# Linkage, Mobility, and Selfishness in the MazF Family of Bacterial Toxins: A Snapshot of Bacterial Evolution

Nikita Chopra[1], Saumitra[1,3], Abhinandan Pathak[1], Rakesh Bhatnagar[2,*], and Sonika Bhatnagar[1,*]

[1]Computational and Structural Biology Laboratory, Division of Biotechnology, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

[2]Laboratory of Molecular Biology and Genetic Engineering, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

[3]Present address: Neuro-Electronic Laboratory, Center for Nanoscience and Engineering, Indian Institute of Science, Bangalore, Karnataka, India

*Corresponding author: E-mail: ecc999@gmail.com; rakeshbhatnagar@mail.jnu.ac.in.

**Accepted:** October 31, 2013

## Abstract

Prokaryotic MazF family toxins cooccur with cognate antitoxins having divergent DNA-binding folds and can be of chromosomal or plasmid origin. Sequence similarity search was carried out to identify the Toxin–Antitoxin (TA) operons of MazF family followed by sequence analysis and phylogenetic studies. The genomic DNA upstream of the TA operons was searched for the presence of regulatory motifs. The MazF family toxins showed a conserved hydrophobic pocket in a multibinding site and are present in pathogenic bacteria. The toxins of the MazF family are associated with four main types of cognate antitoxin partners and cluster as a subfamily on the branches of the phylogenetic tree. This indicates that transmission of the entire operon is the dominant mode of inheritance. The plasmid borne TA modules were interspersed between the chromosomal TA modules of the same subfamily, compatible with a frequent interchange of TA genes between the chromosome and the plasmid akin to that observed for antibiotic resistance gens. The split network of the MazF family toxins showed the AbrB-linked toxins as a hub of horizontal gene transfer. Distinct motifs are present in the upstream region of each subfamily. The presence of MazF family TA modules in pathogenic bacteria and identification of a conserved binding pocket are significant for the development of novel antibacterials to disrupt the TA interaction. However, the role of TAs in stress resistance needs to be established. Phylogenetic studies provide insight into the evolution of MazF family TAs and effect on the bacterial genome.

**Key words:** MazF, toxin–antitoxin systems, multibinding site, phylogenetic studies, linkage, mobility.

## Introduction

Toxin–Antitoxin (TA) systems are widely found in free-living prokaryotes and Archaea. Five distinct types of TA encoding loci are known. The TA systems are classified as Type I–V depending on the chemical nature and mode of action of their two components (Gerdes et al. 2005; Wang et al. 2012). The Type II TA systems consist of two genes encoded by an operonic structure, wherein the upstream gene codes for a labile antitoxin protein and the downstream gene codes for a more stable toxin protein. The Type II TA systems are found to be distributed in both plasmid and chromosomal genetic elements. The toxins of the plasmid-encoded Type II TA systems have been attributed to mediate post segregational killing consequent to plasmid loss, while the chromosome-encoded toxins mediate a wide variety of different cellular functions related to cell stasis, genetic stabilization, stress response, and quorum sensing. (Wang et al. 2012). The Type II proteic antitoxin usually serves two functions; firstly, it binds to and prevents the toxin from mediating its harmful effects. Secondly, the antitoxin alone, or in complex with the toxin, binds to its own operon to bring about transcriptional autoregulation. As the antitoxin protein is highly susceptible to proteolytic degradation by the action of Lon or Clp proteases, a continuous supply of the antitoxin protein in the cell is required to circumvent the lethal action of the toxin. Therefore, the TA systems are also termed as addiction modules. Exceptions to the generalized mode of action of TA systems described are evident in three-component systems (de la Hoz et al. 2000; Hallez et al. 2010). TA modules are important drug targets as it has been hypothesized that the drug-induced disruption of their interaction would set the toxin free to exert its toxic activity in the bacterial cell.

The known TA type II systems have been classified into 12 toxin and 20 antitoxin superfamilies based on sequence

homology and association with toxin/antitoxin-like sequences (Leplae et al. 2011). Toxins having diverse sequence, upstream antitoxin partners, and functions may belong to the same superfamily of toxins (Hayes and Van Melderen 2011). Additionally, they may occur as a TA complex with a number of different antitoxins having divergent DNA-binding folds (Leplae et al. 2011). As an example, the toxins of the HigB/RelE/ParE superfamily adopt a unique RelE-like fold and are associated with antitoxins having divergent folds. The RelE and ParE families of toxins have been unified into one superfamily having a common evolutionary origin despite their divergent activity/cellular targets as well as the lack of significant sequence similarity (Anantharaman and Aravind 2003). Similarly, the toxins of the MazF/PemK/CcdB superfamily share a ribonuclease SH3 fold and have several different types of cognate antitoxins. In previous studies, co-occurrence of PemK-type toxins with a number of different types of antitoxins has also been observed (Makarova et al. 2009).

In general, the antitoxin has a modular structure, wherein the N-terminal DNA-binding domain is responsible for autoregulation, while the C-terminal domain is responsible for neutralization of the toxin. Many of the known antitoxins have a helix–turn–helix (HTH) or a ribbon–helix–helix (RHH) DNA-binding motif at the N-terminus. Functional dimers of CcdA form a β-sheet or ribbon that inserts into the major groove of DNA (Madl et al. 2006), much like the mode of binding seen in the transcriptional repressors MetJ/Arc/CopG (Gomis-Ruth et al. 1998). The mode of binding of RelB to DNA is also expected to be similar to that of CopG (Li et al. 2008). It has been proposed that the RelE2 corepressor of *Streptococcus pneumoniae* serves to bridge the antitoxin dimers bound at two distinct sites of the operator region, leading to greater avidity of binding of the TA complex when compared with the RelB2 alone (Chan et al. 2013). HipB and MsqA antitoxins bind to DNA via an HTH motif, containing a "recognition helix" that penetrates into the major groove of DNA and makes base-specific interactions, whereas additional backbone contacts stabilize the complex (Schumacher et al. 2009; Brown et al. 2011). Though MsqA is a dimer, each of the MsqA recognition helices individually binds to one palindromic half-site of its promoter. Further, DNA recognition by MsqA can be attributed completely to specific residues of the recognition helix, which mediate a direct readout of the promoter DNA sequence (Brown et al. 2011). HipB and MsqA bear significant sequence and structural similarity to the 434 and 434 cro repressors, thus establishing them as members of the Xre-HTH family of transcriptional regulators (Schumacher et al. 2009). Structural studies have shown that the organization of the C-terminal helices of RHH motif is identical to that of the classical HTH domain (Gomis-Ruth et al. 1998). A CopG-like transcription factor from the streptococcal plasmid pMV158 shared structural similarity with both HTH- and RHH-type DNA-binding proteins (Acebo et al. 1998). Mutagenesis studies have shown that even small changes in the strand forming the ribbon are sufficient to induce a packing close to the HTH domain (Cordes et al. 1999). Thus, evolutionary unification of the HTH and RHH domains is possible (Aravind et al. 2005). MazE, PemI, and their homologs have a swapped hairpin β-barrel fold shared by AbrB and SpoVT-type of transition state regulators (Coles et al. 2005). Each monomer of AbrB contains two β-hairpins that interweave with that of the dimer to form two layers of β sheets connected by a short α-helix. Residues from the β1 loop extend into the major groove to make base-specific interactions. Residues in the β2 loop and the α-helix are also critical for DNA-binding ability of AbrB (Sullivan et al. 2008). A bioinformatics approach showed that HTH domain containing antitoxins are found to co-occur with RelE/ParE, Zeta, HipA, GinD, and a large number of other uncharacterized toxins. The RHH domain-containing antitoxins are found together with ParE/RelE and CcdB/MazF type toxins, while the AbrB-type antitoxins are found with Doc, CcdB/MazF, and VapC-type toxins (Leplae et al. 2011).

The exploitation of TA modules presents an effective strategy for the development of novel antibacterials as they are present in most bacterial pathogens, but have no human homolog. Disruption of the preformed TA complex or prevention of formation of the TA complex could thus release the toxin to exert its lethal effect. In case of MazF-type toxins, it may be possible to achieve partial disruption of the TA complex in two ways: 1) disruption of TA interactions at the active site, thus allowing the toxin to cleave free mRNA and 2) causing allosteric activation of the toxin. Also, MazF toxins are proposed to be more impervious to the development of resistance in comparison to CcdB-like toxins. Previous reports have established that a targeted peptide can disrupt the TA interaction (Lioy et al. 2010; Chopra et al. 2011) and the utility of TA disruption for antimicrobial development (Lioy et al. 2010). There is a scarcity of structural information and data on the amino acids and residues that define the binding hotspots between toxin and antitoxin (Williams and Hergenrother 2012). MoxXT from *Bacillus anthracis* (Agarwal et al. 2010; Chopra et al. 2011) and YdcDE from *B*acillus *subtilis* (Pellegrini et al. 2005) are TA systems consisting of a toxin with a SH3-like barrel domain and an antitoxin with a RHH motif at the N-terminus. In an initial Position-Specific Iterative (PSI)-Blast search, a large number of homologs could be identified for MoxT in the nr (nonredundant) sequence database, including MazF, Kid, and YdcE. MoxXT and the previously reported YdcDE are TA systems in which the antitoxin belongs to the CopG family as distinct from the SpovT/AbrB-type antitoxin found in other members of the MazF/PemK family. In the present work, we carried out exhaustive search and filtering to identify MazF family TA modules with simultaneous identification of the different folds of their cognate antitoxin. The various pathogenic bacteria in which they could act as drug targets were identified. Further, we carried out the sequence analysis of the MazF family toxins in order to detect the

presence of conserved residues at the binding sites for the antitoxin.

Evolutionary studies on operons have shown that they may either be transferred by horizontal gene transfer or by direct descent. They may be transmitted either as a whole, as a single gene, or as a subset of genes (Omelchenko et al. 2003). In view of the unique addiction properties of TA operons, their transfer as operons is necessitated. However, diversity in the TA system is evident in the presence of different types of antitoxins paired to the same toxin. Therefore, gene neighborhood analysis was carried out to probe the evolution of the toxins with different types of cognate antitoxins. Another complexity to be considered during evolution of TA modules is their plasmid or chromosomal location. Therefore, we also analyzed the phylogenetic tree obtained in context of chromosomal/plasmid location. Network modeling was carried out to identify the hubs of gene transfer in the set of MazF family toxins. Earlier studies have proposed that the toxin forms a distinct evolutionary module with the C-terminal residues of the antitoxin, while the upstream promoter region of the DNA and the N-terminal DNA-binding domain forms another distinct evolutionary module (Smith and Magnuson 2004). In view of this, we searched for different types of regulatory elements present in the noncoding genomic region upstream of the TA operon in each of the different subfamilies. Our analysis yields insights into the distribution of MazF family TA modules in pathogenic organisms, their multibinding sites, mode of evolution, and conserved upstream motifs of the MazF family TA modules.

