Escherichia coli Strains (*ndk*) Lacking Nucleoside Diphosphate Kinase Are Powerful Mutators for Base Substitutions and Frameshifts in Mismatch-Repair-Deficient Strains

Jeffrey H. Miller,[1] Pauline Funchain, Wendy Clendenin, Tiffany Huang, Anh Nguyen, Erika Wolff, Annie Yeung, Ju-Huei Chiang, Lilit Garibyan, Malgorzata M. Slupska and Hanjing Yang

Department of Microbiology, Immunology, and Molecular Genetics and The Molecular Biology Institute, University of California, Los Angeles, California 90095

Manuscript received January 9, 2002
Accepted for publication June 21, 2002

ABSTRACT

Nucleoside diphosphate (NDP) kinase is one of the enzymes that maintains triphosphate pools. *Escherichia coli* strains (*ndk*) lacking this enzyme have been shown to be modest base substitution mutators, and two members of the human family of NDP kinases act as tumor suppressors. We show here that in *E. coli* strains lacking NDP kinase high levels of mispairs are generated, but most of these are corrected by the mismatch-repair system. Double mutants that are *ndk mutS*, lacking both the NDP kinase and mismatch repair, have levels of base substitutions 15-fold higher and levels of certain frameshifts up to 10-fold higher than those of the respective mutations in *mutS* strains that are NDP kinase proficient. A sequence analysis of the specificity of base substitution mutations generated in *ndk* and *ndk mutS* backgrounds as well as other experiments suggests that NDP kinase deficiency stimulates polymerase errors that lead to A:T → G:C transitions and that the editing capacity of cells may be affected, leading to additional uncorrected mispairs and to A:T → T:A transversions.

NUCLEOSIDE diphosphate (NDP) kinase, a ubiquitous enzyme that maintains deoxynucleotide triphosphate pools by converting deoxynucleotide diphosphates to triphosphates, is highly conserved between bacteria, Drosophila, and humans (see reviews by Lacombe *et al.* 2000 and Lombardi *et al.* 2000). In *Escherichia coli*, the *ndk* gene encodes NDP kinase. A number of studies have shown that deoxynucleotide triphosphate levels are important for genetic stability (see review by Kunz *et al.* 1994). *E. coli* cells lacking NDP kinase are weak mutators, with moderate increases in mutations, as measured by the frequency of rifampicin-resistant (Rif^r) and nalidic acid-resistant (Nal^r) mutants in cell populations (Lu *et al.* 1995). Intracellular pools of dCTP are elevated 20-fold, and dGTP pools are elevated 7-fold (Lu *et al.* 1995). Defects in the Drosophila gene *altered wing direction* (*awd*) encoding NDP kinase lead to developmental abnormalities (Biggs *et al.* 1988, 1990). In humans, the *nm23* family of genes (*nm23*-H1 to *nm23*-H8) encodes NDP kinases that are expressed in a tissue-specific manner and appear to play a role in normal development and differentiation (Lacombe *et al.* 2000; Lombardi *et al.* 2000 and references therein). The *nm23*-H1- and *nm23*-H2-encoded NDP kinases have been identified as tumor suppressors involved in tumor metastasis and pathogenesis (Leone *et al.* 1991, 1993a). The NM23-H2/NDP kinase appears to be a transcription factor for the *c-MYC* oncogene, among others (Postel *et al.* 1993, 2000). Identification of a cleavage activity at a sequence in the c-MYC promoter and partial sequence homologies with certain DNA glycosylases has led to the speculation that the NM23-H2/NDP kinase B might also play a role in DNA repair (Postel *et al.* 2000).

We show here that in NDP kinase (NDK)-deficient strains of *E. coli* very high levels of mispairs are generated, but most of these are corrected by the mismatch-repair system and to some degree by the editing function (ε) of DNA polymerase. These two systems normally safeguard the cell against polymerase errors (Scheuermann and Echols 1984; Modrich and Lahue 1996). However, we find that double mutants lacking both mismatch repair and NDP kinase have extraordinarily high levels of base substitutions and also of frameshifts at repeat-tract sequences that are 15- and 10-fold higher than those of the respective mutations in cells defective for mismatch repair but with normal NDK. We have analyzed the nature of these mutations with genetic reversion systems (Cupples and Miller 1989; Cupples *et al.* 1990) and by DNA sequencing of mutations in *rpoB* that lead to resistance to rifampicin (Rif^r). In doing so we have extended the development of the *rpoB*/Rif^r system (Jin and Gross 1988; Severinov *et al.* 1993) as a useful tool for analyzing chromosomal mutations.

[1]*Corresponding author:* Department of Microbiology, Immunology, and Molecular Genetics, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095. E-mail: jhmiller@mbi.ucla.edu

These and other experiments suggest that NDK deficiency provokes polymerase errors that lead to A:T → G:C transitions. The altered triphosphate pools may hinder the editing capacity of cells, leading to additional uncorrected mispairs. In the absence of mismatch repair (MMR^-), the additional mispairs result in extremely high levels of mutations that are more than an order of magnitude greater than those seen in NDK-proficient MMR^- strains. We consider these results with respect to the tumor suppressor activity of the human NDK protein and to the interaction of different alleles in generating cancer susceptibilities.

