

DNA modification by sulfur: analysis of the sequence recognition specificity surrounding the modification sites

Jingdan Liang, Zhijun Wang, Xinyi He, Jialiang Li, Xiufen Zhou and Zixin Deng*

Laboratory of Microbial Metabolism and School of Life Science & Biotechnology, Shanghai Jiaotong University, Shanghai 200030, China

Received February 15, 2007; Revised March 9, 2007; Accepted March 12, 2007

ABSTRACT

The Dnd (DNA degradation) phenotype, reflecting a novel DNA modification by sulfur in *Streptomyces lividans* 1326, was strongly aggravated when one (*dndB*) of the five genes (*dndABCDE*) controlling it was mutated. Electrophoretic banding patterns of a plasmid (pHZ209), reflecting DNA degradation, displayed a clear change from a preferential modification site in strain 1326 to more random modifications in the mutant. Fourteen randomly modifiable sites on pHZ209 were localized, and each seemed to be able to be modified only once. Residues in a region (5'-c-cGGCCgccc-3') including a highly conserved 4-bp central core (5'-GGCC-3') in a well-documented preferential modification site were assessed for their necessity by site-directed mutagenesis. While the central core (GGCC) was found to be stringently required in 1326 and in the mutant, 'gccc' flanking its right could either abolish or reduce the modification frequency only in the mutant, and two separate nucleotides to the left had no dramatic effect. The lack of essentiality of DndB for S-modification suggests that it might only be required for enhancing or stabilizing the activity of a protein complex at the required preferential modification site, or resolving secondary structures flanking the modifiable site(s), known to constitute an obstacle for efficient modification.

INTRODUCTION

A novel DNA sulfur modification was found in the soil-dwelling, antibiotic-producing bacteria *Streptomyces lividans* and *Streptomyces avermitilis*. Three distinctive features were associated with the modification: (1) Tris-dependent DNA degradation occurred during

normal or pulse-field gel electrophoresis (the Dnd phenotype); (2) the Dnd phenotype did not seem to be inhibited by treatment with formaldehyde, proteinase K and/or lysozyme; and (3) the Dnd phenotype could be repressed if the electrophoretic buffer was supplemented with a small amount of reducing agents, especially those containing sulfur, like thiourea, or if Tris was replaced by Hepes in the electrophoresis buffer (1–6). The Dnd phenotype and its inhibition in the presence of thiourea were not limited to the above-mentioned two *Streptomyces* strains, but found in a large array of gram-positive and -negative bacteria such as *Mycobacterium abscessus* (7), *Salmonella* spp. (1,8–10), *E. coli* strains (1,9,10), *Pseudomonas fluorescences*, *Pseudomonas aeruginosa* (11), *Clostridium difficile* (12,13), *Klebsiella pneumoniae*, *Enterobacter cloacae* and *Serratia marcescens* (10), although in most cases the phenomenon has been interpreted as nuclease contamination in the DNA samples. Gene clusters rendering Dnd phenotype had been identified in *S. lividans*, *S. avermitilis*, and many of the above-mentioned microorganisms (4, and X. He *et al.*, submitted for publication).

The DNA sulfur modification is site specific, as well as post-replicative because single-stranded plasmid replication intermediates do not seem to contain the modification sites, as evidenced from primer extension experiments (5,14). The dsDNA modifications occur at the first guanine residues within a 6-bp palindromic core sequence of cGgccc, as elegantly identified by Dyson *et al.* (14,15). Three independent pairs of a 13-bp direct repeat flanking both sides of the modified site were also recognized to be important for the modification process. Deletion of either the leftmost or the rightmost repeat abolished the modification, whose site lies within the central repeat sequence, while deletion of the left-hand repeat changed the modification to other sites on plasmid pIJ101 (3,14,15).

Recently, we reported that the *dnd* locus (Figure 1A) of *S. lividans* 1326 is directly involved in the DNA

*To whom correspondence should be addressed. Tel: +86 21 6293 3404; Fax: +86 21 6293 2418; Email: zxdeng@sjtu.edu.cn
The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

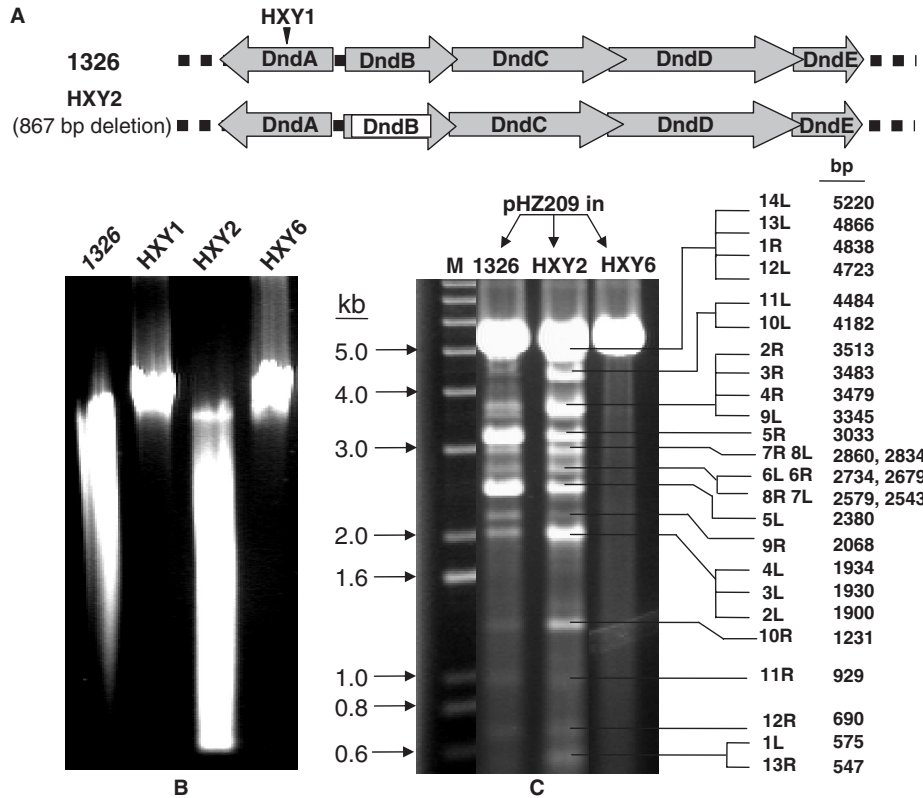


Figure 1. (A) Organization of the *dnd* gene cluster and in-frame deletion of *dndB*. (B) Enhanced Dnd phenotype displayed by *dndB*⁻ mutant HXY2 as compared with that of wild-type 1326. Lack of Dnd phenotype displayed by *dndA*⁻ mutant HXY1 and *dnd*-deletion mutant HXY6 is shown as controls. (C) Preferential cleavage at specific site of EcoRV-linearized pHZ209 originating from 1326, but more random and lack of cleavage of the same DNA originating from HXY2 and HXY6, respectively. All the samples were treated with activated Tris-buffer followed by gel electrophoresis. The designation of the cleaved fragments conforms to those schematically outlined in Figure 3, with their sizes shown. DNA size markers are labeled 'M'. Lagging of the DNA fragments behind the DNA marker could be seen as a result of overloading of the DNA samples.

