Phosphorothioation: An Unusual Post-Replicative Modification on the DNA Backbone

Lianrong Wang[1,2], Shi Chen[1,2] and Zixin Deng[1,2]
[1]State Key Laboratory of Microbial Metabolism, and School of Life Sciences & Biotechnology, Shanghai Jiaotong University, Shanghai
[2]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, and Wuhan University School of Pharmaceutical Sciences, Wuhan China

1. Introduction

DNA molecules are polymers composed of basic repeating subunits of deoxyribonucleotides, which consist of the deoxyribose sugar, phosphate groups, and a nitrogenous base. They appear to fulfill all requirements necessary to maintain the genetic function of DNA. The five elements of nitrogen, phosphorus, carbon, hydrogen, and oxygen had been regarded as the canonical composition of DNA until the discovery of phosphorothioation, with a sixth element, sulfur, identified as an additional naturally occurring constituent on the DNA backbone, as a sequence-selective, stereospecific post-replicative modification governed by the *dnd* gene cluster. Unlike any other DNA or RNA modification system, DNA phosphorothioation is the first-described physiological modification of the DNA sugar-phosphate backbone [1].

The physiological phosphorothioate modification is widespread in bacteria and occurs in diverse sequence contexts and frequencies in different bacterial genomes, implying a significant impact on bacteria [2]. Recently, a counterpart phosphorothioate-dependent restriction system capable of protection against the invasion of unmodified foreign DNA was discovered to maintain the genetic stability of the phosphorothioate modified host [3]. Another type IV endonuclease, ScoA3McrA, was found to be capable of specifically recognizing as well as cleaving phosphorothioate modified DNA [4]. Interestingly, the gene *sco4631*, which code for ScoA3McrA, is unable to coexist with the *dnd* gene cluster in the same host, causing immediate cell death [4]. Here we summarize the discovery of this first reported physiological modification on the DNA backbone, and provide insights and perspectives into the biological functions of the phosphorothioate modification in prokaryotic physiology.

2. Discovery of phosphorothioation as an unusual post-replicative modification on the DNA backbone

The study of the physiological DNA phosphorothioation originated from an observation that an unusual DNA modification in *Streptomyces lividans* renders DNA susceptible to *in*

vitro Tris-dependent double strand cleavage, resulting in a DNA degradation (Dnd) phenotype during conventional and pulsed-field gel electrophoresis [5]. Zhou *et al.* demonstrated that such a Dnd phenotype was not due to nuclease contamination or improper *in vitro* genetic manipulation, but instead, an unusual DNA modification [5]. The modification sites are not randomly distributed in DNA. For instance, both plasmid pIJ101 and pIJ303 from Dnd⁺ *S. lividans* underwent site-specific cleavage during electrophoresis, giving particular fragment profiles [5, 6]. Ray *et al.* then verified that the Dnd phenotype depends on the cleavage activity of an oxidative Tris derivative generated in the electrophoretic buffer adjacent to the anode [7]. In other words, the DNA isolated from *S. lividans* is intact, and the degradation only occurs during electrophoresis in the presence of oxidative Tris. Thiourea can react with the Tris derivative and thus inhibits the DNA scission. Alternatively, non-degradative electrophoresis of the DNA could also be achieved in a different buffer such as Hepes [7]. Based on these observations, it was proposed that the DNA degradation was the consequence of a site-specific modification, which suffered cleavage by oxidative Tris resulting in degradation during electrophoresis [5, 7].

Dyson and Liang *et al.* later revealed that the modification required a conserved consensus sequence, as well as flanking sequences with potential for secondary structure(s) (section 3) [8, 9]. Meanwhile, no Tris-mediated scission was detected in single-stranded plasmid replication intermediates, supporting the post-replicative mechanism. The modifying reagents most probably acted post-replicatively on unmodified double-stranded DNA substrates [8].

The chemical nature of this unusual DNA modification is an intriguing question. Based on the information that two genes involved in this modification are related to sulfur transfer (section 2), Zhou *et al.* were prompted to conduct the ³⁵S labeling experiment. Dnd⁺ strains of *S. lividans*, *Streptomyces avermitilis* NRRL8165, and *Pseudomonas fluorescens* Pf0-1 were selected to propagate in media containing ³⁵SO₄²⁻. Total genomic DNAs were prepared and analyzed on agarose gel followed by Southern blotting. ³⁵S signals were detected in the DNA from three Dnd⁺ strains, but not in Dnd⁻ mutant ZX1 or *Streptomyces coelicolor*. This feeding experiment set up a link between the unusual DNA modification and sulfur [10].

The chemical nature of this unusual DNA modification was eventually found to be a phosphorothioate modification of the DNA backbone by Wang *et al*. In this modification, the non-bridging oxygen of the backbone phosphate group is replaced by sulfur [1]. Sequence specific phosphorothioate d(G$_{PS}$A) and d(G$_{PS}$G) were first detected in *E. coli* B7A and *S. lividans*, respectively. The discovery was based on the inability of nuclease P1 to cleave the phosphorothioate bond. Wang *et al.* fed Dnd⁺ *E. coli* B7A with L-[³⁵S]-cysteine to label the DNA [1]. Enzymatic hydrolyzed and dephosphorylated nucleosides were resolved by liquid chromatography followed by scintillation counting to locate the ³⁵S containing molecules. Mass spectrometric analysis of the ³⁵S containing molecules revealed characteristic *m/z* of 597 accompanied by 446, 348, 152 and 136 fragments (Figure 1). 152 and 136 are characteristic *m/z* of guanine and adenine in positive mode, respectively. This suggests the presence of a G- and A-containing dinucleotide structure for the *m/z* of 597 molecular ion, with loss of guanine yielding the ion at *m/z* of 446. The 16-mass-unit increase over a canonical dG-dA dinucleotide (*m/z* 581) is the exact mass difference between a sulfur and an oxygen atom. The putative dinucleotide species can survive the enzymatic digestion to single nucleosides, indicating nuclease resistance. These features suggested phosphorothioate-containing species shown in Figure 1. Enzymatic digestion with nuclease P1 followed by dephosphorylation with alkaline phosphatase yields phosphorothioate

modified dinucleotides and canonical nucleosides. Wang *et al.* eventually corroborated the phosphorothioate structure in *E. coli* B7A as d($G_{PS}A$) in R_P configuration by using synthetic d($G_{PS}A$) R_P and d($G_{PS}A$) S_P as references [1].

Remarkably, the phosphorothioate modification in *S. lividans* displayed different sequence selectivity as d($G_{PS}G$) R_P. To date, a repertoire of phosphorothioate-containing sequences, including d($C_{PS}C$), d($G_{PS}T$), d($A_{PS}C$), and d($T_{PS}C$), have been discovered in diverse bacterial species [2]. The substitution of sulfur creates a chiral center on the phosphate, resulting in two diastereoisomers, known as the R_P and S_P isomers. However, the physiological phosphorothioate modifications found in bacteria are all in the R_P configuration. DNA phosphorothioation thus represents a sequence-selective and stereo-specific physiological modification of the DNA backbone.
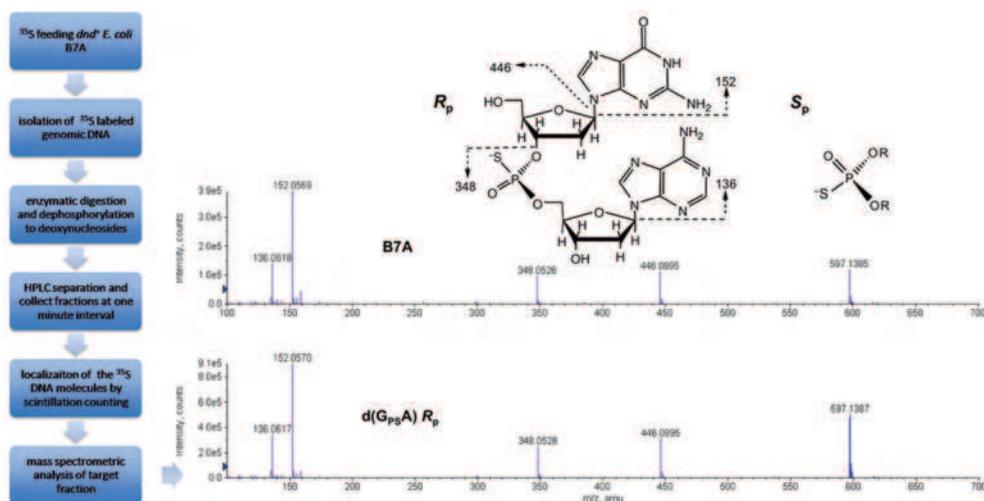

Fig. 1. The flowchart of the localization of phosphorothioate d($G_{PS}A$) R_P and mass spectra of isolated and synthetic d($G_{PS}A$) R_P. The fragmentation of d($G_{PS}A$) R_P is shown in the structural inset, with the [M+H] at *m/z* 597 in positive mode [1].