MATERIALS AND METHODS

Bacterial strains: CC101-108, CC110, and CC111 have been described previously (Cupples and Miller 1989; Cupples *et al.* 1990). Each of these strains is *ara* Δ*(gpt-lac)5 thi*/F′128 *lacIZ proA*$^+$*B*$^+$. The nature of the *lacZ* mutation differs in each case. In CC101-106 the *lacZ* mutation results from a base substitution at coding position 464 in *lacZ*. Reversion to Lac^+ occurs by a different base substitution in each case (see Table 2). Strains CC107, -108, -110, and -111 carry a frameshift mutation in *lacZ*. Each reverts only via a specific frameshift (see Table 4). *ndk*, *mutS*, and *ndk mutS* derivatives of each of the CC101-111 strains were prepared by transduction of each strain with P1*vir* lysates (Miller 1992) grown on strains carrying a mini-Tn*10* in *ndk* or a miniTn*10cam* in *mutS* (J. H. Miller, P. Funchain and A. Yeung, unpublished results) and selecting for either Tet^r or Cam^r cells. Double mutants *ndk::*minTn*10 mutS::*minTn*10cam* were prepared in two successive transduction steps. All genetic procedures were carried out as described in Miller (1992). Derivatives of CC105 and CC107, carrying either pBR329 or pBR329*dnaQ66* (Slupska *et al.* 1998), were prepared by transformation of electrocompetent cells. *dnaQ66* is a weak mutator allele constructed by site-directed mutagenesis and results in a glycine-for-histidine substitution at position 66 in the ε subunit of DNA polymerase III (Slupska *et al.* 1998). Strain J93 (Funchain *et al.* 2000) is *ara* Δ*(lac)RV thi.*

Detection of insertions in *ndk*: We employed mini-Tn*10* derivatives to generate insertions in chromosomal genes. We looked for weak frameshift mutators using the same procedures that are described in detail in Slupska *et al.* (2000) for the detection of rearrangement mutators, except that strain CC107 (see above) was used. The exact position of each insertion was determined by the method of Higashitani *et al.* (1994), as described in detail in Slupska e*t al.* (2000). The junction point of the mini-Tn*10tet* insertion in *ndk* used here was at 229 bp from the start of the coding sequence.

Determination of mutation rates: Rif^r and Nal^r mutants and Lac^+ revertants were determined as described previously (Cupples and Miller 1989; Miller 1992). Briefly, overnight cultures were seeded with ≤100 cells and grown at 37°. Dilutions were plated on selective media and on Luria broth (LB; for determining the viable cell titer). For determining Lac^+ revertants in *mutS* derivatives of CC107 and CC108, cultures were seeded with 10 cells or less. For determinations in *ndk mutS* derivatives of CC107, cultures used for plating were seeded with 1 or 2 cells by diluting a starting culture to the point where a drop delivered to a set of cultures resulted in one-third of the cultures receiving no cells. Lac^+ frequencies were determined with a cell density of 10^9 cells per plate to prevent additional growth of cells leading to additional revertants. When $<10^9$ cells from the actual culture were plated, 10^9 cells of a scavenger (J93; Δ*lac*, see above) were added. Under these conditions 10-fold dilutions of the culture being tested yielded 10-fold fewer revertants. Mutant frequencies were determined, and the median frequency (f) from a set of cultures (the number of cultures varied from 11 to >50) was used to calculate the mutation rate (μ) per replication by the method of Drake (1991), using the formula: $\mu = f/\ln N\mu$, where N is the number of cells in the culture. Ninety-five percent confidence limits were determined according to Dixon and Massey (1969).

Chromosomal DNA isolation and sequencing: Chromosomal DNA was isolated using DNAzol reagent (GIBCO BRL, Rockville, MD) following the protocols supplied by the manufacturer. Portions of the *rpoB* gene were PCR amplified for cloning. The sequence of primers for amplifying the *rpoB* region encoding the main group of mutations (cluster II) was as follows: 5′ GATATGATCAACGCCAAGCCGATTT 3′ and 5′ TTACACCGGAGTCAACGGCAACAGC 3′. The sequence of primers for amplifying the earlier region (cluster 1) was 5′ AATGTCAAATCCGTGGCGT 3′ and 5′ CCAACCGCAGACA AGTCATA 3′. For some of the mutations in cluster II, the amplified fragments were cloned into the pCRII-TOPO vector using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced using vector primers. DNA sequencing was carried out by using [α-^{32}P]dATP and a SequiTherm cycle sequencing kit (Epicentre Technologies, Madison, WI) with reagents and protocols supplied by the manufacturer. About one out of every five sequences yielded a secondary mutation, presumably introduced during the cloning step. In these cases the *rpoB* gene was cloned again from the same Rif^r colony and resequenced. In each case one of the two mutations disappeared. We then confirmed many of the sequences by sequencing a PCR-amplified sample in the absence of cloning. Here, the *rpoB* gene was PCR amplified from genomic DNA using a different set of primers, which allowed us to sequence directly from the PCR product. The following primers were used for PCR: 5′ CGTCGTATCCGTTCCGTTGG 3′ and 5′ TTCACCC GGATACATCTCGTC 3′. The PCR product was purified using the MinElute PCR purification kit (QIAGEN, Valencia, CA). The purified PCR products were then sequenced using the SequiTherm EXCEL II DNA sequencing kit (Epicentre Technologies). The following primers were used for sequencing straight from the PCR product: 5′ CGTGTAGAGCGTGCGGT GAAA 3′ for cluster II and 5′ GAAGGCACCGTAAAAGACAT 3′ for cluster I.

Purification of NDK and NM23-H2: The *E. coli ndk* gene was amplified by PCR from *E. coli* genomic DNA and cloned into the *Nco*I and *Hin*dIII sites of the bacterial expression vector pQE60 (QIAGEN). The construct pQE60*ndk* allows the expression of a recombinant *E. coli* NDK protein that is the same as the native *E. coli* NDK protein. The open reading frame for the human *NM23*-H2 gene was amplified by PCR from the Hela cell line S3 cDNA library (Stratagene, La Jolla, CA). Standard site-directed mutagenesis (Sambrook *et al.* 1989) was used to eliminate a *Nco*I restriction site in the *NM23*-H2 coding region, while maintaining the same protein sequence. The *NM23*-H2 gene was then subcloned between the *Nco*I and *Hin*dIII sites of pQE60. The construct pQE60*NM23*-H2 allows the expression of a recombinant human NM23-H2 protein that is the same as the native human NM23-H2 protein. The *ndk* derivative of *E. coli* strain CC107 (see above), CC107*ndk*, was first transformed with the plasmid pREP4 (QIAGEN), which constitutively expresses the Lac repressor protein encoded by the *lacI* gene, to reduce the basal level of expression from pQE60 plasmids. As a second step, the CC107*ndk* strain was transformed with either pQE60*ndk* or pQE60*NM23*-H2. These new constructs were grown at 37° in

TABLE 1

Mutation rates in *ndk* and *mutS* strains

Locus	Rate (μ) per replication $\times 10^8$			
	Wild type	*ndk*	*mutS*	*ndk mutS*
rpoB (Rif^r)	1.5	27	120	1900
	(1.1–1.7)[a]	(19–44)	(110–140)	(1500–2400)
gyrB (Nal^r)	0.25	4.8	35	310
	(0.01–0.28)	(2.6–6.4)	(27–40)	(250–340)

The mutation rate (μ) was determined in strain CC107 by the method of Drake (1991; see materials and methods).