modification process that incorporates sulfur into specific sites of DNA (4), although the precise chemical nature of the modification remains obscure. Experiments including the use of gene disruption and gain and/or loss of function techniques showed that the five genes in the *dnd* locus are all essential for the DNA sulfur modification. Of these, DndA, DndC, DndD and DndE showed remarkable similarities to cysteine desulfurase, PAPS reductase, chromosome segregation protein and phosphoribosylaminoimidazole carboxylase, respectively. While DndB showed 25% identity and 38% similarity to ABC transporter ATPase from *Sphingomonas* sp. SKA58 (ZP_01303985), and 26% similarity to a DNA gyrase (GyrB) from *Mycoplasma putrefaciens* it is also noticeable that the predicted DndB is likely to be a basic protein (pI: 8.79) under physiological conditions and would conceivably bind nucleic acids. Indeed, this feature was implied by the fact that the Dnd phenotype was significantly aggravated in a *dndB*⁻ mutant, reflecting a relaxed preference for modification sites on DNA. Here, we describe an altered Dnd phenotype in a *dndB*⁻ mutant, and its subsequent use for the convenient and detailed characterization of the consensus sequence on a pIJ101-derived plasmid, pHZ209, containing a well-characterized preferential modification site. This resulted

in a clear demonstration that a consensus sequence with a 4-bp central core is highly conserved and required for modification, but the requirement of its flanking sequences differs remarkably in DndB-deficient strain, HXY2. DndB was implicated as an important mediator for the efficiency of DNA modification.

MATERIALS AND METHODS

Bacterial strains and plasmids

Streptomyces and *E. coli* strains used in this study are listed in Table 1.

General growth conditions and genetic manipulations for *Streptomyces* and *E. coli*

Streptomyces lividans 1326 and its derivatives were grown at 30°C in TSBY liquid medium containing 34% sucrose and SFM agar medium as described in (16). For the plasmid-containing strains, apramycin or thiostrepton were added when needed at a concentration of 30 and 10 µg/ml, respectively. Qiagen Max Plasmid Kits were used to extract plasmid DNA from *Streptomyces*. Protoplast transformation was described in detail in (16).

Table 1. Strains and plasmids used in this study

Strain and plasmid	Description	Source
Strain		
<i>S. lividans</i> 1326	Wild-type, <i>dnd</i> ⁺ , SLP2, SLP3	
HXY1	1326 derivative, <i>dndA</i> disruption mutant	(4)
HXY2	1326 derivative with in-frame deletion of <i>dndB</i>	This work
HXY6	1326 derivative with complete deletion of <i>dnd</i>	This work
ZX1	JT46 derivative with complete deletion of a ~93-kb genomic island including <i>dnd</i>	(4,21)
LJD3	HXY2 derivative with an integrative pSET152, <i>dndB</i> ⁺ , <i>aac(3)IV</i> , <i>tsr</i>	This work
LJD1	HXY2 derivative with an integrative pJTU2020, <i>dndB</i> ⁺ , <i>aac(3)IV</i> , <i>tsr</i>	This work
LJD2	HXY2 derivative with an integrative pJTU1904, <i>dndB</i> ⁺ , <i>aac(3)IV</i> , <i>tsr</i>	This work
DH10B	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) ϕ80dlacZΔM15 Δ <i>lacX74 deoR recA1 endA1 araD139</i> Δ(<i>ara, leu</i>)7697 <i>galU galK λ⁻ rpsL nupG dam dcm hsdS</i> pUZ8002	GIBCO BRL
ET12567 (pUZ8002)		(23)
Plasmid		
pHZ868	pHZ132 derivative carrying the complete <i>dnd</i> gene cluster, <i>bla</i> , <i>tsr</i> , <i>neo</i> , <i>oriT</i> , <i>cos</i> , <i>dnd</i> ⁺ , Amp ^r , Kan ^r	This work
pOJ260	<i>aac(3)IV</i> , <i>oriT</i> , CoEI, Ap ^r	(24)
pJTU155	A 1971-bp XhoI fragment covering part of the C-terminus of <i>dndB</i> gene was cloned into the Sall site of pOJ260. <i>aac(3)IV</i> , Ap ^r	This work
pJTU156	A 1674-bp BamHI-EcoRI fragment internal to a PCR-amplified product covering part of the N-terminus of <i>dndB</i> gene was cloned into the corresponding site of pJTU155, <i>aac(3)IV</i> , Ap ^r	This work
pJTU164	A 1.2-kb fragment carrying <i>oriT</i> inserted into the SpeI site of pJTU156, <i>aac(3)IV</i> , <i>oriT</i> , Ap ^r	This work
pBlueScript II SK(+)	<i>Bla</i> , <i>lacZ</i> , <i>ori</i> ^{T1}	Stratagene
pHZ209	A pIJ101 derivative, <i>tsr</i> , <i>oriT</i> , <i>ori</i> ^{pIJ101}	This work
pSET152	<i>E. coli</i> vector, <i>aac(3)IV</i> , <i>lacZ</i> , CoEI, <i>att</i> ^{ϕC31} , <i>ori T</i> , Ap ^r	(24)
pHZ1904	pSET152 derivative with 8.3-kb DNA carrying the entire <i>dnd</i> gene cluster, <i>aac(3)IV</i> , CoEI, <i>att</i> ^{ϕC31} , <i>ori T</i> , Ap ^r	This work
pJTU2020	pSET152 derivative, <i>dndB</i> ⁺ , <i>aac3(IV)</i> , CoEI, <i>att</i> ^{ϕC31} , <i>ori T</i> , Ap ^r	This work
pGEM-T	<i>E. coli</i> plasmid for DNA sequencing using T7 primer, <i>bla</i> , <i>lacZ</i> , Amp ^r	Promega
pHZ209 derivatives with mutation at specified sites		
pJTU2031	C mutated to A at 2374	This work
pJTU2025	C mutated to A at 2377	This work
pJTU2017	G mutated to A at 2378	This work
pJTU2005	G mutated to A at 2379	This work
pJTU2007	C mutated to A at 2380	This work
pJTU2009	C mutated to A at 2381	This work
pJTU2011	G mutated to A at 2382	This work
pJTU2013	C mutated to A at 2383	This work
pJTU2003	C mutated to A at 2384	This work
pJTU2015	G mutated to A at 2385	This work

pHZ209, and the mutated pHZ209 derivative plasmids were introduced by transformation into 1326, HXY2, HXY6, LJD1, LJD2 and LJD3 by protoplast transformation as described in (16). Isolated DNA (16) was generally stored in autoclaved TE buffer before use.

Cloning of cleaved DNA fragments of pHZ209 after treatment with activated Tris buffer

Escherichia coli cultivation and plasmid manipulation were according to (17). DH10B was used as general *E. coli* host for plasmid transformation.

DNA fragments of 2.4 and 3.0 kb recovered from agarose gels were blunt-ended with Taq DNA polymerase with an extension temperature at 72°C and ligated into pMD18-T. The resultant plasmids carrying the inserts were sequenced by Biosasia Co., Ltd, Shanghai.

