Distinction between the Cfr methyltransferase conferring antibiotic resistance and the housekeeping RlmN methyltransferase

Gemma C. Atkinson¹, Lykke H. Hansen², Tanel Tenson¹, Anette Rasmussen², Finn Kirpekar², and Birte Vester²#

¹ Institute of Technology, University of Tartu, Nooruse 1, Tartu, Estonia
² Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark

# Author to whom correspondence should be addressed; b.vester@bmb.sdu.dk; Tel.: +45-6550-2406; Fax: +45-6550-2467.

Abstract

The cfr gene encodes the Cfr methyltransferase that primarily methylates C-8 in A2503 of 23S ribosomal RNA in the peptidyl transferase region of bacterial ribosomes. The methylation provides resistance to six classes of antibiotics of clinical and veterinary importance. The rlmN gene encodes the RlmN methyltransferase that methylates C-2 in A2503 in 23S ribosomal RNA and A37 in tRNA, but RlmN does not significantly influence antibiotic resistance. The enzymes are homologous and use the same mechanism involving radical S-adenosyl methionine to methylate RNA via an intermediate involving a methylated cysteine in the enzyme and a transient cross-linking to the RNA, but differ in which carbon atom in the adenine they methylate. Comparative sequence analysis identifies differentially conserved residues that indicate functional sequence divergence between the two classes of Cfr- and RlmN-like sequences. The differentiation between the two classes
is supported by previous and new experimental evidence from antibiotic resistance, primer extensions and mass spectrometry. Finally, evolutionary aspects of the distribution of Cfr- and RlmN-like enzymes are discussed.

Introduction

Cfr was first reported in 2000 (1) and RlmN in 2008 (2). Cfr confers resistance to antibiotics binding to the peptidyl transferase center on the ribosome, defining a PhLOPSa phenotype reflecting resistance to phenicol, lincosamide, oxazolidinone, pleuromutilin and streptogramin A antibiotic classes (3) and Cfr also provides resistance to some large macrolide antibiotics (4). The \textit{cfr} gene is thus a health threat when spreading in pathogenic bacteria because many clinically important antibiotics become useless for treatment. The \textit{cfr} gene with only minor sequence differences has now been found worldwide in various bacteria isolated from humans and animals (5, 6), and as summarized in (7)). It is also evident that the \textit{cfr} gene can be horizontally transferred to its hosts as it is always found either on plasmids or together with insertion sequences. Competition experiments involving wild-type and inactivated Cfr indicate only a small fitness cost upon expression of Cfr and no cost related to the C-8 methylation itself (8). Recently, we have cloned three \textit{cfr}-like genes from the order Bacillales and confirmed that they indeed confer resistance like the original Cfr methyltransferase (7). This indicates that there is a natural reservoir of \textit{cfr}-like genes.

The primary product of the Cfr-mediated methylation is 8-methyladenosine (m\textsuperscript{8}A), a new natural RNA modification (9) that has so far not been seen at other sites than A2503 in 23S rRNA. In addition, Cfr provides 2,8-dimethyladenosine, although at a lower
efficiency. It was also established by site-directed mutagenesis that Cfr is a radical S-adenosylmethionine (SAM) enzyme (9) as already suggested by previous sequence comparison (10).

The 2-methyladenosine (m2A) modification at E. coli 23S RNA was discovered in 1995 (11), and the responsible “housekeeping” RlmN methyltransferase was subsequently shown by phylogenetic comparisons to be similar to Cfr (2). The phenotypic effects of RlmN are uncertain, but small effects on antibiotic binding as well as fitness have been presented (2, 12, 13). It has also been suggested that the modification fine-tunes interactions between ribosomes and nascent peptides involved in stalling (14). Recently, it was shown that RlmN is a dual-specificity enzyme that also methylates A37 in tRNA (15) and it was proposed that the loss of A2503 23S RNA modification causes reduced proofreading in protein synthesis.

Both RlmN and Cfr are radical SAM enzymes, a super family that catalyzes a diverse set of reactions that involve cleavage of unreactive C-H bonds by a 5-deoxyadenosyl radical generated by reductive cleavage of SAM (16, 17). A new mechanism involving protein methylation and transitory cross-linking has recently been proposed to explain the mechanism of methylation by Cfr and RlmN (18-20). Also an X-ray structure of RlmN with a Fe-S cluster, both with and without a SAM ligand has been published (21). Cfr can be modeled on this structure showing important differences between the enzymes (21). It has been suggested that the cfr gene evolved from the rlmN gene via gene duplication (22), but the lineage in which the duplication occurred is unknown.

In this study we use bioinformatics to identify RlmN and Cfr homologs, and identify strongly conserved sequence differences between these classes of enzymes. Our
phylogenetic analysis shows that Cfr-like proteins form a distinct, well supported group within the RlmN family. The theoretical differentiation of these enzymes’ function is supported by previously obtained functional evidence together with new findings from gene cloning followed by determination of antibiotic resistance as well as modification analysis by primer extension and mass spectrometry. Our sequence searching and phylogenetic classification also reveals other distinct groups within the RlmN family, including eukaryotic and bacterial clusters of unknown function.

Materials and Methods

Data Set Assembly

PSI-BLAST was carried out against the NCBI RefSeq protein database using S. sciuri Cfr as the query. Three iterations were run with an E value cut off of 0.01. The resulting 5101 sequences were aligned using MAFFT v6.864b (23).