When phosphorothioate linked d($G_{PS}A$) R_P from *E. coli* B7A was treated with activated Tris buffer *in vitro*, the cleavage of the phosphorothioate bond was detected with the observation of dG and dA, whereas regular d(GA) without phosphorothioate bond remained intact. Therefore, the phosphorothioate modification was verified as the molecular basis for the Dnd phenotype during electrophoresis [1].

3. The *dnd* gene cluster is responsible for phosphorothioation

Evidence for a genetic link responsible for phosphorothioation came from the isolation of a mutant of *S. lividans*, ZX1, obtained by NTG ((*N*-methyl-*N*-nitro-*N*-nitrosoguanidine) mutagenesis [10]. In comparison to the wild-type, ZX1 has a ca. 90 kb chromosomal deletion and loses the Dnd phenotype, suggesting that the endogenous genes related to the unusual DNA modification are located in this 90 kb fragment.

A set of 13 overlapping cosmids covering the 90 kb region deleted in ZX1 but present in the wild-type were constructed and aligned as shown in Figure 2. When transformed into ZX1, cosmid 16C3 (ZX1::16C3) could restore the Dnd phenotype of mutant ZX1, indicating that 16C3 harbored genes associated with the DNA modification. By subsequent sub-cloning and Dnd phenotypic tests, a 6,665 bp *dnd* locus containing five *dnd* genes was precisely localized on cosmid 16C3 [11].
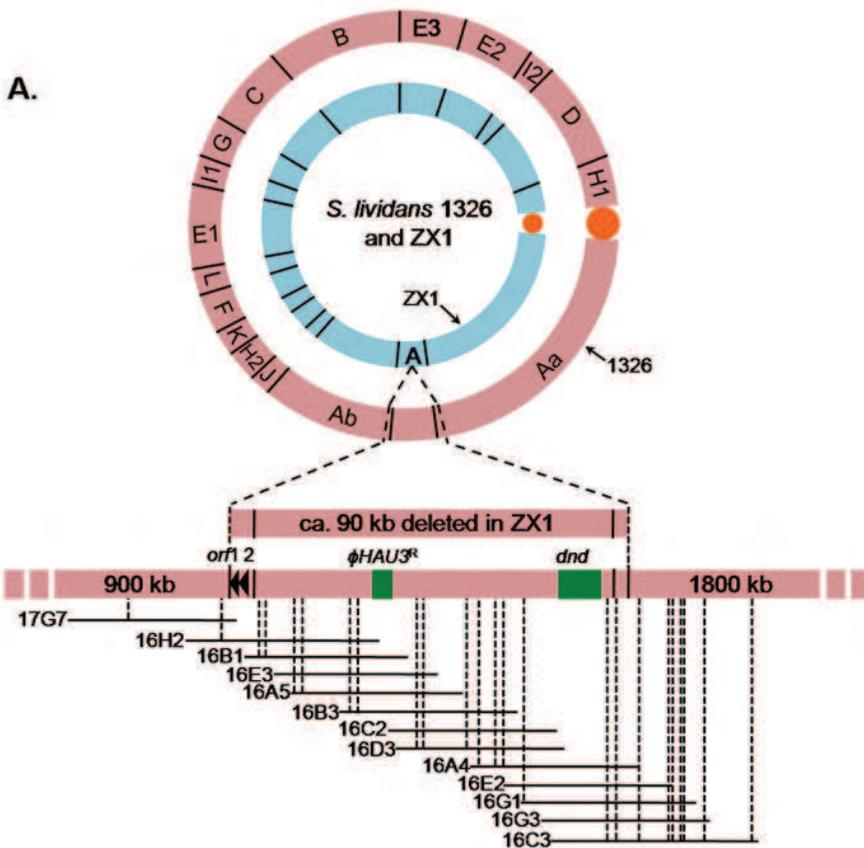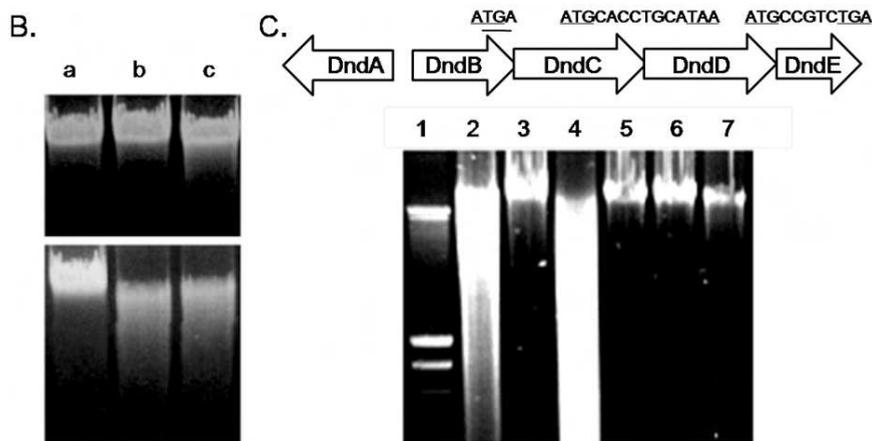
Fig. 2. (A) Physical maps of *S. lividans* 1326 and mutant ZX1. The ca. 90 kb region present in strain 1326 but not in ZX1 is enlarged to show 13 overlapping cosmids [11]. The *dnd* gene cluster and phage ΦHAU3 resistance gene, *ΦHAU3*[R], are shown in green boxes. The positions of two genes immediately flanking the left deletion junction, *orf1* (a P4-like integrase) and *orf2* (a putative transposase) are indicated by black triangles. (B) (top) Thiourea in the electrophoresis buffer can inhibit DNA degradation. (bottom) The Dnd phenotype of ZX1 can be complemented by cosmid 16C3 (lane b) but not 17G7(lane a). Wildtype *S. lividans* 1326 (lane c) is used as a positive control [10]. (C) (top) The five genes *dndABCDE* involved in the DNA phosphorothioate modification in *S. lividans* . (bottom) The disruption of *dndA* (lane 3), *dndC* (lane 5), *dndD* (lane 6) and *dndE* (lane 7) can abolish the Dnd phenotype, whereas the mutation of *dndB* (lane 4) aggravates the degradation. Wildtype *S. lividans* 1326 is used as control (lane 2). Lane 1 is a DNA marker [12]. Figure adapted from [11, 12].

The *dnd* gene cluster in *S. lividans* consists of five genes, *dndABCDE*. *dndBCDE* constitute an operon, which is divergently transcribed from the *dndA* gene (Figure 2C) [12]. The individual disruption of *dndA*, *dndC*, *dndD* or *dndE* abolishes phosphorothioation [10, 12]. *dndA* is predicted to encode a protein of 380 amino acids and homologous to cysteine desulfurase of IscS and NifS proteins in *E. coli.* Purified DndA protein is a pyridoxl 5'-phosphate dependent homodimer and capable of catalyzing L-cysteine to produce elemental S and L-alanine. Cys327 in the C-terminal region of DndA is confirmed to be the active enzymatic center and surrounded by a consensus sequence of ATGSACTS [13]. The mobilized elemental sulfur by DndA could subsequently involve the assembly of a [4Fe-4S] cluster in the DndC protein. DndC possesses ATP pyrophosphatase activity, catalyzing hydrolysis of ATP to AMP and pyrophosphate, and is predicted to have phosphoadenyl sulphate reductase activity [13]. Meanwhile, DndC shares a unique adenylation specific P-loop motif of SGGKDS with SGGFDS in ThiI, an enzyme involved in the formation of 4-thiouridine in tRNAs.