The efficiency of cloning the Taq DNA polymerase blunt-ended DNA fragments into pMD18-T using the above method was low. Thus, cloning of the cleaved DNA

fragments after treatment with activated Tris-buffer was modified as follows: SmaI-digested pJTU2018 (Table 1) was dephosphorylated with CIP before a 2926-bp fragment was purified from an agarose gel and used as a cloning vector. EcoRV-linearized pHZ209 from *S. lividans* 1326 or HXY2 was cleaved by activated Tris-buffer prepared as described below. The cleaved plasmids were precipitated with ethanol and re-suspended in water before treatment with T4 polynucleotide kinase (MBI) according to the company's protocol. The T4 polynucleotide kinase was inactivated by treatment at 70°C for 10 min. Forty units of ExoIII/μg DNA was added to trim the 3' end for 2 min at room temperature before its deactivation at 70°C for 10 min. After re-precipitation with ethanol, the DNA fragments were blunt-ended with KOD DNA pol (Takara) at 68°C for 30 min. DNA fragments extracted with phenol/chloroform and precipitated with ethanol were then ligated with the above-mentioned blunt-ended and dephosphorylated vector

DNA from pJTU2018. The desired plasmids were sequenced by Shanghai DNA Biotechnologies Co., Ltd.

Cleavage of pHZ209 and its mutant derivatives by activated Tris-buffer

pHZ209 and its mutant derivatives were extracted from the wild-type *S. lividans* 1326, or its mutants HXY2 and HXY6, and linearized with EcoRV. Activated Tris-buffer (electrophoresed TAE buffer) for *in vitro* DNA cleavage at the S-modification sites was prepared as described in (3) with slight modifications: the 1× TAE buffer was kept running at a constant 80 V at room temperature, with a total running time of 8 h in an electrophoresis tank.

Five hundred and ninety microliters of buffer was sampled near the anode (3) and added to a 10 µl sample containing 500 ng linearized pHZ209, before 4–15-h incubation at 37°C for adequate cleavage. The resulting samples were precipitated, dried, re-suspended in 20 µl TE buffer and loaded onto a 1% agarose gel for electrophoresis before desired fragments were recovered. The relative DNA band intensities were analyzed using Quantity One software (Bio-Rad). Sequences surrounding the cleaved ends were aligned from 5' to 3' ends with two fixed cleavage sites.

In-frame deletion of the *dndB* gene in *S. lividans* 66

pHZ868, a pHZ132 (18,19) derivative carrying the complete *dnd* gene cluster (Table 1), was digested with XhoI to generate a 1971-bp DNA fragment carrying part of *dndB* encoding the C-terminal amino acids, which was recovered from an agarose gel and ligated into the SalI site of pOJ260, giving rise to pJTU155. Oligonucleotide primers *dndB*-L (5'-CAGAAATTCGAACTTCCCACACTC-3', EcoRI site underlined) and *dndB*-R (5'-CATGGATCCCTTGTTCAAGATCCG-3', BamHI site underlined) were used to amplify a 1.85-kb DNA region with an internal 1674-bp EcoRI–BamHI fragment carrying part of the *dndB* gene encoding the N-terminal amino acids, using pHZ868 as template. The resultant DNA fragment was digested with EcoRI and BamHI before ligation into the corresponding sites of pJTU155 digested with the same enzymes, generating pJTU156. pJTU156 was further digested with SpeI to accept another 1.23-kb XbaI–SpeI *oriT*-carrying fragment from pOJ260 for the final construction of pJTU164. The desired mutant, HXY2, was isolated after pJTU164 was transferred by conjugation into *S. lividans* 1326, and selected for deletion of the internal 867-bp *dndB* region encoding 289 amino acids of the DndB protein (Figure 1A), which was confirmed by PCR and Southern blotting analysis from six mutant candidates.

Constructs carrying a single *dndB* gene and a complete *dnd* gene cluster for genetic complementation of a *dndB* mutant

The *dndB* gene was amplified from total DNA of 1326 using *dndB*-L (TTGCCGAATTCCGCGGTCATCCAGTGCAG, EcoRI site underlined) and *dndB*-R (CGCGGATCCCCTCCGTCAGCTCCTCGAC, BamHI site underlined) as PCR primers. The 1376-bp fragment obtained was digested with EcoRI and BamHI before

ligation into pSET152 (Table 1) digested with the same enzymes. The resultant plasmid, pJTU2020, and pSET152 (Table 1) were transferred from *E. coli* ET12567 carrying pUZ8002 into HXY2, to generate LJD1 and LJD3 (Table 1). Concurrently, conjugation of pHZ1904 from *E. coli* ET12567 carrying pUZ8002 into HXY2 generated LJD2 (Table 1).

Site-directed mutagenesis of the preferentially recognized sequence on pHZ209

Site-directed mutagenesis of a sequence (nucleotides 2374–2385 of pHZ209) including the 4-bp central core and some of its flanking nucleotides from both sides (5'-cGGCCgccg-3') was independently achieved by cloning fusion PCR products with an identical DNA fragment of the same size (1187 bp), but carrying mutations at various positions along 5'-cGGCCgccg-3', all to A, based on a two-step PCR technique as described in (20). Briefly, PCR amplification using a fixed primer (UP1) designed upstream of an EcoRI site leftward to the nucleotides to be mutated (i.e. nucleotides upstream of 2374–2385) and variable primers (DP2 series, depending on the site to be mutated) would generate a Step Ia product (725 bp). Meanwhile, PCR amplification using variable primers (UP3 series, depending on the site to be mutated) and a fixed primer (DP4) designed downstream of an XhoI site rightward to the nucleotides to be mutated (i.e. nucleotides downstream of 2374–2385) would generate a Step Ib product (502 bp). A mixture of the Step Ia and Ib products (with a 40-bp overlap between the 3'-end of Ia and the 5'-end of Ib, to provide mutual priming) was used as templates in a subsequent PCR amplification using two fixed primers (UP1 and DP4) to yield the Step II product (1187 bp). On digestion with EcoRI and XhoI, fragments of identical size (949 bp) but variable mutations along 5'-cGGCCgccg-3' were ligated into the corresponding sites of pBluescript II SK(+), to obtain transient plasmids in *E. coli* DH10B. Their EcoRI–XhoI fragments were recovered from agarose gels and cloned into the corresponding sites of pHZ209 after digestion with the same enzymes (EcoRI and XhoI), so that the fragment of the same size was conveniently replaced. All the replaced fragments in pHZ209 derivatives were sequenced to confirm the desired mutation at specific nucleotide positions before protoplast transformation of *S. lividans* 1326 and its mutants HXY2 and HXY6 to test the effect of each mutation on modification specificity. All the primers used and plasmids obtained are listed in Table 2. The three-step PCR conditions were as follows: 30 rounds of denaturation at 94°C for 5 min, 60°C annealing for 30 s followed by extension at 68°C for 20 or 40 s using KOD-Plus DNA polymerase.