FastTree (24) was used to construct a phylogenetic tree of the alignment. This showed a group corresponding to Cfr-RlmN-like sequences, along with other more distantly related homologs: pyruvate formate lyase activating enzyme and nitrogenase iron-molybdenum cofactor biosynthesis protein, molybdenum cofactor biosynthesis protein A, and coenzyme PQQ synthesis protein. The more distant relatives in the tree and extremely truncated partial sequences were removed to create a dataset of Cfr plus RlmN family sequences, which was realigned.

Consensus sequences were generated with the Python script Consensus Finder (25). The selection was based on the following principles: only two of the plasmid-borne Cfrs are
included as these are almost identical and would be overrepresented in the data set. Those genes, where some functional evidence or other knowledge of the protein is present, are included. In addition, the RlmN sequences are selected to sample broadly across the tree.

Phylogenetic Analyses
Sites were selected for phylogenetic analysis using SeqFIRE (26) with lenient settings (the BLOSUM62 substitution group, a similarity threshold of 50%, the maximum size of a non-conserved block set to 25 and a minimum block size of one amino acid). The resulting Cfr+RlmN dataset dimensions were 214 sites from 1978 sequences. A maximum likelihood tree was constructed with RAxML-HPC2 (27) on the CIPRES portal (28) using the WAG+PROTCAT model and 100 bootstrap replicates. A cut down phylogeny of the Cfr+RlmN family was constructed with RaxML using the same parameters. This dataset was based on the sequences selected for consensus sequences analysis plus additional Cfr-like sequences identified in the NCBI nr database by BlastP searching with *S. sciuri* Cfr and an e value cut-off of E-100 (data set dimensions: 75 taxa, 319 positions, after SeqFIRE streamlining).

Strains used for transformation, expression, MIC analysis and methylation analysis
The *Escherichia. coli* TOP10 strain (Invitrogen) was used for transformation of ligated plasmids. The hyperpermeable *E. coli* AS19 strain (29) was used for MIC analysis as it is much more sensitive to antibiotics relative to other *E. coli* strains. The RlmN minus stain *E. coli* JW2501-1 (30) was used for methylation analysis to facilitate identification of Cfr methylation by avoiding interference from the RlmN methylation at the same position.
Construction of plasmids encoding cfr- and rlmN-like genes

Plasmids were constructed similarly to the pCfrhis plasmid (9, 22), except that no histidine tag was added. The plasmids encoding inducible Cfr-likes or RlmN-likes were constructed by PCR amplification of the genes from genomic DNA or synthetic genes followed by cloning into plasmid pLJ102 (31) for expression of the proteins. The *Brevibacillus brevis* gene GI 226313314 (*nlbb*) was amplified for cloning by a two step PCR amplification. First PCR on genomic DNA using the primers 5' CCACCCATACCACATCTGCTAC 3' and 5' AAATGCCACTCCTTTGCC 3' and then a second PCR adding *Nde*I or *Hind*III sites for cloning with the primers 5' GGATGTGGAGATCATATGCCGTTAACGACATTTAC 3' and 5' CGATTTCCAAAGCTTCCACCCACGGTTTC 3'. The PCR fragment was then cloned as previously described (7) to construct pBbRlmN (Table 2). The construction of pRlmN with RlmN from *E. coli* was done similarly. The synthetic genes were similarly cloned directly from the plasmid provided by Genescript (Piscataway, New Jersey, USA).

pCICs contains a coding sequence for CICs identical to the coding sequence from *Clostridium sporogenes* (GI 187776707) and pCIPa the coding sequence for CIPa from *Paenibacillus sp. Y412MC10* (GI 261407206), but both with another coding usage (listed in supplementary information). Plasmids were retransformed into *E. coli* strains AS19 and JW2501-1. All three plasmid constructs were sequenced at the inserted gene to verify the identity of the cloned genes.

Verification of expression of plasmid coded genes by SDS gel analysis

*E. coli* AS19 cells harboring the plasmids with the cfr-like genes were grown at 37°C to an OD$_{650}$ of 0.2-0.3, followed by addition of IPTG (to 1mM) for induction of the genes. Cells were harvested after 3-3.5 hrs of growth and stored at -80°C. Samples were dissolved in 1 x
SDS/DTT-loading buffer, boiled 5 min. and loaded on standard SDS gels along with standard markers. Gels were run at 180V and then stained with Brilliant Blue G.

**Antibiotic susceptibility testing of strains**

Drug susceptibility testing was done essential as described previously (7) using a microtiter plate format and measuring optical density values at 450 nm with a Victor 3 spectrophotometer (Perkin Elmer). Overnight cultures in LB were diluted to an OD<sub>450</sub> value of 0.01, followed by mixing of 100 µL diluted culture with 100 µL of antibiotic solution in a series with two-fold concentration steps. Expression of the cfr and cfr-like genes was induced at the dilution step by adding 1 mM IPTG to all samples. The plasmids encode a gene for a repressor (LacIq) of the IPTG inducible promoter, which keeps the gene silent without IPTG addition. The tested concentration ranges were: florfenicol, 0.5-32 µg/ml; clindamycin and linezolid, 2-128 µg/ml; tiamulin, 0.25-128 µg/ml; and Synercid, 1-64 µg/ml. The MIC was defined as the drug concentration, at which the growth of the cultures was absent after 24 hrs incubation at 37°C.