## Materials and Methods

### Identification of MazF/PemK/Kid Toxin Family Homologs

#### *Search for Homologous Sequences*

The sequence of MoxT from *B. anthracis* was used for a sequence similarity search against the nr public database using PSI-Blast (Altschul et al. 1997). A cut off e-value threshold of 0.005 was used for PSI-Blast, and the run was iterated till convergence. Distinct representatives from the hits were obtained, and the database was searched with these sequences as queries to perform an exhaustive search for additional distant homologs of MazF protein using the generalized methodology as described earlier (Anantharaman and Aravind 2003; Guglielmini et al. 2008). The putative toxins without a cognate antitoxin partner were discarded. In order to qualify the hits as putative TA systems, sequences were filtered according to the following criteria: 1) the length of the toxin should be 60–150 amino acids; 2) an open reading frame (ORF) of length 40–90 amino acids transcribed in the same direction as the toxin must be present upstream of the putative toxin; 3) the separation or overlap between the putative antitoxin and toxin ORFs must not exceed 30 bases; 4) sequences should have a complete PemK fold. Fold prediction of the obtained toxin sequences was carried out using Batch CD search (Marchler-Bauer et al. 2011) in order to confirm that the fold of the toxin was SH3 barrel as found in MazF. 5) Finally, the toxin should be a member of a two-gene operon. The number of genes present in each operon was predicted using the Database of prOkaryotic OpeRons (DOOR) database (Mao et al. 2009). Entries not found in DOOR were searched for in the microbes online (Dehal et al. 2010) and PropDB databases (Taboada et al. 2012).

Clusters of the toxin sequences were made using Blastclust (L 0.75; S 1.0) using the resources available at http://toolkit.tuebingen.mpg.de/ (last accessed November 24, 2013). The clusters with a single type of cognate antitoxin were combined. The sequence alignment in each individual cluster and combination of clusters was carried out using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004) and manual sequence editing/analysis of individual clusters was done using Bioedit version 7.0.5.3. The secondary structure assignment of the sequences was carried out using STRuctural Alignments of Proteins (STRAP) (Gille and Frommel 2001). The pattern of secondary structures and the placement of gaps in the alignment were inspected manually. Fold prediction of the cognate antitoxin protein was also carried out using a Conserved Domain Database (CDD) or Protein Homology/analogY Recognition Engine V 2.0 (PHYRE2) (Kelley and Sternberg 2009) to identify the DNA-binding motif. Intergenic sequences upstream of the two gene TA operons were collected for motif analysis.

#### *Sequence Coding*

For further analysis, the selected sequences were coded as follows: the genus constitutes the first two letters and species name constitutes the next two letters followed by the GenBank ID (GI) number. Plasmid and chromosomal location was indicated by characters "p" and "c," respectively, after the GI number. Toxins with different cognate partners were encoded separately as follows: ab: AbrB/SpovT, co: MetJ/Arc/CopG, du: Duf(3018), du1: Duf(104), du2: Duf(2281), wi: winged HTH, re: relE, ph: Phd/YefM, and hyp: No conserved domain.

#### *Presence of TA Modules in Strains of Pathogenic Organisms*

The diseases caused by the organisms in which the sequences of the clustered TA modules were present were determined with the help of the Kyoto Encyclopedia of Genes and Genomes (KEGG) Disease resources. The human pathological strains were listed using the PathoSystems Resource Integration Center (PATRIC) database (Gillespie et al. 2011). Using Blast search, it was confirmed that the toxin homologs were present in all the pathogenic strains of the bacteria.

### *Tree Building*

Bayesian and maximum likelihood methods were used to construct the phylogenetic tree using the final alignment for the identified MazF family TA modules. ProtTest v2.4 (Abascal et al. 2005) was run to determine that the Le and Gasquel, 2008 (LG) model was the most appropriate model to evaluate trees. As MrBayes 3.1.2 (Altekar et al. 2004) does not support the LG model of evolution, the parameter was set to nst = mixed with gamma distribution of rates and invariant site categories used for the analysis. The MrBayes tree was built using the Cyberinfrastructure for Phylogenetic Research (CIPRES) science gateway (Miller et al. 2010). Two parallel runs, each consisting of four chains, were run simultaneously for $1.0 \times 10^7$ generations. Every 1,000th tree was sampled, and the first 25% of the generations were omitted from topology and probability reconstruction. In all analyses, convergence of the two parallel runs was observed. Maximum likelihood tree was computed with PhyML (Dereeper et al. 2008) using the Whelan and Goldman, 2001 (WAG) amino acid substitution model. The two methods produced similar trees.

### *Network Modeling*

The split network was created using the Neighbor-Net algorithm (Bryant and Moulton 2004) and uncorrected *P* distances with EqualAngle split transformation settings as implemented in Splitstree4.12.8 (Huson and Bryant 2006). The graph was then visualized using the EqualAngle layout.

### *Search for Conserved Motifs in the Operator Region*

A data set of upstream noncoding DNA sequences in representative members of each TA cluster of the tree was constructed. The presence of motifs in each set was probed using the Multiple Em for Motif Elicitation (MEME) suite (Bailey et al. 2009). Where a number of similar motifs were detected using the MEME suite, an attempt to combine the motifs was made using Glam2, which helps in determining the conserved motif from the DNA or protein sequences (Frith et al. 2008). MEME usually finds the most statistically significant (low *P* value) motifs first. The e-value of a motif is based on its log-likelihood ratio, width, sites, the background letter frequencies, and the size of the training set. Glam2 was run on the data set of upstream sequences to detect the presence of conserved motifs. The sequences were scanned for the presence of 10 shared motifs of length 6–50 bases with any number of repetitions in the data set. For determination of statistically significant *P* values, the MEME run was also performed with the "shuffle sequences" option. The occurrence of the upstream motifs found was also confirmed using Bioprospector (Liu et al. 2001).

## Results

### TA Module Search and Filtering

We carried out exhaustive search and filtering to detect MazF family TA modules in prokaryotic genomes. Over 400 toxin sequences were identified with an exhaustive PSI-Blast search. Of these, singlet toxins with no upstream gene (overlapping or within 30 bases) were discarded. Toxins that could not be annotated with a complete PemK fold were discarded from our data set. Further elimination of TA modules on the basis of their size as well the direction of transcription of the ORF was done in accordance with the method followed in Guglielmini et al. (2008). After extensive operon prediction using the resources as described in the Materials and Methods section, the TA operons present in operons having more than two members were discarded to yield a final data set of 263 TA operons. Two toxins annotated in Genbank as CcdB-type toxins were identified at the 2nd iteration of the PSI-Blast search (GI: 304312769, 58616201). Sequence alignment of MazF family toxins with these identified CcdB toxins showed that they do not share the same secondary structure pattern. As a result, they did not align well with other sequences. Therefore, these CcdB-type toxins were not grouped with the MazF family toxins in the present study.

### Phyletic Distribution of the MazF/PemK Family TA Modules

After filtering out the sequences that did not fulfill the outlined criteria for the MazF-type toxins belonging to a TA operon, the final set of 263 toxin sequences of TA operons with two predicted gene members. The GI numbers, sequence codes, and source organism of the sequences used in this analysis are listed in supplementary table S1 (Supplementary Material online). Of the two-membered TA operons, 95% were found in bacteria, one sequence was from the *Lactobacillus johnsonii* prophage Lj771, while the rest were distributed in Archaea (Euryarchaeotes). Within the bacterial subgroup, 110 sequences (42%) were concentrated in the family Proteobacteria, 67 (25%) were in Firmicutes, while the rest were found to be distributed in Actinobacteria, green sulfur bacteria, Spirochetes, Cytophaga-Flavobacteria-Bacteroides (CFB) group bacteria, green nonsulfur group bacteria, Deinococcales, and others. In the present work, we detected the occurrence of five AbrB-linked two-membered TA modules in Archaea and one in virus. The distribution analysis based on genomic location was also carried out. A total of 215 (80%) of the MazF TA systems were located on the chromosomes, with only one fifth of the sequences being plasmid encoded.

### Clustering of Toxins by Sequence Similarity

The data set of 263 toxin sequences from two gene operons segregated into five clusters having 20 or more members

using Blastclust, with parameters as described in the Materials and Methods section. In subsequent discussions, these clusters will be referred to as Clusters I through V. The gene neighborhood analysis of the selected toxin sequences was carried out to identify their cognate antitoxin partner. Four main different classes of antitoxins were found upstream to the MazF family toxins, namely: AbrB/SpoVT, CopG, Duf(3018), and winged HTH. Duf(3018) is an uncharacterized family of bacterial proteins with unknown function. The TA systems with AbrB/SpovT-type of antitoxin were relatively large in number as compared to antitoxins of other types. The AbrB-linked and Duf(3018) TA systems were found to be highest in number in Proteobacteria, while CopG-linked TA systems were predominantly found in Firmicutes. The single Duf(104) family protein containing TA module was from *Archaeglobus profundus*, while the single Duf(2281) originates from the Cyanobacteria *Trichodesmium erythraeum IMS101*. All the five winged HTH antitoxin containing TA modules were from Archaea. Two putative hybrid TA systems mixing MazF-type toxins with RelB-type antitoxins were also identified (Toxin GI: 83592312 and 213692408) and provide a putative evolutionary link between the MazF and the RelE superfamilies. All the 48 Cluster I toxins showed high within-group sequence conservation, were paired with either CopG family or uncharacterized antitoxins, and had a chromosomal location. Other chromosomal toxins in clusters also shared high sequence similarity. However, the same is not true of the plasmid-borne toxins.