RESULTS

Differences in the Dnd phenotype between wild-type *S. lividans* 1326 and its *dndB* mutant HXY2

An analysis of the Dnd phenotype displayed by HXY2 (with an in-frame deletion in *dndB*) in comparison with its progenitor wild-type strain 1326 indicated

Table 2. Primers and resultant plasmids for site-directed mutagenesis

Primer	Sequence	pHZ209 derivatives
UP1	5' CGATTGTCTGTTGTGCCAG	
DP4	5' GACCTCGATCTGCTCCTGGC	
DP2-2374	CGCGGCGACGGCGGCCGGGTTTCGAAGCCGCGGGAGTAATC	pJTU2031
UP3-2374	GATTACTCCC GCGGCTTCGA <u>A</u> CCCGGCCGCCGTCGCCGCG	
DP2-2377	CGCGGCGACGGCGGCCCTGGGTTTCGAAGCCGCGGGAGTAATC	pJTU2025
UP3-2377	GATTACTCCC GCGGCTTCGACCC <u>A</u> GGCCGCCGTCGCCGCG	
DP2-2378	CGCGGCGACGGCGGGCTGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2017
UP3-2378	GATTACTCCC GCGGCTTCGACCC <u>A</u> GCCGCCGTCGCCGCG	
DP2-2379	CGCGGCGACGGCGGTTCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2005
UP3-2379	GATTACTCCC GCGGCTTCGACCC <u>A</u> CCGCCGTCGCCGCG	
DP2-2380	CGCGGCGACGGCGTCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2007
UP3-2380	GATTACTCCC GCGGCTTCGACCC <u>A</u> CGCCGCCGTCGCCGCG	
DP2-2381	CGCGGCGACGGCTGCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2009
UP3-2381	GATTACTCCC GCGGCTTCGACCC <u>A</u> GCCGCCGTCGCCGCG	
DP2-2382	CGCGGCGACGGTGGCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2011
UP3-2382	GATTACTCCC GCGGCTTCGACCC <u>A</u> CCGCCGTCGCCGCG	
DP2-2383	CGCGGCGACGTCGGCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2013
UP3-2383	GATTACTCCC GCGGCTTCGACCC <u>A</u> GCCGCCGTCGCCGCG	
DP2-2384	CGCGGCGACTGCGGCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2003
UP3-2384	GATTACTCCC GCGGCTTCGACCC <u>A</u> GTCGCCGCG	
DP2-2385	CGCGGCGATGGCGGCCGGGGTTCGAAGCCGCGGGAGTAATC	pJTU2015
UP3-2385	GATTACTCCC GCGGCTTCGACCC <u>A</u> TGCCGCG	

UP: sense-strand primer; DP: antisense-strand primer; underlined bold-type nucleotides and numbers after dashes of the primers: desired site for the introduction of "A" at the specified nucleotide numbers of pHZ209 sequence.

that the mutation does not prevent the genomic DNA of the mutant strain HXY2 from degradation during electrophoresis (Figure 1B). In sharp contrast, the Dnd phenotype was lost in the mutants ZX1 [with deletion of ~93-kb DNA including the complete *dnd* locus (4,21), data not shown] and HXY6 (with a targeted deletion of ~8 kb covering the complete *dnd* locus) (Figure 1B), and those mutant strains with gene disruptions inside *dndA* (ZX64 and HXY1) or *dndD* (LA2), whose DNAs were all stable (4) (data not shown).

Thus, the mutation in *dndB* seemed to be unique among all of the mutants tested.

A brief comparison of the Dnd phenotype between DNA of HXY2 and 1326, however, revealed that the degradation 'smear' from the genomic DNA of HXY2 migrated much faster than that from strain 1326, suggesting that smaller DNA fragments were generated from degradation of HXY2 DNA than from 1326; in other words, the *dndB* mutation appeared to aggravate the Dnd phenotype (Figure 1B).

To unambiguously explore how the *dndB* mutation could affect DNA degradation or modification, we used a pIJ101-derivative plasmid, pHZ209, which contains at least one preferential modification site, with a characterized central core sequence flanked by three direct repeats (14) to test the possible alteration in modification pattern. After passage of pHZ209 into HXY2, 1326 or HXY6 (with deletion of the complete *dnd* locus), DNAs were extracted and treated with activated Tris buffer [general DNA electrophoresis buffer, (22)] before being heavily loaded on an agarose gel for detection of the Dnd phenotype (Figure 1C).

Distinctive bands could be observed from pHZ209 DNA isolated from the in-frame deletion mutant HXY2 (pHZ209/HXY2 in Figure 1C), corresponding to the bands of the same sizes observed from the wild-type 1326 (pHZ209/1326 in Figure 1C), while pHZ209 DNA originating from HXY6 remained stable as a linearized but non-degraded 5.4-kb band on the gel (pHZ209/HXY6 in Figure 1C), indicating that it was not cleaved by activated Tris buffer.

Closer examination of the distinctive banding pattern generated after degradation using plasmid DNA revealed that the DNA modification preferences in the wild-type 1326 and the *dndB* mutant HXY2 were different. The great majority of the pHZ209 DNA in both lanes, pHZ209/1326 and pHZ209/HXY2, remained stable as a 5413-bp non-degraded and linearized pHZ209, as originally loaded in lane pHZ209/HXY6, suggesting that not all of the pHZ209 DNA molecules were modified, in agreement with the early report by Zhou *et al.* (4,5). Two bands (running at ~3.0 and 2.4 kb, Figure 1C) were much stronger than any others degraded from EcoRV-linearized pHZ209 DNA isolated from 1326. They were ~ 43% reduced (as calculated from band scanning) with concurrent increase of the intensity of all the other degradation products when pHZ209 DNA came from HXY2. The difference in the Dnd phenotype between wild-type 1326 and its *dndB* mutant HXY2 was confirmed by complementation by introducing pHZ209 into a HXY2 derivative, using a plasmid carrying either the *dndB* gene alone (pJTU2020) or the complete *dnd* gene cluster (pHZ1904) for integration into the *attB* site of the chromosome to generate LJD1 and LJD2 respectively. A strain (LJD3) receiving a control vector pSET152 without insert had no such effect (Figure 2A).

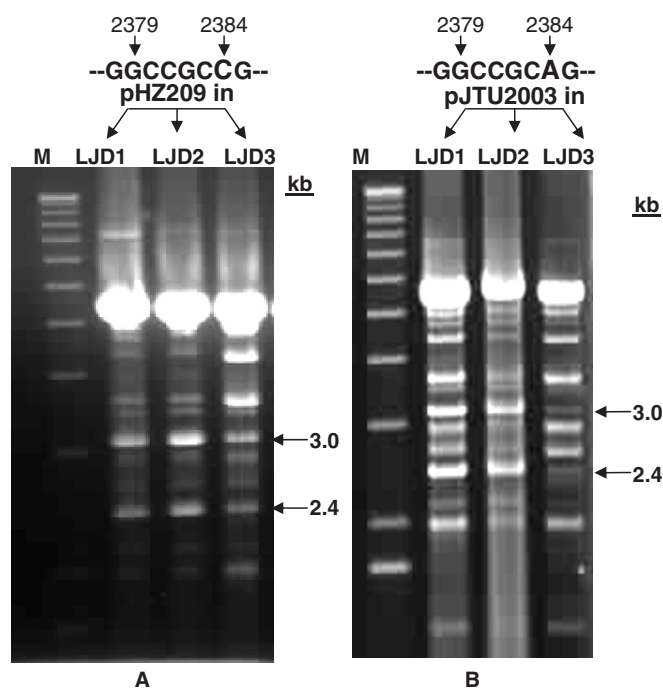


Figure 2. Restoration of modification preference towards EcoRV-linearized pHZ209 (A) and partial restoration of modification preference towards EcoRV-linearized pJTU2003 (in which the nucleotide 'C' at position 2384 of pHZ209 was mutated to 'A' (B), whose DNA came from LJD1 (with integration of the *dndB* gene) and LJD2 (with integration of the complete *dnd* gene cluster), in contrast to the lack of restoration originating from LJD3 (with integration of the vector pSET152 carrying no *dnd* gene). DNA size markers are labeled 'M'.

