Critical Reviews in Biochemistry and Molecular Biology, 43:89–134, 2008

ISSN: 1040-9238 print / 1549-7798 online
DOI: 10.1080/10409230801921262

informa
healthcare

The Bacterial Chromosome

Milton H. Saier, Jr.
Division of Biological Sciences, University of California at San Diego, La Jolla, California, USA

Bacteria represent the vast majority of biological diversity found on Earth. In this review, we focus on selected aspects of their genetic material, those providing insight into structural, functional, dynamic, and evolutionary aspects of their genomes. Bacterial chromosomes are far more dynamic than previously realized, and dozens of mechanisms giving rise to genomic plasticity are now understood. Maturation of the genomics era has provided the tools for unraveling the interwoven details of DNA structure/function relationships that provide a basis for organismal diversity. Some of the most throughly understood processes that underlie the dynamics of genomic structure and function in prokaryotes are examined.

Keywords bacteria, DNA, chromosome, compaction, nucleoid proteins, supercoiling, genetic instability, phase variation, immune evasion

Table of Contents

Address correspondence to Milton H. Saier, Jr., Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116. E-mail: msaier@ucsd.edu

INTRODUCTION

In 1990, Krawiec and Riley (1990) published an authoritative review entitled "Organization of the Bacterial Chromosome." This lengthy review summarized evidence then available concerning structural, dynamic, and functional aspects of bacterial chromosomes with emphasis on the *Escherichia coli* chromosome, by far the best-characterized chromosome of all living organisms. Since the advent of the genomics era (Fleischmann *et al.*, 1995; Saier, 1998), we have gained a much richer view of the bacterial genome, making an update long overdue.

FORCES THAT SHAPE BACTERIAL CHROMOSOMES

Prokaryotic chromosomes are usually small and compact, while eukaryotic chromosomes are large and cluttered with extensive noncoding regions (Lawrence, 2002). Various explanations have attempted to account for this phenomenon. One suggests that bacteria already have approached perfection, having optimized their lifestyles through evolution, while eukaryotes are still imperfect, and consequently have devised approaches that promote rapid evolution. The latter organisms therefore benefit from the presence of increased numbers of unstable genetic elements such as introns, transposons, DNA repeats, and microsatellite DNA (Buschiazzo and Gemmell, 2006) that increase the probability of genetic rearrangements more than in prokaryotes (Noor and Chang, 2006).

Interestingly, if a foreign piece of microsatellite DNA with multiple repeat sequences is put into a bacterium such as *E. coli*, over many generations it gets smaller and smaller, but when transferred to a eukaryote such as the yeast, *Saccharomyces cerevisiae*, the same DNA gets bigger and bigger (Metzgar *et al.*, 2002). This has been suggested to be due to differences in the mismatch repair systems in prokaryotes versus eukaryotes (Metzgar *et al.*, 2002). Moreover, evidence suggests that archaeal genes and genomes are even more compact than those of bacteria (Chung *et al.*, 2001). These observations suggest that fundamentally different mechanistic approaches may have been taken by prokaryotes (cells without true nuclei) and eukaryotes (cells with true nuclei) to deal with stress and optimize fitness. Through natural selection, bacteria and archaea only retain the DNA that benefits them, but eukaryotes carry a lot of "excess baggage." That is, while prokaryotes and archaea streamline their DNA whenever possible, eukaryotes do not (Moran, 2002). The primary aim of biological research is to provide explanations and to understand the various ramifications of these phenomena.

While most chromosomal genes contribute to the fitness of an organism (Welch *et al.*, 2002), some may confer no benefit or may even be deleterious. Such elements can be maintained by "selfish" or "addictive" mechanisms. These elements reduce fitness if lost, as is true for certain bacterial "death modules" (Lewis, 2000). Others are infectious in the sense that they spread faster than they can be removed, either due to rapid replication, or to rapid horizontal transfer. Insertion sequence (IS) DNA, composite transposons, plasmids, phage, pathogenicity islands, integrons, and other potential intra/extra-chromosomal elements provide examples (Manson and Gilmore, 2006; Mazel, 2006; Schmidt and Anderson, 2006; Siguier *et al.*, 2006).

Some essential genes code for products (stable RNAs and proteins) that are required in large amounts. These genes are expressed under the control of powerful promoters and are located near the origin of chromosomal replication, *oriC*. Further, their direction of transcription is the same as that of replication. Because these features are observed in essentially all bacteria, their occurrences cannot be due to chance (Berlyn, 1998; Perna *et al.*, 2001). Their arrangements must provide advantages to the organism and must have been selected throughout evolutionary time.

In rapidly growing bacteria, chromosomal replication requires more time than does division. There may be as many as four duplicated copies of the origin of replication, *oriC*, when only one copy of the replication terminator site (*ter*) exists. This increase in gene dosage would most benefit the organism if genes whose products are needed in large amounts are present near *oriC*. Moreover, if transcription is in the direction opposite from replication, a collision between polymerases will occur, and one or the other polymerase must terminate its effort, giving rise to an incomplete product (Liu and Alberts, 1995). Moreover, polymerase collisions can trigger knotting of the daughter duplexes behind the fork (Olavarrieta *et al.*, 2002). However, codirectional collisions do not have such deleterious effects. RNA polymerase resumes faithful RNA chain elongation after the DNA replication fork passes. Amazingly, during such encounters, RNA polymerase can switch position from its original template strand to the newly synthesized daughter strand (Liu and Alberts, 1995).

Collisions between two identical RNA polymerases can also be detrimental. This often results in disengagement of one polymerase, particularly if both are actively transcribing the DNA. However, when one is active and the other is stalled, both enzyme complexes may remain bound to the DNA. Following the collision, one polymerase forces the other to reverse direction, moving backward along the template (Crampton *et al.,* 2006).

Atomic force microscopy, a direct imaging approach, has been used to discriminate between open promoter complexes, elongating complexes, and collided complexes with single molecule resolution (Crampton *et al.*, 2006). Following collision, an elongating RNA polymerase causes stalled or arrested RNA polymerase to backtrack along its template, possibly to the promoter, where dissociation should occur more easily. Interestingly, measurement of the distances between the two RNA polymerase shows that they are not always at closest approach after 'collision' has caused their arrest, suggesting molecular sensing without direct contact. Convergent transcription has been demonstrated in both bacteria and eukaryotes, and its consequence, transcriptional interference, may have regulatory significance (Callen *et al.*, 2004; Crampton *et al.*, 2006).

GENE CLUSTERS

Both prokaryotes and eukaryotes arrange their coding regions in gene clusters or operons that are transcribed from a limited number of transcriptional initiaton sites (Lawrence, 2002; von Hippel, 1988; Young *et al.*, 2002). Early interpretations suggested that the presence of operons could be accounted for primarily by the increased efficiency resulting from coregulation and cotranscription. A secondary benefit to operonic constructions results from the ease of controlling the ratios of the gene products (Ames and Martin, 1964; Demerec and Hartman, 1959; Jacob and Monod, 1962). In the late 1990s, however, Lawrence and Roth (1996, 1999) suggested another reason for operon construction that included genes coding for proteins that together comprise a metabolic pathway or functional unit. They argued that operons are "selfish" in the sense that only if the genes of a pathway occur in apposition or close proximity to each other can they be readily transferred between organisms with retention of function. Thus, they suggested that horizontal transfer (lateral transfer between organisms) represents a selfish arrangement that promotes the maintenance of these genetic elements in diverse populations of bacteria. This advantage provides evolutionary pressure for gene clustering (Lawrence and Roth, 1996, 1999). Such arguments can partially account for the gain and loss of genetic material over evolutionary time (Lawrence and Roth, 1999; Welch *et al.*, 2002). It may incidentally also benefit the recipient cell. (See the chapter by J. Roth in *The Bacterial Chromosome* (2005) for an in-depth consideration of this type of selfish genetic behavior.)

DNA STRUCTURAL ELEMENTS

Early x-ray structural work on synthetic oligonucleotides led to the suggestion that DNA could exist as double helices with variable numbers of nucleotides (nts) per helical turn. Thus, A-DNA had 11 nts per turn, B-DNA had 10, C-DNA had 9 and D-DNA had 8. Additionally, in contrast to the other forms which have right-handed helices, Z-DNA had 12 nts per turn in a left-handed helix. Further analyses revealed subtypes with non-integral numbers of nts per turn. We now know that all of these designations are artificial. Right-handed DNA has a continuum of helical turn lengths with numbers of nts per turn varying from about 8 to 12 with the average being 10.5. It is a common practice to call this native DNA "B" DNA. There is also intrinsically flexible DNA with a preferred bend orientation. These so-called "permanent" bends are determined by the presence of "A-tracts" (Oussatcheva *et al.*, 2004; see Figure 1A). A-tracts have generalized sequences:

$$\{[A]_m\ [N]_{r-m}\}_{6\text{-}12}$$

(N = any nucleotide, m is the number of As and r is the number of nucleotides per helical repeat. 6–12 indicates the number of sequential A-tracts as determined from genome sequence analyses.) Thus, a regular spacing of A-tracts, one every helical turn, all on the same side of the helix, gives rise to a permanent bend. The approximate bend angle is 13.5° per A-tract (Rivetti *et al.*, 1998).

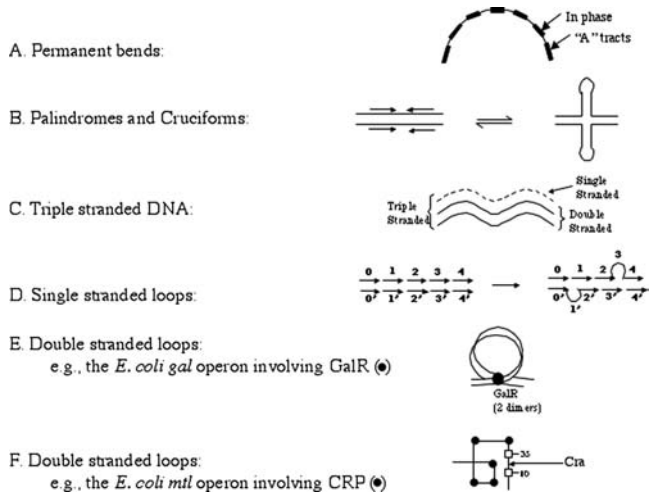

FIG. 1. DNA structural forms. (A) Permanent bends caused by the presence of regularly spaced, in phase A-tracts. (B) Palindromes and cruciforms due to the presence of inverted repeat sequences. (C) Triple stranded DNA due to hydrogen bonding of a third single stranded DNA molecule to a typical double helix. (D) Single stranded loops due to "slipped strand mispairing." (E) Double-stranded loops due to specific high-affinity protein binding: GalR (●) binding in the *E. coli gal* operon. (F) Double-stranded loops: CRP (●) binding in the *E. coli mtl* operon. In this last example, the Cra (catabolite repressor/activator) protein represses while the CRP protein activates gene expression (Ramseier and Saier, 1995).

Another structural anomaly results from the presence of inverted repeats or "palindromes" in the DNA. Inverted repeats can exist in two forms: (1) the basic, more stable double helix, and (2) an alternative, less stable, double stem-loop structure in which bases within the single strands of the palindrome pair with each other (see Figure 1B). Such structures have been shown to exist by using reagents and enzymes (e.g., single strand nucleases) that recognize the single-stranded loops at the top and bottom of the structure shown in Figure 1B. Many DNA binding proteins specifically recognize cruciforms and thereby regulate gene expression.

The basic structure of the DNA is, of course, a double helix, but triple helices can also exist (Figure 1C; Fox and Brown, 2005). The hydrogen bonded base pairing configurations have been determined. These structures are important for the process of homologous recombination, for example, as occur following horizontal gene transfer of single- or double-stranded genetic material from one bacterium to another. In certain instances, they may also play roles in the regulation of gene expression by "antisense" mechanisms (Helene *et al.*, 1997; Neidle, 1997).

Single-stranded loops in double stranded DNA can occur, especially when multiple tandem repeat sequences occur

(Figure 1D). The presence of loops can result in increases and decreases in the numbers of repeat units in the progeny DNA strands (a) following recombination between the loops shown in Figure 1D, (b) as a result of imperfect DNA replication, or (c) during repair processes due to slippage of one of the complementary strands relative to the other (Wells *et al.* 2005). Such changes in the numbers of repeats can affect the structures and antigenic properties of proteins when these elements occur within protein structural genes. They can also change levels of gene expression when they occur in regulatory regions. In humans they are the causes of numerous hereditary neurological diseases (Wells *et al.*, 2005). Such changes in DNA structure are referred to as "slipped strand mispairing" mutations.

DNA duplications, insertions, and deletions provide mechanisms that change the genetic content of a cell. Because slipped strand repeat number alterations occur with high frequency, sometimes a million fold higher than normal mutation rates, the process can give rise to antigenic variation, immune evasion, and drastic and frequent changes in gene expression levels in pathogenic bacteria (Gupta, 2005; van der Woude and Baumler, 2004). They, in part, account for the dynamic nature of chromosomes.

Two examples of double-strand DNA loops, generated by the binding of oligomeric DNA binding proteins, are illustrated in Figures 1E and 1F. One example of double-strand looping that regulates gene expression is the galactose (*gal*) operon of *E. coli* (Figure 1E). The galactose repressor, GalR, is a dimer that binds to a specific consensus sequence in the DNA (Adhya, 1987). There are two binding sites in the *gal* operon, one upstream of the promoter region, and one downstream of this region in the first structural gene, *galE*. The GalR dimer binds to both of them, and then they associate to generate a tetramer bound to a looped DNA molecule. The looped DNA complex is much more efficient in repressing *gal* operon expression than when GalR binds to just one site. A similar looping mechanism in *Bacillus subtilis* appears to account for regulation of the *resABCDE* operon by the dimeric CcpA catabolite repressor, present in many low G+C Gram-positive bacteria (Firmicutes) (Choi and Saier, 2006; Singh *et al.*, 2007).

The mannitol (*mtl*) catabolic operon of *E. coli* (Figure 1F) provides another example of double-stranded loops. Expression at the *mtl* promoter is regulated in response to carbon availability by two carbon-sensing proteins, the cyclic AMP receptor protein (CRP) (also called the catabolite activator protein [CAP]), and the catabolite activator/repressor (Cra) protein (Ramseier and Saier, 1995). The two control proteins regulate *mtl* operon expression but in opposite directions; CRP activates while Cra represses. Cra binds in between and overlapping the –10 and –35 regions of the promoter, blocking RNA polymerase binding, but CRP binds to five different sites upstream of the –35 binding region. These occur at regular intervals, and every CRP dimer bends the DNA 90° when it binds. This gives rise to the close proximity of two of the CRPs at the promoter. CRP thus plays both direct activating roles and indirect structural roles in the activation of this operon.

GENOME SIZE

Bacterial chromosomes may be linear or circular strands of double helical DNA; they comprise most of the genetic material of the cell, the vehicle of inheritance. The sizes of chromosome(s) and the amounts of DNA present in a cell are characteristic of that particular organism or cell type. Table 1 shows the tremendous differences in the amounts of DNA present in the cells of selected prokaryotes and eukaryotes.

Genome sizes in prokaryotes (bacteria and archaea) vary over a range of about 50-fold. Among the smallest genomes are those of the bacterium, *Mycoplasma genitalium* and the archaeon, *Nanoarchaeum equtans*, parasitic and symbiotic organisms, respectively. Their reduced genome sizes correlate with their lifestyles. These organisms grow slowly in constant environments provided by the host organisms, and they consequently lack most proteins concerned with regulation and adaptation. They also lack most biosynthetic and catabolic capabilities since the host provides almost all of their required nutrients, and possibly some essential macromolecules as well!

Recently, the sequenced genomes of two unusual intracellular endosymbiotic bacteria have become available for analysis (Nakabachi *et al.*, 2006; Perez-Brocal *et al.*, 2006). The 420,000 base pair (bp) genome of *Buchnera aphidicola* BCc, one of several aphid endosymbionts of this species, is smaller than any previously sequenced *B. aphidicola* genome. It has lost most metabolic functions including the capability of synthesizing tryptophan and riboflavin, retained by other sequenced aphid endosymbionts. These other strains can synthesize every amino acid not made by the host, although they have lost the capacity to synthesize all others. Thus, host and symbiont cooperate to produce the 20 common amino acids found in proteins since the sap used as a source of nutrition by the aphid is deficient for these metabolites. In addition, *B. aphidicola* BCc is evolving rapidly, probably due to the loss of DNA repair systems. The authors suggest that *B. aphidicola* BCc is losing its symbiotic capabilities because it is being complemented by coexisting symbionts which might be replacing it (Perez-Brocal *et al.*, 2006). *B. aphidicola* BCc may be on the verge of extinction (Andersson, 2006). Indeed, the loss of most DNA repair and protection mechanisms would be expected to enhance mutation rates dramatically.

An even more dramatic example of genome reduction is that of the 160 kbp genome of the γ-proteobacterial endosymbiont, *Carsonella ruddii* Pv (Nakabachi *et al.*, 2006). This organism is present in all species of phloem sap-feeding insects (psyllids) and resides in specialized host cells to which the bacteria are confined. Like organelles, these "bacteriocytes" are maintained in the host by vertical, maternal transmission.

The genome of *C. ruddii* Pv is a single circular chromosome of 160,000 bp with an average G+C content of 16.5%, the lowest yet recorded for any organism. No other symbionts or plasmids

TABLE 1
Approximate genome sizes and gene contents of various prokaryotes and eukaryotes

Organism	Approximate amount of DNA (Mbp)	Estimated # genes	Estimated % coding
Prokaryotes (~50× range in size and # genes)			
Carsonella ruddii Pv	0.16	180	97
Buchnera aphidicola BCc	0.4	400	90
Nanoarchaeum equitans	0.5	500	90
Mycoplasma genitalium	0.5	500	90
Haemophilus influenzae	1.8	1,800	90
Escherichia coli K12 (MG1655)	4.5	4,500	90
Streptomyces coelicolor	9.0	8,000	90
Eukaryotes (~10,000x range in size; 10 to 20x range in # genes)			
Yeast (*Saccharomyces cerevisiae*)	15	6,500	50
Worm (*Carnorhabditis elegans*)	100	20,000	25
Thale cress (*Arabidopsis thaliana*)	160	27,000	21
Rice (*Oryza sativa*)	490	41,000	15
Mammal (*Homo sapiens*)	4,000	~40,000	1
Onion (*Allium cepa*)	16,400	~50,000	0.3
Salamander (*Ambystroma tigrinum)*	31,000	~50,000	0.2
Lungfish (*Protopterus aethiopicus*)	130,000	~50,000	0.05

were detected in the host, rendering it unlikely that through evolution, *C. ruddii* was being replaced by another symbiont. It contains only 182 genes, more than half of which are devoted to translation (35%) and amino acid biosynthesis (18%).

The *Carsonella* Pv genome has a gene density that is higher than that of any other sequenced organism. The protein-coding sequences and RNA genes (one 16S-23S-5S ribosomal RNA operon and 28 tRNA genes for all 20 amino acids) cover 97.3% of the genome. This density is attributable to numerous overlapping genes. Of the 182 genes identified, 164 (90%) overlap with at least one of the 2 adjacent genes, and the average length of all 132 overlaps is 10.7 bases. The majority (92%) are tandem overlaps on the same strand, all of which are out of frame with each other. Moreover, the average length of *Carsonella* genes (826 bp) is notably shorter than that of other bacteria. Indeed, a comparison of 89 orthologous genes conserved in *Carsonella* and in seven bacteriocyte-restricted endosymbionts revealed that the average length of the genes in *Carsonella* is 18% shorter than the average gene lengths of the other endosymbionts. This genome is thus by far the most streamlined studied to date. Its gene inventory seems insufficient for most biological processes that appear to be essential for bacterial life. Possibly the host bacteriocyte compensates.

One of several possible explanations for the absence of genes encoding various bacterium-specific processes is that, as in the case of eukaryotic organelles, some genes were transferred from the genome of a *Carsonella* ancestor to the genome of a psyllid ancestor and are now expressed under the control of the host nucleus. *Carsonella* may be evolving into an organelle! Indeed, its genome size is less than that of plant mitochondria and chloroplasts (Andersson, 2006). Future studies should be directed towards identifying transferred genomes in the host genome.

The human pathogen, *Haemophilus influenzae*, the first organism to have its genome fully sequenced (Fleischmann *et al.*, 1995), has 1.8 Mbp of DNA and about 1800 genes. It has a limited array of biosynthetic, catabolic, and regulatory genes, but cannot grow in a simple medium unless most amino acids and macromolecular precursors are provided. By contrast, *E. coli* with a genome size of 4.5 Mbp and about 4500 genes can make everything it needs to grow if a carbon source and an array of salts are provided. It can make all of the amino acids, nucleobases, sugars and enzyme cofactors that it requires, and it can synthesize all requisite macromolecules.

Streptomyces coelicolor and *Myxococcus xanthus* are large genome organisms that have about twice the amount of genetic material as *E. coli*. They contain genes that allow them to make secondary metabolites such as antibiotics, and they establish programs of differentiation for the formation of multiple cell types. Their larger genomes correlate with their more complex lifestyles and differentiative capabilities. The very largest bacterial genomes are observed for certain cyanobacteria that possess about 15 Mbp of DNA. These organisms are capable of multiple pathways of differentiation. Regardless of the prokaryote, however, roughly 90% of the genome codes for RNA molecules, most of which encode proteins. Only 10% is noncoding, and much of this noncoding DNA is used for regulatory purposes (*e.g.*, promoters, terminators, protein binding sites, etc).