## Sequence Analysis and Structure Comparison

The crystal structure of the MazEF complex shows the presence of two intimate binding sites of interaction between the C-terminal residues of MazE and MazF. The sites have been described previously as Site 1 and Site 2 (Kamada et al. 2003) and are shown in the biological dimer in figure 1 (inset). Site 1 constitutes an interface between the disordered C-terminal

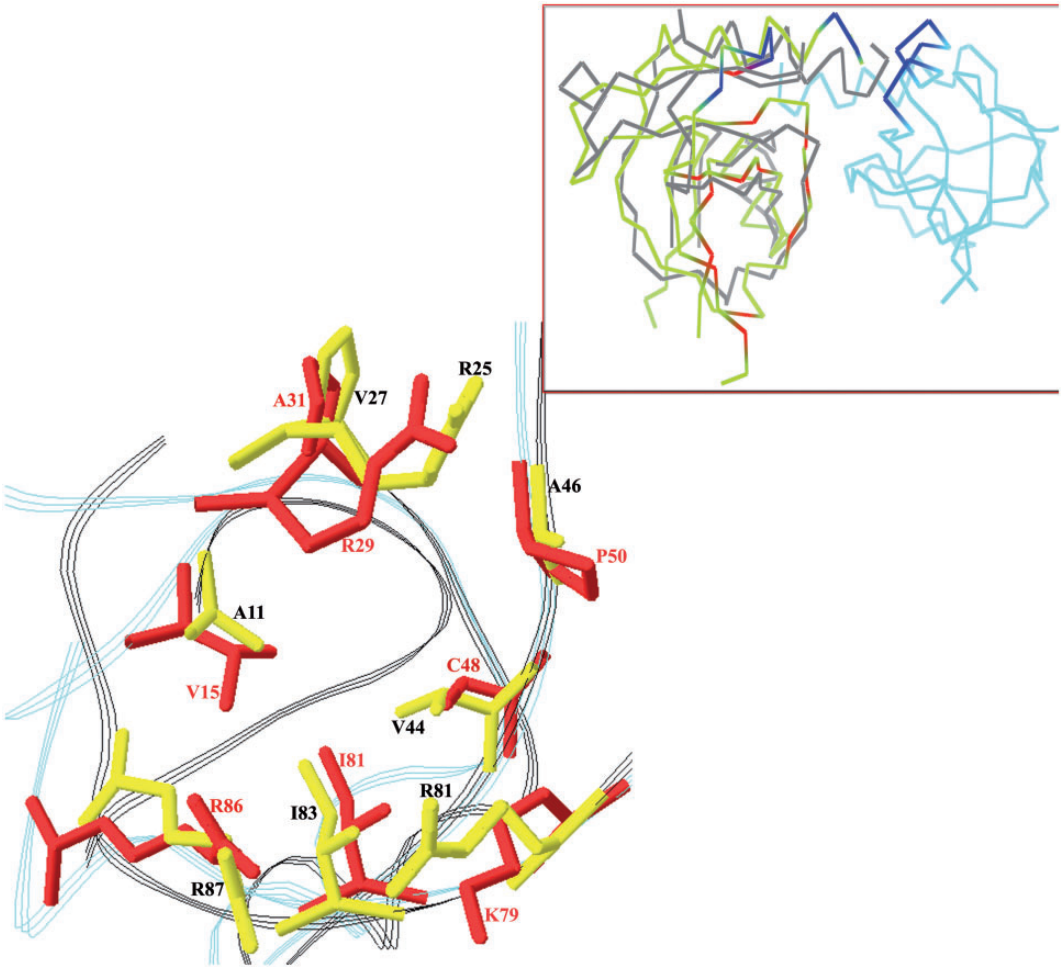


**Fig. 1.**—Superposition of residues at Site 1 of *E. coli* MazF (red) with structurally equivalent residues in *B. subtilis* YdcE (yellow). The inset shows the full length superposition of Chain A of the two toxin proteins with Site 1 residues in red and Site 2 residues in blue.

region of the antitoxin and one of the toxin monomers. Site 1 is the most intimate site of interaction between antitoxin and toxin and thus has the maximum contribution to binding. Specific interaction at Site 1 occurs by the occupation of a pocket in Chain A of MazF by Trp 73 of MazE. The pocket at Site 1 becomes exposed to solvent upon disordering of the S1–S2 loop. Sequence/structural analysis has shown that the hydrophobic pocket at Site 1 as well as a disordered S1–S2 loop is also conserved in CcdB and Kis. Site 2 consists of a crevice constituted by the dimeric interface of the two MazF monomers overlaid by α-helix H2 of MazE. Together, Site 1, disordering of the S1–S2 loop and Site 2 have been proposed as conserved features of TA interaction in the MazEF family (Kamada et al. 2003). Interactions of a small molecule or peptide at either Site 1 or Site 2 can help to disrupt the TA interaction, serving as a potential antibacterial strategy (Williams and Hergenrother 2012). Residues occurring with high probability (>0.9) in binding hotspots between MazE and MazF were predicted by hotspot analysis of the crystal structure of the MazEF using PCRPi (Jubb et al. 2012). The hotspot residues identified coincide with both Site 1 and Site 2. Further analysis was done to identify the evolutionary conservation of sequence and structure of the binding sites suitable for targeting by small molecules.

While YdcE, a toxin obtained from *B. subtilis*, has only 24% sequence identity with MazF, the fold is common (fig. 1, inset). Superposition of the structure of dimer of YdcE (PDB ID:1NE8) (Gogos et al. 2003) with MazF was possible with root mean square deviation (rmsd) of 1.54 Å among 166 residues. As shown in figure 1, the residues of the pocket at Site 1 in MazF are conserved in YdcE. Thus, the pocket at Site 1 is conserved structurally in the toxins of MazF family even in the absence of high sequence identity. Further, the pocket at Site 1 is conserved in both AbrB-linked toxins MazF and Kid (Kamada et al. 2003) as well as CopG-linked toxins YdcE and MoxT. Sequence analysis of the Cluster I–V toxins was carried out to determine the extent of conservation at the various sites. The sequence alignment of MazF toxins as obtained in CopG-linked Cluster I is shown in supplementary figure S1 (Supplementary Material online) with the residues involved in interactions at the two sites marked as 1 and 2, respectively. As shown in this figure, within this cluster (Cluster I), there is a high degree of conservation in the S1–S2 loop as well as at Sites 1 and 2. A consurf rendering showing conservation of the residues of the binding pocket at Site 1 is shown in figure 2. The residues of the pocket (highlighted) in Site 1 were found to be highly conserved and are overlaid by the conserved residues constituting the loop (not highlighted).

All members of Cluster II, Cluster III, and Cluster V, including *Escherichia coli* MazF (PDB ID: 1UB4) and Kid (PDB ID: 2C06), were paired with AbrB/Spovt-type of antitoxin. Multiple sequence alignment of all 3 clusters together with the secondary structure notation derived from experimentally derived structures and catalytic residue information is shown

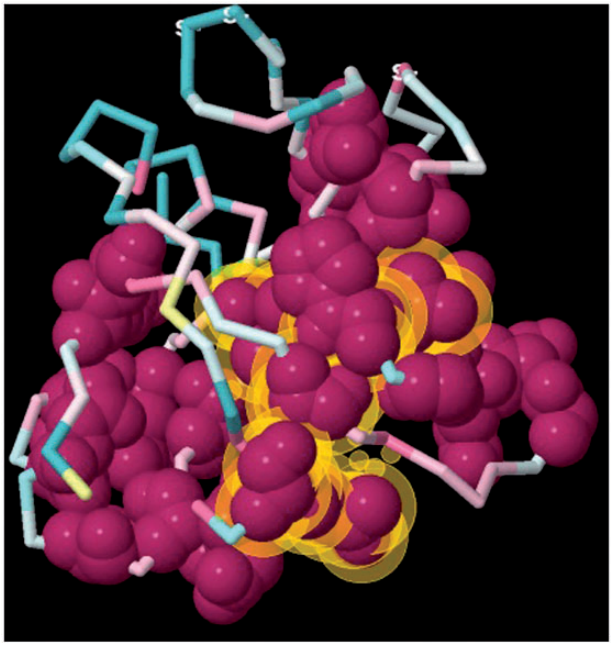

**Fig. 2.**—Consurf rendering of 1NE8 pocket showing the cartoon of the entire protein. The highly conserved residues are shown as purple balls. The pocket residues (11, 87, 83, 81, 44, 46, 25, 27) are highlighted by a yellow halo. The residues overlying the highlighted pocket are constituted by the loop S1–S2 (residues 12–27 in 1NE8).

in supplementary figure S2 (Supplementary Material online). The alignment in the three different panels *A*, *B*, and *C* of the supplementary figure S2 (Supplementary Material online) corresponds to the three clusters and shows the distinct pattern of conservation in the three sets. There are some minor deviations in the side chains lining the hydrophobic pocket at Site 1 in CopG as compared to the AbrB-linked toxins, which may lead to small changes in the size and shape of the pocket. However, it is expected that the compounds that bind to Site 1 in AbrB-linked toxins will be able to bind to the same site in CopG-linked toxins also. At Site 2, the sequence in this cluster is not well conserved either within this cluster or with respect to Cluster 1. Thus, on the basis of sequence conservation, Site 1 is better suited for design of compounds to act by disruption of the TA complex in both groups. Cluster I toxins show high conservation throughout the sequence, Cluster II + III + V toxins show high sequence conservation in the N-terminal domain and at Site 1 in the C-terminal domain. In comparison, though the structure at Site 2 is conserved, the sequence is not conserved in the Clusters II, IV, and V.

All sequences in Cluster IV were paired with Duf(3018)-type of antitoxin partners and had predominantly chromosomal location with few plasmid borne modules. These have not been validated experimentally, and no information about the mode of action of these putative toxins is available at present. However, the toxins may be ribonucleases in view

of the high sequence similarity to toxins in the other clusters identified. The sequence alignment of Cluster IV toxins is shown in supplementary figure S3 (Supplementary Material online). In Cluster IV sequences, sequence conservation of the N-terminal domain is similar to that of the AbrB-linked toxins. The sequence conservation in the C-terminal region differs markedly, showing highly conserved residues interspersed throughout the C-terminal domain of the toxin. Thus, there is a clear difference between the pattern of conservation in the toxin sequences in the three clusters.

In Cluster I, the catalytic residues of *B. anthracis* have been established as H59 and E78 (Agarwal et al. 2010). As the catalytic residues of Kid are known to be R73 and D75 (Kamphuis, Monti, van den Heuvel, Lopez-Villarejo, et al. 2007), the E78 residue could be inferred by sequence similarity D75. However, the R73 is replaced by a nonreactive L in MoxT, therefore H59 was inferred by structural proximity to D75 in the modeled MoxT. Therefore, other charged residues in a structurally favorable position can act as an acid–base pair to mediate catalysis. In accordance with previous studies, sequence analysis (supplementary fig. S1, Supplementary Material online) shows that the catalytic residues E78 and H59 are conserved in majority of the sequences in the CopG-linked subfamily of toxins (catalytic residues marked by asterisk). In Clusters II, III, and V (supplementary fig. S2, Supplementary Material online), the residue at D75 of Kid is found to be identical in MazF and replaced by a polar residue (E or N) in other sequences. However, R73 of Kid is replaced by a nonpolar residue in a majority of the sequences. The difference in the catalytic residues of Kid and MazF has been noted previously. H28 has been proposed as a general base to stabilize the catalytic intermediate in MazF (Li et al. 2006). Accordingly, the residues at analogous to H28 position of the alignment are found to be polar/charged in most of the sequences of Clusters II, III, and V. In Cluster IV Duf-linked toxins, the position analogous D75 of Kid is occupied by D in all the sequences. However, the position analogous to R73 of Kid is replaced by M. The position analogous to H59 of the CopG-linked toxins is also occupied by a nonpolar residue in most of the sequences. Therefore, the catalytic acid of the Cluster IV toxin remains to be determined.