Each pHZ209 molecule seemed to be modified only once at most

One (~3.0 kb) of the two most intense bands (~3.0 and 2.4 kb) originating from degraded pHZ209 in 1326 (Figure 1C) could easily be recovered from an agarose gel, cloned into pGEM-T and sequenced. It was found to be generated by cleavage at a preferential modification site (14), generating an uninterrupted linear band corresponding to 3034 bp (Figure 1C) extending from the preferential modification site to the rightmost EcoRV site [ends cleaved by the activated Tris buffer could easily be distinguished from the other end of the sequenced fragment as it was labeled by an EcoRV site (Figure 3) before electrophoresis on agarose gel in activated Tris buffer]. This suggested that the plasmid molecule is not further modified in other positions once the preferential site is modified. If this site were the only one that could be modified *in vivo*, no other degradation bands should be observed.

This is obviously not the case as we saw many additional bands resulting from the cleavage of pHZ209, in either 1326 or HXY2 (Figure 1C) at other positions. This prompted us to attempt sequencing of as many as possible of the additional fragments with modified termini originating from degraded pHZ209 in HXY2.

A total of 141 clones were obtained after EcoRV-linearized pHZ209 was treated with activated Tris-buffer

to cleave the modified site before insertion into SmaI-digested pBluescript SK(+), and subjected to multiple restriction enzyme analysis (data not shown). Thirty-three clones equal in size to that of linearized pHZ209 (ca 5.4 kb) (three of which were confirmed to be linearized pHZ209 by sequencing from both termini), and 15 clones with inserts larger than 5.4 kb (and so carrying multiple fragments), were disregarded. Forty-eight of the 93 remaining clones, which were obviously smaller than linearized pHZ209 (5.4 kb), were selected for sequencing from both termini. Apart from one clone that seemed to derive from degradation of contaminating chromosome DNA, 41 clones each contained a contiguous piece of pHZ209 DNA, clearly derived from degradation at a single site, and the other six clones contained rearranged DNA fragments resulting from ligation of at least two degraded pHZ209 fragments.

The terminal sequences resulting from ligation of a single DNA fragment among 24 clones could be mapped to 16 positions relative to the unique reference site (EcoRV) on pHZ209. A preliminary assembly of the end positions of 14 sequence reads indeed first revealed an obvious sequence conservation (Figure 3) conforming to the site identified by Dyson *et al.* (14,15), although the two modified nucleotides on opposite strands (Figure 3) seemed to be adjacent rather than separated by two bases within the consensus region as originally identified by experiments involving *in vitro* Tris-activation and PCR (see Discussion). However, 17 independent sequence reads did not conform to the consensus sequence (not shown) but matched totally random nucleotide positions on pHZ209, in sharp contrast to those sequence reads leading to the assembly of the consensus sequence in Figure 3A, some of which had been sequenced more than once using random and independent clones from the same cloning experiment.

In an attempt to discover whether all the sequence reads, or only those conforming to the consensus sequence, represented genuine modifiable sites, we repeated the above cloning experiment using EcoRV-linearized pHZ209 isolated from HXY6 (a strain with complete deletion of the *dnd* gene cluster, hence without modified DNA). Indeed, 24 multiple clones were obtained, but none of the sequence reads from them corresponded to consensus sequences, and none of them matched the sequences from the random and independent clones generated in the earlier experiment, suggesting that they were 'noise' sequences generated by non-specific and physical breakage of DNA fragments during the cloning process involving *in vitro* treatment with activated Tris-buffer and exonuclease III.

In order to confirm that mapped modifiable sites on pHZ209 (Figure 3B) originating from the above assembled consensus sequence were meaningful, thus ruling out non-specific noise, we sequenced DNA fragments from some additional isolatable specific bands from degraded pHZ209 in HXY2 after cloning into pGEM-T in the same way as we had done for one of the two most intense bands (3034 bp) of pHZ209 from 1326 (Figure 1B). Four bands were chosen (4R, 5R, 6R and 9R, Figure 3A), whose sizes added up to 5.4 kb, equivalent to the total size

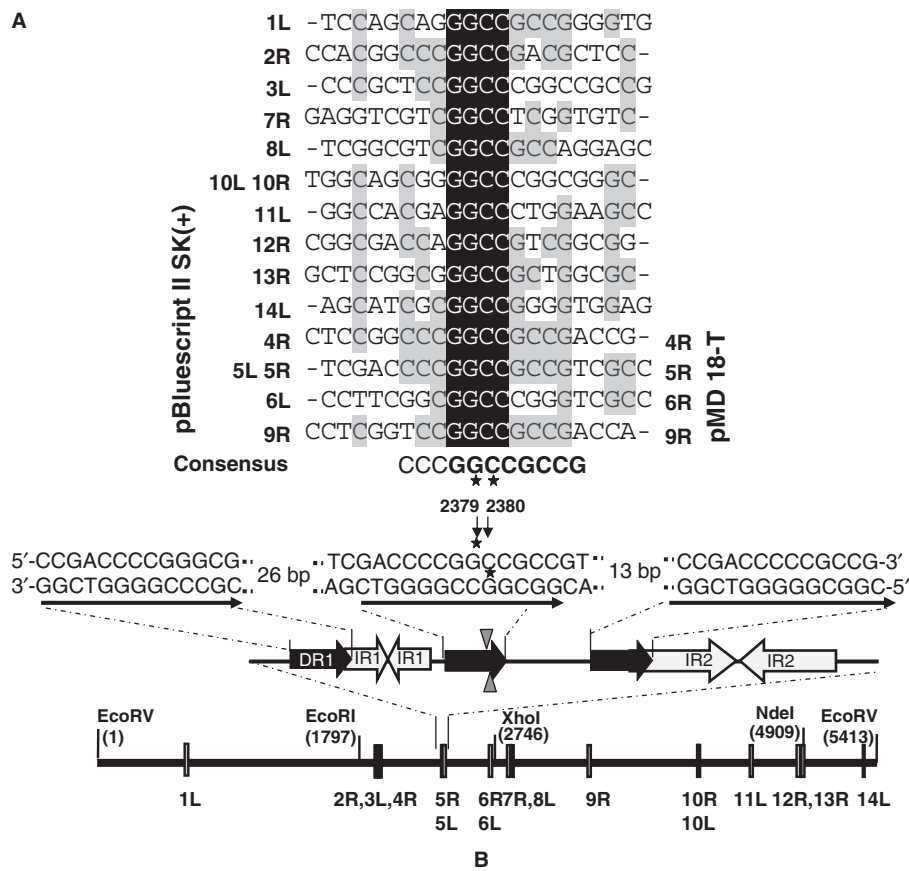


Figure 3. (A) Alignment of the region traversing the modified sites (labeled as asterisks) sequenced by using pBluescript II SK(+) (left) or pMD18-T (right) as vectors for the identification of the central four nucleotides (GGCC) as the highly consensus region, which lies within one (DR2, underlining the corresponding sequence) of the three direct repeat sequences (DR1-3) in a region with a preferentially modified site (5L/R), enlarged to show the details of the region possessing three direct repeats (DR1-3, solid arrows) and two pairs of inverted repeats (IR1-2, framed arrows). Sequence tags conform to the relative positions in Figure 1C and in the linearized pHZ209 schematically presented at the bottom (B), while those labeled L or R are G site at either 2379 or 2380 in the sequenced terminus.

of pHZ209. (Sequencing of other fragment pairs was unsuccessful, mainly because different DNA fragments of the same size were recovered for cloning, which resulted in co-transformation influencing the sequencing reads.) Consistently, the distribution of these sequence reads, which were sure to have resulted from the Dnd phenotype, could be easily localized at positions 4, 5, 6 and 9, respectively in Figure 3B, conforming to the same sites for the generation of the consensus sequence as from the shotgun cloning into pBluescript M13(+) described above (Figure 3A).