Examination of the genomes of eukaryotes (Table 1, bottom) reveals a very different picture (Zimmer, 2007). Instead of the

50-fold range in genome size observed for prokaryotes, eukaryotes exhibit a 100,000-fold range. However, the range of their gene contents (in numbers of genes) varies by only about 20-fold. The yeast, *Saccharomyces cerevisiae*, has a small genome relative in other eukaryotes with a gene content less than that of *M. xanthus* or *S. coelicolor*; however, its physical genome is substantially larger (Table 1). In the worm, *C. elegans*, about 25% of the genome codes for RNAs, and in the plants, thale cress and rice, with slightly larger genomes, 15% to 20% of the DNA codes for RNAs. With mammals such as humans, the size of the genome increases another 100-fold, but the number of genes is nearly the same. For these organisms, only a fraction (often a small fraction) of the genome codes for RNAs. The salamander and the onion have far larger genomes but the numbers of estimated genes present do not change. In these cases, only about 0.2% of the genomes is known to code. The lungfish has an exceptionally large genome, and only about 0.05% of the DNA may code for RNA (Table 1). Finally, single celled protozoans can have larger genomes than humans. For example, that of *Gonyaulax polyhedra* is 30x larger than ours. Thus, while in prokaryotes, genome size correlates with organismal complexity, this is not true for eukaryotes where size correlates better with amounts of noncoding repetitive DNA (Kidwell, 2002).

The degree of volumetric compaction required for eukaryotes can be tremendous. For man, it is about 100,000-fold while for the lily, it is about 10-million-fold. From this consideration, it is clear that eukaryotes need a much more efficient means to compact their DNA than prokaryotes. They use highly structured nucleosomes, complexes of histones, which wrap the DNA around themselves. We shall consider the various mechanisms used by bacteria to achieve this goal.

COMPACTION OF BACTERIAL DNA

State of the *E. coli* Chromosome

Escherichia coli K12 has 4.6 Mbp of DNA with a length of 1.6 mm. This must fit into a cell that is 2 μm long and 1 μm in diameter. DNA supercoiling, macromolecular crowding, polyamines and nucleoid-associated proteins together are believed to result in nucleoid compaction that allows it to occupy only about 25% of the intracellular volume (Dame, 2005; Murphy and Zimmerman, 1997a; Zimmerman, 2006a). The relative contributions of these factors to DNA compaction are unknown.

Early electron micrographs showed *E. coli* nucleoids obtained from lysed cells with large DNA loops emanating from a central core. These loops were inferred to be interwound, suggesting the presence of supercoiling *in vivo* (Kavenoff and Bowen, 1976). In fact, DNA supercoiling and branch formation are effective ways to generate DNA loops. Dynamic, independent, superhelical domains with variable sizes and positioning of their boundaries were later demonstrated using genetic methods (Higgins *et al.*, 1996). Initially, it was believed that these domains were up to several hundred kb in size (Kavenoff and Bowen, 1976), but more recent estimates indicate that these domains are probably closer to 10 to 20 kb in size (Cunha *et al.*, 2001; Postow *et al.*, 2004). Thus, the *E. coli* chromosome may possess as many as ~400 independently supercoiled domains.

At the same time as Kavenoff and Bowen made their images of the *E. coli* nucleoid, another study employing electron microscopy provided evidence for eukaryotic-type compaction into nucleosome-like structures (Griffith, 1976). It was proposed that these structures resulted from the binding of small, histone-like proteins to the DNA (Varshavsky *et al.*, 1977; Loden and van Steensel, 2005).

In rapidly growing bacteria, DNA replication, chromosome segregation and cell division must be accommodated within a 20-minute division time. Moreover, much of the chromosome is probably competent for transcription during the entire cell cycle (Jin and Cabrera, 2006; Zimmerman, 2006b). Transcriptional activity in some regions (*e.g.*, in stable RNA operons) is very high, whereas transcription in other regions is limited (Jin and Cabrera, 2006). However, virtually all genes can be rapidly activated in response to environmental changes. This requires an organization of the chromosomal DNA that is dynamic, allowing proteins involved in DNA transcription continual access.

Nucleoid-Associated Proteins

The major nucleoid proteins under vegetative conditions are HU, H-NS, and Fis, with IHF and StpA playing lesser roles (Table 2). Under stationary phase conditions, the concentrations of HU, H-NS and StpA go down two to threefold while that of Fis goes down >50x relative to active growth conditions. IHF increases nearly 2x, and the stationary phase protein, Dps, which is believed to protect the DNA against oxidative stress (*e.g.*, hydroxyl-free radicals) and nucleases (Grant *et al.*, 1998; Minsky and Kolter, 2005) goes up >100x until it, HU and IHF are the major DNA binding proteins in the nucleoid (Table 2).

As indicated in Table 2, HU and IHF, homologous to each other, bind to 36 bp sites in the DNA, although the minimal binding site for HU may be about 9 bp in length (Bonnefoy and Rouviere-Yaniv, 1991). H-NS and StpA, also homologous to each other, bind to 10 bp sites. Of the proteins tabulated, only IHF and Fis recognize specific sequences (Table 2). All of these proteins exhibit significant sequence-independent affinity for DNA, and they bend, loop or bridge DNA, thereby condensing it (Johnson *et al.*, 2005). Detailed discussions of all aspects of these proteins (structural, functional, regulatory, and mechanistic) can be found in Dame (2005) and Johnson *et al.* (2005). High-resolution 3-D structures are available for most of these architectural proteins, with and without bound DNA (Table 2).

In order to understand the role of architectural proteins in bacterial chromatin folding, the protein contents of isolated nucleoids had to be determined. At least ten proteins have been found associated with the *E. coli* nucleoid in substantial quantities (Ali Azam *et al.*, 1999). Early preparations examined by Varshavsky *et al.* (1977) revealed the presence of two abundant proteins, H-NS (histone-like nucleoid structuring protein) and

TABLE 2
Properties of the most important nucleoid proteins in *E. coli*

Protein	Structure[a]	Preferred DNA target[b]	Binding site size (bp)[c]	Exponential phase		Stationary phase		PDB I.D. # [rmf] and reference
				Copies/cell[d]	% of chromosomal DNA bound[e]	Copies/cell[d]	% of chromosomal DNA bound[e]	
HU	Heterodimer	Kinked, gapped, 3- or 4-way junctions	36	30,000	8	10,000	6	1MUL (Ramstein *et al.*, 2003)
IHF	Heterodimer	WATCAANNNNTTR	36	12,000	4	20,000	11	1ihf (Rice *et al.*, 1996)
H-NS	Homodimer/ oligomer	Curved DNA	10	10,000	1	6,000	1	2jr1 (Rimsky, 2004; Rosselli, K. *et al.* unpublished)
StpA	Homodimer/ oligomer	Curved DNA	10	12,000	1	5,000	<1	N.A.; homologous to H-NS
Fis	Homodimer	GNTYAAWWWTTRANC	21–27	30,000	6	<1,000	<1	1ety (Cheng *et al.*, 2000)
Dps	Dodecamer	None	90?	500	<1	20,000	11	1dps (Grant *et al.*, 1998)

Protein	Growth Phenotype
HU	*hupA* or *hupB*: +; *hupA*/*hupB:* ±; Mutants form filaments and anucleate cells
IHF	Required for growth on several carbon sources
H-NS	Poor growth; - at 20°C
STpA	Non-essential
Fis	Non-essential
Dps	Required for survival in stationary phase and under stress conditions

[a] Quaternary structure, usually determined from the atomic (X-ray) structure of the protein or using biochemical procedures.

[b] Each of the proteins binds DNA nonspecifically with physiologically significant affinities. A preferred binding sequence or DNA structure is noted when known. Y = C or T, R = G or A, W = A or T, and N = any base.

[c] Binding site size is estimated from atomic structures of the DNA complex (HU and IHF), footprinting data (H-NS, StpA, and Fis), or modeling based on the atomic structures of the unbound protein (Fis and Dps). For HU, 14 to 19 bp are actually contacted in the crystal structures, but contacts covering up to 46 bp are supported by the structures. The size and stoichiometry of the Dps-DNA complex is speculative (11).

[d] Fis, HU, and IHF are expressed in dimers per cell, and Dps is expressed in dodecamers per cell. H-NS and StpA are expressed as dimers per cell, but the functional binding form may be a tetramer or higher order complexes. Data were collated from individual studies and Talukder *et al.* (318) from *E. coli* cells cultured in rich media. All values represent approximations as different research groups have reported different values.

[e] Percentage of chromosomal DNA potentially bound by each protein. Calculations are based on three and one chromosome equivalents per cell in exponential and stationary phases, respectively, where the *E. coli* chromosome is 4.6 Mbp. These values are approximate because of the assumptions used in the calculations.

[f] Protein DataBase (PDB (http://pdb.org)) identification number (I.D. #) with reference. Many structures or partial structures of these proteins have been reported. The I.D. # of a representative full length structure is presented. These proteins are from individual studies of *E. coli* cells when available. N.A. = not available.

HU (heat unstable protein). Later, two additional nucleoid proteins were identified, Fis (factor for inversion stimulation) and IHF (integration host factor). Because of their high intracellular abundance and DNA binding and bending properties, these proteins are generally believed to be the most important players in the organization and compaction of vegetative bacterial DNA. Two other proteins that play important roles in vegetative cells are Lrp (leucine responsive protein) and the MukBEF heterotrimeric complex. In stationary phase cells, the DNA-binding protein from starved cells, Dps, plays a dominant role (Ishihama, 1999) (see Table 2).

The expression levels of all nucleoid-associated proteins are dependent on growth phase (Table 2; Talukder *et al.*, 1999), allowing levels of transcription and translation to vary with growth phase. Thus, Fis-dependent stimulation of transcription of stable RNA operons renders expression sensitive to growth (Schneider *et al.*, 2003). Additionally, during the stationary phase, cells can protect the genome and shut off transcription by binding the Dps protein (Almiron *et al.*, 1992; Frenkiel-Krispin *et al.*, 2001). Variation in expression levels of nucleoid-associated proteins can thus provide a means to locally modulate nucleoid structure and function, depending on growth conditions. Interestingly, in order to switch effectively between compact and more relaxed states, some of these proteins have clear-cut roles in compaction, while others act as antagonists of compaction.

H-NS

H-NS, possibly the most important nucleoid structuring protein, consists of an N-terminal dimerization domain and a C-terminal DNA binding domain connected by an unstructured flexible linker involved in oligomerization (Dame *et al.*, 2006; Rimsky, 2004). The fact that oligomerization leads to the simultaneous availability of two or more DNA binding domains suggests that interactions involve multiple DNA binding sites. In fact, it has been demonstrated that H-NS forms bridges between adjacent tracts of double-stranded DNA in register with their helical pitch (Dame *et al.*, 2000, 2005, 2006). H-NS is ubiquitous in bacteria, usually as a homo- or hetero-dimer.

What is the molecular basis underlying DNA bridging? Self association of H-NS may result in the formation of dimers and higher oligomers in which multiples of two DNA binding domains interact with DNA on opposite sides of the multimer (Dame *et al.*, 2006; Esposito *et al.*, 2002). Dimers are stacked on top of each other accounting for bridging by parallel or antiparallel associated H-NS (Dame *et al.*, 2006). Two arginine residues at the N-terminus of H-NS, together with the C-terminal domain, allow direct interaction with the DNA (Bloch *et al.*, 2003).

The DNA binding affinity of H-NS is dependent on structural features of the DNA, features such as flexibility, curvature, sequence composition, and preferential binding to AT-rich, curved DNA. There is insufficient H-NS to cover the DNA (see Table 2), and it is known to bind in patches. Lateral H-NS binding occurs preferentially along A/T-rich tracts, extending into flanking DNA sequences (Rimsky *et al.*, 2001) and accounting for local rigidification (Amit *et al.*, 2003). Such patches probably interact with naked DNA, further extending the bridged DNA areas via cooperative interactions (Dame, 2005; Dame *et al.*, 2006).

Binding of H-NS may change DNA supercoiling, but supercoiling might also promote formation of bridges due to spatial constraints imposed by the supercoiled conformation. Bridge formation is relevant not only to DNA compaction, but also to transcriptional regulation. For instance, transcriptional repression may be controlled by the bridging of two H-NS binding sites, which prevent RNA polymerase from binding to a promoter (Prosseda *et al.*, 2004). Bridging may also physically trap RNA polymerase in a transcription competent but inactive open complex (Dame *et al.*, 2002).

In still another mechanism, repression may occur during elongation (Choi and Saier, 2005) due to downstream bound H-NS that acts as a "roadblock" (Dole *et al.*, 2004, but see Nagarajavel *et al.*, 2007). Since binding of H-NS to areas upstream and downstream of a promoter is needed for effective repression, these regions may form loops (Dole *et al.*, 2004; Nagarajavel *et al.*, 2007). At some promoters, lateral binding of H-NS, initiating at a preferential binding site consisting of curved or flexible DNA that extends into the promoter region, may prove sufficient for repression (Rimsky *et al.*, 2001). Oligomerization allows it to act as a silencer of extended chromosomal regions. Also, the proposed effect of H-NS on DNA supercoiling might provide an indirect means to regulate transcription from a subset of promoters sensitive to changes in supercoiling (Higgins *et al.*, 1988a).

Two enteric bacterial homologs of H-NS, StpA and Sfh, have been studied. *E. coli* StpA, 58% identical to H-NS, increases in amount with increased temperature, with increasing amounts of Lrp, and with decreasing amounts of H-NS (Sonden and Uhlin, 1996). It can compensate for deficiencies in H-NS and can even form active heterodimers with H-NS. StpA functions mainly as an RNA chaperone protein, but like H-NS, it can bind to curved DNA and constrain supercoils.

Shigella Sfh is a second H-NS paralogue that can complement an H-NS deficiency. It is encoded within several conjugation plasmids common to various enteric bacteria (Beloin *et al.*, 2003a). It can influence conjugation, and functions as a stealth protein, aiding lateral plasmid transfer. Specifically, it allows plasmids to be transferred to a recipient bacterium such as *E. coli* without loss of fitness (Doyle *et al.*, 2007). This loss, resembling an H-NS deficiency, may result because the naked incoming DNA binds H-NS in the absence of Sfh, unmasking and decompacting the chromosome. However, the amount of incoming DNA is small compared to the size of the genome, so a relatively small amount of genomic DNA will become unmasked, changing the activity of only a subset of genes.

HU/IHF

HU and its sequence specific homologue IHF (Table 2) consist of two identical or homologous subunits (Swinger and Rice,

2004). Both dimeric proteins have compact globular structures of several intertwined α-helices from which two protruding β-ribbon arms wrap around the minor DNA groove upon binding. Conserved proline residues at the tips of the arms intercalate between base pairs. As a consequence, two kinks are created in the DNA. Positively charged surfaces extend down the sides of both HU and IHF, and these function in DNA binding (Swinger *et al.*, 2003).

Although HU shows little sequence specificity, it binds preferentially to supercoiled DNA and constrains negative supercoiling (Shindo *et al.*, 1992). This is explained when looking into the HU-DNA cocrystal structure, in which, besides bending, underwinding and negative writhe are induced (Swinger *et al.*, 2003). Further, HU binds preferentially to structural distortions in DNA such as nicks, gaps, permanent bends and three- or four-way junctions (Balandina *et al.*, 2002). Measurements indicated that the HU dimer bends native or nicked DNA variably, but on the average between 50° and 80° (Kamashev *et al.*, 1999). Footprinting experiments suggest that the bend angle induced by IHF is similarly variable and depends on the particular site to which IHF binds (Rice *et al.*, 1996). Thus, HU and IHF act as flexible hinges that can accommodate a range of different bending angles (van Noort *et al.*, 2004). These dimers may bind side-by-side on the DNA so that the DNA stiffens and attains a superhelical structure.

Whereas HU largely lacks sequence specificity, IHF-binding occurs with definite specificity. About 1000 IHF binding sites have been identified in the *E. coli* K12 chromosome, most located near promoters. The remaining IHF [10,000 to 20,000 per cell; see Table 2] are available for nonspecific binding and for HU-like compaction of the chromosome. In this capacity, IHF can bind nonspecifically, side-by-side, to form filaments like HU filaments, and it can replace HU when overexpressed (Dame, 2005; Krawiec and Riley, 1990; Rowland *et al.*, 2006).

Fis

In the Fis protein of *E. coli*, a C-terminal helix-turn-helix motif recognizes a poorly conserved 15 bp 'core' binding site (see Table 2), but the protein also has affinity for nonspecific DNA. Fis consists of two identical subunits, and the two helix-turn-helix elements probably insert into adjacent major grooves of the DNA helix. DNA bending is believed to result because the spacing between these elements is too short to fit into a straight helix. Estimates of the DNA bend induced range from 50° to 90° depending on the exact sequence bound and its context (Pan *et al.*, 1996). Nonspecific binding to supercoiled DNA leads to the formation of loops and branches with Fis bound in clusters. Increasing the amount of Fis induces strong compaction due to both bending and bridging (Schneider *et al.*, 2001; Skoko *et al.*, 2006). Bioinformatic analyses have revealed that the *E. coli* chromosome contains up to ~68,000 sites, or one site per 230 bases (Hengen *et al.*, 1997). Most of these sites are probably occupied during exponential growth when Fis levels are highest (see Table 2) (Talukder *et al.*, 1999).

Dps

In stationary phase *E. coli* a cells, Dps, with an N-terminal DNA-binding helix and an iron-binding ferritin-like core (Grant *et al.*, 1998; Stillman *et al.*, 2005), is one of the most abundant nucleoid proteins (see Table 2). Its structural gene is transcribed in a σ^S-dependent process, but expression can be induced in the exponential growth phase by various stresses such as acid, osmotic, and oxidative stresses as well as starvation, mediated by σ^{70} and various transcription factors including Crp and OxyR (Altuvia *et al.*, 1994; Jeong *et al.*, 2006; Park *et al.*, 2005). σ^{70} and σ^S apparently recognize the same promoter with an extended –10 region (Jeong *et al.*, 2006). However, induction under starvation or oxidative stress conditions is also due to stabilization against ClpAP and ClpXP-mediated proteolysis. ClpAP regulates at the translational level (Stephani *et al.*, 2003).

Dps binds DNA nonspecifically, protecting it against oxidation, UV irradiation, thermal shock, acid stress, electrophile exposure, and nuclease cleavage both *in vivo* and *in vitro* (Choi *et al.*, 2000; Ferguson *et al.*, 1998; Martinez and Kolter, 1997; Park *et al.*, 2006). The protein core is a hollow spherical dodecamer with an outer diameter of 90 Å and an inner diameter of 45 Å (Grant *et al.*, 1998), and this characteristic may be important for its protective action.

During the stationary phase, Dps binds to chromosomal DNA, forming a highly ordered and stable nucleoprotein complex, which is in a crystalline phase called a biocrystal. This biocrystalline complex is very different from anything present in the exponential-phase nucleoid complex. Biocrystal formation clearly contributes to the ability of Dps to suppress gene expression and protect chromosomal DNA, but the mode of the Dps-DNA interaction is not yet fully established.

To explain Dps-DNA interactions, two models have been proposed. One model assumes that interaction is mediated through Mg^{2+} bridges. The other model assumes that three adjacent dodecamers define "holes" that are lined by the lysine-rich N-termini. In this latter case, DNA is proposed to thread through the holes, interacting with lysine residues. Evidence for both models has been presented, but other proposed mechanisms for DNA binding by Dps homologs in other bacteria and in archaea leave the field still confusing (Ceci *et al.*, 2004, 2007; Chowdhury and Chatterji, 2007; Grant *et al.*, 1998; Park *et al.*, 2006; Reindel *et al.*, 2005; Stillman *et al.*, 2005). Further research will be required to provide the details.

Lrp

Leucine-responsive proteins (Lrp) seem to have two modes of DNA binding. LrpC of *Bacillus subtilis* can bridge DNA (Tapias *et al.*, 2000) in a fashion similar to H-NS, as revealed by electron microscopy (Beloin *et al.*, 2003b). In this scenario

Lrp-dimers on one tract of DNA are believed to interact with Lrp-dimers on another tract (Beloin *et al.*, 2003b), stabilizing DNA loops. In addition, the protein can assemble into an octameric, globular, 'nucleosome-like" structure around which it wraps the DNA. This has been demonstrated at the promoter of its own structural gene to attain autorepression, but it is also observed non-specifically on supercoiled DNA substrates (Beloin *et al.*, 2003b). Both mechanisms probably act in parallel *in vivo* and are dependent on local sequence context.

MukBEF

The heterotrimeric MukBEF complex is a "structural maintenance of chromosome" (SMC) complex of about 600 kDa. SMC complexes are ubiquitous, from bacteria to humans. They function as core components of sister chromatid condensin and cohesin complexes in eukaryotes (Hirano, 2006; Yu and Koshland, 2005). They adopt a V-shaped structure with two long arms, each of which has an ATP-binding domain at the distal end (Hirano, 2005).

The homodimeric MukB protein of *E.coli*, the major player (Qang *et al.*, 2006), has two DNA binding domains connected by a long coiled-coil with a flexible hinge at its center (Graumann, 2001; Hirano, 2005; Hirano and Hirano, 2005; Matoba *et al.*, 2005). MukBEF can bind DNA simultaneously at more than one site generating MukBEF–DNA complexes that are compacted. Thus, the complex of the MukB, E, and F proteins changes its conformation upon DNA binding (Hirano and Hirano, 2006) and forms a "condenser." MukBEF has been proposed to function in the creation of large topological loops by bridging two DNA helices (Case *et al.*, 2004). It may actively fold, tether, and manipulate DNA strands, thereby assisting sister chromosome segregation (Graumann, 2001; Hirano, 2006).

Detailed studies have shown that the *E. coli* MukB protein, by itself, promotes DNA condensation both *in vivo* and *in vitro* (Petrushenko *et al.*, 2006a). The function of MukEF seems to be to disrupt the MukB-DNA complex, creating a more dynamic chromosomal structure, possibly allowing bridging of distant DNA-bound MukB subunits (Petrushenko *et al.*, 2006a). In the presence of gyrase, it also introduces right-handed knots in the DNA, consistent with a role in condensation (Petrushenko *et al.*, 2006b).

Cooperative and Antagonistic Interactions of Nucleoid Structural Proteins

While proteins clearly shape the bacterial nucleoid, several other factors play a role. Supercoiling of the DNA as mediated by the actions of topoisomerases promotes compaction, and important roles are attributed to polyamines and macromolecular crowding (Odijk, 1998). Transertion, the coupled transcription, translation and membrane insertion of integral transmembrane proteins (Norris, 1995; Norris and Madsen, 2005; Woldringh, 2002) may counteract compaction by pulling DNA loops out from the nucleoid body towards the cell periphery. The resultant shape of the nucleoid stems from a delicate balance of these various forces (Norris *et al.*, 2007; Woldringh *et al.*, 1995).