## TA Modules of the MazF/PemK Family: Presence in Pathological Organisms

It is well known that MazF family toxins are present in pathogenic bacteria (Gerdes et al. 2005). However, due to their occurrence in plasmids, as well as their linkage with mobile genetic elements, their presence within a bacterial species is heterogenous, and their occurrence in clinical isolates is required to be confirmed (Williams and Hergenrother 2012). In this work, we have identified MazF family TA modules with conserved sequence features. Further, we screened the KEGG Disease resources and PATRIC database to confirm the presence of these TA modules in all the known pathogenic strains of these bacteria. TA modules linked to both CopG-type antitoxins and AbrB-type antitoxins were found to be present in some important disease-causing bacteria. Table 1 shows the species distribution of the different MazF-type toxins of the CopG-linked Cluster (I) and the AbrB-linked clusters (II, III, and V) in pathogenic bacteria as well as the disease caused. The presence of homologs of the CopG-linked toxins was detected in all the pathogenic strains of *Clostridia* spp. and *Listeria monocytogenes*. AbrB-linked toxins are found in all the pathological strains of *Enterobacter* spp., *Klebsiella pneumoniae, Salmonella enterica* and *Neisseria meningitidis.* A set of toxins with no identifiable N-terminal domains were found in *Staphylococcus aureus* in Cluster I.

**Table 1**
Pathogenic Organisms Containing MazF Type TA Modules and the Diseases Caused by Them

| | Organism | Disease |
|---|---|---|
| **Cluster I TA modules** | *C. perfringens* | Food poisoning and gas gangrene |
| | *Li. monocytogenes* | Listeriosis |
| | *St. aureus* | Nosocomial infections |
| | *C. botulinum* | Botulism |
| **Clusters II, III, and V TA modules** | *Enterobacter sp.* | Shigellosis |
| | *K. pneumoniae* | Kliebsella pneumonia |
| | *S. enterica* | Enteric fever |
| | *N. meningitidis* | Meningitis |

## Phylogenetic Analysis

Molecular phylogenetic analysis is now a well-established computational technique for studying evolutionary relationships between gene families and genomes. Traditionally, phylogenetic analysis has been used to infer shared ancestry and divergence as well as gene speciation or duplication events. It has also been used to predict functions of novel proteins. However, in view of the large number of bacterial genomes now available, it is now possible to organize different protein families into trees that enable us to analyze these groups in the context of important biological information like interaction partners, function, etc. This will also help to elucidate evolutionary phenomenon like coevolution, gene/domain shuffling, linkage, etc.

Here, we carried out the phylogenetic analysis of the entire set of MazF family of toxins and analyzed the tree in context of their cognate antitoxins and the genomic location. The MrBayes tree depicting the interrelationships between the members of the MazF toxin family is shown in figure 3. As shown in figure 3, the toxins linked to the three dominant types of antitoxins, namely AbrB, CopG, and Duf(3018), cluster together on distinct branches of the tree. Blastclust results

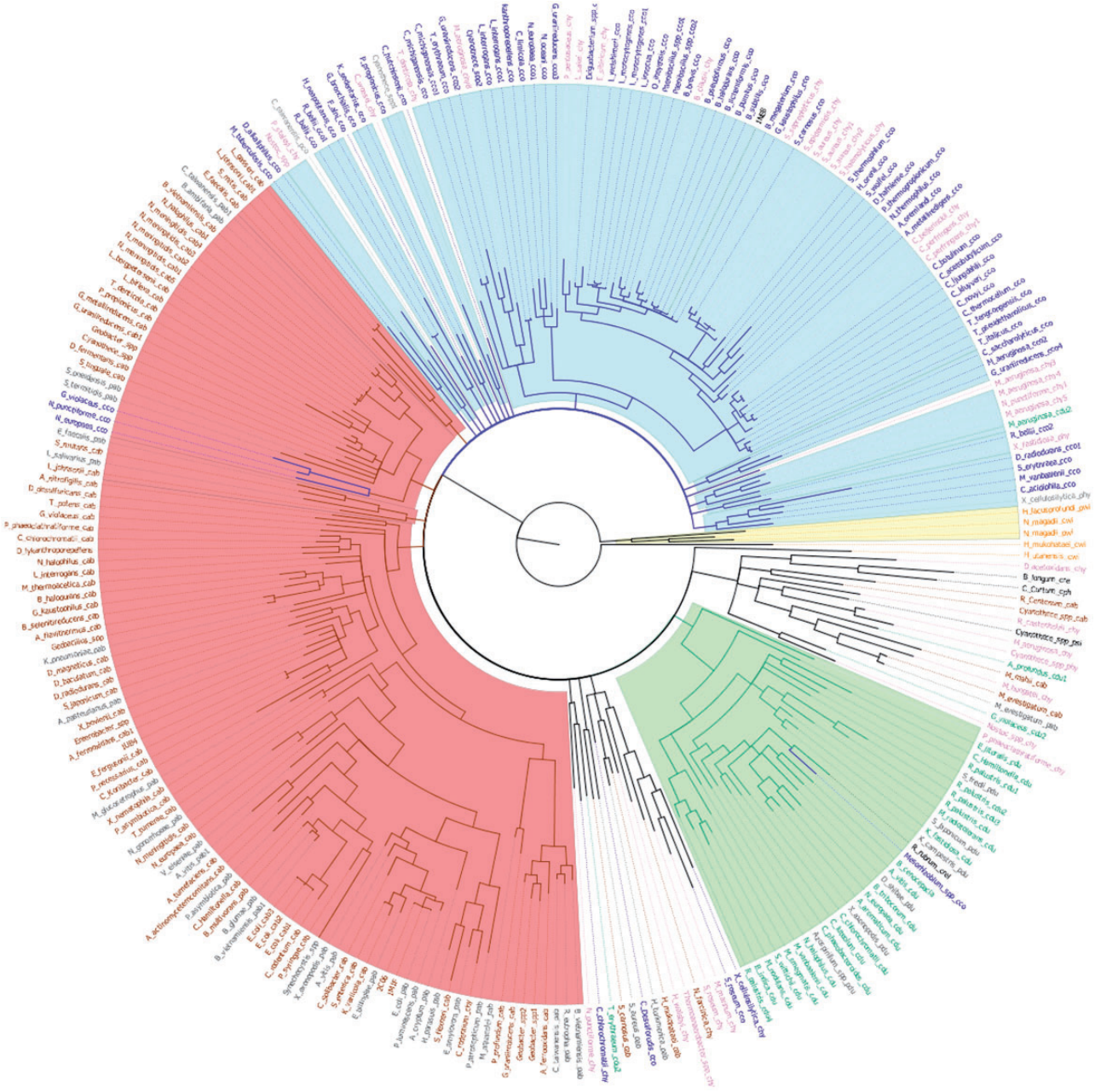

**Fig. 3.**—MrBayes computed tree of the MazF family toxins in two gene operons. All the branches had bootstrap values > 50%. The Abrb, CopG, Duf(3018), and winged HTH-linked toxin clusters are highlighted in red, blue, green, and yellow, respectively. In each highlighted cluster, branches and taxon labels are colored by type of antitoxin as described. However, taxon labels corresponding to plasmid borne TAs in each cluster are colored gray. Taxon labels of toxins for which the antitoxins could not be annotated are highlighted in pink.

show that chromosomal TA modules have better conservation among the toxins in comparison to plasmid borne modules. However, the plasmid borne TA modules are interspersed between the chromosomal sequences on the tree, indicating a possible gene transfer between plasmid and chromosomal TA modules from divergent species. In the AbrB group, a larger number of plasmid borne TAs are interspersed between the chromosomal toxins. The single reported TA module from the *L. johnsonii* prophage LJ771 may be integrated into the prophage genome from the bacterial host. Though another toxin linked to AbrB antitoxin is present in *L. johnsonii*, the closest branch neighbor for the prophage toxin on the inferred tree is a TA module occurring in the *L. salivarius* plasmid.

The two toxins coupled with RelB-type antitoxins were present as a long branch in the midst of the AbrB-linked toxins. However, another RelB-linked toxin clustered closely with CopG-linked toxins. Thus, evolutionary links are indicated between RelBE family with both these subgroups of MazF toxins. The Duf(2281), Duf(104), with Phd antitoxin linked toxins did not occur in any known clusters. The winged HTH-linked toxins occurring in Archaea make a small separate cluster close to the root, having both plasmid and chromosomal members. The toxins linked to antitoxins for which the toxin fold could not be determined were found predominantly interspersed with CopG family toxins. These may represent similar antitoxin N-terminal fold as the CopG cluster in which they occur, but could not be detected by the fold prediction methods used.

### Evolutionary Network of MazF Family TA Modules

Network modeling of the MazF family toxins was carried out as described. The resulting network is shown in figure 4. The Neighbor-Net analysis results support the clustering of different subfamilies of MazF subfamilies, as previously observed in the phylogenetic tree. In cases of higher frequencies of internode gene transfer, the calculated splits give rise to a reticulate network of nodes that helps to detect conflicting phylogenies suggestive of horizontal gene transfer. The figure shows that reticulations are dense within each subfamily. A comparatively lower number of reticulations join distinct subfamilies. The MazF toxins linked to AbrB-type antitoxins form one large, segmented, reticulated cluster with chromosomal toxins (red squares) frequently interspersed by plasmid-borne toxins (gray circles). The chromosomal CopG-linked and Duf-linked toxin subfamilies also occur in clearly demarcated and densely reticulated zones. It was observed that the AbrB-linked toxins occur in several reticulated clusters near the origin as well as further away from it, indicating several ancestral as well as recent instances of divergence. In contrast, the CopG-linked and the Duf-linked toxin branches clustered and were linked by edges near the end of the branch. This suggests that there was more recent divergence in these two groups as compared to the AbrB-linked toxins.

### Motif Discovery

The representative sequence clusters as described before were analyzed for the presence of conserved motifs upstream of their operon. The significance of the motif was confirmed by the fact that it was not found upon using the shuffled upstream sequences. A gapped two block gapped motif consisting of a twelve base inexact repeat separated was found in the AbrB set and is shown in figure 5*a*. The motif was found with high significance in upstream elements of 76 out of the 79 representative AbrB antitoxin-containing TA modules. Four consecutive bases of the ten

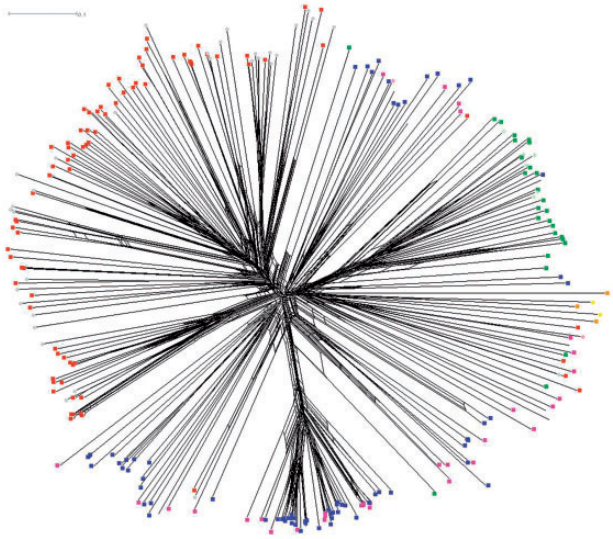

**Fig. 4.**—Split network of MazF-type toxins. Red square: AbrB-linked chromosomal toxins; gray circles: AbrB-linked plasmid-borne toxins; dark blue squares: CopG-linked chromosomal toxins; dark green squares: Duf-linked chromosomal toxins; light green circles: Duf-linked plasmid-borne toxins; dark pink: Chromosomal toxin sequences with hypothetical antitoxins; light pink: Plasmid-borne toxin sequences with hypothetical antitoxins; orange squares: Winged HTH-linked chromosomal toxins; yellow circles: Winged HTH-linked plasmid-borne toxins.

base repeat form a symmetrical palindrome. In a majority of the sequences, the detected motif was present directly upstream of the start site, either on the same or the complementary strand. Detection of the palindrome on the complementary strand is equivalent to an inversion of the recognition sequence in the region upstream of the TA module and may correspond to varying modes of structural recognition by the regulatory proteins.