DndD is homologous to the ATP-binding cassette (ABC) ATP-binding proteins and also shares extensive sequence similarity to the Structural Maintenance of Chromosomes (SMC) family of proteins associated with ATPase and DNA nicking activity. In addition, DndD possesses an ATP/GTP-binding Walker A motif (35-GLNGCGKT-42) and an ABC transporter family signature (556-LSAGERQLLAISLLW-570) [10]. Yao *et al.* located an

spfBCDE gene cluster in Dnd-phenotypic *P. fluorescens* Pf0-1, which has an organization identical to that of *dndBCDE* in *S. lividans* 1326. The *spfBCDE* cluster is essential for the Dnd phenotype in *P. fluorescens* Pf0-1, and the putative SpfBCDE proteins exhibit 51%, 49%, 31% and 39% amino acid sequence homology to DndBCDE, respectively. SpfD, a DndD homolog, possesses an ATPase activity of 6.201 ± 0.695 units/mg and is proposed to provide the energy required in DNA phosphorothioation by hydrolyzing ATP [14].

DndE consists of merely 126 amino acids and shows 46% identity to a phosphoribosylaminoimidazole carboxylase (NCAIR synthetase) from *Anabaena variabilis* ATCC 29413 [10]. NCAIR synthetase is known to act at a condensing carboxylation step in purine biosynthesis [15].

Distinct from the others, the disruption of *dndB* does not abolish the Dnd phenotype, but instead it aggravates DNA degradation (Figure 2C). DndB shows 25% identity and 38% similarity to the ABC transporter ATPase from *Sphingomonas* sp. SKA58, and 26% similarity to a DNA gyrase (GyrB) from *Mycoplasma putrefaciens*. It also shows significant amino acid sequence homology to a group of putative transcriptional regulators. A run of 152 residues is 24% identical and 36% similar to the substrate-binding protein of an ABC transporter of *Streptococcus pneumoniae* TIGE4. In addition, it is noticeable that the predicted DndB is likely to be a basic protein (pI: 8.79) under physiological conditions and would conceivably bind nucleic acids to mediate the modification frequency [9].

In the tRNA sulfur modification system, IscS converts L-cysteine to L-alanine and sulfane sulfur in the form of a cysteine persulfide in its active site. The generated sulfane sulfur is sequentially transferred to ThiI to continue catalyzing the biosynthesis of 4-thiouridine [16]. Assembled by five Dnd proteins, the DNA phosphorothioation system appears to be more complicated than the tRNA sulfur modification system in accomplishing the sequence selective and stereo-specific sulfur substitution.

4. Widespread existence of phosphorothioation in bacteria

Homologous *dnd* clusters are found in phylogenetically diverse bacterial species including *Bacillus*, *Klebsiella*, *Enterobacter*, *Mycobacterium*, *Pseudomonas*, *Pseudoalteromonas*, *Roseobacter*, *Mesorhizobium*, *Serratia*, *Acinetobacter*, *Clostridium*, as well as certain archaea, etc. [12]. Moreover, *dnd* gene homologues are also detected in oceanic metagenomes, including the Sargasso Sea, Roca Redonda, the gulf of Mexico, etc. [2]. In some cases, *dndA* is not found adjacent to clustered *dndBCDE*. DndA is homologous to IscS, which usually has more than one copy in a genome. Therefore, an iscS homologue could be elsewhere in genomes and the cognate proteins may have served as functional homologues of DndA.

Apart from bacteria with a sequenced *dnd* cluster, a large part of bacteria not previously known to possess *dnd* clusters display the Dnd phenotype during electrophoresis. A survey on 74 actinomycetal strains from ecologically differentiated regions identified 5 Dnd+ strains [17]. Genomic DNAs from 50% of the total of 69 tested *Mycobacterium abscessus* isolates degraded during pulse-field gel electrophoresis [18].

To investigate the phosphorothioate modification in diverse bacteria, Wang *et al.* developed a highly sensitive liquid chromatography-coupled electrospray ionization tandem quadrupole mass spectrometry technique (LC-MS/MS) that identifies phosphorothioate modifications at dinucleotide level [2]. Due to the specific resistance of the phosphorothioate bond (R_P) to nuclease P1, DNA harboring phosphorothioate sites generates nucleosides and phosphorothioate-linked dinucleotides upon digestion by nuclease P1 followed by

dephosphorylation. As shown in Figure 3, quantification of phosphorothioate dinucleotides can be achieved using the non-physiological S_P stereoisomer of d(G_{PS}A) as an internal standard. This analytical approach makes it feasible to quantitatively screen all 16 phosphorothioate dinucleotides in DNA samples.
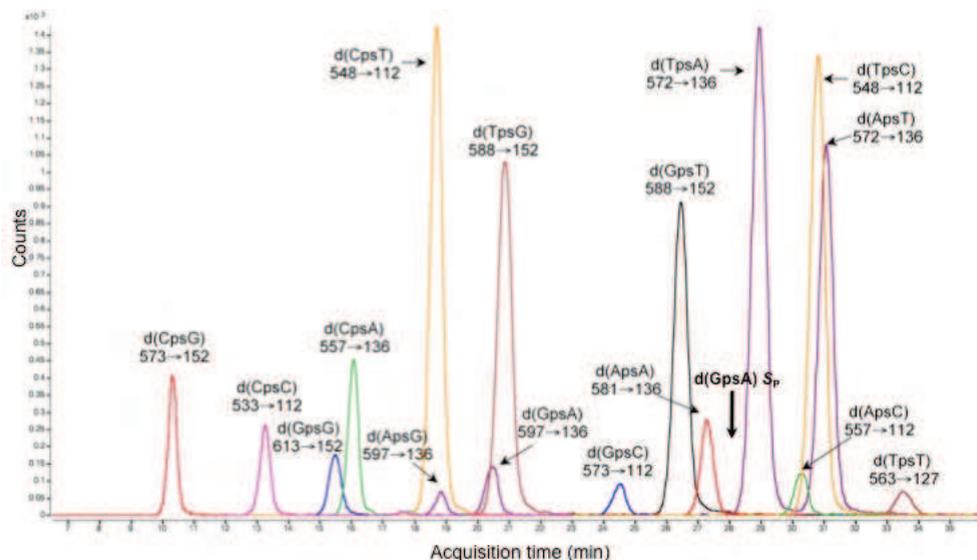

Fig. 3. The LC-MS/MS approach accounting for all 16 phosphorothioate linked dinucleotides in R_P configuration. All of the 16 possible phosphorothioate-linked dinucleotides were resolved by reversed-phase HPLC followed by MS/MS detection in multiple reaction monitoring mode. The ion transitions are labeled under each dinucleotide. Bold arrow indicates the internal standard of d(G_{PS}A) S_P for quantification [2]. Figure adapted from [2].

An extensive study of a collection of bacteria of variable origins and diverse habitats, including marine microbes *Shewanella pealeana* ATCC700345, *Bermanella marisrubri* RED65 and *Hahella chejuensis* KCTC2396, anaerobic *Geobacter uraniumreducens* Rf4, enterotoxigenic *E. coli* B7A and *Salmonella enterica* serovar Cerro 87, and one of the smallest known free-living bacteria *Candidatus* Pelagibacter ubique strain HTCC1002, reveals the common possession of DNA phosphorothioate modifications in these taxonomically unrelated bacterial strains (Table 1). It is conceivable that the *dnd*-associated DNA phosphorothioation is ubiquitous in prokaryotes [2].

The study of representative strains from various habitats, environmental DNA samples and 63 *Vibrio* strains reveals that the phosphorothioate modification occurs in a characteristic manner. In *S. enterica* 87, *E. coli* B7A and *Vibrio* 1F267, the phosphorothioate modification occurs in d(G_{PS}T) and d(G_{PS}A) at the ratio of 1:1. The marine bacteria *B. marisrubri* RED65 and *H. chejuensis* KCTC2396 possess d(G_{PS}A) accompanied by barely detectable d(G_{PS}T). A pair of d(G_{PS}T) and d(G_{PS}G) are simultaneously present in *G. uraniumreducens* Rf4 and *S. lividans* 1326, etc., but at levels that differed by two orders of magnitude. Three phosphorothioate contexts of d(C_{PS}C), d(A_{PS}C) and d(T_{PS}C) occur in *Vibrio* 1C-10, ZF264,

ZF29 and FF75, while the level of total phosphorothioation is almost 10 fold higher than that in other strains (Table 1) [2].