The abundance of nucleoid-associated proteins varies dramatically depending on the growth phase (Table 2; Talukder *et al.*, 1999). As most nucleoid-associated proteins function as transcription factors, either directly or indirectly, the expression of many genes is regulated in a growth phase-dependent manner (Ishihama, 1999). Regulation depends on the antagonistic and cooperative actions of combinations of nucleoid-associated proteins (Grainger *et al.*, 2007). For example, the sequence-specific DNA bending proteins, IHF and Fis, can relieve repression by H-NS at specific promoters (Falconi *et al.*, 2001). Similarly, HU has been proposed to counteract the effects of H-NS by competing for binding sites (van Noort *et al.*, 2004).

More complex arrangements, in which several different proteins engage in cooperative activities, promote transcriptional regulation. Examples include the cooperative action of Fis, IHF, and H-NS in repressing the *E. coli nir* promoter, and the action of IHF, H-NS, and a third regulatory protein, OmpR, in controlling expression at the *S. typhimurium csgD* promoter (Browning *et al.*, 2000; Gerstel *et al.*, 2003). Undoubtedly, cooperative and antagonistic interactions regulate global nucleoid organization as well. Local reorganization of the nucleoid could alter the expression of distant genes, in part explaining the pleiotropic effects of these proteins (Arfin *et al.*, 2000; Blot *et al.*, 2006; Kelly *et al.*, 2004). Open nucleoid structures are expected at sites where heavy transcription takes place, such as within ribosomal RNA operons during exponential growth.

The following model for the bacterial nucleoid has been proposed (Dame, 2005). The nucleoid is organized in loops of 10 to 20 kb. These loops are connected and topologically closed because of the binding of proteins such as H-NS and MukBEF that create cross-links between DNA tracts. The higher the number of DNA bending proteins bound, the more branched and compact it will become. While IHF and Fis recognize specific sequences to bridge and/or bend the DNA, HU preferentially binds cooperatively to A/T-rich regions, possibly resulting in nucleoid rigidification. H-NS probably binds in patches along the DNA that can interact with other patches and/or naked DNA. If this occurs within a supercoiled loop, a particular configuration of that loop may be stabilized. Local disruption of an H-NS-DNA complex (decompaction) usually activates transcription. When growth conditions change, the ratios between H-NS, HU, IHF, Fis and Dps change, giving rise to global changes in compaction and gene expression. All of these proteins can probably act cooperatively and antagonistically with each other in a large variety of homo- and heterotypic interactions, leading to transcriptional activation or repression. While virtually all promoters are influenced by these proteins, each will be differentially controlled by different sets of them, and they will function in different capacities (*e.g.*, bending, looping, and bridging the DNA; promoting or inhibiting binding of other proteins, etc).

DNA SUPERCOILING

DNA supercoiling is found in all living organisms in a species-specific fashion and is subject to tight regulation (Khodursky, 2007; Champion and Higgins, 2007). It is required for compaction of the bacterial chromosome, and it plays important roles in numerous essential genetic processes. These include DNA replication, recombination and transcription among others. Supercoiling provides energy for helix unpairing and drives the formation of nonhelical structures such as cruciforms (Oussatcheva *et al.*, 2004). It also brings distal DNA regions into close proximity by creating inter-wound supercoils. It contributes to the dynamic nature of the genome in response to environmental conditions and growth phase (Travers and Muskhelishvili, 2005a). Thus, supercoiling influences the energetics of structural transitions within chromosomal supercoiled domains as well as the presence of branching structures in the DNA (Thanbichler *et al.*, 2005). It has been argued that supercoiling provides a primary mechanism for regulating ribosomal and transfer RNA synthesis in response to growth rate and growth phase (Travers and Muskhelishvili, 2005b). It also influences expression of genes encoding small translational regulatory RNAs (Wassarman, 2002) and may influence localized mutation rates (Wright, 2004). In *E. coli* there is about one negative supercoil per 20 helical turns. Supercoiling thus provides one of the many crucial means for regulating gene expression (Hatfield and Benham, 2002). In view of these multifaceted interdependencies, it is not surprising that DNA supercoiling appears to be essential for life. All living organisms apparently evolved to maintain supercoiling most suited to their respective environments and growth conditions (Khodursky, 2007).

Properties of Supercoiled DNA

Twisting the two ends of double-stranded helical DNA in opposite directions causes linear DNA to increase its potential energy, just as is the case with a coil of any type, such as a telephone coil. Supercoiling can be of two types: negative, in which case the induced twist is against the direction of the right-handed helix, or positive, in which case the induced twist is in the same direction as the right-handed helix. Some organisms have predominantly positively supercoiled DNA while others have predominantly negatively supercoiled DNA (Napoli *et al.*, 2005). This supercoiling gives rise to a "coiled coil" which will immediately be lost if the ends are not constrained. A single-stranded nick in the DNA will allow one strand to rotate relative to the other until the lowest energy state has been attained. In this "relaxed state," supercoiling is lost. Double strand breaks will similarly result in loss of supercoiling unless both of the two strands are constrained.

Recent research suggests that the *E. coli* chromosome may be divided into as many as 400 independently supercoiled domains, each about 10,000 nucleotides long on the average (Postow *et al.*, 2004). This means that at the boundaries of each of these domains, the DNA must somehow be constrained. This may be important for function since it prevents chromosomes from global loss of supercoiling when nicks or breaks are introduced in the DNA. These boundaries are apparently not usually placed stably at fixed sites on the chromosome although some may be (Higgins *et al.*, 1996; Higgins, 2005). Instead, they are randomly distributed. Permanent bends may provide one type of boundary, but protein association and DNA crosslinking may provide others (Postow *et al.*, 2004).

Enzymes of DNA Supercoiling

Many enzymes influence the degree of chromosomal supercoiling. Such enzymes, for example, include single- and double-stranded nucleases which hydrolyze one or two phosphodiester bond(s) and therefore relax the DNA by allowing free rotation. DNA ligases, which reseal nicks and breaks, and thereby hinder nuclease-promoted relaxation, provide a second example. Enzymes that affect supercoiling also include DNA and RNA polymerases which generate positive supercoiling in front of the polymerization site and negative supercoiling behind it. Thus, in the case of DNA polymerization, as the replisome progresses, the DNA ahead of the fork rotates due to the unwinding of the double helix. This creates positive supercoiling.

By far the most important enzymes determining the extent and direction of supercoiling are the topoisomerases. These enzymes can be thought of as reversible nucleases as they preferentially catalyze transesterification. In *E. coli*, there are four such enzymes, Topoisomerases (Topo) I-IV. They can be of either of two types, ATP-independent type I enzymes that only relax the DNA (Topo I and III in *E. coli*) and ATP-dependent type II enzymes that introduce supercoiling (Topo II and IV in *E. coli*) (Berger, 1998; Tse-Dinh, 1998). Topo II is also called gyrase. Thermophilic prokaryotes of both the bacterial and archaeal domains have a predominance of positively supercoiled DNA. They have an ATP-dependent enzyme called "reverse gyrase" that is responsible for this characteristic (Changela *et al.*, 2003; Forterre *et al.*, 1996, 2000).

Type I topoisomerases, single subunit enzymes, break one strand in the double helix while Type II topoisomerases cleave both strands. Both enzymes bind the DNA strand covalently via a tyrosyl phosphoester bond at the site of the cleaved phosphodiester bond. This reversible reaction is termed "transesterification." For Topo I, no ATP is required; the enzyme breaks one strand of the double helix, and with Topo I bound to it, one strand rotates about the other to relax the supercoils. Then the enzyme reseals the nicked strand and dissociates from the complex.

Topo II (gyrase) binds preferentially to a variety of sites including repetitive extragenic palindromic (REP) sequences in the presence of the HU protein (Yang and Ames, 1988) (see the section entitled "Recurrent Nucleotide Sequences in the *E. coli* Genome"). ATP is required for the action of this enzyme because it allows introduction of negative supercoils with increased potential energy. Topo II is a heterotetramer of $\alpha_2\beta_2$ structure which can wrap about 130 bps of DNA around itself. It breaks both strands, forming a 4 bp staggered double strand

break in the central region of the wrapped DNA. Each of the two catalytic subunits becomes covalently bound to each of the two cleaved strands, and consequently, the enzyme can prevent unwinding. When ATP is hydrolyzed, one strand passes through the other, generating a single negative supercoil. Then the two strands are resealed before gyrase dissociates from the DNA (Wigley, 1995).

While gyrase is essential for the growth of *E. coli*, Topo I is not. Bacteria lacking Topo I are viable but sick, and there is a 30% increase in the negative DNA supercoiling. Mutations arise that can grow normally, and these mutations invariably decrease the gyrase activity so that the degree of supercoiling returns to normal (Krawiec and Riley, 1990). It is therefore the *product* of the topoisomerase reactions (supercoiling) and not the enzymes *per se* that is required for normal cellular function. Interestingly, transcription of the gene encoding Topo I increases with increasing supercoiling, while that of the Topo II genes decreases. This mechanism serves to provide homeostatic control.

Topo III and Topo IV function by the same mechanisms as Topo I and Topo II, respectively. However, their primary functions appear not to be the release and introduction of supercoils. Rather, these enzymes function in a different role, in untangling the DNA after replication in a process called decatenation. This, in essence, involves passage of one DNA double helix that is in close proximity to a second double helix, through the latter. While Topo III facilitates the process, Topo IV is absolutely required. These enzymes, which bind DNA preferentially at double helix contact sites, recognize specific symmetrical sequences around the DNA scission site (Leo *et al.*, 2005).

MACROMOLECULAR (PROTEIN PLUS RNA) CROWDING

Macromolecular crowding is a process whereby some cellular constituents in a crowded cytoplasm press against and constrain the movements and conformations of others. This process constrains the bacterial nucleoid, contributing substantially to compaction (Cunha *et al.*, 2001; Murphy and Zimmerman, 2001). The cellular concentration of protein plus RNA is roughly 300 mg/mL, greater than that of DNA in the nucleoid (Zimmerman and Trach, 1991). Further, polyethylene glycol and polyamines can stabilize condensed DNA (Murphy and Zimmerman, 1997b; Murphy and Zimmerman, 2001). These observations clearly suggest that macromolecular crowding and other molecular interactions play a significant role in nucleoid compaction.

RECURRENT NUCLEOTIDE SEQUENCES IN THE *E. COLI* GENOME

With the completion of the genome sequence of *E. coli* K12 (Blattner *et al.*, 1997), it became clear that certain repetitive sequences occur with much higher frequencies than expected by chance while others occur with much lower frequencies than expected. As an example of the latter, binding sites for the lactose repressor, LacI (O_1, O_2, and O_3), occur in three places within the *lac* operon and only once elsewhere in the rest of the *E. coli* genome. Since LacI is present in very limited amounts and has extremely high affinity for its operators ($K_d \approx 10^{-13}$), it is clear that the presence of additional high affinity sites would sequester LacI, making it necessary for *E. coli* to make larger amounts of this protein for proper control of *lac* operon expression. This consideration presumably provided selective pressure to prevent the appearance or promote the removal of such sites elsewhere in the genome.

Short Oligonucleotide Repeats

Several repeated sequences occur in the *E. coli* genome with high frequency (Table 3; Blattner *et al.*, 1997). The chi site hotspot for recombination, is an octomeric sequence (5′-GCTGGTGG-3′) that mediates binding of the homologous recombination RecBCD complex. It is present in over 16,000 copies in the *E. coli* K12 genome, and these sites are strongly skewed towards the leading strand (Eggleston and West, 1997; Myers and Stahl, 1994; Uno *et al.*, 2006).

The RecBCD complex in *E. coli* aids in the repair of collapsed replication forks (Kuzminov, 1995, 1999). The Chi sequence includes within it, the DnaG primase binding site (CTG), involved in initiation of synthesis of the RNA primers required for Okasaki fragment biosynthesis, used for interrupted DNA replication (Soultanas, 2005). DnaG initiates synthesis of RNA primers, primarily in the lagging strand of the replication fork for DNA replication, initiating synthesis at a CTG trinucleotide (Khopde *et al.*, 2002). It is possible that primase and RecBCD act together at the Okasaki primer initiation site to facilitate strand assimilation by branch migration (Blattner *et al.*, 1997). On the other hand, the CTG trinucleotide sequence occurs in *all* of the most prevalent octanucleotide sequences identified in *E. coli* DNA, and all of these are GC rich, most having few As or Ts except for the one in the CTG trinucleotide sequence (Table 3; Blattner *et al.*, 1997). Interestingly, all such sequences show marked strand skew. Possibly, primase recognizes the flanking regions of its CTG initiation sequence and strongly prefers GC-rich regions. These observations illustrate how "sequence gazing" can lead to new postulates that can be experimentally tested.

The *E. coli* chromosome is extensively methylated by DNA adenine (N-6) methylation (DAM). The DAM methylase recognizes a short palindromic tetranucleotide sequence, 5′-GATC-3′, and methylates the N-6 atom in the purine aromatic ring of the A in the GATC sequence. These sites are clustered around the origin of chromosomal replication, OriC, with increased frequency (Blattner *et al.*, 1997). On the average, GATC occurs near OriC with a 10-fold increased frequency compared to that throughout the chromosome. Methylation, which occurs slowly following DNA replication, allows the cellular replication and repair machineries to distinguish the new from the old strand. This is important for mismatch DNA repair and semi-conservative DNA replication (Palmer and Marinus, 1994).

In contrast to the GATC DAM tetranucleotide sequence, the reverse 5′-CTAG-3′ sequence is rare; it occurs with 5% the

TABLE 3
Functions of repetitive elements in the genome sequence of *E. coli* K12

Repeat sequence	Characteristic	Function
CTG[a]	Most abundant trimer	DnaG primase recognition site; primer initiation
GCTGG[a]	Most abundant pentamer	
CGCTGGCG[a,b]	Most abundant octamer	Both DnaG primase and RecBCD binding sites
GGCGCTGG[a,b]	Second most abundant octamer	
GCTGGTGG[a,b]	Third most abundant octamer (Chi)	
GAT*C	Abundant tetramer	DNA adenine methylase (DAM) binding; adenine methylation
CC*(A/T)GG	Abundant pentamer	DNA cytosine methylase (DCM) binding; 5-methyl-cytosine is a mutational hotspot; target of VSP repair
CTAG	Low abundance tetramer in protein coding regions	Kinks mRNA? Target of very short patch (VSP) DNA repair
(A) GCC(G/T)GATGCC(G/A)CG(C/T)Nx (B) (G/A)CC(C/T)CTTATC(C/A)GGC	REP[c]	Gyrase/HU binding site; mRNA stabilization
127 bp repetitive sequence	ERIC[d] (imperfect palindrome)	Transcribed but not translated (like REP but different in sequence); mRNA stabilization
IS5[e]	Most common transposon	Hopping genes; gene activation; genetic rearrangement
IS1[e]	Second most common transposons	
IS2[e]		

*Target of methylation.

[a]These abundant sequences can all serve as DnaG primase binding sites for initiation of Okazaki fragments because of the internal CTG sequence (underlined).

[b]All three of the most common octamers may allow RecBCD recognition.

[c]REP, repetitive extragenic palindrome.

[d]ERIC, Enterobacterial repetitive intergenic consensus.

[e]IS element, Insertion sequence element; hopping genes, capable of self-mobility.

abundance expected by chance, based on the *E. coli* DNA base composition. It is particularly rare in protein-coding regions of the *E. coli* chromosome, but is much more frequent in stable RNA-encoding genes (rRNA, tRNA, and sRNA genes). Other organisms including archaea also show this bias (Karlin *et al.*, 1997). Possibly CTAG introduces "kinks" in the DNA and/or mRNA which interfere with mRNA translation. Overall, there are <900 of these tetramers compared to the expected >18,000 in the *E. coli* chromosome (Blattner *et al.*, 1997). Thus, we can account for the evolution of the high frequencies of some repeat sequences but the unexpectedly low frequencies of others.

Many short repeat sequences are organism-specific. For example, the genomes of *H. influenzae* (Fitzmaurice *et al.*, 1984) and *N. gonorrhoeae* (Elkins *et al.*, 1991; Goodman and Scocca, 1988) contain distinct decanucleotide sequences which are recognition sites for DNA uptake during transformation. Moreover, *V. cholerae* has 124 bp sequences of imperfect dyad symmetry (>150 copies) that act as recombinational sites for a superintegron (Biskri *et al.*, 2005). Thus, while some repeat sequences can be found over a wide range of distantly related organisms, others are genus-specific.

Large Oligonucleotide Repeats

Among the larger repeat elements found in *E. coli* are the so-called repetitive extragenic palindromic (REP) sequences. REP sequences occur in a wide range of bacteria and archaea (Suyama *et al.*, 2005) although their precise sequences are species-specific (Tobes and Ramos, 2005). They consist of repeated (usually 2 to 4 times) 38 to 40 bp stem-loop (imperfect palindrome) structures spaced from each other by about 25 bps (see for example, Figure 2). They can be associated with short sequence motifs, forming bacterial interspersed mosaic elements (BIME) (Espeli *et al.*, 2001). There are about 200 of these REP elements in the *E. coli* chromosome, but they almost always occur in the 10% of the intercistronic regions of the genome that do not code for proteins. It has been demonstrated that these sequences bind DNA polymerase I, possibly promoting repair, as well as the HU protein and gyrase, promoting supercoiling (Malik *et al.*, 1996; Shanado

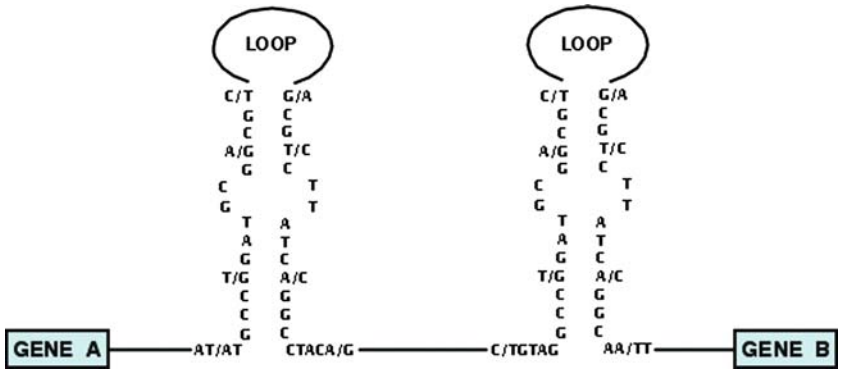

FIG. 2. Two repetitive extragenic palindromic (REP) units exhibiting the consensus sequence for *E. coli*. For clarity, the consensus is shown as a 'stem and loop' structure. Typical REP sequences, containing inverted repeats are found in extragenic positions on bacterial chromosomes (*e.g.*, between gene A and gene B) and exhibit dyad symmetry.

et al., 1998; York and Stodolsky, 1981). When transcribed, they have been reported to stabilize upstream messenger RNA structures by blocking the processive action of $3' \rightarrow 5'$ exonucleases (Higgins *et al.*, 1988b). Multicomponent RNA degradosomes including RNase E are required to degrade the structured mRNAs (Khemici and Carpousis, 2004). Finally, REP sequences may be preferred DNA targets for transposases that allow Insertion Sequence (IS) elements to hop from an initial site in the chromosome to a target site (Tobes and Pareja, 2006). These last observations suggest that these multifunctional REP sequences are hotspots for DNA transposition, thereby promoting genomic plasticity

A set of >20 superficially similar palindromic repeats in *E. coli* that show no sequence similarity with REP sequences are the 127 bp long enterobacterial repetitive intergenic consensus (ERIC) elements (Hulton *et al.*, 1991). ERICs are found in a diverse range of enterobacteria, occurring in these organisms in variable numbers. Although their functions are unknown, they occur with highest frequency adjacent to highly expressed genes (Wilson and Sharp, 2006), suggesting that they may function to stabilize mRNA. ERIC sequences may duplicate using a "master copy mode of generation," and insertion of new copies seems to occur at and involve duplication of the dinucleotide targeting sequence, TA. Surprisingly, two classes of inserts (~70 bps each) occur at different specific sites in ERIC sequences, and these inserts apparently evolved independently of the ERIC sequences (Wilson and Sharp, 2006). These unexplained observations emphasize the dynamic nature of the bacterial chromosome.

Duplicated Genes in Bacterial Genomes

In addition to the well-defined, discrete, repetitive sequences described in the preceding section, the *E. coli* K12 genome contains many large genetic elements that occur multiple times (Table 3; Blattner *et al.*, 1997). Numerous copies of homologous enzymes are found, and these often serve dissimilar functions, or the same function subject to differential regulation. The four intact *rhs* (rearrangement hot spot) loci (*rhsA-D*) appear to encode huge (1400 residue), sequence similar, mosaic, cryptic, cell surface proteins (Feulner *et al.*, 1990). They consist of discrete repetitive elements and are the largest repetitive sequences encoded in the *E. coli* K12 genome. A fifth such sequence is truncated and consequently about half sized. These proteins contain numerous copies of a repeated amino acyl sequence: $GX_3RYXYDX_2GLR$ where X = any residue. RhsA, for example, has 28 such repeats. All five Rhs proteins contain constant and variable regions, and segments of their genes differ markedly in G+C content from each other and from that of the genome as a whole. This fact suggests that all, or at least some, of them have origins in other bacteria, having been acquired by *E. coli* via horizontal (lateral) transfer (Hill *et al.*, 1994).

Evidence has suggested that these proteins may play roles in capsular polysaccharide biosynthesis and export (McNulty *et al.*, 2006). RhsA interacts directly with the multi-protein capsular polysaccharide biosynthetic export complex. This complex, localized to the cell poles, includes both the cytoplasmic enzymes and IM/OM export proteins, forming a complex metabolome (Norris *et al.*, 2007). It has been suggested that RhsA plays a stimulatory role by coupling exopolysaccharide biosynthesis to export (McNulty *et al.*, 2006).