The significant motif found in the upstream sequences of the CopG-linked toxins is shown in figure 5*b*. A single-block ungapped motif was also identified with high significance in 21 out of 25 of the Duf(3018) antitoxin-linked sequences. This motif consists of a six base palindrome on the same strand as shown in figure 5*c* (underlined bases). In most instances, the six base palindromic motif occurs directly upstream of the TA operon, and is present either on the same or the complementary strand. Thus, TA modules with distinct type of antitoxins share unique motifs in their upstream sequences. These may represent regulatory motifs.

## Discussion

In the present work, we have identified 263 TA modules after discarding the singlet or "bad" toxins (Guglielmini et al. 2008). It has been reasoned that these TA modules lack an antitoxin and thus, may also lack the toxic ribonuclease activity. However, the possibility that the antitoxin gene may be

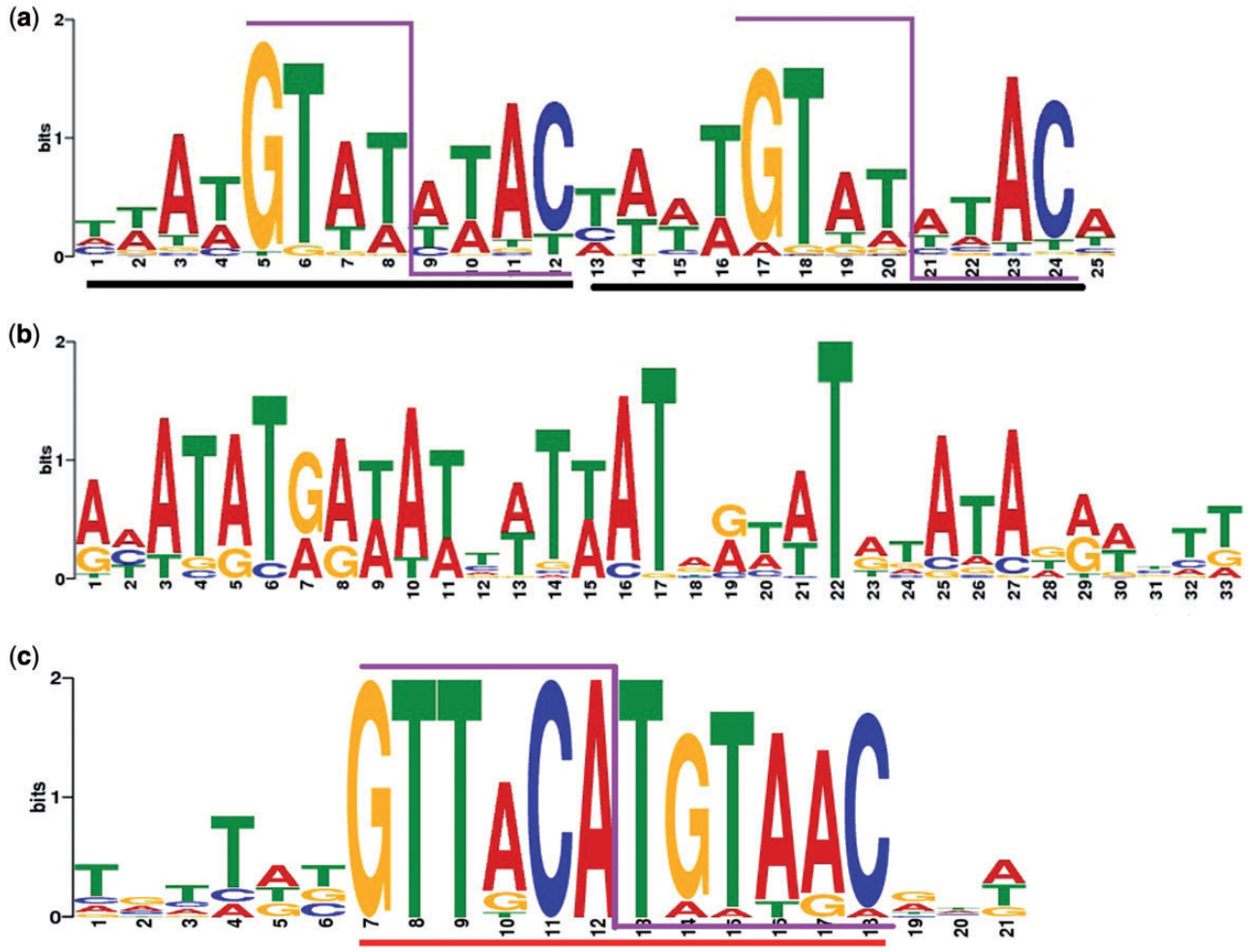


FIG. 5.—Graphical representation of sequence motifs detected in the upstream DNA sequence of TA operons of each subfamily. (*a*) Logo of motif detected upstream of AbrB-linked TA modules. An inexact ten base repeat is underlined in black. The purple lines depict the presence of a four-base palindrome in the motif. (*b*) Logo of motif upstream of CopG-linked TA modules (underlined in red). (*c*) Logo of motif upstream of Duf(3018)-linked TA modules. The 12 base motif is underlined in red and the six base palindromic sequence is shown by purple lines.

present elsewhere in the genome or on extrachromosomal elements and that its proteic product may be supplied "from outside" exists (Nariya and Inouye 2008) and needs to be explored. The presence of 67 predicted TA operons having more than two genes further raises interesting questions about the functionality of the additional genes in the operon and the mechanism by which these mosaic operons are formed. According to earlier reports, the MazF family toxins were thought to be confined to the bacterial domain (Pandey 2005). However, our results are in line with a later study reporting the presence of potential MazF toxins in Archaea (Guglielmini et al. 2008).

The CcdB-type toxins are grouped in the same superfamily as MazF as they share the same fold, even though the biological function/cellular targets are different (Anantharaman and Aravind 2003). While CcdB targets DNA gyrase and causes inhibition of DNA replication, MazF inhibits protein synthesis due to its ribonuclease activity. Previous sequence analysis studies have also not reported any significant sequence homology between MazF and CcdB-type toxins (Guglielmini et al. 2008). The two CcdB-like sequences identified in our database search did not share sufficient sequence similarity with the MazF family toxins and were thus discarded from the data set. However, an approach based on structural similarity may be able to unify the MazF and CcdB families.

## MazF Family Toxins Occur with Four Main Types of Antitoxins. Individual TAs Link MazF Evolutionarily to Other Families

Four main types of antitoxins were identified to occur with the MazF family toxins in the present study namely, AbrB/SpoVT, CopG, Duf(3018), and winged-HTH. The single antitoxin proteins from the Members of Duf(104) (COG2880) have previously been reported as putative antitoxins (Makarova et al.

2009). Structure of one of the members of Duf(104) family AF2212 from *Archaeglobus fulgidus* is known (PDB ID: 2NWT). The Structural Classification of Proteins (SCOP) classification of 2NWT shows that it shares the double split β-barrel fold with AbrB/SpovT antitoxins. Another antitoxin belonged to Duf(2281) (COG 3553), a domain family of bacterial proteins. The nuclear magnetic resonance structure of one member with Duf(2281) domain from *Pseudomonas aeruginosa* has been determined (PDB ID:2JPI). Comparison of 2JPI structure with representative structures in PDB shows that it has high similarity with the structure of transcription initiation factor, Transcription Factor IID (PDB ID: 1QNA).

A TA module having a RelE toxin combined with a MazE-type antitoxin has been identified previously (Schmidt et al. 2007). Similarly, in our data set, a single TA module having a Phd/YefM-type antitoxin (Toxin GI: 256827703) links the MazF family with the FIC domain Doc family (Garcia-Pino et al. 2008). These results are in agreement with previous studies on putative antitoxin partners for MazF-type toxins (Makarova et al. 2009). A recent comprehensive study catalogs Toxin-Antitoxin (TA) partners by sequence similarity as well as "Guilt by association" (GA) technique in prokaryotic genomes. In this study, MazF and CcdB was treated as one superfamily, and five categories of antitoxin partners were identified, of which the GA category constitutes a large group (Leplae et al. 2011). Thus, while MazE and CcdA-type antitoxins are well established as partners for the MazF family, the Duf domain linked TAs as found in the present study may have formed part of the GA category. They have not been reported or validated previously. Validation of the proposed TA function of the Duf-linked proteins is required to establish them as members of the MazEF family.

The observed high sequence conservation of the chromosomal toxins within a subfamily agrees with previous works that view the integration of plasmid TA operons into chromosomal locations as further continuation of their "selfish" behavior (Leplae et al. 2011). Chromosomal integration of the TA modules limits their genetic mobility by horizontal gene transfer and may underlie the observation that chromosomal TA modules, which are mostly inherited by vertical descent, and have higher sequence conservation as compared to plasmid-borne modules.

## A Conserved Pocket in a Multibinding Site of MazF Family Toxins Is a Potential Drug-Binding Hotspot

Targeting protein–protein interactions (PPIs) for development of therapeutic molecules poses the problem of tethering a small molecule to a large, featureless protein surface. However, studies have shown that the regions involved in protein–protein interface formation are inherently flexible or intrinsically disordered, which may allow small molecules to diffuse in and disrupt the complex (Thangudu et al. 2012). Successful targeting of PPIs is also made possible due to the ability of the small molecules to bind to a small fraction of the residues contributing to the interface, also known as binding hotspots, which contribute the most to the free energy of interaction. Further direction to the success of the approach is promised by the recognition that, while one of the protein surfaces is disordered, flat, and featureless, the other consists of one or a series of well-defined pockets. These pockets form multibinding sites and are highly conserved (Jubb et al. 2012).

The MazF family TA systems are proposed to have a conserved mode of binding to the C-terminal of their antitoxin through Site 1 and Site 2 (Kamada et al. 2003). YdcE of *B. subtilis* and *B. anthracis* toxin MoxT have high sequence and structural similarity with MazF. Various peptides based on the sequence of MoxX and MazE were capable of disrupting the MoxXT interaction (Agarwal et al. 2010). A rationally designed peptide could disrupt the MoxXT interaction by 42% at micromolar concentration (Chopra et al. 2011). Thus, it has been shown that some molecules are able to disrupt the interactions of the MazF family TA systems.