Bacterial strain	Total phosphorothioate (per 10^6 nt)	Phosphorothioate contexts (R_P)	Ratio	Putative 4 bp core sequence
E. coli B7A	768 ± 27	d(G_{PS}A), d(G_{PS}T)	1:1	5′-G_{PS}AAC-3′ 3′-CTT$_{PS}$G-5′
S. enterica 87	732 ± 20			
Vibrio 1F267	576 ± 34			
DH10B(pJTU1980)	1078 ± 109			
DH10B(pJTU1238)	1505 ± 103			
P. fluorescens Pf0-1	451 ± 9	d(G_{PS}G)	-	5′-G_{PS}GCC-3′ 3′-CCG$_{PS}$G-5′
S. lividans 1326	474 ± 39	d(G_{PS}G), d(G_{PS}T)	221:1	5′-G_{PS}GCC-3′ 3′-CCG$_{PS}$G-5′
G. uraniumreducens Rf4	520 ± 13		181:1	
Vibrio ZS139	581 ± 19		26:1	
Vibrio 1F230	400 ± 5		126:1	
B. marisrubri RED65	440 ± 23	d(G_{PS}A), d(G_{PS}T)	165:1	5′-G_{PS}ATC-3′ 3′-CTA$_{PS}$G-5′
H. chejuensis KCTC2396	286 ± 9		-	
S. pealeana ATCC700345	489 ± 11	d(G_{PS}A), d(G_{PS}T)	2:1	5′-G_{PS}AAC-3′ 3′-CTT$_{PS}$G-5′ or 5′-G_{PS}ATC-3′ 3′-CTA$_{PS}$G-5′
Virbrio 1C-10	3110 ± 71	d(C_{PS}C), d(A_{PS}C), d(T_{PS}C)	-	5′-C_{PS}CGG-3′ 3′-GGC$_{PS}$C-5′ or 5′-G_{PS}GCC-3′ 3′-CCG$_{PS}$G-5′
Virbrio ZF264	2270 ± 19			
Virbrio ZF29	2242 ± 57			
Virbrio FF75	2626 ± 22			
d(G_{PS}G) is the only phosphorothioate modification detected in *P. fluorescens* Pf0-1; dash indicates that the low frequency of d(G_{PS}T) in *H. chejuensis* KCTC2396, as well as of d(A_{PS}C) and d(T_{PS}C) in *Vibrio* 1C-10, ZF264, ZF29 and FF75 are far less than the major d(G_{PS}A) and d(C_{PS}C), respectively.				

Table 1. Characteristic phosphorothioate modifications in diverse bacterial strains [2].

Wang *et al*. also analyzed the phosphorothioate modification in the environmental seawater from the Sargasso Sea and Oregon coast, leading to the discovery of phosphorothioate modifications of d(G$_{PS}$A), d(G$_{PS}$T), d(G$_{PS}$G) and d(C$_{PS}$C) in these metagenomes [2]. The Sargasso Sea is a low nutrient, low productivity, subtropical ocean gyre [19]. The oceanic water DNA samples represent microbial communities, including uncultured microbes. Interestingly, phosphorothioates d(C$_{PS}$C) and d(G$_{PS}$A) were found throughout the water columns off the Oregon coast (5–40 m) and in the Sargasso Sea (0–200 m), while d(G$_{PS}$G) was found only in deeper zones of the water column in both locations. More phosphorothioate sequence contexts and frequencies might be explored in the future.

5. Recognition sequences of the phosphorothioate modification

Authentic phosphorothioate modification in *S. lividans* requires not only a conserved consensus sequence but also a considerable flanking sequence with the potential to form secondary structures [8]. The investigations by two laboratories have demonstrated that phosphorothioate site selection requires recognition sequences. Dyson *et al.* and Liang *et al.* performed primer extension and cloning assays on the basis of Tris mediated DNA breakage, respectively, to localize the modification sites in *S. lividans*. At that time, the modification sites were proposed to be on closely opposed guanines on either strand of a stringently conserved 4 bp panlindromic core sequence of 5'-GGCC-3' in a region of 5'-c–cGGCCgccg-3' [8, 9]. It is clear now that the modification sites are actually phosphorothioation between two guanosines on both strands in *S. lividans*. Moreover, both groups confirmed that the phosphorothioate modification in *S. lividans* required a substantial portion of DNA sequences containing three 13 bp direct repeats. The central repeat contains the core sequence, while the left-hand and right-hand copies overlap two potential stem–loop structures (Figure 4). Deletion of either the left or right-hand repeat structures abolishes or alters modification within the core sequence [8].

Quantitative characterization of phosphorothioation in bacterial genomes provides an alternative way to predict the 4 bp core sequence for modification (Table 1). The predominant d(G$_{PS}$G) in *P. fluorescens* Pf0-1, *G. uraniumreducens* Rf4, *Vibrio* ZS139 and *Vibrio* 1F230 suggests a conserved palindromic 5'-G$_{PS}$GCC-3' core sequence as it does in *S. lividans*. The 1:1 ratio of d(G$_{PS}$A) and d(G$_{PS}$T) in *E. coli* B7A, *S. enterica* 87 and *Vibrio* 1F267 suggests the asymmetric complementary 5'-G$_{PS}$AAC-3' and 5'-G$_{PS}$TTC-3' core sequence. d(G$_{PS}$A) is the major phosphorothioation in *B. marisrubri* RED65 and *H. chejuensis* KCTC2396, indicating the 5'-G$_{PS}$ATC-3' core sequence.

The quantification shows that the d(G$_{PS}$G) modification of *S. lividans* occurs at the frequency of 474 ± 39 every 10^6 nt, whereas there are 1.1×10^5 d(GG) available on the chromosome (the genomic sequence of *S. lividans* is not available and the statistical calculation of d(GG) is based on *S. coelicolor*) [2]. Further statistical analysis revealed that even 4 nt 5'-GGCC-3' sequences still occur at too high frequency to serve as the consensus sequence. A 6 nt 5'-cGGCCg-3' with 2 bp extension, however, is more consistent with the phosphorothioate frequency on chromosomes. *E. coli* B7A, *P. fluorescens* Pf0-1, *Vibrio* ZS139 and *B. marisrubri* RED65, etc., have phosphorothioate frequencies close to that of *S. lividans*, indicating the consensus sequence in these bacteria is longer than the proposed 4 bp [2, 8].

Another pattern of phosphorothioate modification is represented by d(C$_{PS}$C) in *Vibrio* 1C-10, ZF264, ZF29 and FF75. It leads to the proposal of 5'-C$_{PS}$CGG-3' or 5'-G$_{PS}$GCC-3' as core sequences. The frequency in genomic DNA, 1 site per 333-500 nt, agrees well with a 4 bp