[a] Values in parentheses are 95% confidence limits.

LB medium (Miller 1992) with 100 μg/ml ampicillin and 25 μg/ml kanamycin.

RESULTS

Mutation rates in strains lacking NDK and MMR: *Base substitutions:* We detected inserts of miniTn*10* in the *ndk* gene after a random search for weak frameshift mutators and verified their structure by DNA sequencing (see materials and methods). We prepared double mutants, *ndk mutS*, by P1 transduction and assayed them initially for increases in Rif^r and Nal^r mutant frequencies. Table 1 presents these data in terms of mutation rate per replication, using the method of Drake (1991; see materials and methods). Strains that are *ndk* have rates of *rpoB* (Rif^r) and *gyrB* (Nal^r) mutations modestly higher than those of wild type, as reported by Inouye and coworkers (Lu *et al.* 1995). Strains that are *mutS* and thus lack mismatch repair have greater increases in mutation rates using these assays. However, what is so striking is the increase in *ndk mutS* double mutants. The mutation rates in the double mutant are 16-fold higher than those found in a *mutS* strain, when the Rif^r assay is used, and 9-fold higher when the Nal^r assay is used.

Base substitutions in lacZ: To analyze further the specificity of base substitutions generated in *ndk* and *ndk mutS* strains, we first looked at different derivatives of the series of strains CC101–CC106, which measure each of the specific base substitutions at one site in the *lacZ* gene (Cupples and Miller 1989). Table 2 shows these results. In *ndk* strains, A:T → T:A transverisons are the most prominent, although marginal effects are seen with other base substitutions. A:T → G:C transitions are elevated 10-fold at the one site examined (CC106), but the absolute rate of these mutations is still very low compared with other rates (see below). However, in *ndk mutS* double mutants A:T → G:C transitions are clearly elevated over the levels found in *mutS* strains alone. The level of A:T → G:C and A:T → TA changes in *ndk mutS* double mutants appears too low to account for the mutations in *rpoB* that generate Rif^r mutants (Table 1), even when one takes into account the fact that a number of sites in *rpoB* can be mutated to yield the Rif^r phenotype (Jin and Gross 1988). Therefore, we analyzed the sequence changes that result in the Rif^r phenotype, as described below.

Base substitutions in rpoB: Sequence studies from a number of laboratories have analyzed mutations in *rpoB* leading to Rif^r. Jin and Gross (1988) and subsequently Severinov *et al.* (1993) have compiled the sequence changes that they and others detected so far. Virtually all of these fall into two nearby subregions of the >4-kb *rpoB* gene, permitting sequencing by a single pair of primers. These previous studies identified 34 single-base substitutions at 24 sites (specific nucleotide pairs) distributed among 18 coding positions. Our work, described below, has detected an additional 13 single-base substitutions, 5 new sites, and 3 new coding positions. With these additions, the *rpoB*/Rif^r system now allows the monitoring of 47 base substitutions at 29 sites distributed among 21 coding positions. Table 3 lists the base changes arranged according to the category of substitution. This system tests for A:T → G:C changes at 8 sites, G:C → A:T changes at 12 sites, A:T → TA changes at 7 sites, A:T → C:G changes at 7 sites, G:C → T:A changes at 10 sites, and G:C → C:G changes at 3 sites. (A more complete description of the extended *rpoB*/Rif^r system will be published elsewhere.)

We sequenced the *rpoB* mutations responsible for the Rif^r phenotype in a total of >200 independent mutants from the following four backgrounds: wild type, *ndk*, *mutS*, and *mutS ndk*. The results are shown in Table 3. In the data set from the wild-type strain, all types of mutations are detected, although the A:T → G:C change at one particular site (base 1547) is the most prominent. In the other data sets, this is the predominant mutation detected. The percentage of mutations at this hotspot (A:T → G:C, 1547) varies from 15% of the spontaneous base substitutions detected to 77% in the *ndk* background. Although G:C → A:T changes are seen in the spontaneous and *mutS* backgrounds, none of these appear in the *ndk* or the *mutS ndk* backgrounds. Table 4 shows the surrounding sequence for each of the sites

TABLE 2

Reversion rate to Lac$^+$ in *ndk* and *mutS* strains

Lac$^-$ strain	Base change	Rate (μ) per replication $\times 10^9$			
		Wild type	*ndk*	*mutS*	*ndk mutS*
CC101	AT → CG	0.9 (0.4–1.6)	1.5 (0.9–1.8)	3.5 (1.1–8.5)	0.9 (0.3–2.3)
CC102	GC → AT	1.8 (0.7–2.3)	2.5 (1.5–3.4)	150 (110–210)	160 (130–200)
CC103	GC → CG	<0.1	<0.1	<0.1	<0.1
CC104	GC → TA	3.2 (2.6–6.6)	4.3 (3.1–5.4)	8.9 (6.4–13)	4.0 (2.8–5.2)
CC105	AT → TA	1.1 (1.5–1.7)	59 (36–96)	2.9 (2.0–4.5)	55 (44–80)
CC106	AT → GC	0.13 (0.11–0.17)	1.4 (0.6–3.6)	35 (27–44)	130 (110–180)

Values in parentheses are 95% confidence limits.

at which A:T → G:C substitutions are measured. Both the hotspot site and the second most represented A:T → G:C site have a 5′-G as nearest neighbor to the T that changes to a C. This 5′-G preference has been seen before for T → C transitions in a MMR$^-$ stain. (Schaaper and Dunn 1987; see discussion).