At Site 1 of the MazEF complex, the disordered, extended and essentially featureless C-terminal half of the antitoxin binds via a hydrophobic group to a pocket on the corresponding surface of the toxin. Sequence analysis has shown that Site 1 is also highly conserved among the MazF family TA modules. Also, disruption of Site 2 TA interactions by a small molecule or peptide is expected to cause a decrease in the toxic ribonuclease activity due to proximity of Site 2 to the catalytic site (Agarwal et al. 2010). In view of the contribution of Site 1 to binding, nature of the interface with the disordered antitoxin C-terminus (Kamada et al. 2003), predicted sequence/structure conservation, and possible modulation by a variety of peptides, it represents a multibinding site for further drug design and discovery efforts.

As proposed on the basis of modeling studies, Site 1 may also be targeted for disruption by the bacteria's own quorum sensing mechanisms, resulting in enhanced ribonuclease activity of the toxin (Belitsky et al. 2011). The action of MazEF mediated quorum sensing Extracellular Death Factor (EDF) is yet to be confirmed in independent experiments from other laboratories (Van Melderen 2010; Guglielmini and Van Melderen 2011; Hayes and Van Melderen 2011). However, a recent report in respect of EDF-like peptides is that MazEF mediated cell death in *E. coli* can also be caused by quorum sensing peptides obtained from the supernatant of *B. subtilis* and *P. aeruginosa* (Kumar et al. 2013). While the mediation of programmed cell death by TA modules remains contested, support for the increase in ribonuclease activity of MazF is gained from cross-neutralization experiments between Kis toxin and the antitoxin of CcdAB, which has a similar structural and functional organization as found in the MazEF system. The C-terminal domain of CcdA is found to be sufficient for CcdB inhibition, as has been established for the MazEF family toxins. It was shown that CcdA binding to Kis was accompanied by an increase in ribonuclease activity

(Smith et al. 2012). CcdB and Kid share a similar fold even in the absence of apparent sequence similarity. However, Site 1, disordered loop and mode of toxin binding are similar to that of MazF and Kid. Therefore, Site 1 is indicated as the possible site mediating the interaction of CcdA with Kid. This is also in agreement with the observation that the physical interaction of CcdA with Kid did not cause a decrease in ribonuclease activity. The ribonuclease catalytic site (near to Site 2) could function unhindered by the presence of antitoxin at this site.

The structures of CcdB and Kid show the presence of a hydrophobic tyrosine that stabilizes their closed-loop conformation. However, in the antitoxin-bound disordered-loop conformation of toxin, the F17 that is structurally analogous to the CcdB and Kid tyrosine residues is partially disordered. Therefore, the invariant hydrophobic W73 of MazE may compete with the MazF F17 for interactions with the tryptophan-binding pocket of MazF, causing the closed loop to open up during TA complex formation. Specific occupation of the tryptophan-binding pocket would then serve as a mechanism of antidote recognition. The structure of the MazEF complex shows that the MazF-bound MazE residues at Site 1 (residues 67–73) adopt an extended β-strand like structure. Thus, the presence of the binding hotspot at Site 1 also suggests the mechanism of specific recognition of the extended, disordered, and essentially featureless C-terminal residues 67–73 of MazE at Site 1 of MazF. The recognition of the C-terminal residues of MazE by MazF at Site 1 is additionally aided by the occupation of another small hydrophobic pocket on the surface of MazF by His68 of MazE. Also, Q77 of the A chain of MazF makes a bidentate interaction with the backbone of residues E69 and I71 of MazE (Kamada et al. 2003).

## Roaming of Catalytic Residues in the MazF Family Toxins

Sequence analysis of the catalytic site residues showed that these differ in Clusters I–V. Combined sequence and structure analysis has been used to propose and validate putative toxin active site residues on the basis of spatial proximity to the conserved active site residue (Agarwal et al. 2010). The phenomenon of occurrence of distributive active site residues that are placed in different positions in the binding pocket has been termed as the roaming of catalytic residues and is also seen in the RelB family of toxins (Blower et al. 2011).

## MazF Family TA Modules with Conserved Sequence Features in Pathological Strains of Bacteria

It has been noted that the occurrence of TA modules is variable in different strains of the same organism (Williams and Hergenrother 2012). However, in our study, the presence of homologs of the toxins was detected in all pathogenic strains of the bacteria. Specifically, the presence and conservation of toxins linked to short antitoxins lacking an identifiable N-terminal domain in *St. aureus* is of interest in view of the role of this organism in nosocomial infections and the prevalence of multiple drug resistance in these bacteria (Shopsin and Kreiswirth 2001; Haddadin et al. 2002). One short TA module has been validated as a functional TA pair in *St. aureus* (Fu et al. 2007). Novel antibiotic strategies are also important in *St. aureus* as the community associated nature of the drug-resistant strains raises the possibility of its rapid and widespread colonization in the milieu outside the hospital (Gardam 2000; Miller and Kaplan 2009).

## Linkage and Shuffling in the MazF Family TA Modules

The clustering of the different subfamilies on the branches of the tree indicates the tight evolutionary linkage of each subfamily of TA modules. Thus, the two-gene TA operon is predominantly transferred as a whole, a situation that is perhaps intensified by the addictive nature of the TA interaction. However, as toxin singlets and TA operons consisting of more than two genes were not included in the current analysis, the possibility of their forming a dynamic continuum with the two-gene member TA operons remains to be explored.

The presence of different branches of MazF-type toxins linked to distinct antitoxins indicates that such prokaryotic TA systems may have arisen by gene shuffling wherein; the toxin has acquired a different antitoxin partner during evolution. A new antitoxin gene is incorporated from a distant source accompanied by the loss of the previous antitoxin gene by xenologous displacement (Omelchenko et al. 2003). Such xenologous displacement events are few in the two-membered MazEF family TA operons, and can occur either in plasmid or chromosomal TA modules as observed at the interface of each distinct antitoxin-linked subfamily. The probable mechanism for xenologous displacement is the slow accumulation of mutations at the TA interface. Some of the mutations may render the toxin dysfunctional, leading to decay of the TA system as previously studied for the $ccd_{O157}$ system. Inactivation of the toxin is considered the first and safer (nonlethal) step of TA systems degradation. Subsequent inactivation of the antitoxin may also occur. Study of the $ccd_{O157}$ homologs revealed that 61% of the TA systems possessed both active components, 21% had an inactive toxin, and 2% had both inactive components. Additionally, the plasmid-borne TA modules were found to be under higher selection pressure to retain their function (Mine et al. 2009).

The ability of toxin to successfully pair with different types of antitoxin can be attributed to the modular structure of antitoxin. The N-terminus of the antitoxin, which comprises diverse DNA-binding structural motifs, is mainly involved in mediating transcriptional regulation of the operon by binding to its own upstream DNA. The antitoxin binds to the toxin via its C-terminus, which is unstructured or partially unstructured

in toxin-free state (Nieto et al. 2007) but adopts an extended structure on binding to the toxin. Especially, the C-terminal stretch of the antitoxin comprising Site 1 and Site 2 can be modulated at distinct sequence positions by selected mutations at the TA interface. Thus, the toxin can exist in combination with different cognate antitoxin partners. The two main sites of interaction between TA, namely Sites 1 and Site 2, show a unique pattern of conservation of residues in the three different subfamilies in the present study. Therefore, gene shuffling is accompanied by simultaneous changes in the TA interaction sites in order that the biological complex formation is maintained. Across the entire MazF family, Site 1 shows higher conservation across species and antitoxin types as compared to Site 2. Site 2 coincides with the catalytic site and the interface of the biologically required toxin dimer in MazF family toxins. Yet, Site 2 mutations are preferred by the TA pair in spite of the possible loss of function incurred by making changes at this site. There may be two mechanisms to ensure antitoxin recognition at Site 1 of the toxin. Firstly, as apparent from sequence analysis, large changes at Site 1 of toxin are resisted as they would lead to loss of antitoxin recognition and cause unchained action of the toxin. Secondly, the sequence variability of the antitoxin at the C-terminal end suggests that a large number of peptides with varying sequence and structure can be recognized by the toxin at the invariant Site 1 of MazF family toxins.

Though it has often been shown that cognate antitoxins are unique to their antitoxins even when the antitoxins are from the same family, instances of cross neutralization between TA pairs have been reported (Grady and Hayes 2003; Nieto et al. 2007; Ramage et al. 2009; Fiebig et al. 2010). Instances of cross-neutralization of toxins by homologous antitoxins have been recorded within the CcdB family, which shares structural similarity with MazF (Santos-Sierra et al. 2002; Wilbaux et al. 2007). Point mutations did not adversely affect the ability of a CcdA antitoxin to neutralize its toxin. However, a frameshift mutation in CcdA yielded an inactive antitoxin (Mine et al. 2009). This indicates that mutations at the TA interface are carefully modulated out of sheer necessity of maintaining the TA interface. Functional interactions between ChpB and ParD, two homologous systems in *E. coli*, have also been reported (Santos Sierra et al. 1998). In the MazF family TA modules, functional interactions of MazE and Kid system caused the neutralization of the toxic activity of Kid with an altered stoichiometry of interaction as compared to the MazEF complex (Kamphuis, Monti, van den Heuvel, Santos-Sierra, et al. 2007). Variation in TA stoichiometry may be another mechanism for accommodating changes at the TA interface. Complex formation has also been demonstrated for Kis with CcdB, though with a lower efficiency of neutralization as compared to CcdA,and may suggest their common evolutionary origin (Smith et al. 2012). Functional and physical interactions of the MazEF and the VapBC TA systems have also been reported (Zhu et al. 2010), though the structural basis of such interactions is yet unknown.

### Location and Mobility of the Three Toxin Subfamilies

Our results indicate the passage of TA modules between plasmid and chromosomal DNA of different bacteria akin to that established for antibiotic resistance genes (Ochman et al. 2000) and secondary metabolites. As the AbrB-type toxins are often accompanied by transposon-like elements flanking the AbrB-type modules (Gerdes et al. 2005), this may account for the widespread occurrence of these TA modules on both plasmids as well as chromosomes. Plasmid-chromosomal transfer may also occur by means of conjugative plasmids, that are known to mediate transfer of large fragments of bacterial chromosomes (Wollman et al. 1956).The TA modules are often passengers of the bacterial mobilome, which enjoys horizontal mobility by means of bacteriophages, plasmids, and transposable elements. Horizontal gene transfer between the mobilome and chromosome may be mediated by specialized Gene Transfer Agents. Phylogenetic studies have also shown that replicon fusion is a common event during prokaryotic evolution, further blurring the boundaries between the plasmid and the chromosome (Koonin and Wolf 2008). TA modules have previously been included within the resistome domain of the bacterial mobilome (Makarova et al. 2009), where the term resistome was expanded to include genes responsible for resistance to stress caused by factors other than antibiotics. Indeed, the role of the TA systems in making their host cells resistant to stress by and inducing persister cell formation has been well documented (Jayaraman 2008; Gerdes and Maisonneuve 2012). The occurrence of antibiotic tolerance and persistence is also found to be high in cells overproducing MazF family ribonuclease toxins (Maisonneuve et al. 2011). The frequent transfer of MazF family TA modules between plasmid and chromosomal locations as shown in the present work establishes these selfish elements to be part of the resistome. Also, though the TA module may be "domesticated" into performing the normal cellular functions of the cell by integration into the chromosome, it may also revert to a purely plasmid-borne selfish nature.

