.

Thermal denaturation curves were analyzed with ThermoFluor Analysis software (Three-Dimensional Pharmaceuticals, Inc.).

Genetic knock-in experiments

Genetic knock-in studies were performed by transformation of *S. pneumoniae* with plasmid constructs. These constructs consisted of 300- to 500-bp internal fragments (gene knock-out constructs) of full-length wild type or mutant CFE88 genes (knock-in constructs). Recombinant plasmids were constructed in pEVP3, transformed into *S. pneumoniae*, and assayed for cell viability as previously described (Thanassi et al. 2002).

Electronic supplemental material

Electronic supplemental materials are NMR chemical shift assignments for CFE88, assigned ^1H - ^1H NOE interactions, and a list of residues perturbed by SAH binding.

Acknowledgments

We thank Luciano Mueller, Thomas Dougherty, Michael Pucci, Daniel Davison, John Barrett, and Wesley Cosand for their support of this work.

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **35**: 3389–3402.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
- Bussiere, D.E., Muchmore, S.W., Dealwis, C.G., Schluckebier, G., Nienaber, V.L., Edalji, R.P., Walter, K.A., Lador, U.S., Holzman, T.F., and Abad-Zapatero, C. 1998. Crystal structure of ErmC', an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* **37**: 7103–7112.
- Clubb, R.T., Thanabal, V., and Wagner, G. 1992. A new 3D HN(CA)HA experiment for obtaining fingerprint HN-H α peaks in ^{15}N - and ^{13}C -labeled proteins. *J. Biomol. NMR* **2**: 203–210.
- Cohen, M.J. 1992. Epidemiology of drug resistance: Implications for a post-antimicrobial era. *Science* **257**: 1050–1055.
- Constantine, K.L., Mueller, L., Goldfarb, V., Wittekind, M., Metzler, W.J., Yanchunas Jr., J., Robertson, J.G., Malley, M.F., Friedrichs, M.S., and Farmer II, B.T. 1997. Characterization of NADP $^{+}$ binding to perdeuterated MurB: Backbone atom NMR assignments and chemical-shift changes. *J. Mol. Biol.* **267**: 1223–1246.
- Davies, J. 1994. Inactivation of antibiotics and the dissemination of resistance genes. *Science* **264**: 375–382.
- Eady, E.A., Ross, J.I., and Cove, J.H. 1990. Multiple mechanisms of erythromycin resistance. *Antimicrob. Chemother.* **26**: 461–471.
- Eddy, S.R. 1996. Hidden Markov models. *Curr. Opin. Struct. Biol.* **6**: 361–365.
- Edelman, I., Faigler, S., Mintz, E., Natan, A., and Devereux, J. 1995. Frameshift: A rigorous alignment program for searching protein databases with nucleic acid queries. *Genome Sequence and Analysis Conference*. Hilton Head, SC.
- Farmer II, B.T. and Mueller, L. 1994. Simultaneous acquisition of [^{13}C , ^{15}N] and [^{15}N , ^{15}N]-separated 4D gradient-enhanced NOESY spectra in proteins. *J. Biomol. NMR* **4**: 673–687.
- Farmer II, B.T. and Venters, R. 1995. Assignment of side chain ^{13}C resonances in perdeuterated proteins. *J. Am. Chem. Soc.* **117**: 4187–4188.
- Flockner, H., Domingues, F.S., and Sippl, M.J. 1997. Protein folds from pair interactions: A blind test in fold recognition. *Proteins* **1**: 129–133.
- Friedrichs, M.S., Mueller, L., and Wittekind, M. 1994. An automated procedure for the assignment of protein ^1H , ^{15}N , ^{13}C , ^1H , ^{13}C and ^1H resonances. *J. Biomol. NMR* **4**: 703–726.
- Grzesiek, S. and Bax, A. 1992. Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**: 6291–6293.
- Hajduk, P.J., Dinges, J., Schkeryantz, J.M., Janowick, D., Kaminski, M., Tufano, M., Augeri, D.J., Petros, A., Nienaber, V., Zhong, P., et al. 1999. Novel inhibitors of Erm methyltransferase from NMR and parallel synthesis. *J. Med. Chem.* **42**: 3852–3859.