Most living organisms, both prokaryotes and eukaryotes, have multiple copies of homologous genes encoding the ribosomal RNAs. For example, *E. coli* has seven while *Bacillus subtilis* strains have 9 or 10 copies, depending on the strain. Small-genome organisms have just 1 or 2 copies, and they are usually slow growing. Many bacteria increase their dosage of these genes by carrying multiple copies of their chromosome(s), and this number is greater during active growth than in stationary phase. Thus, *E. coli* has 4 to 8 copies per cell during active growth, but only 1 to 2 in nongrowing, stationary phase cells. Multiple copies allow increased rates of protein synthesis, and hence, more rapid rates of growth.

Finally, as one last example of repeated genetic elements, the *E. coli* K12 genome encodes 10 different types of Insertion

Sequence (IS) elements, each in variable numbers, from 1 copy for several IS elements including IS4, to 11 copies for IS5 (Tables 3 and 5). These transposable "hopping genes" always encode a transposase that catalyzes their transposition as well as terminal inverted DNA repeats that are recognized by the transposase (Blattner *et al.*, 1997; Schneider and Lenski, 2004). IS elemets can be 700 to 2,000 bp long and are genetically compact, encoding only functions concerned with their own transposition. More than 600 IS elements have been identified in bacteria and archaea. They preferentially integrate into short nucleotide target sites and promote DNA rearrangements including inversions, duplications and mobilization (cotransposition) of adjacent genes. Some such "composite transposons" mobilize drug resistance genes and thereby promote dissemination of drug resistance (Mahillon *et al.*, 1999).

Bacterial Retrons

Many proteobacteria, including *E. coli* and *S. typhimurium*, contain hundreds of copies of "multicopy single-stranded 'satellite′ DNA' (msDNA) (Lampson *et al.*, 2001, 2005). They are commonly associated with prophage DNA and resemble in some respects retroviruses and other types of retroelements in eukaryotes. Genetic elements that code for reverse transcriptase and use reverse transcription for their replication or mobility are called 'retroelements.' Interestingly, chromosomal retrons may not be benign; they may be capable of selfishly "reproducing" in the genome, promoting genetic instability by increasing mutational frequencies (Jeong and Lim, 2004). Thus, reverse transcription may serve important functions in promoting stress responsive adaptation.

In *Myxococcus xanthus*, msDNA is 162 nucleotides long, with an attached RNA strand of 77 nucleotides. The 5′ end of the DNA strand is linked to the 2′-OH group of a specific guanosine residue in the middle of the RNA strand, forming a 2′–5′ phosphodiester bond. Thus, msDNA is actually a complex of DNA and RNA, and possibly protein as well.

The locus encoding the msDNA consists of three genes: *msr*, encoding the RNA part, *msd* encoding the DNA part, and the *ret* gene encoding reverse transcriptase. Biosynthesis of msDNA starts with the transcription of all three genes. Depending on the retron element, all three genes are transcribed either into one or two mRNAs, one encompassing the *msr* and *msd* genes, and the second the *ret* gene. The transcript from the *msr-msd* region is folded into a secondary structure using internal repeats. Reverse transcriptase utilizes the 2′-OH group of the branching G residue to initiate cDNA synthesis from the folded mRNA transcript that serves as both a primer and template. Upon completion of cDNA synthesis, the part of the RNA not used as a template for cDNA synthesis remains attached to the cDNA, thus producing the "satellite" msDNA.

Two different types of bacterial retroelements have been described; group II introns which use reverse transcription to mobilize the intron element, and retrons which code for msDNA. Retrons and msDNA have been described in many bacteria including pathogenic proteobacteria, but their functions, if any, are unknown. It has been suggested that retrons may play roles in pathogenicity as is true of many genes introduced into bacterial genomes via phage-mediated integration or transduction (Schumann, 2006).

NUCLEOTIDE CONTENT, CODON USAGE, AND DNA MUTATIONAL INSTABILITY

G+C Content

Bacterial genomes vary over a three-fold range with respect to adenine and thymine (A+T) versus guanine + cytosine (G+C) content. There is a continuum of values ranging from below 25% G+C content to over 75% G+C content. For example, *Mycoplasma* species have slightly less than 25% G+C genomic content while *Micrococcus* species have greater than 75%. Genes encoding stable RNAs (*e.g.*, ribosomal, transfer and small regulatory RNAs) vary the least in G+C content as the product RNA molecules depend on specific nucleotide compositions and sequences to maintain proper base paired secondary structures. Protein-encoding genes vary to a much greater extent, but the greatest variation is observed for noncoding regions and the "wobble" positions (the third nucleotide position in codons; see next section) of protein encoding genes. The genomes of our good friends, *E. coli* and *B. subtilis*, have close to 50% G+C content.

Codon Usage and the Genetic Code

The genetic code in any organism is determined at two levels, that of mRNA codon:tRNA anticodon recognition, and that of tRNA:amino acyl-tRNA synthetase recognition (Ardell and Andersson, 2006). With respect to the latter, tRNAs show extensive variation in tRNA identity variants among bacteria, partly because of variation in overall tDNA base content. It has been shown that cognate tRNAs and their charging synthetases have co-evolved so that the residues in the tRNA binding pocket of the synthetases show covariance with recognition nucleotides in the tRNAs (Ardell and Andersson, 2006).

Protein-encoding genes show G+C contents that generally reflect the overall G+C content of the host genome, but the three codon positions in these genes show tremendously different variations in G+C content (Figure 3). It is apparent that the second codon position shows the least variation in G+C content (the slightest slope) as a function of genomic G+C content while the third position shows the most (the greatest slope). The first position shows an intermediate slope, corresponding to an intermediate degree of G+C content change. Thus, position 1 goes from 41% G+C to 72% G+C (a 31% change in G+C content), position 2 goes from 33% G+C to 45% (a 12% change), and position 3 goes from 10% to 90% G+C content (an 80% change) where the genomic G+C content varies from 25% to 75%. Position 2 is therefore most invariant, while position 3 is

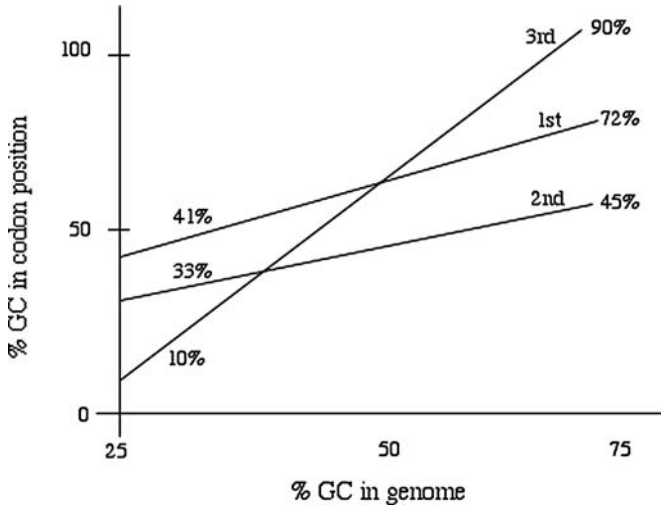

FIG. 3. Relationship of the genomic G+C contents of representative bacteria plotted on the X-axis, versus the G+C contents of each of the three codon positions (the first (*1st*), second (*2nd*) and third (*3rd*) positions) as indicated in the figure, plotted on the Y-axis.

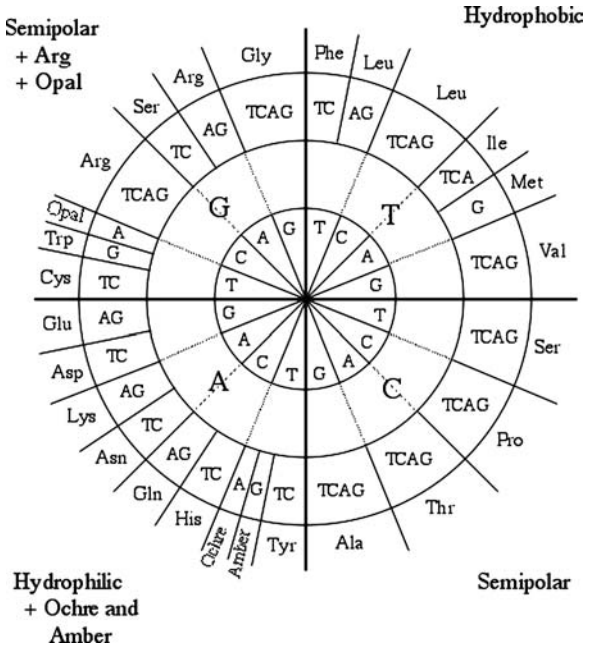

FIG. 4. Wheel representation of codon usage emphasizing the primary importance of the central position (position 2) in determining the *type* of amino acid, the secondary role of position 1 in determining the specific amino acid, and the relatively minor role of the third (wobble) position for amino acid determination. The three letter abbreviations of the amino acids are used. The three chain termination codons are indicated by name (*Ochre*, *Amber*, and *Opal*).

most variable. The relative variability in G+C content at the three codon positions reflects the relative importance of each position in determining the nature of the amino acid encoded as illustrated in Table 4. Codon position 2 is *crucial* for determining the *type of* amino acid encoded; codon position 1 is important for determining the *specific* amino acid of a given type, and the third position, the "wobble" position, is the least important due to the redundancy of the genetic code.

These facts are illustrated in Table 4 and Figure 4. Codons that contain U in the RNA, or T in the anticodon of the DNA in the central position, encode only hydrophobic amino acids (L, I, V, M, and F). By contrast, those that have A at position 2 encode only strongly hydrophilic amino acids (H, K, D, N, E, Q, and Y). When C is at this central position, all encoded amino acids are semipolar or weakly hydrophobic (S, T, P, and A), and when G is at this position, semipolar amino acids (W, S, C, and G) plus one strongly hydrophilic amino acid (R) can be encoded (Table 4, top). The general rule is, therefore, that a central U or A determines strongly hydrophobic or hydrophilic amino acids, respectively, while a central G or C determines mostly semipolar amino acids. As genomes shift from low to high G+C content, their encoded proteins therefore decrease their contents of strongly hydrophobic and hydrophilic amino acids and increase their contents of semipolar amino acids. Table 4 (middle) also

TABLE 4
Dependency of the amino acid (AA) encoded by a particular codon on the nucleotide (Nuc) at the second, first, and third positions of that codon[a]

Position 2:	Nuc:	N<u>U</u>N[b]	N<u>C</u>N	N<u>A</u>N		N<u>G</u>N	
	AA :	LIVMF	STPA	HKDNEQY		WSRCSG	
Position 1:	Nuc:	<u>G</u>UN	<u>C</u>UN	<u>A</u>UN		<u>U</u>UN	
	AA :	V	L	I, M		F, L	
Position 3:[c]	Nuc:	T, C, A or G	T, C, A or G	T, C or A	G	T or C	A or G
	AA :	V	L	I	M	F	L

[a] See Figure 4 for the complete set of assignments in the "universal" genetic code, applicable to most organisms.

[b] N = any nucleotide (Nuc).

[c] Amino acid (AA) assignments when the third codon position is as indicated and the first two codon positions are as indicated for the codon shown above (GUN, CUN, AUN and UUN, respectively).

shows how the first codon position determines the specific amino acid in a subtype of these codons while Table 4 (bottom) provides examples of how the third codon position influences the amino acid selected. In this last regard, it is interesting to note that when different amino acids are specified by a set of codons exhibiting the same nucleotides at positions 1 and 2, the third nucleotide usually specifies the same amino acid when it is a pyrimidine (C or U) versus a purine (A or G) (see Figure 4).

Figure 4 presents a wheel depiction of the genetic code, emphasizing the importance of the central nucleotide in codons instead of the first position as is usually seen in textbooks. It reveals several other interesting facts. For example, some amino acids (W and M) are encoded by a single codon (TGG and ATG, respectively), and interestingly, these amino acids are present in proteins in small amounts. The others are encoded by between 2 to 6 codons each. The amino acids that have 6 codons are L, S, and R, common amino acids in proteins. When the number of codons encoding a particular amino acid is plotted versus the frequency of that amino acid in proteins, the plot shown in Figure 5 is obtained. There is a rough correlation between frequency of occurrence of an amino acid and the number of codons which specify it.

The most frequent initiation codon for protein synthesis is ATG coding for formyl methionine (~85% in *E. coli*). However, with lower frequency, GUG (valine; ~12% in *E. coli*), UUG (leucine; ~2% in *E. coli*) and CUG (leucine; <1% in *E. coli*) can also initiate translation of proteins. When these alternative initiation codons are used, formyl methionine is still inserted as the first amino acyl residue (Tikole and Sankararamakrishnan, 2006). Bacteria with high G+C contents have higher percentages of GUG initiation codons than lower G+C content organisms, and these occur in some bacteria such as *Rhodobacter* species with highest frequency in the first genes of operons (Wu and Saier, 1991). Surprisingly, the codon positions that are invariant in initiation codons are the second and third (wobble) positions. In contrast to amino acid specification, the first position can be varied without loss of initiation capacity.

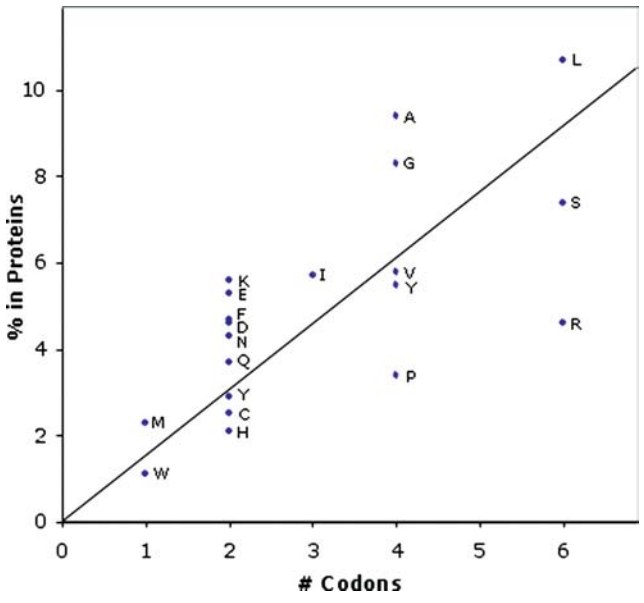

FIG. 5. Plot of amino acid frequency in proteins versus the numbers of codons specifying the amino acids. The one-letter abbreviations of the amino acids are adjacent to the points representing the position corresponding to its relative abundance expressed in percent of the total in proteins on the Y-axis, and the numbers of codons that specify the amino acids are plotted on the X-axis. The amino acid frequencies in representative proteins were taken from Saier (1987).

While amino acyl (aa)-tRNAs are generally synthesized by aa-tRNA synthetases (Ibba and Soll, 2000), genome analyses have shown that many prokaryotes lack asparaginyl- and glutaminyl-tRNA synthetases (Feng *et al.*, 2004). In these cases, a pretranslational amino acyl modification converts Asp-tRNA(Asn) to Asn-tRNA(Asn) and Glu-tRNA(Gln) to Gln-tRNA(Gln). This transamidation pathway is common in archaea and bacteria. In fact, some bacteria rely on this pathway to make asparagine. Such pretranslational modifications are also found in the biosynthesis of formyl methionyl-tRNA(Met) and selenocysteyl(Sec)-tRNA(Sec) (Feng *et al.*, 2004).

Expansion of the Genetic Code

Termination (stop or nonsense) codons can be UAA (ochre), UAG (amber) or UGA (opal). Here, it is the first position which is invariant. To ensure a complete translational stop, two termination codons frequently abut each other at the end of a gene. Moreover, UGA is not a termination signal in all prokaryotes. In some archaea, for example, it is faithfully translated as a tryptophan residue. Because of this difference, the proteins of some archaea cannot be translated in *E. coli*. However, even in *E. coli*, UGA is not always a stop codon. Selenocysteine (Sec), the 21st cotranslationally inserted amino acid found in proteins, is encoded by the UGA codon (Allmang and Krol, 2006). This rare amino acid can be found as an essential constituent at the active sites of certain seleno-oxidoreductases (Burk and Hill, 1999; Gromer *et al.*, 2005). RNA secondary stem loop structures (mRNA hairpins) called selenocysteine insertion sequences (SECIS), downsteam of the UGA codon, recognize specialized partner elongation proteins, SelB homologues (Krol, 2002). They participate in the decision to insert selenocysteine rather than terminate the growing polypeptide chain (Kryukov and Gladyshev, 2004; Lescure *et al.*, 2002; Su *et al.*, 2005).

The twenty-second cotranslationally inserted amino acid found in proteins is pyrrolysine (Pyl) (Krzycki, 2005). It is encoded by the UAG nonsense codon (Lobanov *et al.*, 2006). The insertion of Pyl requires the *pylT* gene, an amber decoding $tRNA^{pyl}$ that is aminoacylated with pyrrolysine by the pyrrolysyl tRNA synthetase PylS (Longstaff *et al.*, 2007). *pylTS* form a gene cluster with *pylBCD*, found in many methanogenic archaea and Gram-positive bacteria. Pyl is found in corrinoid-dependent methylamine methyltransferases of archaea (Krzycki, 2004).

The PylBCD enzyme complex catalyzes biosynthesis of pyrrolysine. When *pylTSBCD* is expressed in *E. coli*, pyrrolysine can be incorporated into proteins, and the activated tRNA *pylT* gene product serves as a UAG nonsense suppressor. These results show that the *pyl* gene cluster represents a "genetic code expansion cassette" (Longstaff *et al.*, 2007).

Analogous cassettes may have served similar functions for some of the other established protein amino acids during evolutionary expansion of the canonical genetic code. Other variations of the standard genetic code include ribosomal programmed −1 and +1 frameshifting, ribosome hopping, and several examples of stop codon readthrough (Cobucci-Ponzano *et al.*, 2005).

Evolution of the Genetic Code

All of this new information provides us with clues as to the evolution of the genetic code (Davis, 2002; Jukes and Osawa, 1993). Initially, there may have been just four amino acids or amino acid types recognized by the translation machinery. These were determined by the nucleotide in the second codon position. A at this position determined one or more strongly hydrophilic amino acid(s). U at this position stipulated one or more strongly hydrophobic amino acid(s), and both C and G determined semipolar amino acids. Over time, increased specificity resulted from recognition of codon position 1 and the consequent specification of a particular amino acid within the type specified by the position 2 nucleotide. Possibly simultaneously or subsequently, the wobble position became recognized in certain instances, allowing further and more precise expansion of the genetic code (Table 4, bottom).

Curiously, when the wobble position is important in specifying an amino acid, the two purines, A and G, almost always specify one amino acid while the two pyrimidines, T and C, specify another (see Figure 4). When a C is in position 2, the wobble position is *never* important, but when an A is in the second position, the wobble position is *always* important. When a G or a T is in position 2, the wobble position is *unimportant* when position 1 is a C or a G but is *important* when position 1 is a T or an A (Figure 4). It is likely that there are thermodynamic explanations for these oberservations.

Selenocysteine and pyrrolysine probably represent the most recent additions to the genetic code. However, molecular genetic manipulations have allowed the cotranslational incorporation of dozens of novel amino acyl residues and their derivatives into proteins (Cropp and Schultz, 2004). The expansive evolutionary process is undoubtedly still in progress, and it would not be surprising if other rare ribosomally inserted amino acids prove to have evolved (or will evolve in the future) for very specific purposes.

Mutator Genes

Sometimes fairly closely related bacteria have very different genomic G+C contents. Moreover, obligatorily pathogenic bacteria with greatly reduced genome sizes tend to have genomes of low G+C contents. So what determines this trait?

For many years, mutator (*mut*) genes, which when mutated give dramatically increased rates of mutation, have been known. The biochemical functions of these genes are now recognized. For example, MutD is the epsilon subunit of DNA polymerase III, a subunit that normally functions in "proofreading," correcting errors in DNA replication. MutR is topoisomerase III, functioning in normal chromosomal segregation as well as homologous recombination and repair (Schofield *et al.*, 1992). The *mutH, L*, and *S* genes encode the DNA adenine methylation (DAM) mismatch repair system (Lobner-Olesen *et al.*, 2005). Thus, mutator genes generally encode nonessential DNA repair proteins. When they are lost, mutation rates increase, giving rise to increased rates of evolutionary divergence.

Some mutator genes influence the genomic G+C content of an organism. Thus, MutT in essence catalyzes GC → AT, so that its loss causes a genome to become more G+C rich. This enzyme, a pyrophosphohydrolase, is a component of the "GO" system responsible for removing an oxidatively damaged form of guanine (7,8-dihydro-8-oxoguanine) from DNA in the nucleotide pool. 8-oxo-dGTP is inserted opposite dA and dC residues in template DNA with almost equal efficiency (Michaels and Miller, 1992). This leads to AT-to-GC transversions; MutT specifically degrades 8-oxo-dGTP to the monophosphate: 8-oxo-dGTP + H_2O → 8-oxo-GMP + pyrophosphate. Thus, only part of the 8-oxo-dGTP becomes incorporated into replicating DNA because the rest is degraded by MutT.

Another mutator gene that influences G+C content is *mutY*, and the enzyme encoded by *mutY* has an effect opposite of that of MutT. MutY catalyzes AT → GC, and its loss consequently lowers the GC content of the genome of the affected organism over time (Lu *et al.*, 2001; Nohmi *et al.*, 2005).

MutY is an adenine glycosylase that is active on G-A mispairs. It removes adenine misincorporated into DNA opposite G or 8-oxoG, the oxidized form of G discussed in the preceding paragraph. It therefore prevents G:C to T:A mutations (Lu *et al.*, 2006). MutY can also correct error-prone DNA synthesis past GO lesions which are due to the presence of 8-oxoG (Lu *et al.*, 2006). The balance between MutT and MutY in an organism can therefore influence the G+C content of the genome over a time period of many generations.

In some organisms, such as *Helicobacter pylori*, the *mutY* gene contains an eight adenine homonucleotide tract. This is subject to slipped strand mispairing, leading to frameshifts that can eliminate *mutY* function (Huang *et al.*, 2006). In this case, expression of *mutY* is "phase variable." Base excision repair may be functional or not functional depending on the need for DNA stability or a high mutation rate.