**Primer extension analysis to verify modification at A2503 23S RNA**

After induction of Cfr and the Cfr-like proteins in *E. coli* JW2501-1 strains, the bacteria were grown for about three hrs. to allow new rRNA to be transcribed, modified and incorporated into ribosomes. Then RNA was extracted with GeneJET RNA Purification Kit (Fermentas). Methylation at A2503 was monitored by primer extension analysis with AMV reverse transcriptase (Finnzymes) and a Cy5-labeled deoxyoligonucleotide primer (5’-GAACAGCCATACCTTG-3’), complementary to nucleotides 2540-2556 of *E. coli* 23S rRNA. The cDNA extension products were separated on 6% polyacrylamide sequencing
gels. The positions of the stops were visualized by fluorescence scan and identified by referencing to dideoxynucleotide sequencing reactions on 23S rRNA that were electrophoresed in parallel.

Mass spectrometric analysis of RNA

23S rRNA sub-fragments of around 50 nucleotides were isolated by hybridizing the rRNA with an oligodeoxynucleotide complementary to the region around A2503, followed by digestion with mung bean nuclease as described in (32). The oligodeoxynucleotides used had the sequence GCC CCA GGA TGC GAC GAG CCG ACA TCG AGG TGC CAA ACC TCC CCG CC for *Thermus thermophilus*, GCC CCA GGA TGC GAT GAG CCG ACA TCG AGG TGC CAA ACC TCC CCG TCG for *Bacillus subtilis*, and GCC CCA GGA TGT GAT GAG CCG ACA TCG AGG TGC CAA ACA CCG CCG TCG for *E. coli*.

After sub-fragment purification, the RNA was digested with RNase T1 for MALDI Time-of-Flight mass spectrometric analysis as previously reported (33). Briefly, 1–2 pmol rRNA sub-fragment were RNase T1 digested to completion and analyzed directly using 3-hydroxypicolinic acid as matrix. Mass spectra were recorded in positive ion mode with a reflectron Time-of-Flight mass analyzer on a PerSeptive Voyager-DE STR instrument (Applied Biosystems).

MALDI tandem mass spectrometry of the *E. coli/pBbRlmN A*₂₅₀₃*UG*₂₅₀₅ 23S rRNA RNase T1 digestion product was performed on a MicroMass MALDI Q-TOF Ultima instrument (Waters, Manchester, UK) in positive ion mode as previously described (34).

The *E. coli/pClPa*, *T. thermophilus* and the *B. subtilis* rRNA sub-fragments were reduced to nucleosides by enzymatic digestion, and separated by liquid chromatography on a graphitized carbon column using an Agilent LC/Chip Cube system. Detection was done on-
line with an Agilent XCT Ultra 6340 ion trap mass spectrometer, where tandem MS analysis was done to the MS$^4$ level. Experimental details were identical to the ones previously described (9).

Results and discussion

A natural chromosomal host for Cfr has not yet been identified, as the cfr gene is hitherto found on plasmids together with transferable sequences or in a few cases on chromosomes, but flanked by transferable sequences. Until recently, all cfr genes found were very similar with only one or two amino acid changes. It is now clear that more divergent cfr-like genes are present on the chromosome of some bacteria in the Bacillales order (7). This study also pointed to the difficulty in predicting without any experimental evidence whether a gene codes for the worrisome Cfr methyltransferase conferring antibiotic resistance or the harmless RlmN methyltransferase.

Alignment and phylogenetic analysis of Cfr and RlmN like sequences

Sequences similar to Cfr were retrieved from the NCBI RefSeq protein database using PSI-BLAST with *Staphylococcus sciuri* Cfr as the query. The resulting dataset contained Cfr, the close relative RlmN and more distant relatives such as the pyruvate formate lyase activating enzyme, nitrogenase iron-molybdenum cofactor biosynthesis protein, molybdenum cofactor biosynthesis protein A, and coenzyme PQQ synthesis protein. The dataset was reduced to a family of Cfr-like and RlmN-like proteins, which were aligned and subjected to maximum likelihood phylogenetic analysis. A phylogenetic tree of 1978 sequences is presented in Supplementary figure S1. This tree places the Cfrs from clinical and veterinary samples in a strongly supported (85% Maximum likelihood bootstrap
support) group within the RlmN family, together with 22 sequences from Bacillales and Clostridia and a single Enterococcus. There is, however, no statistical support for the Cfr-like group’s specific placement within the RlmN family. Therefore, it is not possible to infer the source of Cfr, beyond the fact that it evolved from an RlmN, or an RlmN-like protein.

To investigate the differences between Cfr and RlmN on the amino acid level, we selected two sets of bacterial genes from the alignment for deducing consensus sequences. Criteria for selection (see Materials and Methods) were employed to maximize the likelihood of only sampling “true” Cfr and RlmN sequences and to sample broadly across the RlmN family phylogeny. Alignments showing Cfr-like and RlmN-like sequences, and their consensus sequences are presented in Supplementary figure S2 (the whole reduced dataset) and Figure 1 (a further reduced alignment of representative sequences). The alignment clearly shows the proteins are alignable across their full lengths, with some specific differences. 57 amino acids are strongly conserved for both classes (indicated by grey shading in Figure 1). 13 amino acids are specific and conserved for each class and these are boxed in Figure 1. These differences are likely to be associated with important functional differences between the Cfr and RlmN families. The reduced alignment of the RlmN family was used to create a cut-down version of the RlmN plus Cfr family tree, presented in Figure 2, that also include newly added Cfr-like sequences from the NCBI nr database. The tree shows a clear distinction between Cfr-likes and RlmN-likes and an early divergence of eukaryotic Cfr-like sequences. Also to be noted is the divergence of clostridial Cfr-like proteins that will be discussed below.
We also looked at eukaryotic Cfr-like and RlmN-like sequences from the 1978 sequence alignment to determine their relationship to Cfr and RlmN. A sister group to the Cfr-like group in the phylogenetic tree contains a small group of genes from alveolate eukaryotes including Plasmodium parasites. This relationship of Alveolate Cfr-like and Cfr clades has full support in the phylogenies (100% bootstrap percentage, Supplementary figure S1 and Figure 2). Alignment of these eukaryotic sequences with the Cfr and RlmN bacterial consensus sequences (Supplementary figure S2) shows that four of the 13 strongly differentially conserved sites are identical to the Cfr consensus (position 45, 154, 216 and 282 in S2) and three (position 223, 320 and 327 in S2) are identical to RlmN consensus while the remaining six (position 53, 99, 158, 244, 310 and 324 in S2) are different from both Cfr and RlmN or not well conserved. In general, there is a good correlation between conservation in Cfr and RlmN and the eukaryotic look-alikes, especially around the C-terminal G.DIdAACGQL sequence. Although the sequence conservation suggests that these proteins are RNA methyltransferases with a similar function to Cfr and RlmN, their specific function and molecular target are not predictable from this analysis.