The common consensus sequence generated from either a shotgun cloning of a mixture of fragments [into pBluescript SK(+)] or from the targeted cloning of specific fragments (into pGEM-T) clearly suggested that each degraded DNA fragment originated from specific cleavage of the plasmid, which had been modified at different modifiable positions, and that each pHZ209 molecule is modified not more than once, even though multiple sites are available for modification. To further prove this, we analyzed the specific banding patterns of the two specifically degraded fragments from pJTU2031 linearized with three restriction enzymes on which the recognition

sites are all unique (Figure 4). [pJTU2031 is a pHZ209 derivative plasmid with 'C' mutated to 'A' at nucleotide 2374 (Figure 5), resulting in a reduced intensity of the two specific bands compared with those caused by preferential modification in pHZ209]. Modification at position 10 (L/R), for example, would generate DNA fragments of 4182 and 1231 bp if pJTU2031 (or pHZ209) was linearized with EcoRV. The banding profile was assumed to change into one with 3027 and 2386, 3976 and 1437, 4687 and 726 bp bands if pJTU2031 were linearized with EcoRI, XhoI or NdeI, respectively (Figure 4), as expected (the intensity of the shifted bands varied because occasionally some heterologous fragments of the same sizes migrated together, as detected before in Figure 1). In fact, nearly all of the degraded DNA fragments seen in Figure 1 could be attributed to cleavage at one specific site of the many on the EcoRV-linearized pJTU2031 (or pHZ209), thus establishing a map with S-modifiable sites (Figure 3). In other words, the sizes of any pair of fragments generated by cleavage of linearized pJTU2031 (or pHZ209) at any modifiable site in Figure 3 correlated well with the corresponding bands in Figure 1 (and/or Figure 4), confirming the conclusion that each pJTU2031

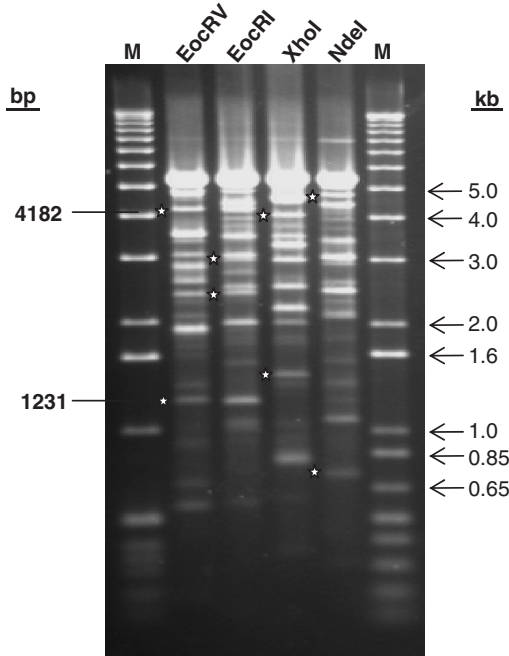


Figure 4. Banding patterns of pHZ2031 respectively linearized using four different but unique restriction enzymes (EcoRV, EcoRI, XhoI and NdeI) before treatment with activated Tris-buffer. Two fragments adding up to the size of linearized pHZ2031 (5.4 kb) were always seen at their expected positions after cleavage at a common modification site, as exemplified by cleavage at specific modification site 10 (Figure 3), indicated as two white asterisks in each respective gel panel. (EcoRV: 4182 and 1231; EcoRI: 3027 and 2386; XhoI: 3976 and 1437; NdeI: 4687 and 726).

(or pHZ209) molecule could be modified either at none or, at most, only one of the multiple modifiable sites, which is also consistent with the routine observation that the cleavage of un-linearized pHZ209 will only generate a concentrated full-size linear fragment together with some un-cleavable circular pHZ209 (not shown), as detected on agarose gels.

Analysis of the consensus sequence by site-directed mutagenesis

A site-directed mutagenesis of the identified consensus sequence surrounding the cleaved ends (Figure 3), which agreed well with the identification of the conserved region using DNA from chromosome and plasmid DNA (14,15), was performed to evaluate the significance of each of the highly conserved nucleotide residues (GGCC), and its flanking but probably more flexible nucleotide residues in determining the site-specificity of DNA modification. We again chose to mutate the site surrounding the preferential recognition sequence 5'-cGGCCgccg-3' (Figure 3), whose specificity changes could easily be monitored by the intensity changes of the bands of 3.0 and 2.4 kb after DNA cleavage (Figure 5). Each of the 9 bp in the putative consensus sequence (5'-cGGCCgccg-3') was independently changed to an 'A' by three steps of PCR-based site-directed mutagenesis (see Materials and methods section) before each of the mutated pHZ209 derivatives was independently introduced into 1326 and HXY2, respectively. pHZ209 DNA was then isolated from each transformant in order to detect possible small