LONG-TERM GENOMIC DYNAMICS

rrn Genes

In their 1990 review, Krawiec and Riley (1990) considered the dynamic nature of bacterial chromosomes. For example, they

cited the occurrence of ribosomal RNA (*rrn*) genes which occur in multiple copies in an organism. In *E. coli* there are seven copies; however, in different strains of *B. subtilis*, there can be either 9 or 10 copies, depending on the strain. Usually these genes occur in operons coding for the 16S-23S-5S rRNAs in this order, but (1) a locus may have 1, 2, or all 3 rRNA genes, (2) the order and spacing can vary, (3) the percent identities vary for the different paralogous genes in an organism and (4) transfer RNA-encoding genes may or may not be present in the spacer regions (Blattner *et al.*, 1997; Kunst *et al.*, 1997). All of these facts attest to the occurrence of chromosomal rearrangements that have occurred over relatively recent evolutionary time.

Alternative Alleles

Another example of long-term genome dynamics results from the study of alternative alleles in closely related organisms such as the enteric bacteria. Shakeri-Garakani *et al.* (2004) analyzed a set of "facultative" (nonessential) genes involved in sugar metabolism in many strains of *E. coli* and its close relatives. In about 60% of the *E. coli* strains studied, the galactitol (*gat*) catabolic genes occurred at position 46.8 min on the chromosome. In the closely related *Salmonella* and *Klebsiella* strains examined, the *gat* genes, although orthologous to the *E. coli* *gat* genes, mapped to a different chromosomal location, at 70.7 min. At this same site on the *E. coli* chromosome, this locus was usually occupied by a gene cluster for the metabolism of galactosamine and N-acetyl galactosamine. Finally in 90% of the *Klebsiella* strains and 5% of the *E. coli* strains examined, genes for the catabolism of arabinitol or ribitol occurred at the 46.8 min site. When the arabinitol or ribitol genes were present in *E. coli* at this locus, the *gat* genes were absent, and in all *Salmonella* strains, none of these genes was present at the corresponding locus. These observations presumably reflect horizontal transfer of whole operons from other bacteria, recombined into specific, preferential, chromosomal loci in this group of enteric bacteria (Shakeri-Garakani *et al.*, 2004).

IS Elements

Another basis for chromosomal instability deals with the presence of insertion sequence (IS) elements which are small transposons, capable of jumping from one location in the organismal DNA to another (Schneider and Lenski, 2004). Each of these sequences in general codes for the transposase that catalyzes transposition using either a replicative or a nonreplicative mechanism (Mahillon & Chandler, 1998; Mahillon *et al.*, 1999). Ten types of IS elements were discovered in the *E. coli* K12 genome, present in from 1 to 11 copies each (Blattner *et al.*, 1997; see Table 5). These elements, which belong to 6 different IS element families, range in size between 768 nts and 1443 nts, have been implicated in adaptive mutation (Bongers *et al.*, 2003; Hall, 1998; Twiss *et al.*, 2005). Transposon hopping to a preferred target site can activate an operon or inactivate a gene, depending on where it lands and in what orientation it resides. Moreover, the occurrence of multiple copies of one such element provides the opportunity for large-scale chromosomal rearrangements due to homologous recombination.

TABLE 5
Insertion sequence (IS) elements present in the *E. coli* K12 genome

Insertion sequence	# Copies, *n*	Family	Length (nts)	Acc #
IS1	7	IS1	768	X52534
IS2	7	I53	1,331	M18426
IS3	5	I53	1,258	X02311
IS4	1	I54	1,426	J01733
IS5	11	IS5	1,195	J01735
IS150	1	IS3	1,443	X07037
IS186	3	IS4	1,341	M11300
IS30	3	IS30	1,221	X00792
IS600	1	IS3	1,264	X05952
IS911	2	IS3	1,250	X17613

References: Blattner *et al.*, 1997; Mahillon & Chandler, 1998.

Evolutionary Significance of Repeated Genetic Elements to Genomic Plasticity

Another indication of long-term genetic plasticity is the presence of repeat elements and gene duplications with extensive sequence similarity. The presence of such repeats shows that duplication or lateral transfer from another organism occurred relatively recently in evolutionary time. Among the largest of the duplicated genes are the so-called "recombination hot spot" (*rhs*) genes discussed earlier in this article, encoding cell surface proteins, probably involved in adhesion, pathogenesis and capsular polysaccharide production (see section entitled "Duplicated Genes in Bacterial Genomes"). These five composite genes (*rhsA-E*) in *E. coli* K12 differ in sequence, but they all possess giant cores consistent with a cell surface ligand-binding function (Hill, 1999; Hill *et al.*, 1994). Within this core are several repeat sequences, one of which is a peptide motif $GX_3RYXYDX_2GRLX_7$ that is repeated up to 28 times. This N-terminal core is well conserved between the different *rhs* loci, but the C-termini are highly variable. They are mosaic proteins with internal segments derived from different sources, suggestive of horizontal transfer. The internal regions of these genes are hot spots for recombination and for insertion sequence integration, and different such loci are found in other *E. coli* strains (Bhattacharya, 1999; Wang *et al.*, 1998).

Many other examples of gene duplication and transfer in addition to those cited above have been documented (Blattner *et al.*, 1997; Welch *et al.*, 2002). In fact, about one-third of all genes in *E. coli* have paralogs. Two copies of the enzymes of glycolysis are often found, in agreement with the fact that this pathway functions in two capacities, for sugar synthesis and sugar catabolism. Some amino acid biosynthetic genes are present in triplicate, and in the case of the enzymes for leucine,

TABLE 6
Eleven prophage genomes present within the *Escherichia coli* K12 genome

Prophage	Location	Genome Size (nts)	G+C Contents	ORFs,* *n*
KpLE1	2464404-2474619	10,216	0.44	15
CP4-44	2064181-2077053	12,873	0.52	11
CP4-57	2753978-2776007	12,873	0.46	25
e14	1195443-1210646	15,024	0.45	23
Qin	1630450-1646830	16,381	0.44	35
DLP12	564025-585326	21,302	0.43	29
Rac	1409966-1433025	23,060	0.47	31
CP4-6	262182-296489	34,308	0.53	37
KpLE2	4494108-4534178	40,071	0.50	39
CPZ-55	2556791-2563352	6,561	0.48	9
PR-X	2165324-2166023	699	0.47	3

*ORFs =

isoleucine and valine biosynthesis, each of these is regulated differently (Yang *et al.*, 2005). Most spectacular is the occurrence of 80 paralogs of the ATP-binding cassette (ABC)-type energizers (ATPases) in *E. coli* K12. In fact, most bacteria have multiple paralogs of these proteins.

Gene Transfer Between Organisms

An indication of genome plasticity is the presence of at least 11 cryptic prophage regions in the *E. coli* K12 genome (Blattner *et al.*, 1997; see Table 6 and the Prophase Database [http://bicmku.in:8082/]). Each contains multiple genes, some of them useful to the host. Among these are two component cell lysis cassettes that can potentially be used for programming cell death (Srividhya and Krishnaswamy, 2007). Their presence is a clear indication that phage are major players for the promotion of genomic variability.

Sequencing efforts have revealed high occurences of prophage elements in bacterial genomes. Prophages sometimes contribute 10% to 20% of bacterial genomes and account for substantial inter-strain genetic variability. These elements are defective and cannot form infective particles, but they can provide fitness factors. The Prophage Database provides a comprehensive tabulation of prophage elements and phage remnants. It contains hundreds of prophage entries and their encoded proteins. Genome level data provide details of prophage locations, genome sizes, GC contents, taxonomies, and functionalities. Protein level data suggest possible functional and domain associations.

Several different *E. coli* strains have been examined for prophage. *E. coli* K12 has at least 11 such elements (see Table 6), *E. coli* 0157·H7 Sakai has 23, *E. coli* EDL933 has 20, and *E. coli* CFT073 has only 8. These prophage vary in G + C content from 42% to 55%, with from 3 to 84 ORFs per prophage genome. It is therefore clear that there is tremendous variation in prophage content. The prophage usually have inserted between genes or in transfer RNA-encoding genes, but they also occur in regulatory and (occasionally) metabolic genes. They fall into several different types (lambda-like, P2-like, P4-like, P22-like, Mu-like, CP4-like, PhiC31-like, *etc*).

Besides phage-mediated gene transfer between bacteria, other mechanisms of genetic exchange include: (1) conjugation between male (F^+) and female (F^-) bacteria, (2) spontaneous zygogenesis or Z-mating transformation with naked DNA, and (3) gene transfer via phage-like gene transfer agents (Gratia, 2005; Lang and Beatty, 2000, 2001; Lang *et al.*, 2002; Matson *et al.*, 2005). Perhaps phage mediated gene transfer is the most important means of gene transfer in prokaryotes. In nature there may be 10x as many bacteriophage as bacteria (Breitbart and Rohwer, 2005; Edwards and Rohwer, 2005; Rohwer, 2003).

Genes gained by horizontal transfer seem to encode proteins and stable RNAs of virtually all types (Lawrence, 2005). These include genes for central carbohydrate and nitrogen metabolism (e.g., glyceraldehyde-3-phosphate dehydrogenase and glutamine synthase), complete biosynthetic pathways (e.g., cytochrome c biogenesis and coenzyme B_{12} biosynthesis), and components of the transcription and translation assemblies (e.g., elongation factors, ribosomal proteins and even ribosomal RNAs). Indeed, lateral transfer seems to be a common adaptive mechanism promoting niche invasion. Yet not all genes are transferred and incorporated with equal proficiency. Thus, genes for transport and drug resistance seem to have a particularly high propensity for transfer between closely related organisms although not between the three domains of living organisms (Saier, 2003).

Two different bacteria may have relied on transfer for adaptive purposes to tremendously differing degrees. For example, the genome of *Thermotoga maritima* has undergone extensive lateral transfer with archaea (Logsdon and Faguy, 1999; Nelson *et al.*, 1999). Species of *Neisseria* show extensive evidence of horizontal transfer while *Haemophilus* species do not, even though both are competent for DNA uptake. Evidence for gene

transfer from *Haemophilus* to *Neisseria* is available (Kroll *et al.*, 1998; Sandberg *et al.*, 2001).

Genomes in equilibrium, not undergoing significant changes in genome size, must gain and lose DNA at similar rates, dependent on selection pressures operative over many generations. Bacteria tend to lose genes that do not provide appreciable benefit, and since large regions of a chromosome may be deleted in a single event, even beneficial genes may be lost. The genomes of the various sequenced *E. coli* strains and of other bacteria have short nucleotide repeats that occur with variable frequencies, reflecting specific evolutionary pressures dependent on the biological functions of these repeats as discussed in the section entitled "Recurrent Sequences in the *E. coli* Genome." While single nucleotide mutation and duplication provide mechanisms for increasing the numbers of repeats, genetic deletion and mutation provide mechanisms for decreasing these numbers. Thus, we can account for the variable frequencies of specific repeat sequences.

COMPARATIVE *E. COLI* GENOME ANALYSES

As noted in the section entitled 'Codon Usage', *E. coli* K12 genome analyses (Blattner *et al.*, 1997) revealed that translational start sites, ATG, GTG, and TTG, occur with relative frequencies of 85%, 12%, and 2%, respectively, and values are similar in other *E. coli* strains. Translational stop sites, UAA (ochre), UAG (amber) and UGA (opal) occur with relative frequencies of 64%, 28%, and 8%, respectively. Moreover, overlapping stop and start codons, ATGA, TAATG and TGATG, occurring in about 10% of *E. coli* genes, occur with frequencies of 61%, 26%, and 13%, respectively. Overlapping genes allow for translational coupling, so the ribosome does not fall off, but merely terminates translation and initiates translation of a new gene product in an efficiently coupled process that allows for maintenance of a constant ratio of the products of the overlapping genes. This increases efficiency and obviates the need for a new ribosome-binding site (Inokuchi *et al.*, 2000).

In a remarkable set of studies, Perna *et al.* (2001) and Welch *et al.* (2002) compared the genome sequences of different *E. coli* strains, the nonpathogenic *E. coli* K12 lab strain, MG1655, an entero-hemorrhagic strain, EDL933, the haemorrhagic *E. coli* strain, 0157:H7, and a uropathogenic strain, CFT0739. Surprisingly, these strains share only about 2000 genes, or less than 40% of the total. Most of these "core" genes, encode central metabolic "housekeeping" functions. The variable genes, present in some, but not all four strains, encode proteins or complete protein complexes that are important for virulence in the latter three strains, not required for the nonpathogenic K12 strain. These include toxins, autotransporters that export virulence factors to the cell surface, permeases of the ATP-binding cassette (ABC) and phosphotransferase system (PTS) superfamilies, Fe^{3+} uptake systems, switch recombinases (for phase variation; see section entitled "Short-Term Genomic Rearrangements") and type III protein secretion systems (Barabote and Saier, 2005; Kim *et al.*, 2006; Nguyen *et al.*, 2000). The occurrence of frequent gain and loss of genetic material over a relatively short period of evolutionary time (*i.e.*, since these *E. coli* strains diverged from each other) was documented (Welch *et al.*, 2002). Moreover, only 10% of the virulence genes are shared by the two pathogenic strains, each of which contains in its genome about 60 "islands" of foreign DNA. Even in three different uropathogenic strains, the islands often proved to be different or in different positions of the genome. Additionally, *E. coli* K12 MG1655 possesses many genes not found in the other strains examined. Thus, major differences reflected insertions and deletions rather than just single nucleotide polymorphisms (Kudva *et al.*, 2002; Taylor *et al.*, 2002). These differences may reflect intraspecies recombination (Mau *et al.*, 2006; Milkman *et al.*, 2003) and undoubtedly account for the broad host ranges, dissimilar disease states, and varied niches characteristic of these different *E. coli* strains. As a result of extensive DNA gain/loss, the genomes of various strains of a single bacterial species may differ in size by well over a million base pairs.

M. Riley, K. Rudd and others have carried out additional analyses of two *E. coli* K12 strains (MG1655 and W3310) as well as certain pathogenic *E. coli* strains with fully sequenced genomes. These investigators provided much greater gene annotation accuracy than had been available previously (Riley *et al.*, 2006; Rudd, 2000). Rudd (2000), for example, revised over 700 gene start sites, corrected over 60 frame shifts, most caused by sequencing errors, reconstructed over a dozen sequences interrupted by IS element insertion, and identified almost 100 pseudogenes. Serres and Riley (2005) identified multimodular proteins and provided detailed annotations. Riley *et al.* (2006) extended the functional assignments of many gene products and provided literature documentation.

Work reported by Ochman and Davalos (2006) showed that *E. coli* genomes harbor hundreds of inactivated and otherwise functionless pseudogenes. Such bioinformatic analyses are required for evaluation of the metabolic potential of a cell. However, they also provide clues as to the evolutionary histories of these strains and emphasize the dynamic nature of bacterial chromosomes (Mau *et al.*, 2006). These conclusions are generally applicable to many closely and distantly related bacterial types although each strain exhibits characteristic gene complements, frequencies of horizontal transfer and genetic rearrangements, and occurrences of prophage and other foreign elements (Casjens *et al.*, 2000; Deng *et al.*, 2003; Palmer *et al.*, 2000; Serres and Riley, 2005; Tsuru *et al.*, 2006; Wei *et al.*, 2003).

REDUCED GENOME SIZES IN OBLIGATORY PATHOGENS AND SYMBIONTS

The genome of *Mycoplasma genitalium* (Fraser *et al.*, 1995) is of a reduced size (580,000 bp with a mere 470 recognized genes). This human pathogen causes urethritis and grows slowly with a 24-hour generation time. The genome encodes all gene products required for DNA replication, transcription and translation. There are only one set of ribosomal RNA genes and 33 tRNA genes, in contrast to *E. coli* which has 7 sets of *rrn* genes

and 86 tRNA genes. It largely lacks small regulatory RNA genes. *E. coli* has over one hundred of these genes that control transcription, translation, RNA stability, and protein stability and activity (Kaberdin and Blasi, 2006; Majdalani *et al.*, 2005; Vogel and Sharma, 2005). *Mycoplasma* have minimal gene complements for DNA-repair and nucleotide salvage.

M. genitalium has two sugar transporters, both of the phosphotransferase system, one for glucose, and one for fructose, the only two sugars present inside human tissues in appreciable amounts. It has all the enzymes of glycolysis as well as the pyruvate dehydrogenase complex, but it lacks other energy-generating pathways. It cannot make vitamins but can turn vitamins provided by the human host into cofactors. It has minimal capacity for phospholipid metabolism but has chaperone proteins to facilitate proper protein folding as well as an abbreviated but complete general secretory pathway for the export of proteins (Fraser *et al.*, 1995).

What *M. genitalium* lacks are the enzymes for cell wall and outer envelope biosynthesis, as well as biosynthetic enzymes for making amino acids, nucleic acid bases, fatty acids, and vitamins. These macromolecular building blocks and cofactor precursors must be provided by the host. *M. genitalium* lacks oxidative metabolism, a Krebs cycle, the pentose phosphate pathway and gluconeogenesis. It has almost no transcriptional regulation, and protective mechanisms against environmental stress conditions are absent. Presumably this loss resulted from its adaptation to the homeostatic environment provided by the host organism.

Two other related pathogenic organisms with fully sequenced small genomes are *Treponema pallidum*, the causative agent of syphilis, and *Borellia burgdorferi*, the cause of Lyme's disease (Fraser *et al.*, 1997, 1998). The genomes of these two spirochetes are both about 1 Mbp in size, but they are very different. While the *T. pallidum* chromosome is circular, that of *B. burgdorferi* is linear with a central origin of replication and covalently closed, 26 bp AT-rich inverted terminal telemers. However, *B. burgdorferi* strains also contain as many as 21 linear and circular plasmids, none of which is present in *T. pallidum* (Casjens *et al.*, 2000). This is the largest number observed for any bacterium to date. While 93% of the chromosome of both organisms codes for RNAs, only 71% of the plasmids code for anything. The latter code primarily for dozens of *Borrelia*-specific virulence-related surface lipoproteins not found in any other group of organisms. Since genetic rearrangements occur with high frequency on these plasmids, the occurrence of a high percentage of pseudogenes is not entirely surprising (Casjens *et al.*, 2000). The occurrence of spatially distinguishable genus-specific genes seems to be a characteristic of many organisms from the three domains of life and probably has evolutionary significance (Palmer *et al.*, 2000).

The tendency of obligate pathogens to have reduced genome sizes was examined in 2001 when about 200 fully sequenced bacterial genomes were available for analysis (Moran, 2002). The following conclusions resulted: (1) Obligate parasites and symbionts (a) have the smallest genomes, (b) show more rapid protein evolution, and (c) usually have DNA of higher A+T content. (2) They have only ~80 essential, universal genes, but have many species-specific genes. Of course, more than 80 genes are required for life, but nonhomologous replacements and unrecognized orthologues could account for this small number. (3) Reduced genome bacteria have LESS biosynthesis, LESS regulation, and FEWER energy-yielding pathway enzymes. Interestingly, *Buchnera*, an endosymbiont of aphids, an insect that can only make a few of the 20 natural amino acids, has lost the capacity to make all of the amino acids the aphids make, but retains biosynthesis for all aphid-essential amino acids (Nakabachi *et al.*, 2005; Perez-Brocal *et al.*, 2006). (4) DNA recombination/repair enzymes are lost, and there are fewer tRNAs (sometimes only one per amino acid). These facts give rise to HIGHER mutation rates and HIGHER A+T contents (since U substitutes for T, and the conversion of C to U occurs due to deamination of C). Increased A+T occurs at all positions, but mostly in non-coding regions and the third codon "wobble" position as expected. (5) Beneficial genes may be lost due to genetic drift and the pressure to minimalize genome size. Further, (6) some "non-essential" genes are retained. In *M. genitalium*, 129 of the 484 recognized genes are not required for growth, and 50 *Buchnera* genes shared by *E. coli* are of unknown function (Fraser *et al.*, 1995; Moran, 2002).

Why does a bacterium "choose" to reduce its genome size when it becomes an obligate pathogen or symbiont? Does less DNA promote faster, more efficient cell replication? Or does the pathogen strive to minimize transcription and translation? Analyses reveal that genes are *inactivated* before being lost, and they are *not more tightly packed*. Deletions are of all sizes: some are small intragenic deletions, while others encompass many genes (Moran, 2002). In comparing *Mycobacterium leprae* with *M. tuberculosis*, 2000 genes proved to be lost by the former, relative to the latter, but 1000 pseudogenes (partial or nonfunctional genes) remain (Cole *et al.*, 1998, 2001). They are not expressed in *M. leprae*, but are in *M. tuberculosis*. *M. leprae* is apparently still "in transition." Fifty percent of its DNA does not code for proteins. The primary pressure for genome minimalization is energy saved by minimizing translation. This is not surprising when it is considered that 10,000 ribosomes in a typical bacterial cell use a major fraction of the cell's ATP. Minimizing transcription and replication provide secondary and tertiary reasons. These three bases for genome minimalization correlate with the energy expended for their completion.

COMPARISONS BETWEEN PROKARYOTES AND EUKARYOTES

Noncoding and Duplicated DNA

The first eukaryotes for which fully sequenced genomes became available were the yeast, *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996) and the worm, *Caenorhabditis elegans* (Chervitz *et al.*, 1998). As is typical of all eukaryotes, both show increased numbers and sizes of genes as well as greatly

increased amounts of noncoding regions, relative to prokaryotes (Chung *et al.*, 2001). *C. elegans* has a genome including 5 autosomes and an X chromosome, a total of 79 Mbp of DNA. It has about 20,000 genes, each with an average of 5 introns. It thus has nearly 20x as much DNA as *E. coli* with about 4x the number of genes, and 8x the DNA of *S. cerevisiae* with 3x the number of genes. It has just about half as many genes as do humans. Among its 20,000 genes are 659 tRNA genes with 44% on the X chromosome. Large numbers of large ribosomal RNAs occur in tandem at the end of chromosome I, probably reflecting recent duplications. Numerous 5S rRNA genes are found in tandem on chromosome IV. This arrangement can be rationalized from an evolutionary standpoint, but it differs strikingly from that in prokaryotes where the three ribosomal RNA genes most frequently occur at a single locus.