The X-ray model structure of RlmN with a Fe-S cluster and a SAM ligand (21) presented in Figure 3 makes it possible to relate the site-specific differential conservation of Cfr/RlmN to structural features. It shows the position of the 13 Cfr/RlmN specific amino acids from Figure 1 in the RlmN structure model from Boal et al. (21). The active center, if defined as the Fe-S/SAM site, is positioned at the end of a funnel-like structure and eight of the Cfr/RlmN specific amino acids sites (G91, V119, V123, V174, P181, R274, E278, and M281) are located at or very close to this binding pocket (Figure 3). Three of these (V174, R274, and E278) are particularly note-worthy, as they line the funnel with side chains...
oriented to interact with incoming molecules, and may contribute to m$^2$A/m$^8$A specificity.

Three sites (F40, W48, and G200) are positioned far away from - but on the same site as -
the “entrance” to the active site and on the outside of the enzymes. These sites might thus
be implicated in target binding/target positioning, although further structural and mutational
analyses is required to confirm this. There are three well conserved sequence insertions in
Cfr relative to RlmN (dotted line boxes in Figure 1) and the structural location of these
insertions is shown in blue in Figure 3. Two of these insertions (alignment positions 18 and
313-317, Figure 1) are at the rim of the entrance to the binding pocket and may allow an
alternative positioning of the entering target adenine approving for methylation at the C-8
position. Another potentially important position is the differentially conserved V123 in
RlmN (alignment position 127), which is oriented away from the binding pocket, and is
replaced by cysteine in Cfr (Figure 1). The residue shown in green in the structure is
specifically changed in Clostridia at an otherwise conserved position, and will be discussed
below.

Verification of the dispersal of Cfr methyltransferase genes amongst bacteria

A reliable classification of the Cfr and RlmN classes of enzymes is desirable to identify cfr
resistance genes in various organisms and to get a better understanding of the Cfr and
RlmN functions and evolution. The predicted Cfr sequences are presented in the tree in
Figure 2. The clades with Cfr and Cfr-likes with evidence as resistance determinants are
marked “True” Cfr, and those with direct investigation of the gene function are marked
with check marks. The Cfr-likes from Bacillus amyloliquefaciens, Bacillus clausii and
Brevibacillus brevis provide PhLOPSa resistance and are thus true Cfr-likes (7). The rest of
the Cfr-like group comprises genes from Enterococcus, Paenibacillus and Clostridium that
diverge before these previously experimentally confirmed Cfr (Figure 3). We hypothesized that the well supported (97% bootstrap support) group of Cfr-like enzymes are all true Cfrs.

To functionally test our hypothesis based on phylogenetic classification of Cfr sequences, we cloned genes from Paenibacillus and Clostridium (see details in Table 1), expressed them in *E. coli* and investigated resistance. The first experiments to isolate the genes starting from genomic DNA from the hosts failed, and therefore synthetic genes coding for the same proteins were used and cloned in a plasmid behind an inducible promoter as previously done with the *S. sciuri cfr* gene (9, 22). The plasmids named pClPa (with *cfr*-like Paenibacillus gene) and pClCs (with *cfr*-like Clostridium gene) (Table 1) were transformed into *E. coli* AS19 (29) that shows an increased sensitivity to most antibiotics.