Frameshifts: We originally detected the miniTn*10* insertion in *ndk* during a search for frameshift mutators. Table 5 provides mutation rates for frameshifts using strains CC107, CC108, CC110, and CC111. These strains revert from the addition or loss of a single base pair from a run of six G's or six A's (see Table 7 and Cupples *et al.* 1990). The biggest effect is seen in strain CC107, which reverts from *lacZ* to wild type by the addition of a G to a run of six –G's. In MMR$^-$-deficient strains (*mutS*) frameshifts at repeat-tract sequences such as that found in *lac* in CC107 occur at very high levels as reflected in Table 7. However, even these high levels are elevated an additional 8- to 10-fold in *ndk mutS* double mutants. The levels reported in Table 5 (see also materials and methods) are so high that they can be verified by direct visualization of Lac$^+$ colonies on indicator plates without the use of any selection (data not shown). Strain CC108, which reverts to Lac$^+$ by the loss of a G from a run of six G's, shows smaller increases in a *ndk* and a *mutS ndk* strain. One of two frameshifts at runs of A's is also affected. Thus, CC110 shows an increase, but the absolute rate of frameshifts is low compared to that seen in CC107 and CC108. CC111 does not show a significant increase in the *mutS ndk* strain *vs.* the *mutS* background.

Effect of *mutD* alleles: We tested the effect of a partially impaired polymerase editing function, ε, on an NDP kinase-deficient strain. We previously constructed a set of single-base changes in a *dnaQ* insert on the plasmid pBR329 that result in varying levels of mutator activity (*mutD*; Slupska *et al.* 1998). These effects are seen when the plasmid with the altered *dnaQ* gene is carried by a strain with a normal chromosomal copy of *dnaQ*. Those that generate weak mutators have reduced activity of the editing function. Table 6 shows the effects of one of these alleles, *mutD66*, which results in a glycine-for-histidine substitution at position 66 in the ε subunit. In otherwise wild-type strains, pBR*mutD66* results in a moderate increase in mutations as measured by the Rif^r assay. However, a double mutant that is *ndk*/pBR*mutD66* has significantly increased mutation rates for *rpoB* (Rif^r). We also tested whether the increase in A:T → T:A transversions in a *ndk* strain (Table 2) might be the result of a secondary effect of saturating the *mutD*-encoded (ε) editing function, since A:T → T:A transversions are prominent in a *mutD* strain under certain conditions (Schaaper 1988). Therefore, we measured Lac$^+$ revertants in strain CC105, which reverts only by an A:T → T:A tranversion, in *ndk*, *ndk*/pBR, and WT/pBR*mutD66* strains. As can be seen in Table 6, the *ndk*/pBR*mutD66* strain displays a significant increase in reversion of the *lacZ* mutation in CC105 over that seen in *ndk* or *ndk*/pBR*mutD66* strains alone. Although it is possible that the NDP kinase deficiency generates more polymerase errors that lead directly to A:A (or T:T) mispairs, the data in Table 6 are consistent with the increase in A:T → T:A transversions occurring indirectly as a result of inactivating the editing function (see discussion).

Complementation of *E. coli ndk* with human *nm23-H2*: We expressed native NDK protein from both *E. coli* genomic and human cDNA (see materials and methods), using the inducible bacterial expression vector pQE60 (QIAGEN) and the plasmid pREP4 (QIAGEN) that expresses the Lac repressor protein. Here, expression from pQE60 is increased in the presence of isopro-

TABLE 3

Distribution of mutations leading to Rif^r

Site (bp)	AA change	bp change	Occurrence of mutations			
			Wild type	*ndk*	*mutS*	*ndk mutS*
443[a]	Q148R	AT → GC	3	1	0	0
1532	L511P	AT → GC	0	0	4	3
1534	S512P	AT → GC	6	2	8	7
1538	Q513R	AT → GC	0	0	3	0
1547	D516G	AT → GC	9	46	29	22
1552[a]	N518D	AT → GC	0	0	2	0
1577	H526R	AT → GC	0	0	0	0
1598	L533P	AT → GC	0	0	0	0
1535	S512F	GC → AT	2	0	0	0
1546	D516N	GC → AT	2	0	6	0
1565	S522F	GC → AT	3	0	0	0
1576	H526Y	GC → AT	3	0	0	0
1585[b]	R529C	GC → AT	0	0	0	0
1586	R529H	GC → AT	2	0	1	0
1592	S531F	GC → AT	1	0	1	0
1595	A532V	GC → AT	0	0	0	0
1600[a]	G534S	GC → AT	0	0	3	0
1691	P564L	GC → AT	4	0	0	0
1721	S574F	GC → AT	0	0	0	0
2060[b]	R687H	GC → AT	0	0	0	0
443[a]	Q148L	AT → TA	6	4	0	0
1538	Q513L	AT → TA	1	2	0	0
1547	D516V	AT → TA	0	0	0	1
1568	E523V	AT → TA	0	0	0	0
1577[a]	H526L	AT → TA	0	0	0	0
1598[a]	L533H	AT → TA	1	0	0	0
1714	I572F	AT → TA	3	0	0	0
1525[a]	S509R	AT → CG	1	0	0	0
1532[a]	L511Q	AT → CG	0	1	0	0
1538	Q513P	AT → CG	0	0	0	0
1577	H526P	AT → CG	1	0	0	0
1687[b]	T563P	AT → CG	0	0	0	0
1714[a]	I572L	AT → CG	7	1	0	0
1715[a]	I572S	AT → CG	0	2	0	0
1537	Q513K	GC → TA	0	0	0	0
1546	D516Y	GC → TA	1	0	0	0
1576[a]	H526N	GC → TA	1	0	0	0
1578	H526Q	GC → TA	0	0	0	0
1585[b]	R529S	GC → TA	0	0	0	0
1586	R529L	GC → TA	0	0	0	0
1592	S531Y	GC → TA	0	0	0	0
1595	A532E	GC → TA	0	0	0	0
1600[a]	G534C	GC → TA	2	0	0	0
1601	G534D	GC → TA	0	0	0	0
1574	T525R	GC → CG	0	1	0	0
1576[a]	H526D	GC → CG	3	0	0	0
1578	H526Q	GC → CG	0	0	0	0
Total			62	60	57	33

The DNA sequence change in *rpoB* was determined in each case.