consensus sequence which has the theoretical frequency of once every 256 bp (4^4). Thus, the four *Vibrio* strains might have recognition mechanisms that are distinct from the former group [2, 8].
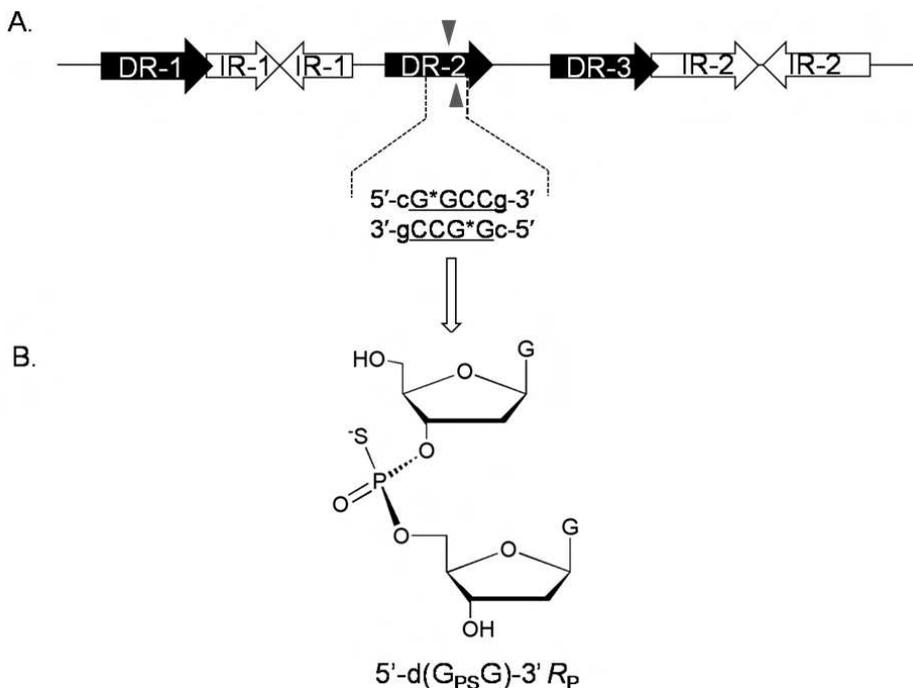

Fig. 4. (A) DNA phosphorothioate modification occurs within a highly conserved 4 bp core sequence, 5'-G$_{PS}$GCC-3' in *S. lividans*. In plasmid pIJ101, the modification sequence lies in the middle (DR-2) of the three direct repeats (DR1–3). The two inverted repeats (IR-1 and IR-2), overlapping direct repeat sequences of DR-1 and DR-3, have the potential to form stem-loop structures. (B) The chemical structure of d(G$_{PS}$G) R_P in *S. lividans* [9, 20]. Figure adapted from [20].

Apart from d(C$_{PS}$C), phosphorothioate modified d(A$_{PS}$C) and d(T$_{PS}$C) co-occur in the four *Vibrio* strains at low frequencies of 1-6 per 10^6 nt (Table 1). A similar situation holds for *S. lividans*, *G. uraniumreducens* Rf4 and *B. marisrubri* RED65, etc., in which low levels of d(G$_{PS}$T) are detected. To explain the low phosphorothioate frequencies, the *dnd* cluster from *S. enterica* 87 was inserted to a low- and high-copy vector of pACYC184 and pBluescript SK+, respectively, generating pJTU1980 and pJTU1238. Both plasmids still confer host *E. coli* DH10B with d(G$_{PS}$T) and d(G$_{PS}$A) modifications in a close 1:1 ratio. Moreover, the total phosphorothioate frequencies on chromosomes of DH10B(pJTU1980) and DH10B (pJTU1238) increased 1.5- and 2-fold in comparison to that of the original host *S. enterica* 87. Remarkably, three more phosphorothioate modifications of d(C$_{PS}$A), d(T$_{PS}$A), and d(A$_{PS}$A) at low levels appeared due to the increased expression of the *dnd* cluster. The low-frequency phosphorothioate modifications might result from relaxed DNA target recognition by Dnd proteins [2].

6. Phosphorothioation dependent restriction-modification system

After unveiling the chemical nature of the DNA phosphorothioate modification, an immediate question is what role this novel post-replicative DNA backbone modification plays. In bacteria, site-specific DNA modifications are often, but not always, associated with a sequence-specific endonuclease. The endonuclease is capable of making subtle distinctions between DNA molecules to prevent the invasion of foreign DNA from phage and plasmids that lack the specific DNA modification. For example, DNA methylation has been regarded as the classic restriction modification system. Because of the known resistance of phosphorothioate linkages to a variety of nuclease activities, as well as the post-replicative and site-specific nature of the modification, phosphorothioation of DNA could possibly function as a type of host defense mechanism, akin to restriction and modification systems [1].

Soon after the chemical nature of the Dnd modification was addressed, the *dnd* cluster was found to constitute a host-specific phosphorothioation-restriction system along with an adjacent *dptFGH* cluster in *S. enterica* 87 [3]. A 15 kb DNA fragment from *S. enterica* 87 conferred both host-specific phosphorothioation (*dptBCDE*) and restriction (d*ptFGH*) in *E. coli*. The two clusters are divergently transcribed. In addition to four phosphorothioation-related genes, three genes are responsible for restriction activity in this DNA fragment, confirmed by gene deletion experiments. With at least 7 genes, phosphorothioation-restriction components seem to form a large complex. The *dptFGH* restricts the invasion of non-phosphorothioate-modified pUC18 but not pUC18 with phosphorothioation. When transformed by pUC18 plasmid, *S. enterica* 87 reproducibly yielded about 100 times fewer colonies with non-phosphorothioate pUC18 than with phosphorothioate pUC18. Plasmids from *E. coli* that had escaped restriction were no longer restricted in *S. enterica* 87 [3]. This observation is similar to the phenomenon leading to the discovery of restriction and modification systems in 1950s [21]. Interestingly, once the modification cluster *dptBCDE* is disrupted, *dptFGH* loses the restriction function. The restriction genes *dptFGH* require the phosphorothioation genes *dptBCDE* to confer the restriction activity of *S. enterica* 87 to ensure that the attack on invasive DNA occurs only when the host DNA is already protected by phosphorothioation [3].

On the basis of subunit composition, sequence recognition and cofactor requirement, the DNA phosphorothioate modification is close to Type I restriction modification systems but far more complicated. Homologous phosphorothioation-restriction genes were identified in 19 diverse bacteria strains (Figure 5), including phosphorothioate tested *E. coli* B7A, *S. pealeana* ATCC700345, *B. marisrubri* RED65, *H. chejuensis* KCTC2396, as well as *E. cereus* E33L, *Vibrio cholera* MZO-2, etc. *E. coli* B7A was confirmed to possess a similar phosphorothioation-restriction system by transformation experiments. Plasmids from *dnd*-XTG102 transformed *E. coli* B7A with 100-fold lower efficiency than phosphorothioate modified plasmid DNA from wild-type *S. enterica* 87. However, which restriction genes are responsible for DNA cleavage site selection and DNA sequence specificity is not clear. Many bacteria possess only the homologous *dnd* cluster without simultaneous *dptFGH* across their genomes, suggesting that the phosphorothioate modification may act not only as a sort of protective system against infection by bacteriophages, but also as an epigenetic signal for new biological function(s) that need to be explored [3].

The quantification of phosphorothioation is also supportive for a restriction-modification system. Analysis of the quantitative data revealed that the levels of phosphorothioation

were classified into three distinct levels: 2-3 per 10^3 nt, 3-8 per 10^4 nt, and 1-6 per 10^6 nt [2]. Along with defined sequence contexts, the first two frequency ranges are consistent with a restriction-modification system with a 4-nt or 5-6 nt consensus sequence, respectively [22].
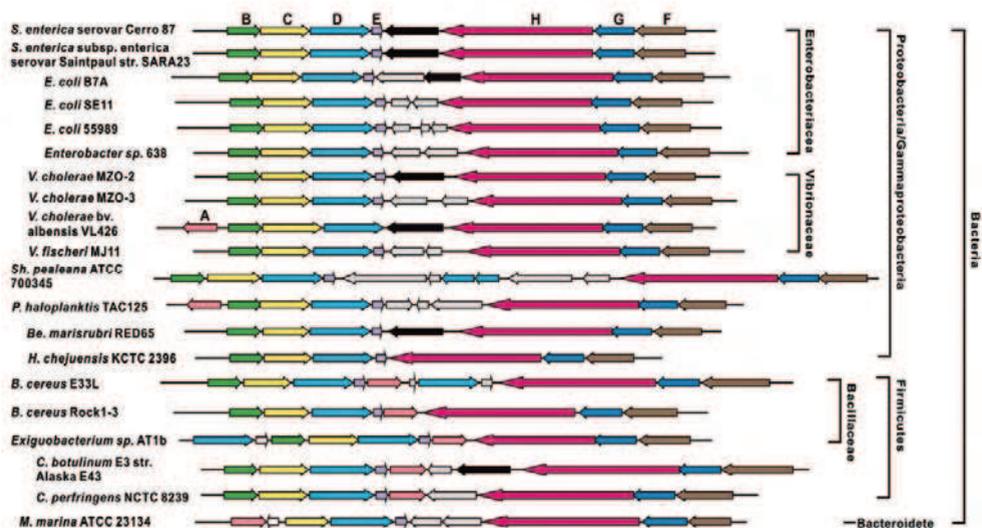

Fig. 5. Alignment of phosphorothioate modification (*dptABCDE*) and restriction (*dptFGH*) gene clusters from 20 bacterial strains. Colored arrows indicate homologous ORFs. Light gray arrows represent diverse ORFs without predicted functions that are not homologous to each other [3]. Figure adapted from [3].