In the expression system used protein production is strongly dependent on addition of IPTG as previously shown by primer extension analysis for Cfr (22). Also, there is no significant growth effect from IPTG induction as well as only a very minor effect of the presence of the plasmid (data not shown). Expression was investigated by SDS gel analysis and as shown in Figure 4A, strong protein bands appear with expression of ClPa and ClCs – the Cfr-likes from Paenibacillus and Clostridium (Table 1). The ClPa protein appears at a lower mass than expected, but its identity was verified by peptide mass fingerprinting (data not shown). To establish if the Cfr-like proteins ClPa and ClCs confer a resistance pattern similar to the Cfr methyltransferase, minimal inhibitory concentrations (MICs) were determined with the five antibiotics florfenicol, clindamycin, linezolid, tiamulin and Synercid as done previously (7). These represent the five antibiotic classes in the PhLOPSa phenotype conferred by the Cfr methyltransferase (3). The MICs are summarized in Table 2 together with controls of strains without plasmid, with the parent pLJ102 plasmid or with
pCfrhis expressing his-tagged Cfr from Staphylococcus (9). ClPa confers some resistance to the PhLOPSa antibiotics, verifying that it has a Cfr-type function, although it apparently is less effective than the Staphylococcus plasmid-coded Cfr (Table 2). The Clostridia ClCs does not mediate MIC changes. The RNA methylation at A2503 in 23S rRNA was also checked by primer extension as in our previous study (7) to prove the relationship between modification and phenotype. This was done by transforming the plasmids into JW2501-1, an E. coli RlmN minus strain (30) (as the inherent m^2A methylation mediated by RlmN causes a minor primer extension stop at A2503 that interferes with detection of the m^8A methylation from Cfr-like enzymes). The analysis is presented in Figure 4B and shows a clear stop at A2503 of 23S RNA from ClPa containing E. coli JW2501-1 in line with the resistance observed in the MIC experiment. The stronger stop in the Cfr-encoding strains is consistent with the higher resistance observed in these strains (Table 2). Finally, the exact identity of the methylation by ClPa was checked by mass spectrometric analysis. The chromatographic retention time as well as the MS^4 behavior of the methylated A-nucleoside were identical to an 8-methyladenosine standard (data not shown), in accordance with our previous report on the identification of m^8A2503 (9); thus confirming that the clpa gene from Paenibacillus encodes a m^8A2503 methyltransferases. An independent MALDI Time-of-Flight mass spectrometric analysis revealed a sub-stoichiometric methylation of A2503 – less than 50% judged by MS signal intensities (data not shown) - in accordance with the low antibiotic resistance observed (Table 2) and the relative weak extension stop (Figure 4B). Hence, the Paenibacillus clpa gene encodes an m^8A2503 methyltransferases, but the enzyme does not exhibit high efficiency when functioning in the ectopic E. coli system.
There is no stop in the ClCs-containing strain, consistent with no observed MIC changes. Thus, the Clostridia ClCs protein does not methylate *E. coli* 23S RNA. Indeed, the tree in Figure 2 shows that the Clostridial sequences branch off early in the Cfr-like clade, suggesting that the Cfr-type methylation function may have evolved after the divergence of the clostridial sequences. However, we cannot conclude that ClCs does not have a Cfr-like function altogether, as this protein may be able to methylate *Clostridium* 23S RNA, but not *E. coli* 23S RNA due to sequence differences in RNA or r-proteins or in the ribosome assembly process. All the Clostridia Cfr-like sequences in Supplementary figure S2 show a sequence change from N to D at position 289/307 (Cfr/RlmN numbering) of an otherwise conserved amino acid (alignment position 322) and is mapped at the entrance to the active site as shown in green in Figure 3. A future study will be directed to investigate the significance of this particular change for methylation of A2503 in 23S RNA in *E. coli*.

**Verification of the dispersal of RlmN methyltransferase genes amongst bacteria**

RlmN was identified as the enzyme responsible for the m$^2$A2503 23S RNA modification in *E. coli* in 2008 (2). A similar protein from the clinical *S. aureus* strain Newman, exhibits 36% sequence identity to *E. coli* protein RlmN and a primer extension stop at A2503 of 23S rRNA from this strain (*E. coli* numbering used for all organisms) is absent when the corresponding gene is inactivated (2), suggesting it is also an RlmN. More recently, *Deinococcus radiodurans* has also been shown to have the m$^2$A2503 modification (35). *T. thermophilus* was reported to have a methylation at A2503 in 23S RNA (36), and knockout of the rlmN-like gene abolishes A2503 methylation (A. Rasmussen, H. Park and F. Kirpekar, unpublished), and *Bacillus* species are also methylated at A2503. (B. T. Porse and F. Kirpekar, unpublished observation). To verify that these methylations are in fact
m\(^2\)A, we purified approx. 50 nucleotide long 23S rRNA sub-fragments around position A2503 from *T. thermophilus* and *B. subtilis*. These sub-fragments were digested to nucleosides and analyzed by liquid chromatography/ion trap mass spectrometry as previously described (9). The chromatographic retention time as well as the MS\(^5\) behavior of the methylated A-nucleosides were identical to m\(^2\)A2503 obtained from *E. coli*, and distinctly different from an 8-methyladenosine standard (data not shown), which strongly suggests that the RlmN-like genes in *T. thermophilus* and *B. subtilis* encode m\(^2\)A2503 methyltransferases.

Our phylogenetic analysis shows that most bacteria have an *rlmN*-like gene, and all those that have a *cfr*-like gene also have an *rlmN*-like gene. To verify the diverse function we investigated RlmN function in Cfr-containing *Brevibacillus brevis*. We recently investigated the *B. brevis cfr*-like gene *cIBb* (GI: 226314089) and observed antibiotic resistance as well as stop at position 2503 with primer extension on 23S RNA indicating methylation (7). Here, we cloned the *Brevibacillus brevis* rlmN-like gene *nlbb* (Figure 1, Table 1) in a similar way and investigated its function. As a control we also cloned the *E. coli* rlmN gene in the same way. NlBb coded by *nlbb* was investigated as described above for ClPa and ClCs. As expected the protein NlBb was expressed (Figure 4A), had no effect on PhLOPSa MICs (Table 2), but provided a primer extension stop at A2503 23S RNA similar to the one provided by *E. coli* RlmN expressed the same way (Figure 4B). Finally, it was verified by MALDI tandem mass spectrometry that the A\(\text{2503UG2505}\) fragment of 23S RNA from the RlmN minus *E. coli* JW2501-1 expressing BbRlmN contains an additional methyl group on A2503 (data not shown).
The origin and the dissemination of Cfr- and RlmN-likes

Even though the plasmid-borne Cfr has been found worldwide and in different bacteria, we have yet no indications of the direct origin of the Cfr that has been found in clinical and veterinary samples. According to our previous study (7), and the present phylogenetic analysis and the CIPa data, there are natural Cfr-likes in Bacilliates but apparently not widespread beyond that. The function of the clostridial Cfr-likes remains an unresolved question. It is worth taking into account that the databases only represent a tiny part of the diversity of bacterial life, and this information is biased from the interest of academia and industry. It is likely that more Cfr and Cfr-like sequences will be revealed with future genome sequencing.