[a] Mutations not described previously in Jin and Gross (1989), Severinov *et al.* (1993), or Rangarajan *et al.* (1997).

[b] Mutations that lead to a temperature-sensitive phenotype between 30° and 42° and may not have been detected at 37°.

TABLE 4

Sequence surrounding sites of A:T → G:C substitutions

Site (bp)	5′ sequence 3′
1532	5′ CCAGC**T**GTCTC 3′ 3′ GGTCG**A**CAGAG 5′
1534	5′ AGCTG**T**CTCAG 3′ 3′ TCGAC**A**GAGTC 5′
1538	5′ TAAAC**T**GAGAC 3′ 3′ ATTTG**A**CTCTG 5′
1547	5′ TCTGG**T**CCATA 3′ 3′ AGACC**A**GGTAT 5′
1552	5′ GTTGT**T**CTGGT 3′ 3′ CAACA**A**GACCA 5′
1577	5′ GTTTG**T**GCGTA 3′ 3′ CAAAC**A**CGCAT 5′
1598	5′ CGCAC**T**CGGCC 3′ 3′ GCGTG**A**GCCGG 5′

The sequences in *rpoB* have been oriented to show the base 5′ to the T undergoing the T → C change.

pyl thiogalactoside (IPTG). Table 7 shows the mutation rate for *rpoB* mutations leading to Rif^r in strains lacking a functional chromosomal *ndk* gene. With the pQE60 lacking an insert, the mutation rate is high in the presence and absence of IPTG, since the NDK deficiency leads to a modest mutator phenotype (see Table 1). With the pQE60 carrying the *E. coli ndk* gene, the mutator effect is eliminated. The residual expression of the *ndk* gene on pQE60 in the absence of IPTG is still enough to provide sufficient NDK function. With the pQE60 carrying the human *nm23*-H2, the mutator effect is only partially overcome in the absence of IPTG, but completely nullified in the presence of IPTG, since the increased expression provides enough NDK function. This experiment shows that the human enzyme can provide NDK function to *E. coli*.

DISCUSSION

Understanding pathways of mutagenesis is relevant to the study of repair systems and to certain human diseases. As a tool, mutators, strains with a higher than normal mutation rate, have played an important role in the elucidation of DNA repair systems and the characterization of new pathways of mutagenesis, both in bacteria and in higher cells (see review by Miller 1998). For instance, the definition of *mutH*, *mutL*, *mutS*, and *mutU* (now termed *uvrD*) has been instrumental in understanding the mismatch-repair system in *E. coli* (Radman and Wagner 1986; Modrich and Lahue 1996; Modrich 1997). Subsequent studies identified the human counterpart to this system and showed its involvement in the inherited form of colon and ovarian cancer (Fishel *et al.* 1993; Leach *et al.* 1993; see also Kolodner 1996; Modrich 1997). The importance of this area of research is underscored by the fact that some tumor lines are mutators (Aaltonen *et al.* 1993; Leach *et al.*, 1993; Peltomäki *et al.* 1993; Rampino *et al.* 1997), and several authors have postulated that an early step in the origin of a cancer cell is the creation of a mutator phenotype (Nowell 1976; Loeb 1991), since this would accelerate the accumulation of the five to seven mutations needed to result in a complete tumor cell. Thus, up to 30% of colon cancer lines are mutators lacking the mismatch-repair system. Detecting new pathways of mutagenesis or mutagenesis enhancement can thus open up new avenues of research into the causes of cancer.

It is well established that perturbing the dNTP pool balance can result in increased mutation and is relevant to human disease (see review by Kunz *et al.* 1994). In *E. coli*, the *ndk* gene encodes NDP kinase that plays a

TABLE 5

Reversion rate to Lac$^+$ in *ndk* and *mutS* strains

		Rate (μ) per replication $\times 10^8$			
Lac$^-$ strain	Base change	Wild type	*ndk*	*mutS*	*ndk mutS*
CC107	6 G's → 7 G's (+1)	8.9 (7.4–12)	74 (59–120)	3,600 (3,000–3,900)	28,000 (24,000–42,000)
CC108	6 G's → 5 G's (−1)	7.3 (6.0–9.4)	23 (13–29)	1,900 (1,800–2,100)	4,500 (3,300–5,000)
CC110	6 A's → 7 A's (+1)	0.5 (0.2–0.8)	0.5 (0.4–1.0)	38 (34–115)	160 (110–230)
CC111	6 A's → 5 A's (−1)	2.8 (2.3–4.2)	7.9 (5.9–9.5)	320 (270–370)	400 (220–470)

Base additions (+1) or deletions (−1) that lead to the Lac$^+$ phenotype are scored. Values in parentheses indicate 95% confidence limits.

TABLE 6

Mutation rates in *ndk* and *mutD* (*dnaQ*) strains

Locus	Wild type	Rate (μ) per replication $\times 10^8$			
		ndk	*ndk*/pBR	WT/pBR*mutD66*	*Ndk*/pBR*mutD66*
rpoB (Rif^r)[a]	1.5 (1.1–1.7)	27 (19–44)	40 (17–58)	32 (26–49)	470 (400–760)
lacZ (Lac$^+$)[b] AT → TA	0.11 (0.05–0.17)	5.9 (3.6–9.6)	2.1 (1.3–3.3)	6.2 (5.3–9.6)	140 (100–190)

Wild-type and *ndk* strains are shown. In some cases these strains carry either pBR329 or pBR329 with a mutated *dnaQ* gene (pBR329*mutD66*). Values in parentheses are 95% confidence limits.