7. Cleavage of phosphorothioate DNA by type IV restriction endonuclease ScoMcrA

When the bi-functional plasmid pIJ699 was isolated from *S. lividans* and *E. coli*, only pIJ699 from *S. lividans* degraded during electrophoresis, indicating that the Dnd phenotype selectively occurred in certain bacteria [2]. This is consistent with the observation that *dnd* is not present in *S. coelicolor* but is in *S. lividans*, although chromosomes of the two strains share an almost identical DNA banding pattern upon enzymatic digestion [5, 6]. Most of the *S. coelicolor* and *S. lividans* DNA sequenced is similar or even identical. Interestingly, a type IV restriction endonuclease (ScoA3McrA) coded by gene *sco4631* in *S. coelicolor* cuts foreign DNA containing phosphorothioates. The search for a phosphorothioate-cutting enzyme in *S. coelicolor* originated from the restriction to the *dnd* gene cluster. Liu *et al.* tried to introduce the *dnd* gene cluster from *S. lividans* into its close relative *S. coelicolor*. However, they unexpectedly failed, while the same gene cluster with a single base insertion for a frame-shift mutation in *dndE* gene generated exoconjugants [4]. This implied restriction towards phosphorothioate modification by *S. coelicolor*.

Comparison between the genome sequence of *S. coelicolor* and the *dnd+* of *S. avermitilis*, revealed an endonuclease ScoA3McrA in *S. coelicolor* that is absent in *S. avermitilis*. *S. coelicolor* lost its restriction to the *dnd* gene cluster after disruption of ScoA3McrA. After

integration of a vector containing ScoA3McrA into the genome, a *dnd* mutant *S. lividans* HXY6 confers restriction toward the *dnd* gene cluster. These knock-out and knock-in experiments confirmed the role of ScoA3McrA as the determinant of restriction of phosphorothioate in *S. coelicolor*. Moreover, *in vitro* in presence of Mn^{2+} and Co^{2+}, the purified ScoA3McrA protein cleaved *in vivo* phosphorothioated DNA as well as a synthesized 118 bp double strand DNA oligonucleotide bearing one phosphorothioate on each strand. ScoA3McrA specifically cleaves both the top and the bottom strand, and on both sides of the S-modification at multiple cleavage sites 16-28 nt away from the phosphorothioate sites. Liu *et al.* proposed that expression of the *dndA-E* gene cluster in *S. coelicolor* resulted in phosphorothioation of the host DNA. ScoA3McrA would then cleave the phosphorothioated host DNA near the modified sites and result in cell death as a cell suicide process. Purified ScoMcrA also cleaved Dcm-methylated DNA or Dcm-containing oligos 12-16 bp away from a C5mCWGG Dcm methylation site [4]. ScoA3McrA thus builds an interesting link between phosphorothioation and methylation.

8. Phylogenetic relationship and evolutionary path of *dnd* genes

The phylogeny of Dnd from 12 bacteria shows strong correlation between phosphorothioate modifications and four Dnd proteins (Figure 6). With the exception of *Candidatus* Pelagibacter ubique, the other 11 strains are well classified based on DNA phosphorothioate sequence contexts and frequencies. Results suggest the diversification of DNA phosphorothioate modifications depends on Dnd protein sequence but not on the phylogenetic descent of the bacteria strains. Furthermore, using phylogenetic analysis based on Dnd proteins and 16S rRNA, Wang *et al.* found the Dnd phylogenies do not follow their corresponding species tree. This is clearly seen in three *Vibrio* isolates (ZS139, 1F230, and 1F267) which are phylogenetically incoherent in all four DndBCDE proteins. The phylogenetic differentiation of the *Vibrio* isolates suggests horizontal gene transfer of *dnd* clusters facilitated by genomic islands in evolution [2, 17].

Sequence analysis reveals that the ca. 90 kb fragment containing the *dnd* cluster in *S. lividans* is indeed a genomic island with precise length of 92,770 bp [17]. The G+C content of the genomic island is 67.8%, lower than the average for *S. coelicolor* of 72.1%, indicating the *dnd* system may have originated from elsewhere. Genomic islands are discrete DNA segments, which differ among closely related strains. It explains why the *dnd* cluster occurs in *S. lividans* and *S. avermitilis*, but not in *S. coelicolor*, a close relative of *S. lividans* even at genomic sequence level. Genomic islands play a role in the evolution, diversification and adaption of microbes as they are involved in the dissemination of variable genes, including antibiotic resistance and virulence genes, as well as catabolic genes [23].

Active genomic island transfer has been reported in some cases. For instance, the PAPI-1 pathogenicity island in *P. aeruginosa* was shown to transfer from a donor strain into *P. aeruginosa* strains [24]. ICE*Hin1056*, an integrative and conjugative element from *Haemophilus influenzae*, proceeds conjugative transfer between two *H. influenzae* strains. Moreover, ICE*clc* of *Pseudomonas* sp. strain B13 can self-transfer to *P. putida*, *Cupriviadus necator* or *P. aeruginosa* at similar frequencies [25]. He *et al.* demonstrated that the 93 kb genomic island in *S. lividans* was capable of spontaneous excision from the chromosome at a level of 0.016%-0.027%. However, exposure to MNNG (*N*-methyL-*N'*-nitro-*N*-nitrosoguanidine) can increase the excision frequency by at least five fold. The excised island loses its capabilities of inter and even intra-species transmission between *Streptomyces*

strains. This genomic island may have lost genes required for its transfer during evolution in order to maintain a relatively stable inheritance with the host [17].
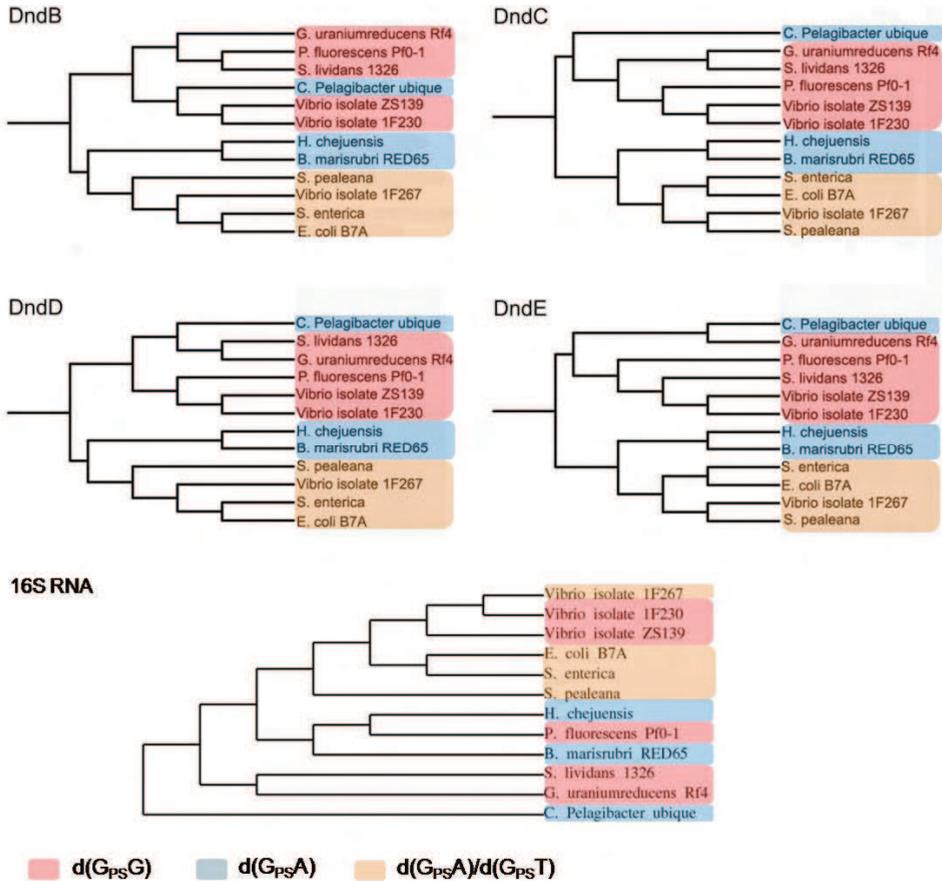

Fig. 6. Correlation between DNA phosphorothioate sequence contexts and Dnd proteins. DNA phopshorothioate modification follows the Dnd protein phylogenies but not species phylogenies (16S RNA tree), supporting horizontal rather than vertical gene transfer for *dnd* genes [2].