The rlmN-like genes seem to be widespread in bacteria, with additional divergent paralogous groups of unknown function. At present, Cfr is the only paralogous subgroup within the RlmN-like family tree with known function. The phylogenetic analysis also revealed two particularly taxonomically limited groups (highlighted by yellow boxes in Supplementary figure S1) that are strongly supported. One group comprises Aquificiae, one Deferrribacterium and one delta-proteobacterium, while the other contains members of beta- and gamma-proteobacteria. The long branches separating these groups from the main RlmN backbone are suggestive of functional divergence, similarly to what has occurred in the evolution of Cfr. In addition to the Cfr-like eukaryotic clade, there are other eukaryotic subgroups throughout the RlmN-like tree, most widespread in plants and algae. Genes for RlmN-like proteins are also found in two species of archaea: Nitrosopumilus maritimus and Candidatus Nitrosoarchaeum limnia. The multiple paralogous groups of RlmN-like microbial sequences suggest horizontal gene transfer has played a significant role in the
distribution of these proteins. Gene duplication and differential lineage sorting in the RlmN family tree is also likely to have contributed. With the new knowledge about the dual specificity of RlmN mediating m^2A2503 on 23S rRNA and m^2A37 on tRNA (15), one might speculate that the main substrate is tRNA for some RlmN-like enzymes. It is also possibly that Cfr has one or more additional molecular targets. Another interesting question is which part of the enzymes accounts for the specificity. We need more studies to obtain knowledge about the specificity of the various enzymes to clarify their exact function, target and dissemination.

Concluding remarks

It is not obvious where the cfr gene circulation on plasmids and transposons came from or how it evolved. Our analyses suggest that cfr-like genes are limited to a small subset of bacterial species. However, their presence in known pathogens, their mobility and their action against multiple antibiotics even in heterologous systems makes them a matter of concern for antibiotic resistance. The similar rlmN genes, that do not confer significant resistance, are abundant but it remains to be established if they all have the same targets or whether there is diversity towards 23S rRNA, tRNA and even other RNAs. The Cfr- and RlmN-specific conserved sites provide a very good indication of whether a gene is Cfr-like or RlmN-like. The classification can then be verified by phylogenetic analysis, as we have carried out in this study.

Supplementary material

Figure S1: Large Phylogenetic tree

Figure S2: Consensus from selected Cfr-like and RlmN-like protein sequences
Acknowledgements

The Danish Council for Independent Research - Natural Sciences and the Lundbeck Foundation is gratefully acknowledged for financial support. G.C.A. is supported by the European Social Fund through Estonian Science Foundation programs Mobilitas (MJD99) and ETF (I9020). T.T. is supported by the European Regional Development Fund through the Center of Excellence in Chemical Biology. We also thank Prof. Peter Højrup for mass spectrometry analysis of ClCs, Wanida Skibsted for construction of pRlmN and Jesper Foged Havelund for help with the mass spectrometry analysis of methylation by CIPa.

References


15. Benitez-Paez A, Villarroya M, Armengod ME. 2012. The *Escherichia coli* RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy. RNA. 18:1783-1795.


methyl transfer by SN2 displacement and radical generation. J Am Chem Soc. 503
133:19586-19589.

20. **Yan F, Fujimori DG.** 2011. RNA methylation by radical SAM enzymes RlmN and
Cfr proceeds via methylene transfer and hydride shift. Proc Natl Acad Sci U S A. 507
108:3930-3934.

21. **Boal AK, Grove TL, McLaughlin MI, Yennawar NH, Booker SJ, Rosenzweig AC.**
2011. Structural basis for methyl transfer by a radical SAM enzyme. Science. 510
332:1089-1092.

Insights into the structure, function and evolution of the radical-SAM 23S rRNA
methyltransferase Cfr that confers antibiotic resistance in bacteria. Nucleic Acids Res. 514
38:1652-1663.

23. **Katoh K, Toh H.** 2010. Parallelization of the MAFFT multiple sequence alignment

24. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2--approximately maximum-
likelihood trees for large alignments. PloS one. 5:e9490.

25. **Atkinson GC, Baldauf SL.** 2011. Evolution of elongation factor G and the origins of

2012. SeqFIRE: a web application for automated extraction of indel regions and
conserved blocks from protein multiple sequence alignments. Nucleic Acids Res. 523

27. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic


centered in regions of RNA-RNA contact. The Journal of biological chemistry. 281:22108-22117.


Figure legends

Figure 1: **Cfr and RlmN consensus sequences**

Cfr and RlmN consensus sequences together with examples of sequences used for determining the consensus - the full list is presented in Supplementary figure S2. The black line open boxes show the 13 positions that are selectively conserved in the designated Cfr and RlmN alignment, meaning that >70% has a specific amino acid in all Cfrs and another specific amino acid in RlmNs. Grey shadings mark the 57 positions that are totally conserved according to the same criteria. The dashed boxes show the sites of insertions that might play a role in Cfrs/RlmNs distinction. The black lined triangle marks the position of an N to D change in Clostridia at an otherwise conserved position. Dots above the sequence indicate the CxxxCxxC motif cysteines involved in binding of the Fe-S cluster.