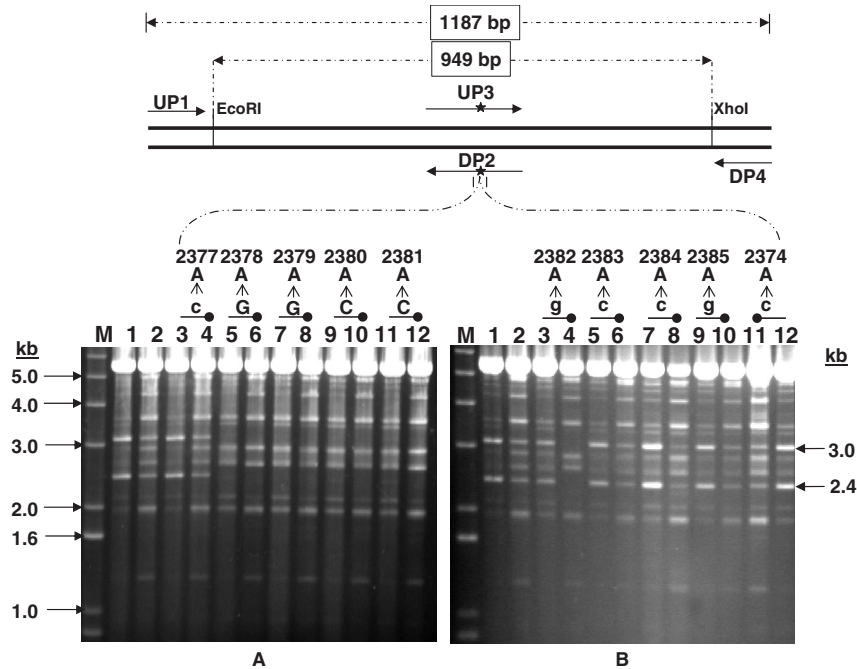


Figure 5. Site-directed mutagenesis of a region traversing the consensus sequence by a two-step PCR technique involving use of two variable oligonucleotide primers (DP2 and UP3) for introducing mutations at variable positions (indicated as a star in the primer region) with two fixed primers at both ends. The 949-bp DNA with site-specific mutations (all changed to A, see text and Table 2) at variable sites (specific nucleotide numbers indicated above each gel lane) could be obtained to replace the corresponding region on pHZ209 after digestion of the respective 1187-bp PCR products with EcoRI and XhoI. The individual constructs with desired replacements (confirmed by sequencing) were introduced into 1326 (heads with no dots) and HXY2 (heads with dots) before their respective DNA samples were isolated for assay in gel A and B. M stands for markers whose sizes are shown at the left of the gel.

changes in the Dnd phenotype reflected by variations in DNA banding patterns.

Both the 3.0- and 2.4-kb bands remained unchanged no matter whether pHZ209 originated from 1326 or HXY2 when the first residue 'c' of the 5'-cGGCCgccg-3' was mutated to 'A' (Figure 5A, lanes 3 and 4) as compared with their respective controls (Figure 5A, lanes 1 and 2) without mutation. In sharp contrast, each independent change from any one of the following four residues from 'GGCC' to 'A' resulted in complete disappearance of the two DNA bands, no matter whether pHZ209 DNA came from 1326 (Figure 5A, lanes 5, 7, 9 and 11) or HXY2 (Figure 5A, lanes 6, 8, 10 and 12). Thus, each of the four residues GGCC was deduced to be equally essential for modification specificity both in 1326 and in HXY2.

When each of the last four residues 'gccg' was independently mutated to 'A', the relative intensity of the two corresponding DNA bands did not seem to change if the mutated plasmids came from 1326 (lanes 3, 5, 7 and 9 in Figure 5B). In contrast, when the plasmid DNA originated from HXY2, the two bands almost disappeared when the first 'g' of the last four residues 'gccg' was mutated to 'A' (Figure 5B, lane 4), but was ~50% reduced for the corresponding change in the third 'c' (Figure 5B, lane 8), while changes in the second 'c' and the fourth 'g' had little effect (lane 6 and 10 in Figure 5B), as judged by their scanned band intensities. Thus, while gccg is concluded as non-essential for modification specificity in 1326, the first (g) and the third (c) of the four nucleotides (gccg) is either essential or important in HXY2. As a control, a pHZ209 derivative (pJTU2003) with a 'c' at the third position of the 'gccg' mutated to 'A' in HXY2 (Figure 2B) could be fully complemented when pHZ1904 carrying the complete *dnd* gene cluster, was integrated into the chromosome of the host strain HXY2, forming LJD2 (Figure 2B), or partially complemented by pJTU2020 carrying the *dndB* gene only, forming LJD1 (Figure 2B).

A similar mutation three nucleotides left of the central core (site 2374 in Figure 5B, lane 11) also caused a ~40% reduction of the relative modification frequency in HXY2.

DISCUSSION

The consensus sequence (5'-cGGCCgccg-3') identified in this work agrees well with that deduced from analysis of preferential modification sites (14,15), although the modified site positions varied by one base pair inward from each side, still at G on both strands. A combination of earlier reports (14,15) and our current work indicates that the base-pairing between C and modified G, if it still occurs, is probably not as strong as normal G-C pairing, and therefore, the primer extension reactions (14,15) used earlier to localize preferential modification sites could have run off one base before meeting the modified/cleaved site(s), in all cases.

It is obviously difficult to predict exactly how the five products of the *dnd* gene cluster interact with each other, and with the nucleotides of the highly conserved central

core or the flanking regions. The absolute requirement of the central core (GGCC), in which modification sites were located at the two central bases on complementary strands, indicates unambiguously that the modifying activity is concentrated on these four nucleotide residues (GGCC) both for preferential and random modification in wild-type 1326 and in the *dndB*⁻ mutant HXY2. Conceivably, these four nucleotides must be in close and direct contact with one of the major modifying enzymes. The preferential recognition specificity, totally unaffected in 1326 but variably abolished or reduced in HXY2, after mutation of the four nucleotides flanking the right, and two separated nucleotides flanking the left of the central core (GGCC), strikingly implicates the necessity of DndB in determining the recognition specificity. This agreed well with the detection of the significant amino acid sequence similarity of DndB to a DNA gyrase, which suggests that the DndB protein could in some way affect DNA topology, and thus the efficiency and/or specificity of S-modification on certain sites flanked by the sequences with a potential to form secondary structures. This could also explain why the DndB protein is not required at 'simple' recognition sites without flanking sequence complexity. We assume that the change from preferential modification sites, as detected in wild-type 1326, to a relatively random distribution of modification sites, as detected in its *dndB* mutant HXY2, is mediated by the DndB protein encoded by *dndB*. It is likely that random modification does not need DndB but a close association of DndB with the modifying enzymes could stabilize the contact between GGCC and the modifying complex.

It is not clear why a 5.4-kb circular plasmid molecule could only be modified once, although multiple modification sites are available. The presence of numerous modifiable sites on the chromosome suggests that modification frequency does not depend on the number of molecules, large or small, but more likely, by the length of the DNA flanking both sides of the modified site occupied by modifying enzymes. We do not know whether it is the binding of the modifying enzyme complex to a specific GGCC-containing site, or the result of modification, which prevents further binding of the modifying enzyme complex to a neighboring GGCC-containing site on either chromosome or plasmid, but it is hard to think that the result of modification could have contributed to this phenomenon. The fact that the Dnd phenotype could only be observed when a plasmid carrying the *dnd* gene cluster was present either in the integrated form or on a low copy number (<10) plasmid, but not at high copy number (Li, A., unpublished data) implied that the expression of the *dnd* gene cluster in 1326 could be tightly controlled, and thus the dosage of the modifying enzyme complex could be limited in the cell. We propose that a protein complex formed at one specific site which involved nicking, winding and unwinding of the supercoiled double helix could exclude nearby site(s), and additionally, the number of sets of the Dnd protein complex is restricted to a cell-tolerated number by an unknown mechanism. To test the former hypothesis, it would be

interesting to define the approximate length of the sequences flanking the consensus GGCC, within which modifiable sites were insensitive to modification, by analyzing plasmids of variable size. This work is now under consideration.