Large duplications found in the *C. elegans* genome include (a) a pair of adjacent 108 kb direct repeats containing 6 genes with only 16 nucleotide differences, and (b) a 23.5 kb inverted repeat with only 8 nucleotide differences, as well as (c) hundreds of other chromosomal duplications (Chervitz *et al.*, 1998). The largest paralogous protein families include (1) 650 homologous integral membrane receptors, (2) 410 protein kinases, (3) 240 zinc finger proteins, and (4) 170 collagenases. It has become clear that eukaryotic genome expansion resulted largely from intragenomic duplication events. The proliferation of paralogues of similar sequences in multicellular organisms may have provided a mechanism allowing differentiation with isoforms being selectively expressed in the various cell types and at different times during development. While eukaryotes thus have larger numbers of genes, prokaryotes have larger numbers of functional gene types in spite of their individually smaller genome sizes (Bardy *et al.*, 2003; Koonin, 2007; Saier *et al.*, 1990). This latter fact presumably reflects their adaptability to a far greater number of environmental niches.

Gene Type Segregation

Borrelia burgdorferi, the Lyme disease bacterium, sequesters its housekeeping genes on its more stable chromosome while *Borrelia*-specific surface lipoprotein genes can be found on its many plasmids where more frequent recombination occurs. *Streptomyces* species (*S. coelicolor* and *S. avermitilis*) also segregate their genes according to gene type. These organisms possess large (8 to 10 Mbp) linear chromosomes with the housekeeping genes localized to the central portions of their genomes and the nonessential and genus-specific genes present on the chromosomal arms (Paradkar *et al.*, 2003). The internal housekeeping regions are highly conserved while the ends are much more dynamic (Chen *et al.*, 2002). They contain large tracts of repetitive DNA and undergo rearrangements and deletions with high frequencies (Chen *et al.*, 2002). These organisms, like *Borrelia*, seem to take advantage of differential mutation rates for different parts of their genomes in order to achieve variable rates of evolution, depending on the gene type (Chater and Chandra, 2006).

Examination of the *C. elegans* genome reveals that spatial separation of housekeeping genes from nematode-specific genes is similarly observed. The former are localized to the centers of the chromosomes, while the latter are found at their ends where meiotic recombination occurs with much higher frequency (Chervitz *et al.*, 1998). Repeated DNA segments are also more frequent at the ends, accounting in part for increased rates of recombination. These features correlate with slow rates of evolution for housekeeping genes and higher rates of evolution for the spirochete-, *Streptomyces*- or worm-specific genes. Parallels between eukaryotes and prokaryotes with respect to segregation of gene types may be a general feature allowing more rapid evolution when rapid change provides selective benefit.

Eukaryotic Operons

Analyses of the *C. elegans* genome have revealed that about one quarter of the genes are in multicistronic operons. This also proved to be true, but to a lesser degree, for vertebrates (Lawrence, 2002). For example, for the vertebrate α-globulin gene cluster and the β-globulin gene cluster, all genes in each cluster are homologous, encoding proteins that are expressed at different developmental times or tissues via temporal *vs.* spatial regulation, respectively. They probably arose by repeated gene duplication events. These genes are controlled by "locus control regions" (LCRs) where the gene position relative to the LCR is an important determinant of expression level. A global LCR regulates multiple genes (Lawrence, 2002). These clusters may have arisen by tandem gene duplication events. A similarity with the *C. elegans* genome, where genes of a type "flock" together, must reflect the evolutionary process that gave rise to multiple paralogs.

In bacteria, operons are not always stable over evolutionary time. They split up and fuse, showing different operon structures in related bacteria (Xie *et al.*, 1999). However, *cis*-acting genes (Yanofsky, 1981) cannot be separated. For example: (1) leader peptides for the *trp*, *leu*, *thr*, and *tnaA* operons in *E. coli* are responsible for antitermination and cannot be separated from the rest of the operons without loss of function; (2) genes upstream of antibiotic resistance genes (*cmlA* and *cat*) encode peptides that interact with ribosomes during their translation, controlling transcription of the downstream genes. The same is true for eukaryotes, except that the upstream ORFs often control *translation* of the downstream mRNA (co-cistrons), rather than transcription.

Other Parallels between Prokaryotes and Eukaryotes

Many other characteristics of prokaryotes can be found in eukaryotes (Lawrence, 2002). One such phenomenon is the occurrence of ribosomal frameshifting during translation (-2, -1, or +1) which yields two or more proteins with the same N-terminus but different C-termini, and sometimes different functions. For example, the *dnaX* gene of *E. coli* encodes both the τ and γ subunits of DNA polymerase, which influence the fidelity of chromosome replication (Pham *et al.*, 2006). τ is shorter than

γ because γ is frameshifted so as to avoid a nonsense codon. Also, the presence of overlapping genes can be observed. For example, a 5-codon gene determining erythromycin resistance is present in an *E. coli* 23S rRNA gene (Tenson *et al.*, 1996, 1997), and *comS*, involved in competence regulation in *B. subtilis*, is present within the *srf* gene, responsible for the biosynthesis of surfactin (Cosby *et al.*, 1999; D'Souza *et al.*, 1995).

Related genes within genes, and genes that fully or partially encode multiple gene products, are common in eukaryotes (Lawrence, 2002). Several examples will be mentioned. (1) In eukaryotes, transcripts containing introns are alternatively spliced yielding multiple proteins with common regions and unique regions. It has been estimated that in humans, each gene gives rise to 10 to 20 gene products (Godovac-Zimmerman *et al.*, 2005, Gudlillo *et al.*, 2006; Humphrey-Smith & Blackstone, 1997). (2) Exon exclusion yields proteins differing in the middle, but the same at the ends. These proteins can differ dramatically in size (O'Neil *et al.*, 1998). (3) The acetylcholine transporter and the enzyme choline acetyl transferase of animals share the same untranslated leader and the first exon. This allows coordinate synthesis of the two proteins. (4) In eukaryotic polycistronic messages, translational coupling can occur as detailed above. As in prokaryotes, this allows coordinated synthesis of the gene products. (5) Genes may overlap, and genes are found in introns. Some of this may arise accidentally, but in other cases, it is of functional importance. (6) 25% of the *C. elegans* genome is in polycistronic operons, promoting increased translational efficiency. The operons often encode proteins catalyzing related functions, but usually not all of the proteins required for a particular function are present within such gene loci. Some of these similarities between prokaryotic and eukaryotic genomes may be coincidental or represent examples of convergent evolution, but others may have resulted from the direct descent of prokaryotic mechanisms to eukaryotes.

SHORT-TERM GENOMIC REARRANGEMENTS

Phase Variation: General Features

Phase variation is a common phenomenon in bacteria, many other organisms, and viruses. There are on/off switches that are usually all-or-none control mechanisms, but there are also up/down switches that lead to different levels of gene expression. Finally, there are switches that allow changes in protein structure, not between just two possibilities but between hundreds of such structures (Schumann, 2006). Sometimes there are even more structures found than can be explained the genetic material of the organism undergoing such changes! We shall see how this is possible.

Many bacterial properties prove to be subject to switch mechanisms, particularly cell surface antigenic macromolecular synthesis in pathogens. These include flagellar antigen production, fimbrial synthesis, lipopolysaccharide biosynthesis in Gram-negative bacteria, capsular polysaccharide generation in both Gram-negative and Gram-positive bacteria, and envelope lipoprotein production (Bahrani-Mougeot *et al.*, 2002; Haagmans and van der Woude, 2000; Moore *et al.*, 2004; Nicholson and Low, 2000; Saier and Jacobson, 1984; Weyand *et al.*, 2001). Immune evasion provides a primary basis for the selection of on/off, up/down, or alternative structure switches. But intracellular processes can also be subject to epigenetic change. Thus, drug resistance, bioluminescence, DNA restriction/modification, complement degradation, transport system activity, and the production of various intracellular enzymes have been shown to be subject to these processes (Henderson *et al.*, 1999; Iqbal *et al.*, 2004; Srikhanta *et al.*, 2005). Pleiotropic regulatory genes controlling the production of multiple cellular traits can also be the targets of this type of regulation (Merickel *et al.*, 1998; Srikhanta *et al.*, 2005). As a consequence, gross cellular physiological characteristics such as colony morphology, propensity for biofilm formation, cell size and growth rate can vary (Deziel *et al.*, 2001; Saier and Jacobson, 1984).

Many mechanisms of gene switching are recognized. These include slipped-strand mispairing recombination, DNA inversion, DNA insertion/deletion, cassette sorting, and DNA methylation. Some of the better-characterized mechanisms will be described below.

DNA Deletions/Insertions

Gene Amplification

Some bacterial genes have been shown to be amplified under selective pressure. Amplification can affect gene dosage, gene expression or the gene product itself (Kashi and King, 2006). Many examples have been described in the literature (Krawiec and Riley, 1990). Weakly expressed drug resistance genes have been tandemly duplicated by homologous recombination, giving rise to multiple copies of the genes coupled with progressive increases in drug resistance. For example, in *B. subtilis*, under conditions where low concentrations of tetracycline had been present during extended periods of growth, 80 to 100 copies of a tet^R marker could be detected (Ives and Bott, 1990).

In *Streptomyces* species, a single resistance gene can be amplified over 500 fold yielding large ($\sim$7–8 kb) imperfectly repeated arrays of extra, reiterated sequences comprising $>$40% of the chromosomal DNA (Fishman and Hershberger, 1983; Fishman *et al.*, 1985). These hypervariable repeat sequences can give rise to mutation rates as high as 0.01 mutations per cell per generation (Hugouvieux-Cotte-Pattat *et al.*, 1989; Leblond *et al.*, 1989). When the drug used for selection is removed, these arrays of repeats remain for many generations, but sporulation is accompanied by excision of the supernumerary copies of the amplified DNA (Hornemann *et al.*, 1989).

Many novel, remarkable examples of gene amplification have more recently been observed. For example, Nicoloff *et al.* (2006) described the amplification of genes encoding a multidrug efflux pump in *E. coli*, the AcrAB-TolC system. When the bacteria were exposed to tetracycline, a multiple antibiotic resistance (Mar) phenotype resulted. A large (149 kb) tandem duplication

occurred that encompassed the genes that encode the drug efflux system. Additionally, an insertion sequence element, IS186, had hopped into the nearby *lon* gene encoding an ATP-dependent protease that degrades the MarA transcriptional activator. Inactivation of Lon resulted in higher MarA levels and greater activation of the expression of the *acrAB* genes. The duplication event was believed to be due to homologous recombination between two IS186 elements, the one in the *lon* gene and a second IS186 element 149 kb away. Both the increased gene dosage and the increased expression of both copies of the *acrAB* genes were responsible for the increased resistance to the drugs (Nicoloff *et al.*, 2006).

Another example involved the lactose catabolic (*lac*) operon in *E. coli* (Hersh *et al.*, 2004). In this case, amplification of the *lac* operon gave rise to a phenomenon often referred to as "adaptive" or "directed" evolution, whereby mutation rates increase in response to stress, giving rise to mutations that specifically reduce the stress (Denamur and Matic, 2006; Gilad *et al.*, 2006; Hastings and Rosenberg, 2002; Schmidt and Anderson, 2006). In this case, amplification of a leaky *lac* allele under starvation conditions in the presence of lactose generated about 25 tandem repeats, providing enough β-galactosidase for growth. The authors reasonably suggest that stress-induced genetic change may not only underlie microbial evolution, it may also contribute to pathogenesis, drug resistance and cancer (Hersh *et al.*, 2004). While such examples of amplification mutagenesis are well documented, it is not clear that this is a true example of adaptive mutation (see Roth and Andersson, 2004).

Intragenic Duplication/Deletion

The cell surface M protein of *Streptococcus pyogenes* (the flesh-eating bacterium; Almengor *et al.*, 2006), a potential macromolecular candidate for vaccine development (Batzloff *et al.*, 2006), varies in size from 40 kDa to 80 kDa (see Figure 6; Hollingshead *et al.*, 1989). This coiled-coil fibrillar protein extends about 60 nm from the streptococcal cell wall and is anchored to the membrane via a C-terminal α-helical hydrophobic segment. Recombination-mediated mutations within the M-protein gene occur with frequencies as high as 10^{-3}/cell/generation, probably by a slipped strand mispairing mechanism. The M-protein inhibits phagocytosis, and the variable N-terminal region, with 4 types of repeats (A-D), specifically binds two components of the human immune system, the complement inhibitor, the C4b-binding protein, and IgA-Fc (Areschoug *et al.*, 2004). The multiple tandem repeats of all four repeat units can vary in number, conferring upon the protein its characteristic antigenic variability (see Figure 6; Fischetti *et al.*, 1988). This turns out to be a common mechanism used by many pathogenic bacteria, trypanosomes and viruses to evade the immune systems of their hosts (Wertz and Moudy, 2004).

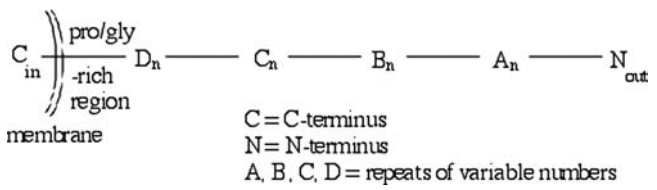

FIG. 6. Schematic depiction of the linear structure of the M protein of *Streptococcus pyogenes*. The C-terminus is in the cytoplasm (C_{in}). The polypeptide chain passes through the membranes as a hydrophobic α-helix, and a proline/glycine-rich region follows. Then the four repeats D, C, B and A occur, each in variable numbers, with all repeats of a unit occurring in tandem. The N-terminus (N_{out}) of the protein can extend 60 nm from the cell surface.

Developmentally-Regulated Genomic Deletions

The cyanobacterium, *Anabaena*, normally grows in linear chains of photosynthetic vegetative cells. However, when nitrogen is limiting, a few distantly spaced cells along the chain undergo spatially and temporally regulated differentiation with the formation of large, terminally differentiated, nitrogen-fixing "heterocyst" (Adams, 2000; Golden and Yoon, 2003; Zhang *et al.*, 2006). The exact spacing of the heterocysts, each separated by 8 to 12 vegetative cells, is species-specific and determined by a diffusible, inhibitory, morphogenic peptide (Adams, 2000; Callahan and Buikema, 2001; Jones *et al.*, 2003; Wolk, 1996).

One of the essential sets of genes expressed in heterocysts, but not in photosynthetic cells, allows nitrogen fixation. In this gene set, three genes occur in an interrupted operon, p–*nifHD*'–*D*"*K*, where p is a promoter, NifD and NifK comprise the N_2 fixing enzyme nitrogenase, and NifH is the electron donor that reduces nitrogenase. The unusual feature of this system is the splitting of *nifD* into two fragments, *nifD*' and *nifD*", separated from each other by an 11 kb sequence. This 11 kb sequence does not encode an intron to be excised at the mRNA level.

In order for nitrogenase to be synthesized, the 11 kb DNA sequence must be precisely excised, fusing *nifD*' to *nifD*" and yielding the functional tricistronic operon, *nifHDK* with all three genes cotranscribed from the *nifH* promoter. The byproduct, resulting from this site-specific recombination event, is a circular 11 kb DNA fragment (Golden *et al.*, 1991). An excision recombinase gene, *xisA*, is located on the 11 kb element. Transcription and excision have been shown to be independently regulated (Golden *et al.*, 1991).

Two other genes in *Anabaena* are interrupted by DNA sequences, and each requires a developmentally regulated excisase to produce the active operon (Carrasco *et al.*, 2005; Ramaswamy *et al.*, 1997). Each DNA excision event, catalyzed at a specific stage during heterocyst development, is mediated by a distinct site-specific recombinase (Carrasco *et al.*, 2005). These recombinases are members of the distantly related tyrosine and serine recombinase families, which together with other enzymes of related function, comprise a huge protein superfamily.

A second example of a programmed deletion allowing expression of a crucial developmental gene occurs during sporulation in *Bacillus* species (Stragier *et al.*, 1989). Two essential sporulation genes are separated by a 42 kb DNA element encoding a site-specific recombinase that recognizes a 5 bp repeat sequence flanking the DNA element (Kunkel *et al.*, 1990). When, as a part of the differentiation program, this so-called "SKIN" element is cleaved out as circular DNA, the two fused partial genes form a functional gene, *sigK*, which encodes the terminal RNA polymerase sigma factor, σ^K. σ^K transcribes genes that mediate cell death in one of the two cells in the sporangium. Programmed death takes place only in the mother cell, not the forespore which upon maturation will form the dormant spore. Since the mother cell houses and contributes to maturation of the forespore by making σ^K-dependent gene products, deletion of this DNA is essential for completion of sporulation. Because the mother cell is programmed to die, the process can be irreversible without loss of genetic information to the progeny of the spore (Kunkel *et al.*, 1990).

When the excised DNA SKIN element was sequenced (Takemaru *et al.*, 1995), it was found to contain 60 recognizable genes. Analyses of these genes led to the probability that this element derived from an ancestral temperate bacteriophage; it is a cryptic remnant similar to two previously known elements, phi105 and the defective phage, PBSX (Mizuno *et al.*, 1996; Takemaru *et al.*, 1995). One can therefore hypothesize that a phage genome inserted into the *sigK* gene and became a permanent resident of the chromosome, to be excised only in response to developmental control signals. This event may have occurred late in the evolutionary process giving rise to the program of sporulation since (1) not all sporulating *Bacillus* species have the SKIN element, and (2) when researchers permanently excised it from the genome of *B. subtilis*, using molecular genetic techniques, the program of differentiation was not interrupted (Qiu *et al.*, 2004; Stragier *et al.*, 1989). The presence of the SKIN element probably provides a safeguard or back-up mechanism to ensure that the *sigK* gene is not expressed too early before σ^K is needed. This ensures proper timing of events during development including the terminal step, mother cell death.

DNA Inversions

The Principles of Recombination

In the preceding sections, DNA rearrangements involving a variety of genetic events such as insertions, duplications and deletions were discussed. Often these processes are catalyzed by recombinases, employing either site-specific or homologous recombination. Such events require that regions of homology flank the inserted/deleted segment. For site-specific recombination, these repeats are usually short and precise, but for homologous recombination, they are longer and need not be so precise. In either case, these repeated elements must be *direct* repeats for *insertion/deletion*. If instead, these repeat elements are *inverted*, relative to each other, *inversion* of the intervening DNA segment upon recombination will occur. However, the mechanism of recombination is the same. These processes are illustrated in Figures 7A and B, respectively.

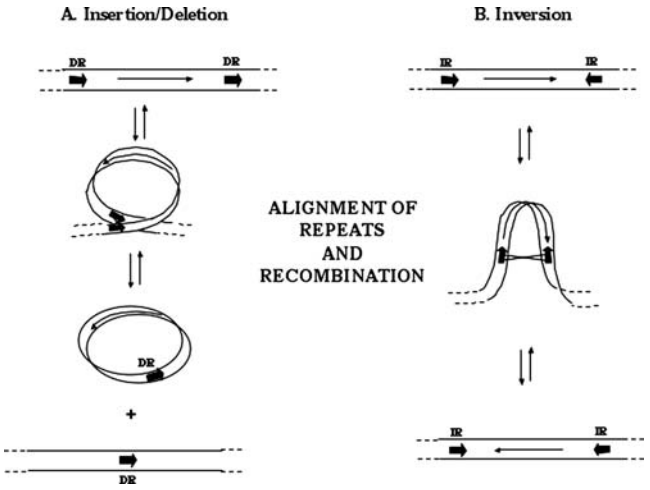

FIG. 7. Recombination giving rise to (A) insertion/deletion and (B) inversion of the region between the two repeats. The repeats are indicated by the filled arrows. DR, direct repeats; IR, inverted repeats. The arrows between the repeats indicate the relative direction of the nucleotide sequences.

Inversion of the Hin System in Salmonella

An example of phase variation involving DNA inversion, and the first such system to be studied in detail, is the flagellar switch mechanism in *Salmonella* strains (Craig and Kleckner, 1987; Glasgow *et al.*, 1989; Plasterk *et al.*, 1985). Although this bacterium encodes a single flagellum, there are two strongly antigenic flagellin monomers, each of which can comprise the long helical flagellar filament that extends from the bacterial surface. Either one flagellin (H1) or the other flagellin (H2) is expressed at any one time, but the two types interconvert at a rate of 10^{-3}–10^{-5}/cell/generation, about 100,000-fold higher than normal mutation rates. The relevant gene arrangement on the *Salmonella* chromosome is shown in Figure 8A. The invertible loop is about 1000 nucleotides long and bears the *hin* gene that codes for the H-invertase, a site-specific serine recombinase expressed from its own promoter. This loop also bears a promoter for the operon containing the *H2* and *rh-1* genes (Simon and Silverman, 1983).

Two identical dimeric Hin recombinases, each bound to one of the two inverted repeats, form a tetramer, bringing the two sites together (Lee *et al.*, 1998). The Hin dimer interface recognizes the Fis nucleoid protein which acts as an inversion enhancer, stimulating inversion many fold. Thus, a Hin dimer, bound to the recombination site, forms a stable complex with Fis bound to the recombinational enhancer (Lee *et al.*, 1998). When the promoter faces the *H2* gene, the *H2* and *rh-1* gene products are made. The H2 flagellin is incorporated into the

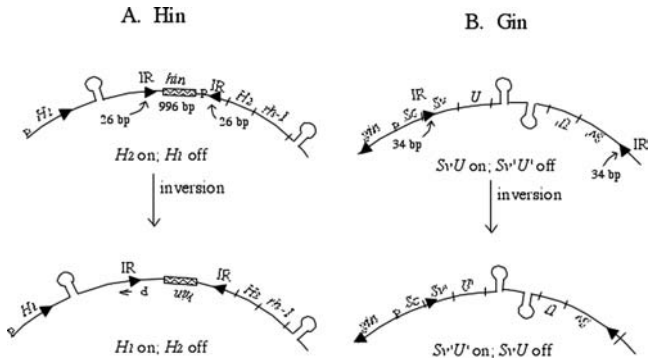

FIG. 8. Phase variation in (A) the flagellar Hin (H-loop inversion) system and (B) the phage Mu Gin (G-loop inversion) system. (A) *H*1 and *H*2 are the two flagellin structural genes; IR, the inverted repeat of 26 bps. The Hin loop is 996 bps long. The *hin* gene, with its own promoter on the invertible element, codes for the H-invertase. The *rh-1* gene codes for the putative repressor of *H*1 expression (see text). The stem loop structures are transcriptional terminators. (B) The two tail fiber genes, *S* and *U* are partially and fully present on the 3000 bp invertible loop, respectively. *Sc* is outside of the invertible loop and is therefore constant while *Sv* is within this loop and is therefore variable. U, also on the loop, is variable, depending on the orientation of the loop, relative to the promoter (p, on the left). The *gin* gene is fully encoded outside of the invertible loop. See text for further explanation.

flagellum while the Rh-1 protein shuts off expression of the *H*1 gene.