key role in maintaining the triphosphate pool (see Lu *et al.* 1995). This enzyme is highly conserved from bacteria to Drosophila to humans and appears to be involved in a number of important cellular processes. For instance, in Drosophila, defects at the NDP kinase-encoding *awd* locus result in abnormal development (Biggs *et al.* 1988, 1990). In humans, a family of NDP kinase-encoding genes (*nm23*-H1 to *nm23*-H8) is expressed in a tissue-specific manner and is also postulated to play a role in normal development and in some cases a direct involvement in the proliferation of tumors (Lacombe *et al.* 2000; Lombardi *et al.* 2000). *In vivo* experiments have shown that *nm23*-H1 suppresses metastasis of human breast carcinoma cells (Leone *et al.* 1993a), and point mutations have been found in *nm23*-H1 and *nm23*-H2 in several tumor cell lines (Leone *et al.* 1993b; Chang *et al.* 1994; Hamby *et al.* 1995). The *nm23*-H2-encoded NDP kinase is also a regulator of at least the *c-MYC* oncogene (Postel *et al.* 1993, 2000) and makes a sequence-specific cleavage at the *c-MYC* promoter sequence (Postel *et al.* 2000). On the basis of sequence homologies with certain glycosylases, it has been postulated that the NM23-H2/NDP kinase may function as a repair protein in humans (Postel *et al.* 2000). The NM23-H1/NDP kinase has been shown to regulate Rac1 GTPase by interaction with Tiam1 (Otsuki *et al.* 2001).

We have examined the involvement of the *E. coli ndk* gene in mutagenesis. Earlier studies showed a moderate mutator activity for base substitutions of NDP kinase-deficient strains in *E. coli* (Lu *et al.* 1995). We detected minitransposon inserts in the *E. coli ndk* gene as part of a search for frameshift mutators (see materials and methods) and found that in the absence of mismatch repair NDP kinase-deficient strains (*mutS ndk*) generate extremely high levels of both base substitutions and frameshifts that reflect mutation rates 16- and 8-fold higher, respectively, than those seen in MMR-deficient (*mutS*) strains alone and rates 75-fold higher than those seen in MMR-proficient *ndk* strains (see Tables 1 and 5). What can account for this very high level of mutator activity? Is this due simply to the effects of altered nucleotide triphosphate pool sizes on polymerase errors or to the lack of a repair activity provided by the NDP kinase? We tested purified native *E. coli* NDK protein, as well as purified native human NM23-H2 protein for glycosylase activity, using a complete collection of single-base mispairs, as well as several frameshift mispairs, and failed to detect any glycosylase activity on these mispairs (H.-J. Yang and J. H. Miller, unpublished results). On the other hand, several lines of evidence point to the combined effects of increased mispairs caused by the change in nucleotide triphosphate pools and the subsequent partial saturation of the *dnaQ/mutD*-encoded editing subunit (ε) of DNA polymerase.

Using a reversion system that detects specific base substitutions or frameshifts in the *lacZ* gene (Cupples and Miller 1989; Cupples *et al.* 1990), we found that in otherwise wild-type backgrounds, only the A:T → T:A and the A:T → G:C changes are significantly increased in *ndk* strains (see Table 2). In a MMR$^-$ (*mutS*) background, *ndk* strains display a much greater increase of A:T → G:C transitions. No such effect of *ndk* was found for G:C → A:T transitions in a *mutS* background. However, the *lacZ* system utilizes only one site for each of the six base substitutions. Because of the possibility of surrounding sequence effects on mutation rate, we decided to investigate a wider spectrum of base substitution mutations by sequencing a portion of the *rpoB* gene in Rif^r mutants. This *rpoB*/Rif^r system has been developed and improved by several groups (Jin and Gross 1988; Severinov *et al.* 1993 and references therein)

TABLE 7

Complementation of *E. coli ndk* with human *nm23*-H2

Plasmid in CC107 *ndk*	IPTG	Mutation rate (μ) $\times 10^8$ (95% confidence limits)
pREP4/pQE60	−	47 (35–56)
pREP4/pQE60	+	30 (24–75)
pREP4/pQE60 *ndk*	−	0 (0–5.6)
pREP4/pQE60 *ndk*	+	0 (0–5.9)
pREP4/pQE60 *nm23*-H2	−	10 (6.7–19)
pREP4/pQE60 *nm23*-H2	+	0 (0–2.3)

and has been extended by the work reported here and elsewhere (Garibyan *et al.* 2002). This is a system that now offers a number of interesting advantages for mutational analysis. First, the mutations responsible for the Rif^r phenotype are virtually all within a small segment of the *rpoB* gene that can be amplified and sequenced with two primer pairs. Second, the locus is on the chromosome, eliminating any variables in using mutational sites on an extrachromosomal element such as an F′ plasmid. Third, only mutants in the culture at the time of plating are scored, since Rif^r colonies are not generated by events taking place after plating. The current extended system measures all six possible base substitutions distributed among 47 single-base substitution events at 29 sites (base pairs) with different sequence contexts (see Tables 3 and 4). The region of the *rpoB*-encoded β-subunit of RNA polymerase that contains the binding site for rifampicin is very highly conserved among several microorganisms, including pathogens such as certain mycobacteria (see review by Musser 1995). The Rif^r mutations in *M. tuberculosis* and *M. smegmatis* are at sites equivalent to their counterparts in *E. coli* (Musser 1995; Karunakaran and Davies 2000).

The mutations detected by DNA sequencing of Rif^r mutants (Table 3) show that whereas spontaneous mutations are distributed among many sites, 49 of 60 (82%) of the mutations seen in an *ndk* strain and 32/33 (97%) of the mutations in an *ndk mutS* strain are A:T → G:C transitions. All of the mutations seen in the sequenced sample from the *mutS* strain are transitions, of which 46/57 (81%) are A:T → G:C, and 11/57 (19%) are G:C → A:T). Although G:C → A:T transitions are found in the *mutS* spectrum, they are absent from the *ndk* and the *mutS ndk* spectrum. This propensity for increased A:T → G:C transitions in *ndk* and *mutS ndk* strains is consistent with increased pool sizes of dCTP and dGTP found in ndk strains (Lu *et al.* 1995). *In vitro* studies have shown that when the concentrations of dGTP and dCTP were increased by 20-fold, the rate of stable misincorporation of dCMP and dGMP opposite adenine and thymine nucleotides increased (Roberts and Kunkel 1988). One particular hotspot site in *rpoB* (Table 3) represents 51% of all the mutations in the *mutS*, 77% of those in the *ndk*, and 67% of those in the *mutS ndk* spectrum. This and the second most prominent site in these spectra, also A:T → G:C, have a 5′ G as the nearest neighbor to the T that is changed to a C (Table 4). This same preference for a 5′ G nearest neighbor has been reported for *lacI* mutations in a *mutS* strain (Schaaper and Dunn 1987) and also for *lacI* mutations in *mutD5* grown under conditions where the MMR system is saturated (Schaaper 1988). This latter result indicates that the preference for the 5′ G emanates from the polymerase errors themselves rather than being imprinted by the editing function or MMR.