The proposed biochemical pathway leading to DNA modification by S involves five putative proteins encoded by a well-characterized *dnd* gene cluster. Of these, DndA was characterized as a PLP-containing homodimer that specifically catalyzes formation of L-alanine and elemental S using L-cysteine as substrate, DndC as an ATP pyrophosphatase catalyzing hydrolysis of ATP to pyrophosphate (PPi), and the function of DndC was found to be mediation of the formation of an [4Fe-4S] iron-sulfur cluster protein, whose reconstitution was activated by DndA (25). Instead, the exact biochemical functions of DndD and DndE, have not been demonstrated. While DndD, a bioinformatically predicted SMC-like ATPase with a distinguishable myosin-tail consisting of a coiled-coil region and a flexible hinge, could function as an ATP-modulated DNA cross-linker and energy generator by ATP hydrolysis, DndE was expected to involve in determining the sulfur-existing status. From the present work, it seems reasonable to assume that DndC, forming a [4Fe-4S] iron-sulfur cluster protein, might provide a platform for the orchestrated assembly of the Dnd protein complex, although some member(s) could be associated or not at specific stages of their required activities. The merit of the present work is thus, in essence, to have provided information for additional experiments aimed at demonstrating the interaction(s) between different proteins or between variable combinations of protein(s) and specific target DNA, including those mediating secondary structure formation and/or binding tightly or loosely under different condition(s) using, e.g. immuno-precipitations and gel shift mobility assays. This work is now in active progress.

ACKNOWLEDGEMENTS

We are very grateful to Prof. Sir David Hopwood, FRS for his continuous support and encouragement throughout this study for many years, and critical and patient editing of the manuscript. The authors wish to thank the Ministry of Science and Technology (973 (2003CB114205) and 863 programs), the National Science Foundation of China, the Ministry of Education of China, and the Shanghai Municipal Council of Science and Technology for research supports. Funding to pay the Open Access publication charge was provided by grant 2003CB114205 from the Ministry of Science & Technology of China.

Conflict of interest statement. None declared.

REFERENCES

1. Koort, J.M., Lukinmaa, S., Rantala, M., Unkila, E. and Siitonen, A. (2002) Technical improvement to prevent DNA degradation of

- enteric pathogens in pulsed-field gel electrophoresis. *J. Clin. Microbiol.*, **40**, 3497–3498.
2. Evans, M., Kaczmarek, F.S., Stutzman-Engwall, K. and Dyson, P. (1994) 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*, **140**(Pt. 6), 1367–1371.
3. Ray, T., Weaden, J. and Dyson, P. (1992) Tris-dependent site-specific cleavage of *Streptomyces lividans* DNA. *FEMS Microbiol. Lett.*, **75**, 247–252.
4. Zhou, X.F., He, X.Y., Liang, J.D., Li, A.Y., Xu, T.G., Kieser, T., Helmann, J.D. and Deng, Z.X. (2005) A novel DNA modification by sulphur. *Mol. Microbiol.*, **57**, 1428–1438.
5. Zhou, X.F., Deng, Z.X., Firmin, J.L., Hopwood, D.A. and Kieser, T. (1988) Site-specific degradation of *Streptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. *Nucleic Acids Res.*, **16**, 4341–4352.
6. Evans, M. and Dyson, P. (1993) Pulsed-field gel-electrophoresis of *Streptomyces lividans* DNA. *Trends Genet.*, **9**, 72.
7. Zhang, Y., Yakrus, M.A., Graviss, E.A., Williams-Bouyer, N., Turenne, C., Kabani, A. and Wallace, R.J.Jr (2004) Pulsed-field gel electrophoresis study of *Mycobacterium abscessus* isolates previously affected by DNA degradation. *J. Clin. Microbiol.*, **42**, 5582–5587.
8. Murase, T., Nagato, M., Shirota, K., Katoh, H. and Otsuki, K. (2004) Pulsed-field gel electrophoresis-based subtyping of DNA degradation-sensitive *Salmonella enterica* subsp. *enterica* serovar Livingstone and serovar Cerro isolates obtained from a chicken layer farm. *Vet. Microbiol.*, **99**, 139–143.
9. Liesegang, A. and Tschape, H. (2002) Modified pulsed-field gel electrophoresis method for DNA degradation-sensitive *Salmonella enterica* and *Escherichia coli* strains. *Int. J. Med. Microbiol.*, **291**, 645–648.
10. Silbert, S., Boyken, L., Hollis, R.J. and Pfaller, M.A. (2003) Improving typeability of multiple bacterial species using pulsed-field gel electrophoresis and thiourea. *Diagn. Microbiol. Infect. Dis.*, **47**, 619–621.
11. Romling, U. and Tummler, B. (2000) Achieving 100% typeability of *Pseudomonas aeruginosa* by pulsed-field gel electrophoresis. *J. Clin. Microbiol.*, **38**, 464–465.
12. Klaassen, C.H., van Haren, H.A. and Horrevorts, A.M. (2002) Molecular fingerprinting of *Clostridium difficile* isolates: pulsed-field gel electrophoresis versus amplified fragment length polymorphism. *J. Clin. Microbiol.*, **40**, 101–104.
13. Corkill, J.E., Graham, R., Hart, C.A. and Stubbs, S. (2000) Pulsed-field gel electrophoresis of degradation-sensitive DNAs from *Clostridium difficile* PCR ribotype 1 strains. *J. Clin. Microbiol.*, **38**, 2791–2792.
14. Dyson, P. and Evans, M. (1998) Novel post-replicative DNA modification in *Streptomyces*: analysis of the preferred modification site of plasmid pIJ101. *Nucleic Acids Res.*, **26**, 1248–1253.
15. Boybek, A., Ray, T.D., Evans, M.C. and Dyson, P.J. (1998) Novel site-specific DNA modification in *Streptomyces*: analysis of preferred intragenic modification sites present in a 5.7 kb amplified DNA sequence. *Nucleic Acids Res.*, **26**, 3364–3371.
16. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. John Innes Foundation, Norwich.
17. Sambrook, J. and Russel, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
18. Bao, K., Hu, Z.H., Zhou, X.F., Zhou, Q. and Deng, Z.X. (1997) A bifunctional cosmid vector for the mobilized conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Prog. Nat. Sci.*, **7**, 568–572.
19. Deng, Z.X. and Bai, L.Q. (2006) Antibiotic biosynthetic pathways and pathway engineering – a growing research field in China. *Nat. Prod. Rep.*, **23**, 811–827.
20. Kuwayama, H., Obara, S., Morio, T., Katoh, M., Urushihara, H. and Tanaka, Y. (2002) PCR-mediated generation of a gene disruption

- construct without the use of DNA ligase and plasmid vectors. *Nucleic Acids Res.*, **30**, E2.
21. Zhou,X.F., He,X.Y., Li,A.Y., Lei,F., Kieser,T. and Deng,Z.X. (2004) *Streptomyces coelicolor* A3(2) lacks a genomic island present in the chromosome of *Streptomyces lividans* 66. *Appl. Environ. Microbiol.*, **70**, 7110–7118.
22. Ray,T., Mills,A. and Dyson,P. (1995) Tris-dependent oxidative DNA strand scission during electrophoresis. *Electrophoresis*, **16**, 888–894.
23. Paget,M.S., Chamberlin,L., Atrih,A., Foster,S.J. and Buttner,M.J. (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.*, **181**, 204–211.
24. Bierman,M., Logan,R., O'Brien,K., Seno,E.T., Rao,R.N. and Schoner,B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*, **116**, 43–49.
25. You, D.L., Wang, L.R., Yao, F., Zhou,X.F. and Z. Deng (2007). 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*. DOI:10.1021/bi602615k.