Besides *S. lividans*, 11 additional *dnd+* bacteria were analyzed by He *et al.* Remarkably, all *dnd* clusters lie on mobile genetic elements based on the characteristic features of G+C content, dinucleotide bias, direct repeats, and possession of intergrase and/or transposase (Figure 7). Ten of them lie within chromosomal genomic islands and one on a large plasmid. This indicates the dissemination of *dnd* genes in evolution and explains the ubiquitous occurrence of *dnd* clusters in taxonomically unrelated bacteria. It is still unclear how the *dnd* clusters evolved and disseminated across different bacterial species. He *et al.* suggested that the *dnd* cluster might be organized into a functional locus on a conjugative plasmid or other mobile element in very ancient times followed by extensive dissemination and diversification over the eons [17].
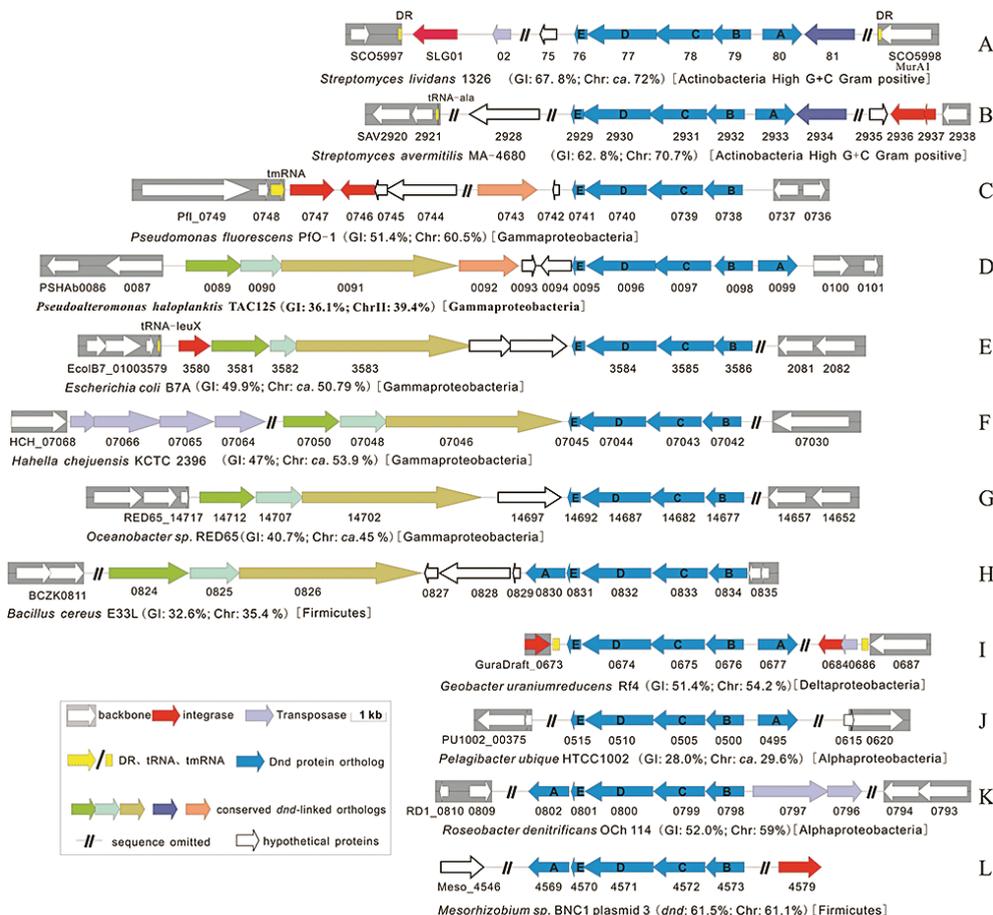
Fig. 7. Twelve *dnd* gene clusters on mobile elements. Blue arrows represent the *dnd* gene homologues. The characteristic elements of genomic islands, including integrase, transposase, direct repeats, insertion hotspots of tRNA, tmRNA sites are shown in red, purple and yellow colors [17]. Figure adapted from [17].

Ou *et al.* organized available data from experimental and bioinformatics analysis of the DNA phosphorothioation to assemble a *dndDB* database [26]. It contains detailed phosphorothioation-related information including the Dnd phenotype, *dnd* gene clusters, genomic islands harboring *dnd* genes, and Dnd proteins and conserved domains. The *dndDB* database provides a useful tool to effectively combine and interlink the genetics, biochemistry and functional aspects of *dnd* systems and related genomic islands.

9. Discussion

Chemically synthesized phosphorothioate internucleotide bonds had been in use for decades prior to the discovery of the physiological phosphorothioate modification in bacteria. Enzymes like snake venom phosphodiesterase, nuclease S1 and nuclease P1

recognize R_P and S_P phosphorothioate isomers differently, hydrolyzing one isomer more efficiently than the other. Therefore, phosphorothioate isomers have been utilized widely to elucidate the stereochemical action of different enzymes [27]. Other nucleases such as DNase I, DNase II, staphylococcal nuclease and spleen phosphodiesterase are unable to hydrolyse the internucleotidic linkage of either phosphorothioate diastereomer [28]. The significantly increased resistance of phosphorothioate linkage to nuclease hydrolysis inspired the extensive application of phosphorothioate oligonucleotide analogues in antisense therapy to treat a broad range of diseases, including viral infections, cancer and inflammatory diseases. It has been more than a decade since the approval of the first antisense drug Vitravene in 1998 by the FDA. The synthetic 21-mer oligonucleotide with phosphorothioate linkage is used in the treatment of cytomegalovirus retinitis (CMV) in immunocompromised patients, including those with AIDS [29].

Phosphorothioates can be introduced into oligonucleotides and DNA by both chemical synthesis and enzymatic polymerization. Currently, phosphorothioate-modified oligonucleotides are available via the oxathiaphospholane method, in which nucleoside 3'-O-(2-thio-1,3,2-oxathiaphospholane) derivatives are used as monomers. This method generates a racemic mixture of R_P and S_P stereoisomers at a close 1:1 ratio. The access to stereospecific phosphorothioate bearing oligonucleotides is still severely limited despite the considerable efforts that have been made [30]. Interestingly, the S_P diastereomer of dNTPαS can be accepted as a substrate by *E. coli* DNA polymerase I, and may be employed in polymerization reactions to produce phosphorothioate linkages of the R_P configuration [31]. This is consistent with the physiological configuration of phosphorothioation. However, the phosphorothioate modification modified by the *dnd* genes is post-replicative, requiring conserved core sequences and flanking sequences.

The desulfurization of phosphorothioate to a phosphate bond is an easy process. However, the reverse phosphorothioation is thought to be energetically uphill. It agrees well with the observed role of DndD which acts as an ATPase. Friz Eckstein proposed that the phosphorothioate modification might first require the activation of target phosphodiester bonds by alkylation, acylation, adenylation or phosphorylation followed by the successive substitution by a nucleophilic sulfur [32]. It is still unclear how the five Dnd proteins cooperate together to use L-cysteine and SO_4^{2-} as sulfur sources and transfer the sulfur to the DNA backbone sequence selectively and R_P specifically. Although the biochemical activity of several Dnd proteins has been assayed, additional insights are still needed to elucidate the role of each Dnd protein in the DNA phosphorothioation pathway and the interaction between Dnd proteins and target DNA regions. The *dnd* gene cluster is widespread in diverse and distantly related bacteria, however, a complete set of *dnd* homologs has not yet been found in eukaryotes.