- Hanessian, S. and Sgarbi, P. W. 2000. Design and synthesis of mimics of S-adenosyl-L-homocysteine as potential inhibitors of erythromycin methyltransferase. *Bioorg. Med. Chem. Lett.* **10**: 433–437.
- Healy, M.D. 2003. Finding homologs to nucleic acid or protein sequences using the FrameSearch program. In *Current Protocols in Bioinformatics*. (eds. Baxevanis et al.) pp. 3.2.1–3.2.23. John Wiley & Sons, NY.
- Higgins, D.G., Thompson, J.D., and Gibson, T.J. 1996. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**: 383–402.
- Hillisch, A., Pineda, L.F., and Hilgenfeld, R. 2004. Utility of homology models in the drug discovery process. *Drug Disc. Today* **9**: 659–669.
- Homans, S.W. 2004. NMR spectroscopy tools for structure-aided drug design. *Angew. Chem. Int. Ed.* **43**: 290–300.
- Kay, L.E., Keifer, P., and Saarinen, T. 1992. Pure absorption gradient-enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**: 10663–10665.
- Kay, L.E., Xu, G.Y., Singer, A., Muhandiram, D.R., and Forman-Kay, J.D. 1993. A gradient-enhanced HCCH-TOCSY experiment for recording sidechain ^1H and ^{13}C correlations in H $_2\text{O}$ samples of proteins. *J. Magn. Reson. B* **101**: 333–337.
- Kay, L.E., Xu, G.Y., and Yamazaki, T. 1994. Enhanced-sensitivity triple resonance spectroscopy with minimal H $_2\text{O}$ saturation. *J. Magn. Reson. A* **109**: 129–134.
- Leclercq, R. and Courvalin, P. 2002. Resistance to macrolides and related antibiotics in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **46**: 2727–2734.
- Levitt, M. 1992. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* **226**: 507–533.
- Logan, T.M., Olejniczak, E.T., Xu, R.X., and Fesik, S.W. 1992. Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.* **314**: 413–418.
- Maravic, G., Feder, M., Pongor, S., Flogel, M., and Bujnicki, J.M. 2003. Mutational analysis defines the roles of conserved amino acid residues in the predicted catalytic pocket of the rRNA:m ^6A methyltransferase ErmC'. *J. Mol. Biol.* **332**: 99–109.
- Matsuo, H., Kupce, E., Li, J., and Wagner, G. 1996. Use of selective C α pulses for improvement of HN(CA)CO-D and HN(CACO)NH-D experiments. *J. Magn. Reson. B* **111**: 194–198.
- Metzler, W.J., Constantine, K.L., Friedrichs, M.S., Bell, A.J., Ernst, E.I., Lavoie, T.B., and Mueller, L. 1993. Characterization of the three-dimensional solution structure of human profilin: ^1H , ^{13}C and ^{15}N NMR assignments and global folding pattern. *Biochemistry* **32**: 13818–13829.
- Metzler, W.J., Wittekind, M., Goldfarb, V., Mueller, L., and Farmer II, B.T. 1996. Incorporation of $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -[Ile, Leu, Val] into a perdeuterated, ^{15}N -labeled protein: Potential in structure determination of large proteins by NMR. *J. Am. Chem. Soc.* **118**: 6800–6801.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Nikaido, H. 1994. Prevention of drug access to bacterial targets: Permeability barriers and active efflux. *Science* **264**: 382–388.
- Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. 1997. CATH: A hierarchic classification of protein domain structures. *Structure* **5**: 1093–1108.
- Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., et al. 2001. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screening* **6**: 429–440.
- Persson, B.C., Gustafsson, C., Berg, D.E., and Bjork, G.R. 1992. The gene for a tRNA modifying enzyme, m $^{\text{U}}54$ -methyltransferase, is essential for viability in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **89**: 3995–3998.
- Pozzi, G., Masala, L., Iannelli, F., Manganelli, R., Havarstein, L.S., Piccoli, L., Simon, D., and Morrison, D.A. 1996. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: Two allelic variants of the peptide pheromone. *J. Bacteriol.* **178**: 6087–6090.