Originally, Rh-1 was thought to be a transcriptional repressor, but more recent studies indicated that it prevents *H*1 expression by several complicated mechanisms (Aldrich *et al.*, 2006; Bonifield and Hughes, 2003). It blocks both translation of the *H*1 gene and use of the H1 protein for flagellar construction by binding to the *H*1 mRNA upstream of the translational start codon. To explain their experimental observation, Aldrich *et al.* (2006) proposed that the two mRNAs that encode the H1 and H2 flagellins compete for occupancy of a unspecified site that directs the flagellins toward flagellar assembly, and that Rh-1 influences this competitive process. This proposal suggests a more complicated mechanism of flagellar assembly than was previously considered. It might serve to prevent assembly of flagella with a mixture of subunit types.

Inversion of the Gin System in Phage Mu

Bacteriophage Mu is capable of switching its bacterial host range by a mechanism that is analogous to flagellar phase variation. In one of the two possible states, termed (+), the phage is capable of infecting *E. coli* but none of a variety of other Gram-negative bacteria. Phage Mu is capable of switching its host specificity with a frequency of about 0.03 times per phage particle per generation. In the alternative state, termed (–), it is incapable of infecting *E. coli*, but it exhibits specificity toward *Citrobacter freundii* and certain other closely related Gram-negative bacteria. The availability of several Mu phage-like genomes reveals that not all such phage have this capability, suggesting that the mechanism may have been relatively volatile over evolutionary time (Morgan *et al.*, 2002).

Host range variation results from inversion of a segment of the phage genome. The invertible strand, termed the G segment, is about 3000 base pairs long and represents about 8% of the total phage DNA. Inversion results from intramolecular recombination between inversely repeated DNA sequences, 34 base pairs long, which flank the G segment (Figure 8B). As for flagellar phase variation, a single enzyme, encoded by the *gin* ("*G*-loop *in*version") gene, catalyzes inversion of this segment and thus functions as the "flippase." The *gin* gene is adjacent to and to the right of the G segment (Figure 8B).

Two pairs of homologous genes are fully or partially encoded on the G loop. Two of these genes, denoted *S* and *U*, are read off of one DNA strand in one direction, while the other two genes, designated *S′* and *U′*, are read off of the other strand in the opposite direction. These genes code for the proteins that comprise the phage tail fibers and function in adsorption of the phage to a cell surface macromolecule, probably lipopolysaccharide, in the host bacterium. These proteins confer upon the phage its host range specificity. The promoter is outside of and to just one side of the G-loop segment. The situation is therefore the converse of that of the *H*2 operon in *Salmonella* where the promoter and gene encoding the inverting enzyme are within the invertible segment. Here these two genetic elements map outside the G loop while the structural genes (*S* and *U*, or *S′* and *U′*) of the operons undergo inversion.

The actual situation is somewhat more complex. While the *U* and *U′* genes are fully encoded within the G loop, the *S* and *S′* genes are only partially included within the 3000 base pair invertible segment, and these variable parts of the genes are termed *Sv* and *Sv′*, respectively (Figure 8B). A constant upstream region of the *S* and *S′* genes is located outside of the G segment and adjacent to it. This common *S* region (*Sc*) is joined by G inversion to either one or the other of the two variable parts of the *S* gene (*Sv* or *Sv′*). The *S* or *S′* protein synthesized therefore represents the product of a genetic hybrid between the *Sc* region and either the *Sv* or the *Sv′* gene fragment, respectively.

It is of considerable evolutionary interest to note that the G-loop is found in another *E. coli* phage, P1. Moreover, the *hin* gene in *Salmonella* and the *gin* gene in phage Mu are homologous, and these two flippases can substitute for one another catalytically. This suggests that these genetic elements may once have been transposable as well as invertible, and that they arose from a common ancestral nucleotide sequence (an inversion sequence) that could insert into bacterial and phage chromosomes. The evolutionary advantages of these switch mechanisms to the bacteria and the phage are clear. Flagellar phase variation allows *Salmonella* to escape the host immune

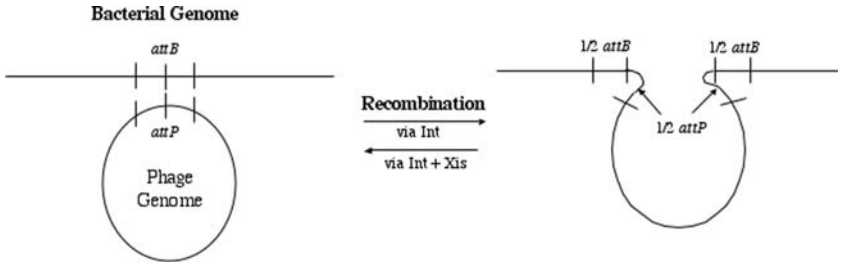

FIG. 9. Schematic depiction of the recombinational event that takes place when the circular phage lambda chromosome (*bottom*) integrates into the *E. coli* chromosome (*top*). The recombinational sites (direct repeats) are *attB* (attachment site on the bacterial chromosome) and *attP* (attachment site on the phage chromosome). Following integration to give the lysogenic state, each of the two recombined sites consists of half of *attB* and half of *attP*. Consequently, all of the four potential recombination sites are different in sequence. The different sequences of the four sites, together with the structures of integrase and excisase, can provide recombination site specificity. Other proteins also play roles (see text).

system, while G loop inversion allows the phage to expand its host range without loss of specificity. It is of interest to note that phase variation in viruses and microorganisms frequently influences the potential for interactions with environmental agents that promote survival.

The Phage Lambda Intasome: A Model System

In phage lambda, the site-specific integrase, Int, promotes integration of the phage genome into the bacterial chromosome. Moreover, Int, together with an excisase, Xis, catalyzes excision (Figure 9). Because two proteins with different directional properties are involved in insertion and excision, the greater degree of resultant complexity allows for irreversible as well as reversible transitions. Moreover, the relative rates of integration versus excision can be responsive to environmental conditions and the presence of accessory proteins such as integration host factor (IHF). For integration, Int, a tyrosyl recombinase, catalyzes recombination between the phage and bacterial chromosomes while bound to a DNA core site via a C-terminal catalytic domain and an N-terminal "arm" site. The process is facilitated by the presence of accessory DNA binding proteins (Radman-Livaja *et al.*, 2006).

For excision, Xis regulates construction of higher order nucleoprotein complexes termed "intasomes." The intasome can preferentially integrate or excise the λ genome, depending on conditions and the protein composition. While Int catalyzes site specific recombination for both integration and excision, Xis modulates the directionality of recombination by stimulating excision and inhibiting reintegration (Abbani *et al.*, 2007). Directionality results from cooperative assembly of the nucleoprotein complex onto a DNA regulatory element which the proteins distort to favor and stabilize the excisive conformation over the integrative conformation.

The recently solved crystal structure of Xis bound to its control region reveals the presence of cooperatively associated Xis protomers, bound head to tail. Each protein binds via nonbase-specific contacts, and the polymer introduces DNA curvature with slight positive writhe. This complex apparently serves to correct disparate segments of the DNA double helix, bridged by integrase within the intasome (Abbani *et al.*, 2007).

Several Xis protomers may normally bind head to tail before excision occurs (Sun *et al.*, 2006). Individual steps in the multistep reaction have been proposed (Mumm *et al.*, 2006). The same or similar processes are presumably responsible for integration of the numerous prophage found in bacterial genomes such as that of *E. coli* K12.

The E. coli Type 1 Fimbrial (fim) Switch

The phage lambda intasome provides a model for understanding the molecular details of the recombinational events that accompany inversion of the fimbrial (*fim*) loop in *E. coli*, catalyzed by FimB and FimE, one of the most complicated of all well characterized recombinational switch mechanisms. In fact, there is still substantial controversy about the detailed mechanism. In this case, a 314 bp invertible loop contains only a promoter. Switching is controlled by the products of two homologous genes, *fimB* which reversibly catalyzes ON ↔ OFF, and *fimE,* which essentially irreversibly catalyzes ON → OFF (Henderson *et al.*, 1999; see also Stentebjerg-Olesen *et al.*, 2000). Curiously, these fimbrial recombinases, FimB and FimE, are not homologous to Hin or Gin, but instead are homologous to lambda phage integrase, Int. Other proteins play regulatory roles in switching: the H-NS and the IHF nucleoid histone-like proteins, the Leucine responsive protein, Lrp, and the stationary/starvation phase stress sigma factor, σ^S. Availability of certain charged tRNAs also influences the switch (Ritter *et al.*, 1997).

The unique directionality of FimE, which catalyzes ON → OFF versus the lack of directionality of FimB, is based on (a) differences in the structures (amino acid sequences) of the homologous FimB and FimE recombinases, (b) the specific DNA sequences of the recombination FimB and FimE binding half sites, the centers of which are the cleavage sites for recombination, (c)

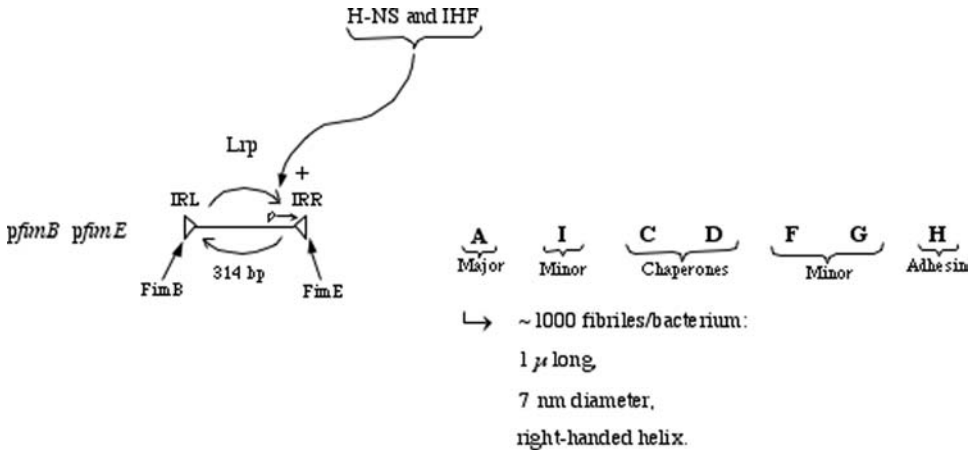

FIG. 10. Gene structure of the *fim* locus in *E. coli*. H-NS and IHF act synergistically to enhance promoter strength when the promoter is in the "on" position in front of and to the left of the *fimA* gene. The *fimB* and *fimE* genes are in front of (to the left of) the invertible segment of 314 bp which is flanked by the inverted repeat left (IRL) and the inverted repeat right (IRR). Following recombination, the two half sites recombine, yielding different sites as noted in Figure 9 for the lambda recombination sites. FimB preferentially binds half sites that flank/overlap IRL while FimE preferentially binds half sites that flank/overlap IRR. FimE catalyzes a unidirectional switch because it can't bind effectively to IRL, due to low affinity and because it is not made in the "off" configuration. The latter is due to a *cis* DNA effect. Note that all four half sites have a conserved 5′CA doublet that is required for binding of either FimB or FimE. Fimbrial assembly requires the major FimA subunit, FimC and FimD, the chaperone and usher proteins, respectively, that fold and export the subunits in preparation for assembly on the external surface of the outer membrane, the minor fimbrial subunits, FimI, F and G, and the adhesin, FimH, at the end of the fimbrium that binds to glycolipids and glycoproteins in the host cell plasma membrane, thereby facilitating infection.

the presence of *cis*-active sites that flank the *fim* recombination sites, and (d) differences in the levels of expression of the two recombinases (Kulasekara and Blomfield, 1999; Leathart and Gally, 1998).

The actions of FimB and FimE are modulated by temperature and medium conditions. First, the synthesis and activity of FimB is optimal at 37°C while FimE is optimal at lower temperatures (Olsen *et al.*, 1998). H-NS plays a role in regulating *fimB* and *fimE* expression (Olsen *et al.*, 1998). As a consequence, *fim* is more "on" at body temperature, and "off" outside the host. Second, the rate of catalysis of the irreversible switch, ON → OFF, by FimE is much faster (∼0.7/cell/generation) than the reversible switch ON ↔ OFF, catalyzed by FimB (∼0.001/cell/generation). This renders fimbrial synthesis OFF most of the time, especially at lower temperatures where FimE exhibits maximal activity. Third, Lrp, which is most effectively potentiated by leucine binding, binds to three DNA binding sites in a control region within the invertible loop (Figure 10; Lahooti *et al.*, 2005). When it binds to sites 1 and 2, it induces a conformation with increased frequency of inversion, but binding to site 3 promotes a conformation with a decreased frequency of inversion. Fouth, the presence of either leucine or alanine, bound to Lrp, may favor a complex with sites 1 and 2 occupied by selectively decreasing binding to site 3 (Roesch and Blomfield, 1998). Therefore, the presence of these amino acids, indicative of the environment of the host animal, increases the rate of switching (Lahooti *et al.*, 2005). IHF and Lrp both bend the DNA creating loops, the latter probably by wrapping and bridging mechanisms. Looping helps to align the two inverted repeats for recombination. These proteins also promote essential protein-protein and protein-DNA interactions. Finally, an additional means of regulation involves IS1 insertion that inactivates *fimE* (Stentebjerg-Oleseu *et al.*, 2000).

H-NS enhances the activity of the *fimA* promoter primarily when this promoter is in the "on" configuration. The H-NS binding site is downstream of the FimE binding site but extends into the switch region that overlaps one of the two IHF binding sites, site 1. H-NS and IHF may act synergistically to enhance promoter activity. These regulatory features of the *fim* locus are illustrated schematically in Figure 10.

In addition to the proposed direct effects of H-NS, this histone-like protein binds to the promoter regions of both *fimB* and *fimE* and represses expression of *fimB* while activating *fimE* expression. Therefore, *hns*-negative strains show increased *fimB* and decreased *fimE* expression. This dependency allows switching to be responsive to the cytoplasmic concentrations of H-NS, and hence to temperature and growth phase. H-NS modulates expression in a way that favors a fimbriated state at mammalian body temperature (Olsen *et al.*, 1998). Both local transcription and H-NS modulate the recombinational event by both direct and indirect mechanisms (O'Gara and Dorman, 2000).

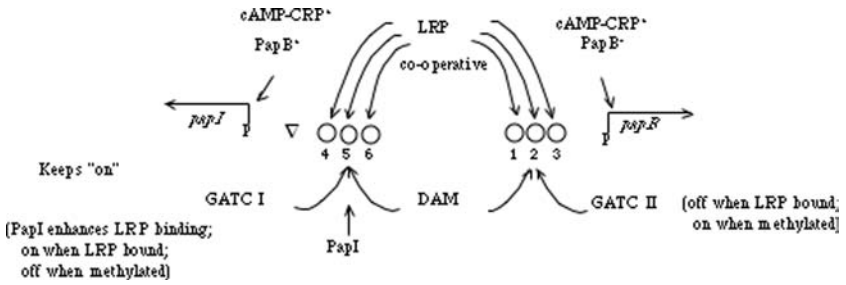

FIG. 11. Proposed mechanism for the switch that controls *papB* operon expression. The *papI* and *papB* operons are transcribed divergently, each from its own promoter. Between these two promoters are six (two sets of 3) Lrp binding sites, each with a DNA adenine methylation (DAM) site in Lrp binding sites 2 (GATCII) and 5 (GATCI). Lrp binding to each set of these three binding sites is cooperative, but methylation of site 2 or site 5 diminishes affinity and prevents cooperative Lrp binding. Normally, one triplet is methylated while the other binds Lrp.

More recent studies have shown that the *fim* switching mechanism is even more complicated than described above. At least two additional recombinases, found in many *E. coli* strains, have been shown to catalyze inversion, preferentially from off to on (Bryan *et al.*, 2006; Xie *et al.*, 2006). In fact, as many as 10 FimB/FimE pairs can be encoded within the genome of a single pathogenic *E. coli* strain. These homologous recombinases apparently act independently of FimB and FimE and may have other primary functions in the cell.

Over 50 distinct fimbriae have been identified in *E. coli*, based on the genome analyses of many pathogenic strains. Usually only a few are present in any one strain, but each fimbrium recognizes a different set of glycolipids and/or glycoproteins in the animal host cell membrane, and the complement of these complex adhesins determines in part the host range of that strain. Moreover, when more than one fimbrium is present in a single strain, regulators of one can interact and influence expression of another by influencing the switching process (Holden *et al.*, 2001). This type of fine-tuning "crosstalk" undoubtedly has physiological significance and may result from the tremendous energy investment required for synthesis of the hundreds of hair-like fimbriae that can cover the surface of a bacterial cell. It may also help to ensure variability or sequential expression of the various adhesins throughout the stages of an infection process (Holden *et al.*, 2001; Xia *et al.*, 2000).

The E. coli Pyelonephritis-Associated Fimbrial (pap) Switch

Another fimbrium in *E. coli* is the pyelonephritis-associated pilus (Pap). *pap* and *fim* fimbrial adhesin switches exhibit regulatory crosstalk in uropathogenic *E. coli* strains (Holden *et al.*, 2006). One might therefore anticipate that the switch mechanisms for these two fimbriae would prove to be similar. Careful examination of these two switch systems do, in fact, reveal some superficial similarities that may be suggestive of a common origin, but if so, the two mechanisms have diverged so extensively that the differences now outweigh the similarities (Henderson *et al.*, 1999).

pap genes, like those for many other *E. coli* fimbriae, can be found within pathogenicity islands in the bacterial chromosome (Kariyawasam *et al.*, 2006). The mechanism of *pap* switching, illustrated in Figure 11, involves differential DNA methylation using S-adenosyl methionine as the methyl donor (Urig *et al.*, 2002). In contrast to most other switch mechanisms, DNA methylation gives rise to a change in phenotype (*e.g.*, on or off) without a change in genotype (*i.e.*, DNA sequence). Methylation is probably the most common mechanism used in eukaryotes to control expression of differentiation-specific genes (Bender, 2004; Goll and Bestor, 2005; Huang *et al.*, 1999; Schob and Grossniklaus, 2006).

The methyltransferase involved in *pap* gene control is the DNA adenine methyltransferase, DAM (see section entitled "Short Oligonucleotide Repeats"). DAM binds to the tetranucleotide sequence, GATC, to methylate the adenine at position N6 in the purine ring (Peterson and Reich, 2006). The newly synthesized DNA strand slowly become methylated following DNA replication; it takes about 20 minutes to methylate half of the available substrate adenines, allowing post replicative mismatch repair with the enzyme apparati distinguishing the old strand from the new strand, as discussed previously in this review. However, DAM methylation has been utilized by parasites such as phage to estimate the optimal timing for phage replication, and also for the purpose of controlling gene expression, including switching of *pap*.

Dam sites are usually processively methylated in a "random walk" process, methylating many sites in one strand of the hemi-methylated double DNA helix during a single processive run. In the process, the enzyme keeps its orientation with respect to the DNA, methylating target GATC sites on just one DNA strand (Hermann *et al.*, 2004). However, methylation sites can be protected from methylation by the binding of other proteins (Humeny *et al.*, 2003; Urig *et al.*, 2002). In the *pap* system, a simple competitive binding mechanism allows switching to be controlled at two alternative sites, GATCI and CATCII, by a nonprocessive mechanism (Peterson and Reich, 2006).

Switching is controlled by methylation at GATCI and II, both of which are in the regulatory region between the two divergently transcribed *papI* and *papBA* operons (see Figure 11). Regulation results from the competitive action of DAM and Lrp. When GATCI is methylated, GATCII is unmethylated and Lrp binds. Then *pap* is off. With the opposite arrangement, with GATCII methylated and GATCI unmethylated with Lrp bound, *pap* is on. Leucine and alanine promote Lrp binding (Berthiaume *et al.*, 2004).

The promoter for the *pap* structural genes is upstream of *papB*, overlapping the GATCII site (Figure 11). Lrp, bound to sites 1, 2, and 3 overlapping GATCII, blocks *papB* transcription (phase off). Methylation of GATCI keeps transcription off. Transcription occurs if Lrp binds to sites 4, 5, and 6 overlapping GATCI (phase on). Methylation of GATCII keeps transcription on. Moreover, PapI binding to the Lrp-DNA complex bound at sites 4, 5, and 6 stabilizes the complex, keeping *pap* on via enhancement.

Other proteins are also involved. H-NS, PapB, Crp and the phosphorylated form of CpxR (CpxR-P) participate in regulation. PapB oligomers bind the DNA and autoregulate *papB* operon expression. PapB also activates *papI* transcription by binding upstream of the *papI* promoter. Crp is required for transcription of both the *papB* and *papI* operons. H-NS is responsible for thermal regulation of *pap* expression by blocking methylation at 23°C but not at 37°C because it binds effectively to both GATCI and GATCII at 23°C but not at 37°C. Finally, the environmentally responsive CpxAR sensor kinase/response regulator inhibits *pap* phase variation because CpxR-P competes with Lrp for binding to both sets of triplet binding sites, inhibiting *pap* transcription. In contrast to Lrp, CpxR-P binding is methylation insensitive. CpxAR-dependent repression of *pap* expression can be observed in response to alkaline conditions (Hernday *et al.*, 2004). Methylation is believed to be the primary switch that prevents cooperative binding of Lrp to either of the two alternative triplet binding sites, 1, 2, and 3 or 4, 5, and 6. The switch probably depends on the cooperative and antagonistic actions of H-NS, PapI and CpxR-P for proper function.