Most of the commonly found structural changes in DNA are due to methylation of particular bases. In some viral DNAs, certain bases may be hydroxymethylated or glucosylated [33-35]. DNA phosphorothioation apparently is an unprecedented physiological modification, which renders DNA susceptible to Tris derivative leading to the characteristic Dnd phenotype.

It is such a surprise to find out that nature can synthesize a phosphorothioate-containing DNA backbone using the *dndABCDE* genes. Particularly, it is interesting that the modification occurs in a sequence-selective and stereo-specific manner. The discovery of physiological DNA phosphorothioation has revolutionized our view of the composition and

structure of DNA, opening a new window that will stimulate research into novel aspects of DNA.

10. References

[1] Wang, L., et al., *Phosphorothioation of DNA in bacteria by dnd genes.* Nat Chem Biol, 2007. 3(11): p. 709-10.

[2] Wang, L., et al., *DNA phosphorothioation is widespread and quantized in bacterial genomes.* Proc Natl Acad Sci U S A. 108(7): p. 2963-8.

[3] Xu, T., et al., *A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella.* Nucleic Acids Res. 38(20): p. 7133-41.

[4] Liu, G., et al., *Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA.* PLoS Genet. 6(12): p. e1001253.

[5] Zhou, X., et al., *Site-specific degradation of Streptomyces lividans DNA during electrophoresis in buffers contaminated with ferrous iron.* Nucleic Acids Res, 1988. 16(10): p. 4341-52.

[6] Evans, M., et al., *Characterization of a Streptomyces-lividans-type site-specific DNA modification system in the avermectin-producer Streptomyces avermitilis permits investigation of two novel giant linear plasmids, pSA1 and pSA2.* Microbiology, 1994. 140 (Pt 6): p. 1367-71.

[7] Ray, T., A. Mills, and P. Dyson, *Tris-dependent oxidative DNA strand scission during electrophoresis.* Electrophoresis, 1995. 16(6): p. 888-94.

[8] Dyson, P. and M. Evans, *Novel post-replicative DNA modification in Streptomyces: analysis of the preferred modification site of plasmid pIJ101.* Nucleic Acids Res, 1998. 26(5): p. 1248-53.

[9] Liang, J., et al., *DNA modification by sulfur: analysis of the sequence recognition specificity surrounding the modification sites.* Nucleic Acids Res, 2007. 35(9): p. 2944-54.

[10] Zhou, X., et al., *A novel DNA modification by sulphur.* Mol Microbiol, 2005. 57(5): p. 1428-38.

[11] Zhou, X., et al., *Streptomyces coelicolor A3(2) lacks a genomic island present in the chromosome of Streptomyces lividans 66.* Appl Environ Microbiol, 2004. 70(12): p. 7110-8.

[12] Xu, T., et al., *DNA phosphorothioation in Streptomyces lividans: mutational analysis of the dnd locus.* BMC Microbiol, 2009. 9: p. 41.

[13] You, D., et al., *A novel DNA modification by sulfur: DndA is a NifS-like cysteine desulfurase capable of assembling DndC as an iron-sulfur cluster protein in Streptomyces lividans.* Biochemistry, 2007. 46(20): p. 6126-33.

[14] Yao, F., et al., *Functional analysis of spfD gene involved in DNA phosphorothioation in Pseudomonas fluorescens Pf0-1.* FEBS Lett, 2009. 583(4): p. 729-33.

[15] Nakamura, Y., et al., *Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1.* DNA Res, 2002. 9(4): p. 123-30.

[16] Kambampati, R. and C.T. Lauhon, *IscS is a sulfurtransferase for the in vitro biosynthesis of 4-thiouridine in Escherichia coli tRNA.* Biochemistry, 1999. 38(50): p. 16561-8.

[17] He, X., et al., *Analysis of a genomic island housing genes for DNA S-modification system in Streptomyces lividans 66 and its counterparts in other distantly related bacteria.* Mol Microbiol, 2007. 65(4): p. 1034-48.

[18] Zhang, Y., et al., *Pulsed-field gel electrophoresis study of Mycobacterium abscessus isolates previously affected by DNA degradation.* J Clin Microbiol, 2004. 42(12): p. 5582-7.

[19] Morris, R.M., et al., *SAR11 clade dominates ocean surface bacterioplankton communities.* Nature, 2002. 420(6917): p. 806-10.

[20] Chen, S., L. Wang, and Z. Deng, *Twenty years hunting for sulfur in DNA.* Protein Cell. 1(1): p. 14-21.

[21] Boyer, H.W., *DNA restriction and modification mechanisms in bacteria.* Annu Rev Microbiol, 1971. 25: p. 153-76.

[22] Wilson, G.G. and N.E. Murray, *Restriction and modification systems.* Annu Rev Genet, 1991. 25: p. 585-627.

[23] Juhas, M., et al., *Genomic islands: tools of bacterial horizontal gene transfer and evolution.* FEMS Microbiol Rev, 2009. 33(2): p. 376-93.

[24] Qiu, X., A.U. Gurkar, and S. Lory, *Interstrain transfer of the large pathogenicity island (PAPI-1) of Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2006. 103(52): p. 19830-5.

[25] Gaillard, M., et al., *The clc element of Pseudomonas sp. strain B13, a genomic island with various catabolic properties.* J Bacteriol, 2006. 188(5): p. 1999-2013.

[26] Ou, H.Y., et al., *dndDB: a database focused on phosphorothioation of the DNA backbone.* PLoS One, 2009. 4(4): p. e5132.

[27] Sayers, J.R., et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide.* Nucleic Acids Res, 1988. 16(3): p. 803-14.

[28] Eckstein, F., *Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them?* Antisense Nucleic Acid Drug Dev, 2000. 10(2): p. 117-21.

[29] Marwick, C., *First "antisense" drug will treat CMV retinitis.* JAMA, 1998. 280(10): p. 871.

[30] Oka, N., et al., *Solid-phase synthesis of stereoregular oligodeoxyribonucleoside phosphorothioates using bicyclic oxazaphospholidine derivatives as monomer units.* J Am Chem Soc, 2008. 130(47): p. 16031-7.

[31] Eckstein, F., *Nucleoside phosphorothioates.* Annu Rev Biochem, 1985. 54: p. 367-402.

[32] Eckstein, F., *Phosphorothioation of DNA in bacteria.* Nat Chem Biol, 2007. 3(11): p. 689-90.

[33] Hattman, S., *Unusual modification of bacteriophage Mu DNA.* J Virol, 1979. 32(2): p. 468-75.

[34] Hattman, S., *Specificity of the bacteriophage Mu mom+ -controlled DNA modification.* J Virol, 1980. 34(1): p. 277-9.

[35] Swinton, D., et al., *Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function.* Proc Natl Acad Sci U S A, 1983. 80(24): p. 7400-4.

DNA Replication-Current Advances

Edited by Dr Herve Seligmann

The study of DNA advanced human knowledge in a way comparable to the major theories in physics, surpassed only by discoveries such as fire or the number zero. However, it also created conceptual shortcuts, beliefs and misunderstandings that obscure the natural phenomena, hindering its better understanding. The deep conviction that no human knowledge is perfect, but only perfectible, should function as a fair safeguard against scientific dogmatism and enable open discussion. With this aim, this book will offer to its readers 30 chapters on current trends in the field of DNA replication. As several contributions in this book show, the study of DNA will continue for a while to be a leading front of scientific activities.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Lianrong Wang, Shi Chen and Zixin Deng (2011). Phosphorothioation: An Unusual Post-Replicative Modification on the DNA Backbone, DNA Replication-Current Advances, Dr Herve Seligmann (Ed.), ISBN: 978-953-307-593-8, InTech, Available from: http://www.intechopen.com/books/dna-replication-current-advances/phosphorothioation-an-unusual-post-replicative-modification-on-the-dna-backbone

INTECH
open science | open minds