- Schluckebier, G., Zhong, P., Stewart, K.D., Kavanaugh, T.J., and Abad-Zapatero, C. 1999. The 2.2 Å structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: Implications for the reaction mechanism. *J. Mol. Biol.* **289**: 277–291.
- Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H., and Kay, L.E. 1996. Assignment of ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and HN resonances in an ^{15}N , ^{13}C , ^2H labeled 64 kDa Trp repressor-operator complex using triple-resonance NMR spectroscopy and ^2H -decoupling. *J. Am. Chem. Soc.* **118**: 6570–6579.
- Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* **147**: 195–197.
- Sonnhammer, E.L., Eddy, S.R., and Durbin, R. 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* **28**: 405–420.
- Spera, A. and Bax, A. 1991. Empirical correlation between protein backbone conformation and C^α and C^β ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**: 5490–5492.
- Spratt, B.G. 1994. Resistance to antibiotics mediated by target alterations. *Science* **264**: 388–393.
- Thanassi, J.A., Hartman-Neumann, S.L., Dougherty, T.J., Dougherty, B.A., and Pucci, M.J. 2002. Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res.* **30**: 3152–3162.
- Venters, R.A., Huang, C.-C., Farmer II, B.T., Trolard, R., Spicer, L.D., and Fierke, C.A. 1995. High level $1\text{H}/^{13}\text{C}/^{15}\text{N}$ labeling of proteins for NMR studies. *J. Biomol. NMR* **5**: 339–344.
- Venters, R.A., Farmer II, B.T., Fierke, C.A., and Spicer, L.D. 1996. Characterizing the use of perdeuteration in NMR studies of large proteins: ^{13}C , ^{15}N , and ^1H assignments of human carbonic anhydrase II. *J. Mol. Biol.* **264**: 1101–1116.
- Vuister, G.W. and Bax, A. 1993. Increased resolution and improved spectral quality in four-dimensional $^{13}\text{C}/^{13}\text{C}$ separated HMQC-NOESY-HMQC spectra using pulsed field gradients. *J. Magn. Reson. ser. B* **101**: 210–213.
- Walsh, C. 2003. Where will new antibiotics come from? *Nat. Rev. Microbiol.* **1**: 65–70.
- Wang, Y., Liu, M., Dougherty, B.A., Healy, M.D., Davison, D.B., Mazzucco, C.E., Krystek, S.R., and Bassolino, D.A. 2003. New fungal nucleic acids and polypeptides, useful in diagnosing, preventing, treating or ameliorating fungal infection or infectious disease, eliciting an immune response or in drug screening. Patent WO2003091418-A2 (Publication date: 06 November 2003; Filing date: 25 April 2003; 2003WO-US013081).
- Weisblum, B. 1995. Erythromycin resistance by ribosome modification. *Antimicrob. Chemother.* **39**: 577–585.
- Wishart, D.S. and Sykes, B.D. 1991. Relationship between nuclear magnetic resonance chemical shifts and secondary structure. *J. Mol. Biol.* **222**: 311–333.
- Yu, L., Petros, A.M., Schnuchel, A., Zhong, P., Severin, J.M., Walter, K., Holzman, T.F., and Fesik, S.W. 1997. Solution structure of an rRNA methyltransferase (ErmAM) that confers macrolide-lincosamide-streptogramin antibiotic resistance. *Nat. Struct. Biol.* **4**: 483–489.
- Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C.-Y., Powers, R., and Montelione, G.T. 1997. Automated analysis of protein NMR assignments using methods from artificial intelligence. *J. Mol. Biol.* **269**: 592–610.