The phylogenetic tree of the MazF family TA modules shows that the single RelB antitoxin-containing module is closely evolutionarily linked to the AbrB-linked toxins, while the archaeal-winged HTH-containing modules form a separate cluster near the root of the tree. Due to their notable structural similarity, the winged HTH motif is also an elaboration of the structural scaffold present in HTH. The HTH domain is thought to be an ancient conserved domain found in the last universal common ancestor of all extant life forms (Aravind et al. 2005). Therefore, the winged HTH-linked archaeal MazF Toxins may either represent the putative ancestral sequences for the MazF

family toxins or can be attributed to archaeobacterial gene transfer (Koonin and Wolf 2008).

### Bacterial Genome Stabilization by Transfer of a TA Module via a Bacteriophage

The bacterial prophage is an important agent of change during prokaryotic evolution. Bacteriophages and plasmids often integrate into chromosomes reversibly or irreversibly. Reversible integration can mobilize chromosomal genes, while irreversible integration serves to domesticate the mobile elements into resident genes (Ochman et al. 2000; Sundin 2007). The LJ771-borne TA module helps in the stabilization of the prophage (Denou et al. 2008). Thus, both the TA module and the prophage become incurably integrated in the chromosome of *L. johnsonii.* Once the TA operon is incorporated in the prophage, the selfish and addictive nature of the TA module ensures its own propagation.

### AbrB-Linked Toxins Form a Large Hub of Gene Transfer in an Evolutionary Network of MazF Family Toxins

In view of the pervasive horizontal gene transfer in prokaryotic genomes, the validity of tree thinking is questionable. A more accurate representation can be achieved with the help of directed network graphs, with weighted edges corresponding to the intensity of the flow of genes. The MazF family TA evolution can be conceptualized as a network graph with the chromosomal modules at the center, plasmid borne modules at the periphery, and distinct zones of toxins, representing the different antitoxin-linked subfamilies. The concept of distinct horizontal gene transfer highways and hubs in such a graph has already been introduced (Gogarten and Townsend 2005). Thus, the highly mobile AbrB-linked toxins occurring in superintegrons (Gerdes et al. 2005) form a hub of gene transfer between replicons as reflected in the widespread occurrence of the AbrB-linked TA subfamily in divergent bacteria. The presence of specific mechanisms that permit such mobility in the chromosomal CopG-linked and Duf-linked toxins remains to be investigated.

### Unique DNA Motifs Are Present in the Upstream Region of Each Toxin Subfamily

Many of the known TA modules are negatively autoregulated at the level of transcription. The autoregulation is mediated by binding of the antitoxin to the promoter region and is enhanced in presence of the toxin. Several antitoxins have been shown to bind to characteristic palindromic repeats of DNA in their upstream ORF (Madl et al. 2006; Mattison et al. 2006; Schreiter et al. 2006; Li et al. 2008). Further, the number, length, sequence and distance between the palindromes can modulate the strength of the protein–DNA interaction. We found that unique motifs were present in the upstream region of the subfamilies identified.

As the promoter region has been identified in both the MazEF and the Kis Kid TA modules, these sequences were scanned for the presence of the pattern found. Though the binding site identified by footprinting analysis is much larger, the MazEF module can be expressed by two promoters, namely P2 and P3 occurring at a distance of 13 base pairs (Marianovsky et al. 2001). Of these, P2, the upstream promoter is active in exponentially growing cells and has ten times the activity of the P3 promoter. Autoregulation by MazEF is attributed to the presence of an unusual alternating palindrome in its upstream DNA. The alternating promoter consists of three parts, namely a, b, and c. The middle a part is complementary to both the upstream c and downstream b parts of the palindrome. The MazEF complex may bind to either the c or the b arms of the palindrome, leading to inhibition of its own expression. The downstream b arm also overlaps with the strong P2 promoter (Marianovsky et al. 2001). Examination of the sequence of the MazEF promoter showed that the 10 base repeat containing a 4 base palindrome found by Glam2 is present both in the c and the b arms of the MazEF palindrome. However, the region identified in the motif finding by Glam2 corresponds to the downstream b arm of the palindrome. The Kis Kid promoter also contains two repeats, I and II. Motif I contains a perfect palindrome, whereas Motif II contains an imperfect palindrome. The binding of the Kid and Kis Kid complex is tighter to promoter I as compared to promoter II (Monti et al. 2007). The region containing the motif shown in figure 5*a* coincides with the promoter I of the Kis Kid upstream DNA. Thus, in both MazEF and Kis Kid systems, the promoter sequence established to be important for tight binding to the TA complex has been recognized. The significance of such regulatory motifs has been probed in conjunction with the varying stoichiometry of interaction in the Kis Kid module. It has been proposed that the expression of the Kis Kid TA system is tightly controlled by the molar ratio of the toxin and antitoxin, with additional complexity being provided by cooperative interactions between the two promoter regions (Monti et al. 2007). A palindromic pattern or repeat is not obvious in the motif found upstream of the CopG-linked toxins (fig. 5*b*). However, two regions of the upstream sequence consisting of a long and a short palindrome have been established to be important for MoxXT binding in *B. anthracis* (Chopra et al. 2011). The region recognized in MoxXT promoter by Glam2 contains the long palindrome. Thus, the motifs conserved in the upstream region of the different subfamilies of MazF family TA modules may correspond to regulatory sequences recognized by the distinct N-terminal domain of the antitoxin in each subfamily. Conservation of specific motifs in the upstream region of each TA subfamily is indicative of the fact that distinct upstream regulatory elements of DNA form an evolutionary module along with the N-terminal DNA-binding domain of

the antitoxin. Though there is a strong correlation of the unique motifs with each subfamily of MazF marked by unique N-terminal domain of the antitoxin, there remains the possibility remains that the motifs identified may represent binding sites for the TA complex, transcription factors or other cis-regulatory factors that modulate the action of the TA modules.

## Conclusion

Toxin proteins of the MazF/PemK/Kid family are found to occur in a majority with three types of cognate antitoxin partners, namely AbrB, CopG, and Duf(3018). The modular nature of antitoxin protein wherein the N-terminus has the DNA-binding motif, whereas the extended C-terminus binds to the toxin groove, facilitates such multiple antitoxin partners. AbrB was found to be the most prevalent type of antitoxin in the MazF family. Five populated clusters of MazF toxins were obtained and sequence analysis was used to identify an evolutionarily conserved toxin-binding hotspot for antitoxin/peptides. As disruption of TA modules is a potential drug target, a search for homologous two-member TA modules was made in all the strains of the pathogenic organisms. MazF family TA modules were present in the pathological strains of *C. perfringens, Li. monocytogenes, St. aureus,* and *C. botulinum*. The specific effect of disruption of TA interaction in these organisms requires further investigation.

Phylogenetic analysis showed that the toxin sequences with distinct cognate antitoxin partners clustered together on the branches of the Bayes tree, indicating that the operons are usually inherited as a whole. However, fewer gene shuffling events mediate the acquisition of new antitoxins. Acquisition of a new antitoxin needs to be accompanied by accumulation of mutations at the TA interface. The requirement for specificity in the C-terminal domain is perhaps aided by the disordered, extended nature of the C-terminal residues of the antitoxin sequence and structure plasticity. The plasmid borne TA modules were found to be interspersed with the chromosomal TA modules of the same subfamily presenting evidence of the exchange between TA modules of the plasmid and chromosomes of different bacteria. Especially, the plasmid–chromosomal exchange is frequent in the AbrB family of toxins, and may be attributed to the presence of upstream transposon-like elements in AbrB family TA operons. Interchange of genes between Archaea, divergent bacteria, closely related bacteria, the bacterial mobilome and the chromosome lends to the view that these exist in dynamic and stable equilibrium with each other. Network modeling of evolution in the MazF family TA modules shows that the chromosomal AbrB-linked TA modules are present as hubs of gene transfer. We found the conservation of specific upstream motifs in the AbrB and CopG family. The motifs identified could be correlated with the known operator regions of AbrB and CopG family TA modules. Taken together, phylogenetic analysis of MazF family TA modules unifies many observed phenomenon and presents a snapshot into the mechanisms of bacterial evolution. Further, while the presence of a possible multibinding hotspot outlines an opportunity for inducing TA disruption, the emergent role of TA modules in persister cell formation indicate that caution is required while pursuing this path.

## Supplementary Material

Supplementary tables S1 and figures S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

This work was supported by Indian Council of Medical Research grant number 64/5/2011-BMS. S.B. acknowledges Dr Aparup Das, NIMR, New Delhi, and Pooja Khurana, NSIT, for illuminating discussions about the methodology and interpretation of the tree. N.C. acknowledges financial support from CSIR.

## Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105.

Acebo P, et al. 1998. Structural features of the plasmid pMV158-encoded transcriptional repressor CopG, a protein sharing similarities with both helix-turn-helix and beta-sheet DNA binding proteins. Proteins 32: 248–261.

Agarwal S, Mishra NK, Bhatnagar S, Bhatnagar R. 2010. PemK toxin of *Bacillus anthracis* is a ribonuclease: an insight into its active site, structure, and function. J Biol Chem. 285:7254–7270.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407–415.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Anantharaman V, Aravind L. 2003. New connections in the prokaryotic toxin–antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome Biol. 4:R81.

Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev. 29:231–262.

Bailey TL, et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:W202–W208.

Belitsky M, et al. 2011. The *Escherichia coli* extracellular death factor EDF induces the endoribonucleolytic activities of the toxins MazF and ChpBK. Mol Cell. 41:625–635.

Blower TR, Salmond GP, Luisi BF. 2011. Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin–antitoxin partners. Curr Opin Struct Biol. 21:109–118.

Brown BL, Wood TK, Peti W, Page R. 2011. Structure of the *Escherichia coli* antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation. J Biol Chem. 286:2285–2296.

Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 21:255–265.