Figure 2: **Maximum likelihood phylogenetic tree of a subset of Cfr- and RlmN-like sequences.**

Maximum likelihood bootstrap percentage (MLBP) support is indicated on branches. Only branches with >50% MLBP support are labeled. The scale bar below the tree shows the evolutionary distance expressed as substitutions per site. Numbers in taxon names are
NCBI GI numbers. Underlined genes are investigated in this study. ✓ indicates function/modification as predicted and X indicates function that was tested and not verified.

Figure 3: **Model structure of RlmN with marking of Cfr/RlmN specific amino acids**

X-ray model structure of RlmN (21) with a Fe-S cluster in purple, and a SAM ligand in red.

The 13 Cfr/RlmN specific amino acids from Figure 1 are marked yellow and numbered according to the RlmN sequence with alignment numbers according to figure 1 in parentheses. The blue markings are the positions of three insertions in Cfrs relative to RlmNs (marked with dashed boxes in figure 1). The green residue is the site of an N to D substitution in *Clostridia* at an otherwise conserved position.

Figure 4: **Gels with protein expression profiles and RNA primer extension stops**

A. Analysis of whole cell extracts by SDS-PAGE showing expression of Cfr, ClCs, ClPa, RlmN and BbRlmN. The extracts are from *E. coli* AS19 harboring the plasmids listed on top of the gel. Lanes marked with + induc. contain samples with IPTG induced expression.

M is size markers: 100, 70, 55, and 35 kDa. from top. B. Primer extension analysis of reverse transcriptase stops on 23S rRNA from *E. coli* JW2501-1 strains harboring the plasmids expressing Cfr, ClCs, ClPa, RlmN and BbRlmN. The region shown is limited to the nucleotides flanking A2503 that is methylated by Cfr and RlmN. Lanes 1-8 show primer extension reactions on total RNA from cells harboring the indicated plasmids and induced by IPTG. Lanes 9-12 marked C, U, A, and G refer to dideoxynucleotide sequencing reactions. The arrow points to the A2503 stop (one position below the sequencing position as the extension stops before the modified nucleotide) mediated by methylation from Cfr, ClPa, RlmN and BrRlmN.
Table 1. Origin of cfr-like and rlmN-like genes and modification activities at A2503 in 23S RNA (E. coli numbering).

<table>
<thead>
<tr>
<th>Gene label</th>
<th>Protein name</th>
<th>Host organism - (Ec) if modification investigated in E. coli</th>
<th>Modification</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfr</td>
<td>Cfr</td>
<td>Staphylococcus (Ec)</td>
<td>m^8A2503 and m^2A2503</td>
<td>(1, 9)</td>
</tr>
<tr>
<td>clba</td>
<td>ClBa</td>
<td>Bacillus amyloliqufaciens (Ec)</td>
<td>m^8A2503*</td>
<td>(7)</td>
</tr>
<tr>
<td>clbc</td>
<td>ClBc</td>
<td>Bacillus clausii (Ec)</td>
<td>m^8A2503*</td>
<td>(7)</td>
</tr>
<tr>
<td>clbb</td>
<td>ClBb</td>
<td>Brevibacillus brevis</td>
<td>m^8A2503*</td>
<td>(7)</td>
</tr>
<tr>
<td>clcs</td>
<td>ClCs</td>
<td>Clostridium sporogenes</td>
<td>none</td>
<td>this study</td>
</tr>
<tr>
<td>clpa</td>
<td>ClPa</td>
<td>Paenibacillus sp.</td>
<td>m^8A2503</td>
<td>this study</td>
</tr>
<tr>
<td>rlmN/yfgB</td>
<td>RlmN</td>
<td>E. coli</td>
<td>m^8A2503</td>
<td>(2, 11, 37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mA2503***</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m^2A2503</td>
<td>(36), this study</td>
</tr>
<tr>
<td></td>
<td>Bacillus subtilis subtilis</td>
<td>( m^2A_{2503} )</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>* based on phenotype and primer extension, no information on additional ( m^2A_{2503} ) modification.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deinococcus radiodurans</td>
<td>( m^2A_{2503} )</td>
<td>(35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( nabh )</td>
<td>NlBb</td>
<td>Brevibacillus brevis</td>
<td>( m^2A_{2503}**** )</td>
<td>this study</td>
</tr>
<tr>
<td>( NBRC 100599 (Ec) )</td>
<td>**** C-2 methylation is assumed as it is an adenine methylation but not ( m^8A ).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** the gene sequences have been modified to suit codon usage in *E. coli* for optimal expression.

*** Methylation assignment based on primer extension stop.
Table 2: MICs of *E. coli* AS19 strains in the presence or absence of plasmids expressing *cfr*, *rlmN* or -like genes.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Gene</th>
<th>Florfenicol</th>
<th>Clindamycin</th>
<th>Linezolid</th>
<th>Tiamulin</th>
<th>Synercid</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
<td>1</td>
<td>16-32</td>
<td>8</td>
<td>0.5</td>
<td>2-4</td>
</tr>
<tr>
<td>pLJ102</td>
<td></td>
<td>1</td>
<td>16-32</td>
<td>8</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>pCfrhis</td>
<td><em>cfr</em></td>
<td>16-32</td>
<td>&gt;128</td>
<td>128</td>
<td>32-64</td>
<td>64</td>
</tr>
<tr>
<td>pClPa</td>
<td><em>clpa</em></td>
<td>4</td>
<td>128</td>
<td>16-32</td>
<td>2</td>
<td>4-8</td>
</tr>
<tr>
<td>pClCs</td>
<td><em>clcs</em></td>
<td>1</td>
<td>32</td>
<td>4-8</td>
<td>0.25-0.5</td>
<td>2-4</td>
</tr>
<tr>
<td>pBbRlmN</td>
<td><em>nlbb</em></td>
<td>1</td>
<td>16-32</td>
<td>8</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>pRlmN</td>
<td><em>rlmN</em></td>
<td>1</td>
<td>16-32</td>
<td>8</td>
<td>0.5</td>
<td>2</td>
</tr>
</tbody>
</table>