The A:T → T:A substitutions seen in *ndk* strains (see Tables 2 and 3) would be difficult to explain by the 20-fold increase in dCTP pools or the 7-fold increase in dGTP pools reported by Inoue and co-workers for *ndk* strains (Lu *et al.* 1995). On the other hand, the A:T → T:A transversion is one of the signatures of a *mutD* strain, being the most prominent base change in the spectrum of *lacI* mutations in a *mutD* strain under certain conditions (Schaaper 1988). The experiment depicted in Table 6 shows that a *mutD* allele (*mutD66*) causing a partial mutator phenotype (Slupska *et al.* 1998) results in a mutator effect much greater in strains that are *ndk* than that seen in either *ndk* or *mutD66* strains alone. This suggests that the defect in *ndk* strains may hinder proofreading. It is possible that an increase in the pool size of the next correct nucleotide might increase the rate of polymerization and thus inhibit proofreading (Mendelman *et al.* 1990; Roberts *et al.* 1991).

The finding that strains lacking NDP kinase are hypermutable in mismatch-repair backgrounds provides an example of synergistic effects of different alleles, in this case *ndk* and *mutS*. Given the involvement of both NDP kinase deficiency (see Lacombe *et al.* 2000 and Lombardi *et al.* 2000) and mismatch-repair deficiency (Fishel *et al.* 1993; Leach *et al.* 1993) in human cancers, we can ask the question of whether alleles giving weaker effects by themselves might still produce strong effects in the presence of certain other alleles or conditions. Thus, certain alleles of *mutS* might result in high mutation rates only in the presence of altered pool sizes caused by growth conditions or by genetic inactivation of NDP kinase. The challenge is to be able to detect mutations that result in cancer susceptibilities only in certain genetic backgrounds or only under certain conditions. One example is the finding that a single-nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 carriers (Levy-Lahad *et al.* 2001). An example of how environmental challenges can unmask susceptibilities by hidden alleles is provided by the variation in benzo[a]pyrene diol epoxide susceptibilities in controls and lung carcinoma patients (Wu *et al.* 1998).

We thank Tomas Lindahl and Edith Postel for helpful discussions. J.H.M. was supported by a grant from the National Institutes of Health (ES-0110875).

LITERATURE CITED

Aaltonen, L. A., P. Peltomäki, F. S. Leach, P. Sistonen, L. Pylkkänen *et al.*, 1993 Clues to the pathogenesis of familial colorectal cancer. Science **260:** 812–816.

Biggs, J., N. Tripoulas, E. Hersperger, C. Dearolf and A. Shearn, 1988 Analysis of the lethal interaction between the prune and Killer of prune mutations of Drosophila. Genes Dev. **2:** 1333–1343.

Biggs, J., E. Hersperger, P. S. Steeg, L. A. Liotta and A. Shearn, 1990 A Drosophila gene that is homologous to a mammalian gene associated with tumor metastasis codes for a nucleoside diphosphate kinase. Cell **63:** 933–940.

Chang, C. L., X. X. Zhu, D. H. Thoraval, D. Ungar, J. Rawwas *et al.*, 1994 Nm23–H1 mutation in neuroblastoma. Nature **370:** 335–336.

Cupples, C. G., and J. H. Miller, 1989 A set of *lacZ* mutations in

Escherichia coli that allow rapid determination of each of the six base substitutions. Proc. Natl. Acad. Sci. USA **86:** 5345–5349.

Cupples, C. G., M. Cabrera, C. Cruz and J. H. Miller, 1990 A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of specific frameshift mutations. Genetics **125:** 275–280.

Dixon, W. J., and F. J. Massey, Jr., 1969 *Introduction to Statistical Analysis.* McGraw-Hill, New York.

Drake, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins *et al.*, 1993 The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell **75:** 1027–1038.

Funchain, P., A. Yeung, J. L. Stewart, R. Lin, M. M. Slupska *et al.*, 2000 The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics **154:** 959–970.

Garibyan, L., A. Nguyen, T. Huang, T. Nguyen, A. Diep *et al.*, 2002 Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. DNA Repair (in press).

Hamby, C. V., C. E. Mendola, L. Potla, G. Stafford and J. M. Backer, 1995 Differential expression and mutation of NME genes in autologous cultured human melanoma cells with different metastatic potentials. Biochem. Biophys. Res. Commun. **211:** 579–585.

Higashitani, A., N. Higashitani, S. Yasuda and K. Horiuchi, 1994 A general and fast method for mapping mutations on the *Escherichia coli* chromosome. Nucleic Acids Res. **22:** 2426–2427.

Jin, D. J., and C. A. Gross, 1988 Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. J. Mol. Biol. **202:** 45–58.

Karunakaran, P., and J. Davies, 2000 Genetic antagonism and hypermutability in *Mycobacterium smegmatis.* J. Bacteriol. **182:** 3331–3335.

Kolodner, R., 1996 Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev. **10:** 1433–1442.

Kunz, B., A. Kohalmi, S. Kohalmi, T. A. Kunkel, C. K. Mathews *et al.*, 1994 Deoxyribonucleoside triphosphate levels: a critical factor in the maintenance of genetic stability. Mutat. Res. **318:** 1–64.

Lacombe, M. L., L. Milon, A. Munier, J. G. Mehus and D. O. Lambeth, 2000 The Human Nm23/nucleoside diphosphate kinases. J. Bioenerg. Biomembr. **32:** 247–258.