Chan WT, Moreno-Cordoba I, Yeo CC, Espinosa M. 2013. Toxin–antitoxin loci in *Streptococcus pneumoniae*. In: Gerdes K., editor. Prokaryotic toxin–antitoxins. Berlin: Springer-Verlag. p. 315–339.
Chopra N, Agarwal S, Verma S, Bhatnagar S, Bhatnagar R. 2011. Modeling of the structure and interactions of the B. anthracis antitoxin, MoxX: deletion mutant studies highlight its modular structure and repressor function. J Comput Aided Mol Des. 25:275–291.
Coles M, et al. 2005. AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. Structure 13:919–928.
Cordes MH, Walsh NP, McKnight CJ, Sauer RT. 1999. Evolution of a protein fold in vitro. Science 284:325–328.
de la Hoz AB, et al. 2000. Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. Proc Natl Acad Sci U S A. 97:728–733.
Dehal PS, et al. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 38: D396–D400.
Denou E, et al. 2008. The role of prophage for genome diversification within a clonal lineage of *Lactobacillus johnsonii*: characterization of the defective prophage LJ771. J Bacteriol. 190:5806–5813.
Dereeper A, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465–W469.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Fiebig A, Castro Rojas CM, Siegal-Gaskins D, Crosson S. 2010. Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin–antitoxin systems. Mol Microbiol. 77:236–251.
Frith MC, Saunders NF, Kobe B, Bailey TL. 2008. Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput Biol. 4: e1000071.
Fu Z, Donegan NP, Memmi G, Cheung AL. 2007. Characterization of MazFSa, an endoribonuclease from *Staphylococcus aureus*. J Bacteriol. 189:8871–8879.
Garcia-Pino A, et al. 2008. Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. J Biol Chem. 283: 30821–30827.
Gardam MA. 2000. Is methicillin-resistant *Staphylococcus aureus* an emerging community pathogen? A review of the literature. Can J Infect Dis. 11:202–211.
Gerdes K, Christensen SK, Lobner-Olesen A. 2005. Prokaryotic toxin–antitoxin stress response loci. Nat Rev Microbiol. 3:371–382.
Gerdes K, Maisonneuve E. 2012. Bacterial persistence and toxin–antitoxin loci. Annu Rev Microbiol. 66:103–123.
Gille C, Frommel C. 2001. STRAP: editor for STRuctural Alignments of Proteins. Bioinformatics 17:377–378.
Gillespie JJ, et al. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect Immun. 79:4286–4298.
Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol. 3:679–687.
Gogos A, Mu H, Bahna F, Gomez CA, Shapiro L. 2003. Crystal structure of YdcE protein from *Bacillus subtilis*. Proteins 53:320–322.
Gomis-Ruth FX, et al. 1998. The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. EMBO J. 17:7404–7415.
Grady R, Hayes F. 2003. Axe-Txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium*. Mol Microbiol. 47:1419–1432.
Guglielmini J, Szpirer C, Milinkovitch MC. 2008. Automated discovery and phylogenetic analysis of new toxin–antitoxin systems. BMC Microbiol. 8:104.
Guglielmini J, Van Melderen L. 2011. Bacterial toxin–antitoxin systems: Translation inhibitors everywhere. Mob Genet Elements. 1:283–290.
Haddadin AS, Fappiano SA, Lipsett PA. 2002. Methicillin resistant *Staphylococcus aureus* (MRSA) in the intensive care unit. Postgrad Med J. 78:385–392.
Hallez R, et al. 2010. New toxins homologous to ParE belonging to three-component toxin–antitoxin systems in *Escherichia coli* O157:H7. Mol Microbiol. 76:719–732.
Hayes F, Van Melderen L. 2011. Toxins–antitoxins: diversity, evolution and function. Crit Rev Biochem Mol Biol. 46:386–408.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.
Jayaraman R. 2008. Bacterial persistence: some new insights into an old phenomenon. J Biosci. 33:795–805.
Jubb H, Higueruelo AP, Winter A, Blundell TL. 2012. Structural biology and drug discovery for protein–protein interactions. Trends Pharmacol Sci. 33:241–248.
Kamada K, Hanaoka F, Burley SK. 2003. Crystal structure of the MazE/MazF complex: molecular bases of antidote–toxin recognition. Mol Cell. 11:875–884.
Kamphuis MB, Monti MC, van den Heuvel RH, Lopez-Villarejo J, et al. 2007. Structure and function of bacterial kid–kis and related toxin–antitoxin systems. Protein Pept Lett. 14:113–124.
Kamphuis MB, Monti MC, van den Heuvel RH, Santos-Sierra S, et al. 2007. Interactions between the toxin Kid of the bacterial parD system and the antitoxins Kis and MazE. Proteins 67:219–231.
Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 4:363–371.
Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36: 6688–6719.
Kumar S, Kolodkin-Gal I, Engelberg-Kulka H. 2013. Novel quorum-sensing peptides mediating interspecies bacterial cell death. MBio 4: e00314-13.
Leplae R, et al. 2011. Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. Nucleic Acids Res. 39:5513–5525.
Li GY, et al. 2006. Characterization of dual substrate binding sites in the homodimeric structure of *Escherichia coli* mRNA interferase MazF. J Mol Biol. 357:139–150.
Li GY, Zhang Y, Inouye M, Ikura M. 2008. Structural mechanism of transcriptional autorepression of the *Escherichia coli* RelB/RelE antitoxin/toxin module. J Mol Biol. 380:107–119.
Lioy VS, Rey O, Balsa D, Pellicer T, Alonso JC. 2010. A toxin–antitoxin module as a target for antimicrobial development. Plasmid 63:31–39.
Liu X, Brutlag DL, Liu JS. 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput. 127–138.
Madl T, et al. 2006. Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. J Mol Biol. 364:170–185.
Maisonneuve E, Shakespeare LJ, Jorgensen MG, Gerdes K. 2011. Bacterial persistence by RNA endonucleases. Proc Natl Acad Sci U S A. 108: 13206–13211.
Makarova KS, Wolf YI, Koonin EV. 2009. Comprehensive comparative-genomic analysis of Type 2 toxin–antitoxin systems and related mobile stress response systems in prokaryotes. Biol Dir. 4:19.
Mao F, Dam P, Chou J, Olman V, Xu Y. 2009. DOOR: a database for prokaryotic operons. Nucleic Acids Res. 37:D459–D463.
Marchler-Bauer A, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39: D225–D229.
Marianovsky I, Aizenman E, Engelberg-Kulka H, Glaser G. 2001. The regulation of the *Escherichia coli* mazEF promoter involves an unusual alternating palindrome. J Biol Chem. 276:5975–5984.
Mattison K, Wilbur JS, So M, Brennan RG. 2006. Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin–

antitoxin heterodimers containing pin domains and ribbon–helix–helix motifs. J Biol Chem. 281:37942–37951.
Miller LG, Kaplan SL. 2009. *Staphylococcus aureus*: a community pathogen. Infect Dis Clin North Am. 23:35–52.
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE); 2010 Nov 14; New Orleans, LA. p. 1–8.
Mine N, Guglielmini J, Wilbaux M, Van Melderen L. 2009. The decay of the chromosomally encoded ccdO157 toxin–antitoxin system in the *Escherichia coli* species. Genetics 181:1557–1566.
Monti MC, et al. 2007. Interactions of Kid–Kis toxin–antitoxin complexes with the parD operator–promoter region of plasmid R1 are piloted by the Kis antitoxin and tuned by the stoichiometry of Kid–Kis oligomers. Nucleic Acids Res. 35:1737–1749.
Nariya H, Inouye M. 2008. MazF, an mRNA interferase, mediates programmed cell death during multicellular Myxococcus development. Cell 132:55–66.
Nieto C, et al. 2007. The yefM-yoeB toxin–antitoxin systems of *Escherichia coli* and *Streptococcus pneumoniae*: functional and structural correlation. J Bacteriol. 189:1266–1278.
Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.
Omelchenko MV, Makarova KS, Wolf YI, Rogozin IB, Koonin EV. 2003. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. Genome Biol. 4:R55.
Pandey DP. 2005. Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res. 33: 966–976.
Pellegrini O, Mathy N, Gogos A, Shapiro L, Condon C. 2005. The *Bacillus subtilis* ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. Mol Microbiol. 56:1139–1148.
Ramage HR, Connolly LE, Cox JS. 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin–antitoxin systems: implications for pathogenesis, stress responses, and evolution. PLoS Genet. 5: e1000767.
Santos-Sierra S, Pardo-Abarrio C, Giraldo R, Diaz-Orejas R. 2002. Genetic identification of two functional regions in the antitoxin of the parD killer system of plasmid R1. FEMS Microbiol Lett. 206:115–119.
Santos Sierra S, Giraldo R, Diaz Orejas R. 1998. Functional interactions between chpB and parD, two homologous conditional killer systems found in the *Escherichia coli* chromosome and in plasmid R1. FEMS Microbiol Lett. 168:51–58.
Schmidt O, et al. 2007. prlF and yhaV encode a new toxin–antitoxin system in *Escherichia coli*. J Mol Biol. 372:894–905.
Schreiter ER, Wang SC, Zamble DB, Drennan CL. 2006. NikR-operator complex structure and the mechanism of repressor activation by metal ions. Proc Natl Acad Sci U S A. 103:13676–13681.
Schumacher MA, et al. 2009. Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. Science 323: 396–401.
Shopsin B, Kreiswirth BN. 2001. Molecular epidemiology of methicillin-resistant *Staphylococcus aureus*. Emerg Infect Dis. 7: 323–326.
Smith AB, et al. 2012. A common origin for the bacterial toxin–antitoxin systems parD and ccd, suggested by analyses of toxin/target and toxin/antitoxin interactions. PLoS One 7:e46499.
Smith JA, Magnuson RD. 2004. Modular organization of the Phd repressor/antitoxin protein. J Bacteriol. 186:2692–2698.
Sullivan DM, et al. 2008. Insights into the nature of DNA binding of AbrB-like transcription factors. Structure 16:1702–1713.
Sundin GW. 2007. Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. Annu Rev Phytopathol. 45:129–151.
Taboada B, Ciria R, Martinez-Guerrero CE, Merino E. 2012. ProOpDB: prokaryotic operon database. Nucleic Acids Res. 40: D627–D631.
Thangudu RR, Bryant SH, Panchenko AR, Madej T. 2012. Modulating protein–protein interactions with small molecules: the importance of binding hotspots. J Mol Biol. 415:443–453.
Van Melderen L. 2010. Toxin–antitoxin systems: why so many, what for? Curr Opin Microbiol. 13 781–785.
Wang X, et al. 2012. A new type V toxin–antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. Nat Chem Biol. 8: 855–861.
Wilbaux M, Mine N, Guerout AM, Mazel D, Van Melderen L. 2007. Functional interactions between coexisting toxin–antitoxin systems of the ccd family in *Escherichia coli* O157:H7. J Bacteriol. 189: 2712–2719.
Williams JJ, Hergenrother PJ. 2012. Artificial activation of toxin–antitoxin systems as an antibacterial strategy. Trends Microbiol. 20: 291–298.
Wollman EL, Jacob F, Hayes W. 1956. Conjugation and genetic recombination in *Escherichia coli* K-12. Cold Spring Harb Symp Quant Biol. 21: 141–162.
Zhu L, Sharp JD, Kobayashi H, Woychik NA, Inouye M. 2010. Noncognate *Mycobacterium tuberculosis* toxin–antitoxins can physically and functionally interact. J Biol Chem. 285:39732–39738.

**Associate editor:** John McCutcheon