The tabulated MIC values are given in units of micrograms/mL and are the average of at least three independent experiments. An interval is given when no clear distinction between the values was obtained. Only greater than two-fold differences are considered significant.
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Organism</th>
<th>Protein Type</th>
<th>Eukaryotic Cfr-like</th>
<th>RlmN-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>26554015_Mycoplasma_penetrans_HF2__hypothetical_protein_MYPE5630</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>329930082_Paenibacillus_sp._HGF5__23S_rRNA_m2A2503_methyltransferase</td>
<td>Paenibacillus sp.</td>
<td>23S rRNA methyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>226950274_Clostridium_botulinum_A2_str._Kyoto__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>168184510_Clostridium_botulinum_Bf__florfenicol/chloramphenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Florfenicol/chloramphenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>153939955_Clostridium_botulinum_F_str._Langeland__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>225314089_Brevibacillus_brevis_NBRC_100599__chloramphenicol/florfenicol_resistance_protein</td>
<td>Brevibacillus brevis</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>409735992_Clostridium_botulinum_CFSAN001628__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170758262_Clostridium_botulinum_A3_str._Loch_Maree__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170754770_Clostridium_botulinum_B1_str._Okra__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148380815_Clostridium_A_str._ATCC_3502__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium acidurici</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>168179276_Clostridium_A2_str._Lyric__chloramphenicol/florfenicol_resistance_protein</td>
<td>Clostridium botulinum</td>
<td>Chloramphenicol/florfenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15805960_Deinococcus_radiodurans_R1__radical_SAM_protein_putative</td>
<td>Deinococcus radiodurans</td>
<td>Radical SAM protein putative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10856377_Trehonobacter_urbicola_ATCC_34045__radical_SAM_protein</td>
<td>Trehonobacter urbicola</td>
<td>Radical SAM protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22561972_Brachyspira_hyodysenteriae_WA1__putative_FeScluster_redox_enzyme</td>
<td>Brachyspira hyodysenteriae</td>
<td>FeScluster redox enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21542321_Mycobacterium_tuberculosis_TG2__hypothetical_protein_MtbT9_18783</td>
<td>Mycobacterium tuberculosis</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>285114380_Streptomyces_avermectins_MT2__hypothetical_protein</td>
<td>Streptomyces avermectins</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>215610039_Streptomyces_viridochromogenes_MT2__hypothetical_protein</td>
<td>Streptomyces viridochromogenes</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15597036_Pseudomonas_aeruginosa_PA01__ribosomal_LSU_methyltransferase_N</td>
<td>Pseudomonas aeruginosa</td>
<td>Ribosomal LSU methyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25070091_Pseudomonas_erythropolis_RS21__radical_SAM_protein</td>
<td>Pseudomonas erythropolis</td>
<td>Radical SAM protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10856377_Trehonobacter_urbicola_ATCC_34045__radical_SAM_protein</td>
<td>Trehonobacter urbicola</td>
<td>Radical SAM protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>342608315_Paenibacillus_beijerinckii_C454__florfenicol/chloramphenicol_resistance_protein</td>
<td>Paenibacillus beijerinckii</td>
<td>Florfenicol/chloramphenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>308130334_Streptococcus_subsp._thermophilus_Tf2__hypothetical_protein</td>
<td>Streptococcus subsp._thermophilus</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312973242_Escherichia_coli_182770__UPF0063_protein_yfgB</td>
<td>Escherichia coli</td>
<td>UPF0063 protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19059606_Deinococcus_radiodurans_R1__ribosomal_LSU_large_subunit_methyltransferase_N</td>
<td>Deinococcus radiodurans</td>
<td>Ribosomal LSU large subunit methyltransferase N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>154539058_Pseudomonas_erythropolis_RS21__radical_SAM_protein</td>
<td>Pseudomonas erythropolis</td>
<td>Radical SAM protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>240490081_Bacillus_thermohaliotrophicum_TM4__hypothetical_protein_TM4_18643</td>
<td>Bacillus thermohaliotrophicum</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>342608315_Paenibacillus_beijerinckii_C454__florfenicol/chloramphenicol_resistance_protein</td>
<td>Paenibacillus beijerinckii</td>
<td>Florfenicol/chloramphenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19059606_Deinococcus_radiodurans_R1__ribosomal_LSU_large_subunit_methyltransferase_N</td>
<td>Deinococcus radiodurans</td>
<td>Ribosomal LSU large subunit methyltransferase N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>154539058_Pseudomonas_erythropolis_RS21__radical_SAM_protein</td>
<td>Pseudomonas erythropolis</td>
<td>Radical SAM protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>240490081_Bacillus_thermohaliotrophicum_TM4__hypothetical_protein_TM4_18643</td>
<td>Bacillus thermohaliotrophicum</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>342608315_Paenibacillus_beijerinckii_C454__florfenicol/chloramphenicol_resistance_protein</td>
<td>Paenibacillus beijerinckii</td>
<td>Florfenicol/chloramphenicol resistance protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312903419_Enterococcus_faecalis_TX0635__radical_SAM_enzyme__Cfr_family</td>
<td>Enterococcus faecalis</td>
<td>Radical SAM enzyme</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.4 substitutions