Leach, F. S., N. C. Nicolaides, N. Papadopoulos, B. Liu, J. Jen *et al.*, 1993 Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. Cell **75:** 1215–1225.

Leone, A., U. Flatow, C. R. King, M. A. Sandeen, I. M. Margulies *et al.*, 1991 Reduced tumor incidence, metastatic potential, and cytokine responsiveness of nm23-transfected melanoma cells. Cell **65:** 25–35.

Leone, A., U. Flatow, K. VanHoutte and P. S. Steeg, 1993a Transfection of human nm23–H1 into the human MDA-MB-435 breast carcinoma cell line: effects on tumor metastatic potential, colonization and enzymatic activity. Oncogene **8:** 2325–2333.

Leone, A., R. C. Seeger, C. M. Hong, Y. Y. Hu, M. J. Arboleda *et al.*, 1993b Evidence for nm23 RNA overexpression, DNA amplification and mutation in aggressive childhood neuroblastomas. Oncogene **8:** 855–865.

Levy-Lahad, E., A. Lahad, S. Eisenberg, E. Dagan, T. Paperna *et al.*, 2001 A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 carriers. Proc. Natl. Acad. Sci. USA **98:** 3232–3236.

Loeb, L. A., 1991 Mutator phenotype may be required for multistage carcinogenesis. Cancer Res. **51:** 3075–3079.

Lombardi, D., M. L. Lacombe and M. G. Paggi, 2000 nm23: unraveling its biological function in cell differentiation. J. Cell. Physiol. **182:** 144–149.

Lu, Q., X. Zhang, N. Almaula, C. K. Mathews and M. Inouye, 1995 The gene for nucleoside diphosphate kinase functions as a mutator gene in *Escherichia coli.* J. Mol. Biol. **254:** 337–341.

Mendelman, L. V., J. Petruska and M. F. Goodman, 1990 Base mispair extension kinetics. Comparison of DNA polmerase α and reverse transcriptase. J. Biol. Chem. **265:** 2338–2346.

Miller, J. H., 1992 *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Miller, J. H., 1998 Mutators in *Escherichia coli.* Mutat. Res. **409:** 99–106.

Modrich, P., 1997 Strand-specific mismatch repair in mammalian cells. J. Biol. Chem. **272:** 24727–24730.

Modrich, P., and R. Lahue, 1996 Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. **65:** 101–133.

Musser, J. M., 1995 Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin. Microbiol. Rev. **8:** 496–514.

Nowell, P. C., 1976 The clonal evolution of tumor cell populations. Science **194:** 23–28.

Otsuki, Y., M. Tanaka, S. Yoshii, N. Kawazoe, K. Nakaya *et al.*, 2001 Tumor metastasis suppressor nm23H1 regulates Rac1 GTPase by interaction with Tiam1. Proc. Natl. Acad. Sci. USA **98:** 4385–4390.

Peltomäki, P., R. A. Lothe, L. A. Aaltonen, L. Pylkkänen, M. Nyström-Lahti *et al.*, 1993 Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Res. **53:** 5853–5855.

Postel, E. H., S. J. Berberich, S. J. Flint and C. A. Ferrone, 1993 Human c-myc transcription factor PuF identified as nm23–H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. Science **261:** 478–480.

Postel, E. H., B. M. Abramczyk, M. N. Levit and S. Kyin, 2000 Catalysis of DNA cleavage and nucleoside triphosphate synthesis by NM23–H2/NDP kinase share an active site that implies a DNA repair function. Proc. Natl. Acad. Sci. USA **97:** 14194–14199.

Radman, M., and R. Wagner, 1986 Mismatch repair in *Escherichia coli.* Annu. Rev. Genet. **20:** 523–538.

Rampino, N., H. Yamamoto, Y. Ionov, Y. Li, H. Sawai *et al.*, 1997 Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science **275:** 967–969.

Rangarajan, S., G. Gudmundsson, Z. Qui, P. L. Foster and M. F. Goodman, 1997 *Escherichia coli* DNA polymerase II catalyzes chromosomal and episomal DNA synthesis *in vivo.* Proc. Natl. Acad. Sci. USA **94:** 946–951.

Roberts, J. D., and T. A. Kunkel, 1988 Fidelity of a human cell DNA replication complex. Proc. Natl. Acad. Sci. USA **85:** 7064–7068.

Roberts, J. D., D. C. Thomas and T. A. Kunkel, 1991 Exonucleolytic proofreading of leading and lagging strand DNA replications errors. Proc. Natl. Acad. Sci. USA **88:** 3465–3469.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schaaper, R. M., 1988 Mechanisms of mutagenesis in the *Escherichia coli* mutator mutD5: role of DNA mismatch repair. Proc. Natl. Acad. Sci. USA **85:** 8126–8130.

Schaaper, R. M., and R. L. Dunn, 1987 Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. Proc. Natl. Acad. Sci. USA **84:** 6220–6224.

Scheuermann, R. H., and H. Echols, 1984 A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **81:** 7747–7751.

Severinov, K., M. Soushko, A. Goldfarb and V. Nikiforov, 1993 Rifampicin region revisited. New rifampicin-resistant and streptolydigin-resistant mutants in the beta subunit of *Escherichia coli* RNA polymerase. J. Biol. Chem. **268:** 14820–14825.

Slupska, M. M., A. G. King, L. I. Lu, R. H. Lin, E. F. Mao *et al.*, 1998 Examination of the role of DNA polymerase proofreading in the mutator effect of miscoding tRNAs. J. Bacteriol. **180:** 5712–5717.

Slupska, M. M., J. H. Chiang, W. M. Luther, J. L. Stewart, L. Amii *et al.*, 2000 Genes involved in the determination of the rate of inversions at short inverted repeats. Genes Cells **5:** 425–437.

Wu, X., J. Gu, C. I. Amos, H. Jiang, W. K. Hong *et al.*, 1998 A parallel study of *in vitro* sensitivity to benzo[a]pyrene diol epoxide and bleomycin in lung carcinoma cases and controls. Cancer **83:** 1118–1127.

Communicating editor: H. Ochman