Nucleases, Recombinases and Topoisomerases: A Unified Mechanism of Action

We have seen that invertases which invert DNA segments using a pair of inverted repeats, and integrases/excisases which integrate or excise DNA elements using a pair of direct repeats, are merely site-specific recombinases, two classes of which have been discussed in previous sections. Seryl and tyrosyl recombinases are distantly related to each other, and they use similar mechanisms. These enzymes are homologous to and use essentially the same mechanisms as do topoisomerases; they also share mechanistic features with nucleases (Cheng *et al.*, 1998; Wigley, 1998).

The generalized mechanism used by recombinases and topoisomerases involves reversible transesterification where the enzyme forms a covalent bond between the aromatic or aliphatic hydroxyl group in the nucleophilic tyrosyl or seryl residue, respectively, at the active site of the enzyme and the 3′ nucleotide (Li *et al.*, 2005). This frees the 5′ hydroxyl group in the DNA. The core recombination sites include two inverted elements, usually 11 to 13 bps long, separated by a 6 to 8 bp spacer. Each inverted element is recognized by a functionally monomeric enzyme.

Three-dimensional x-ray structures of tyrosyl recombinases and their homologs with and without bound DNA have revealed that each monomer in these di- or tetrameric enzymes binds the DNA in a crescent shape. The small N-terminal "arm" domain and the large C-terminal active site domain together provide a crevice for the nucleic acid (Kamtekar *et al.*, 2006; Subramanya *et al.*, 1997; Tirumalai *et al.*, 1997; Yang and Steitz, 1995). The enzyme may wrap the DNA around an α-helix, bending it into a higher energy, more reactive form that is more susceptible to nucleophilic attack (Subramanya *et al.*, 1997). Large movements of both protein and DNA are probably required to achieve strand exchange (Li *et al.*, 2005). A 180° rotation of one dimer within the tetramer relative to the other may reposition the DNA duplexes for strand exchange (Li *et al.*, 2005).

The generalized nucleophilic displacement mechanism for tyrosyl topoisomerases and site-specific recombinases is shown in Figure 12. Two arginyl residues (R_1 and R_2) provide positive charges, neutralizing the developing negative charges in the transition states and intermediates. Two histidyl residues (H_1 and H_2) function by general acid/base catalysis, and the deprotonated aromatic hydroxyl group in the tyrosyl residue serves as a potent nucleophile, attacking and then forming a covalent bond with the electropositive phosphorous atom. The reaction is fully reversible, allowing reformation of a phosphodiester bond in the original (*e.g.*, for topoisomerases) or a dissimilar (*e.g.*, for recombinases) DNA strand. Reformation of the DNA strand allows dissociation of the enzyme from the nucleic acid.

This mechanism, and the roles of the arginines and histidines, are essentially the same in recombinases, topoisomerases and nucleases. In nucleases, there is no covalent bond formed between the enzyme and the DNA, and the arginines coordinate the nonbridged oxygens, stabilizing the pentavalent phosphorous intermediate via charge neutralization. The histidines stabilize the transition state, provide general acid/base catalysis (*e.g.*, for deprotonation of the aromatic hydroxyl in tyrosine for nucleophilic attack and reprotonating it during strand rejoining). They also protonate the 5′-hydroxyl that is freed when the enzyme forms a covalent bond with the 3′ hydroxyl and deprotonate it in the reverse reaction (Figure 12).

The following functional characteristics provided evidence that topoisomerases, recombinases and nucleases might share a common origin before sequence data were available to substantiate this possibility (Champoux, 1994): (1) Topoisomerases have site-specific nuclease activity. (2) Many recombinases have topoisomerase activity. (3) Removal of the active site tyrosyl residue in recombinases converts them into nucleases. (4) Restriction endonucleases can be converted into topoisomerases by point mutations, and (5) some DNAases are also RNAases.

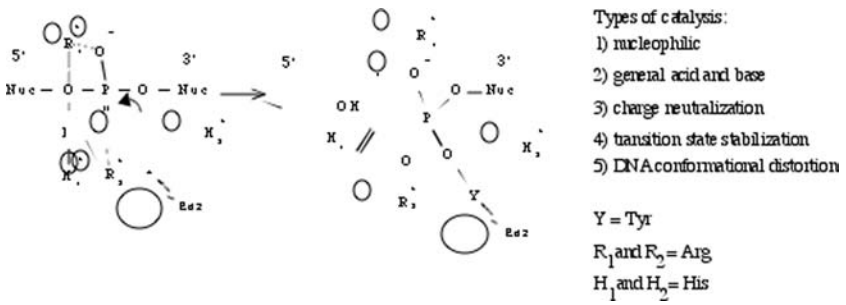

FIG. 12. Mechanism of nucleophilic catalysis for tyrosyl topoisomerases and recombinases. The conserved arg-his-arg-his [or trp in some enzymes] (RHRH(W)) motif surrounds and activates the phosphodiester bond for cleavage. $R_1 + R_2$ = two arginyl residues that function in charge neutralization; $H_1 + H_2$ = two histidyl residues that function in general acid/base catalysis. $Y\text{-}O^-$ is the nucleophile that attacks at the electropositive phosphorous atom to break the phosphodiester bond in the nucleic acid. The reverse reaction with the same or a different strand reseals the nucleic acid and releases the enzyme.

Thus, topoisomerases, recombinases, RNAases, DNAases, and restriction endonucleases are all variants on a single mechanistic theme. This results from the fact that they all share a common ancestry and have retained essential mechanistic features during their functional divergence from each other (Wigley, 1998).

Phase Variation Involving Multiple Mechanisms

The Opacity (opa) Locus in Neisserial Species

Cell surface proteins contribute to the unique cell-type specific properties of a species or cell type (Boyle and Finlay, 2003). In Neisserial species (*N. gonorrhoeae*, the causative agent of gonorrhea, and *N. meningitidis*, a cause of meningitis), cell surface "opacity" proteins, which confer on colonies an opaic appearance, contribute substantially to the antigenic and biological characteristics of the cell. There are 3 to 4 *opa* loci in *N. meningitidis* and 11 to 12 in *N. gonorrhoeae* (Aho *et al.*, 1991; Connell *et al.*, 1990). Opa proteins cover the surfaces of these two species with the specific antigens expressed varying continuously (Gray-Owen, 2003).

The opacity (Opa) proteins are outer membrane antigens that enhance the rate of natural transformation (Hill, 2000). Opa proteins are the "receptors" that bind DNA via ionic (DNA^-–Opa^+) electrostatic interactions. They also promote adhesion and invasion of epithelial and endothelial cells (Dehio *et al.*, 1998; Gray-Owen, 2003; Hauck and Meyer, 2003; Merz and So, 2000). In this capacity, Opa variants can mediate entry and transcytosis into host epithelial and endothelial cells, suppression of lymphocyte function and killing of neutrophiles (Gray-Owen, 2003). Surprisingly, in epidemics, there is greater antigenic variability than can be accounted for by the *opa* gene repertoire of a cell (Achtman, 1995).

There are several mechanisms promoting *opa* variability, and these involve both phase and antigenic variability (Henderson *et al.*, 1999). Phase variation (on/off) results from reading frame alterations. At the beginnings of the genes encoding these proteins, is a pentanucleotide repeat region $(CTCTT)_n$ (see Figure 13). If n=3, 6, 9, 12, *etc.*, the *opa* gene codons will be in frame, and the protein will be synthesized. However, if n=1, 2, 4, 5, 7, 8, 10, 11, *etc.*, the reading frame will be shifted, and although the mRNA will be made, no functionl protein will be synthesized. This repeated region can code for leucine (the CTC and CTT codons) and serine (the TCT codon). It therefore codes for a fairly hydrophobic transmembrane α-helix, containing twice as many leucine as serines, that serves as part of a signal sequence for secretion of the protein across the cytoplasmic membrane. Slipped strand mispairing occurs with a frequency of 10^{-3}/cell/generation, giving rise to phase variation, and hence to variable surface antigenic properties of the cells in a population. This helps the bacteria to escape the immune responses of the host organism to infection.

Because there are multiple *opa* loci, and these exhibit sequence similarity, recombination can occur between them, giving rise to hybrid proteins of altered sequence. Because there are

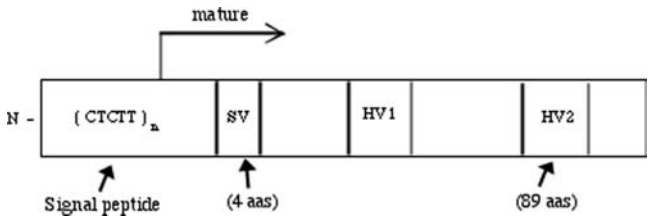

FIG. 13. Generalized structure of the opacity (Opa) proteins of Neisserial species. At the beginning of the gene encoding an Opa protein is the (CTCTT)n pentanucleotide repeat that is responsible for phase variation via slipped strand mispairing. This hydrophobic sequence is cleaved off following secretion yielding the mature protein. Then follows a semi-variable (SV) region and two highly variable regions (HV_1 and HV_2), all flanked by constant regions where homologous recombination can occur. Because there are 3 to 4 *opa* genes in *N. meningitidis* and 11 to 12 *opa* loci in *N. gonorrhoeae*, recombination between the genes can change the variable regions (see text).

multiple chromosomes per cell, recombination can occur within a single locus, involving sister chromosomes. This can yield variant proteins of altered lengths (Meyer, 1990). Finally, Neisserial species are capable of natural transformation, so they can take up single stranded DNA and incorporate it into their chromosomes, again by homologous recombination. Via this process, *opa* DNA from dead bacteria that have released their DNA can be incorporated into the live bacterium's chromosome, changing the sequence and structure of the Opa proteins. All of these mechanisms contribute to antigenic variation and host immune evasion (Hauch and Meyer, 2003).

The Pilus (pil) Locus in Neisserial Species

One of the best-studied examples of high frequency RecA-dependent and RecA-independent phenotypic switching is that of the type IV pilin (Pil) of *Neisseria gonorrhoeae* (Carbonnelle *et al.*, 2006; Hill and Grant, 2002; Mehr and Seifert, 1998). The structural subunit of the gonococcal type IV pilus is expressed from the *pilE* (E for 'expressed') locus. Transcriptionally inactive alleles of *pilE* (called *pilS*; S for 'silent') are located elsewhere on the chromosome. Recombinational deletions may occur at the 5′ end of *pilE*, giving rise to ON → OFF variation (Swanson *et al.*, 1986). In addition, nonpiliated populations of bacteria are generated which produce so-called L-pilin or S-pilin. L-pilin variants result from unequal recombination between *pilS* and *pilE*, leading to the presence of multiple tandem copies of *pilS* in the expression site. These L-pilin subunits cannot form pilus structures and accumulate in the periplasm and/or the outer membrane. Reversion to the piliated state may occur by deletion of the extra pilin gene copies (Seifert, 1996). The S-pilin state is manifested by the formation of soluble, truncated forms of pilin, which are secreted from the bacterium rather than assembled on the surface (Haas *et al.*, 1987).

In addition, a mechanism termed "cassette sorting" occurs. As noted above, there is one expression site (*pilE* with promoter) and several silent variants (*pilS* without promoters). The 5′ regions are constant and provide a region for recombination. The 3′ regions are variable and strongly antigenic. The variable regions of silent genes can be transferred to the expression site yielding active, but altered, antigens. All of these mechanisms allow escape from immune surveillance and consequently benefit the bacteria when in the animal host. They may provide other benefits to the bacterium when outside of the host, allowing variable Opa function as the structures of these proteins change.

CONCLUSIONS AND PERSPECTIVES

Seventeen years after publication of the Krawiec and Riley review on the bacterial chromosomes, with hundreds of prokaryotic genomes and dozens of eukaryotic genomes fully sequenced, we can tremendously expand our understanding of this subject and provide mechanistic details that were not available previously. Given the amount of relevant information available, such an undertaking is a daunting one; in fact, thousands of publications and several books have since appeared on this topic. In undertaking the updating of this topic, it was clear that these efforts would of necessity be selective rather than comprehensive. The reader is referred to the 1990 review of Krawiec and Riley for the early, classical literature. The books, *The Bacterial Chromosome* (Higgins, 2005) and *Dynamics of the Bacterial Chromosome* (Schumann, 2006) as well as many references cited in the text offer more in-depth discussions of several of the topics presented in this review.

Following discussion of some very general aspects of prokaryotic chromosomes, we began in earnest with a consideration of DNA compaction. This led us to discuss nucleoid proteins, supercoiling and macromolecular crowding, the three most important compaction mechanisms. Each nucleoid protein exhibits unique properties that allow it to contribute to compaction differently from the others. These proteins respond primarily to (1) sequences, but also to (2) permanent bends, (3) DNA curvature, (4) palindromes and cruciforms, (5) repeat sequences, (6) supercoiling, and (7) several other DNA structural elements, depending on the protein under consideration. Some coat the DNA while others crosslink it. Protein-protein interactions allow formation of double-stranded DNA loops. The chromosome is best characterized as a dynamic structure, always changing its configuration and protein associations. Rapid equilibria open the DNA for potential transcription, and in contrast to eukaryotes, most of the genome is subject to expression throughout vegetative growth. Even in stationary phase, the chromosome appears to be dynamic although less so.

Prokaryotes often arrange their genes in operons and clusters, where all genes required for a specific function are present at a single locus. While this confers regulatory advantages to the bacterium, it also allows lateral transfer of function between bacteria. We referred to this arrangement as "selfish" as it facilitates propagation of these genes between organisms, benefiting the genetic element, and either helping or harming the host organisms.

Repeated DNA sequences are generally present in increased or decreased frequencies, relative to their expected occurrences based on chance because of specific functional or structural considerations. Many that occur at very high or low frequencies are binding sites for proteins with catalytic or regulatory functions.

Most genes in bacteria code for stable or unstable RNAs, the latter usually coding for proteins. Consideration of codon usage provided insight into the evolutionary process whereby the triplet code may have evolved to its present level of complexity. It was suggested that some exceptions to the "universal" genetic code represent relatively recent adaptations, allowing for expansion and specificity. We could rationalize changes in G+C content and codon usage patterns based in part on the mechanisms of repair as revealed when analyzing mutator genes.

A major part of this review focused on long- and short-term rearrangements that demonstrated the dynamic nature of prokaryotic genomes. Various *E. coli* strains share only

about 40% of their genetic contents, and explanations for the differences could be provided. These include: the presence of (1) pathogenicity islands, (2) alternative alleles, (3) mobile transposons including insertion sequence (IS) elements, (4) phage genomes, (5) horizontal transfer occurring by at least four mechanisms, and (6) numerous tandemly repeated sequences that could vary with high frequency. In fact, mutation, inversion, duplication and deletion as well as incorporation of foreign DNA appear to occur so frequently that it may be a rare event that two daughter cells resulting from chromosome duplication and cell division are genetically identical. Thus, the bacterial chromosome is dynamic over evolutionary time, but also from day to day.

In contrast to eukaryotes, prokaryotes appear to reduce their genome sizes whenever functions are provided by the continual proximity of another organism or cell type. Pathogens and symbionts frequently reduce their genetic compositions when their partner organisms provide essential nutrients and homeostatic conditions. Biosynthesis, regulation, and stress-alleviating mechanisms are preferentially lost. A three-pronged explanation appears to result from energy saved due to less (1) translation, (2) transcription and (3) replication, with the importance of these three processes being in this order.

Genes in an organism can be divided into two types: housekeeping genes, required for every day cell function, and genus-specific "differentiation" genes, required only under specific conditions. The latter appear to be subject to more rapid mutational rates, causing greater rates of sequence divergence. This presumably facilitates rapid evolutionary adaptation. Different organisms appear to have accomplished this goal using different mechanisms.

With the advent of the meta genomics era, we can expect that a diversity of mechanisms will soon become recognized for maintaining and varying the genetic material of a cell, although such possibilities will require experimental verification. As it is estimated that less than 1% of prokaryotic species are currently known, and that entire kingdoms are out there, just waiting to be discovered, we can expect that in another 17 years, our knowledge of the bacterial chromosome will expand far beyond that resulting from studies conducted over the past few decades. However, this presumption depends upon humankind making the right choices. We must strive to expand our knowledge and to use the information gained for the equitable benefit of all. Science and microbes can help us solve the immense problems facing humankind, but we must be sure they are not used to our detriment. Let us hope that we can achieve this goal.

ACKNOWLEDGMENTS

I would like to thank Dave Ardell, Monica Riley, Remus Dame, and Mike Cox for critically reading and contributing ideas to this manuscript. I also thank Mary Beth Hiller, Van Duong, and Dorjee Tamang for assistance in its preparation. This work was supported by National Institutes of Health grant GM 077402.

REFERENCES

Abbani, M.A., Papagiannis, C.V., Sam, M.D., Cascio, D., Johnson, R.C., and Clubb, R.T. 2007. Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc Natl Acad Sci USA* 104:2109–2114.

Achtman, M. 1995. Epidemic spread and antigenic variability of *Neisseria meningitidis. Trends Microbiol* 3:186–192.

Adams, D.G. 2000. Heterocyst formation in cyanobacteria. *Curr Opin Microbiol* 3:618–624.

Adhya, S. 1987. The galactose operon, pp. 1503–1512. In: Neidhardt, F.C., ed. *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, Vol. 2, Washington, DC: ASM Press.

Aho, E.L., Dempsey, J.A., Hobbs, M.M., Klapper, D.G., and Cannon, J.G. 1991. Characterization of the *opa* (class 5) gene family of *Neisseria meningitidis*. *Mol Microbiol* 5:1429–1437.

Aldridge, P.D., Wu, C., Gnerer, J., Karlinsey, J.E., Hughes, K.T., and Sachs, M.S. 2006. Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proc Natl Acad Sci USA* 103:11340–11345.

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* 181:6361–6370.

Allmang, C., and Krol, A. 2006. Selenoprotein synthesis: UGA does not end the story. *Biochimie* 88:1561–1571.

Almengor, A.C., Walters, M.S., and McIver, K.S. 2006. Mga is sufficient to activate transcription in vitro of sof-sfbX and other Mga-regulated virulence genes in the group A Streptococcus. *J Bacteriol* 188:2038–2047.

Almiron, M., Link, A.J., Furlong, D., and Kolter, R. 1992. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* 6:2646–2654.

Altuvia, S., Almiron, M., Huisman, G., Kolter, R., and Storz, G. 1994. The *dps* promoter is activated by OxyR during growth and by IHF and σ^S in stationary phase. *Mol Microbiol* 13:265–272.

Ames, B.N., and Martin, R.G. 1964. Biochemical aspects of genetics: the operon. *Annu Rev Biochem* 33:235–258.

Amit, R., Oppenheim, A.B., and Stavans, J. 2003. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys J* 84:2467–2473.

Andersson, S.G.E. 2006. The bacterial world gets smaller. *Science* 314:259–260.

Ardell, D.H., and Andersson, S.G. 2006. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res* 34:893–904.

Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Riehle, M.M., Paegle, E.S., and Hatfield, G.W. 2000. Global gene expression profiling in Escherichia coli K12. The effects of integration host factor. *J Biol Chem* 275:29672-29684.

Areschoug, T., Carlsson, F., Stalhammar-Carlemalm, M., and Lindahl, G. 2004. Host-pathogen interactions in *Streptococcus pyogenes* infections, with special reference to puerperal fever and a comment on vaccine development. *Vaccine 6* (Suppl 1): S9–S14.

Bahrani-Mougeot, F.K., Buckles, E.L., Lockatell, C.V., Hebel, J.R., Johnson, D.E., Tang, C.M., and Donnenberg, M.S. 2002. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* determinants in the murine urinary tract. *Mol Microbiol* 45:1079–1093.

Bahrani-Mougeot, F., Gunther, N.W., IV, Donnenberg, M.S., and Mobley, H.L.T. 2002. Uropathogenic *Escherichia coli*. In

Donnenberg, M.S., ed. *Escherichia coli. Virulence Mechanisms of a Versatile Pathogen*, Boston, MA: Academic Press.

Balandina, A., Kamashev, D., and Rouviere-Yaniv, J. 2002. The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J Biol Chem* 277:27622–27628.

Barabote, R.D., and Saier, M.H., Jr. 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol Mol Biol Rev* 69:608–634.

Bardy, S.L., Ng, S.Y., and Jarrel, K.F. 2003. Prokaryotic motility structures. *Microbiology* 149:295–304.

Batzloff, M.R., Pandey, M., Olive, C., and Good, M.F. 2006. Advances in potential M-protein peptide-based vaccines for preventing rheumatic fever and rheumatic heart disease. *Immunol Res* 35:233–248.

Beloin, C., Deighan, P., Doyle, M., and Dorman, C.J. 2003a. *Shigella flexneri* 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol Genet Genomics* 270:66–77.

Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hegarat, F., and Le Cam, E. 2003b. Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J Biol Chem* 278:5333–5342.

Bender, J. 2004. DNA methylation and epigenetics. *Annu Rev Plant Biol* 55:41–68.

Berger, J.M. 1998. Type II DNA topoisomerases. *Curr Opin Struct Biol* 8:26–32.

Berlyn, M.K.B. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. *Microbiol. Mol. Biol. Rev*. 62:814–984.

Berthiaume, F., Crost, C., Labrie, V., Martin, C., Newman, E.B., and Harel, J. 2004. Influence of L-leucine and L-alanine on Lrp regulation of foo, coding for F1651, a Pap homologue. *J Bacteriol* 186:8537–8541.

Bhattacharya, P. 1999. Evidence for integration of IS1 and IS10 in recombinant plasmid containing rearrangement hot-spot(rhs) common-shared block(CSB) in *Escherichia coli* K-12. *Indian J Exp Biol* 37:1022–1024.

Biskri, L., Bouvier, M., Guerout, A.M., Boisnard, S., and Mazel, D. 2005. Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J Bacteriol* 187:1740–1750.

Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science* 277:1453–1474.

Bloch, V., Yang, Y., Margeat, E., Chavanieu, A., Auge, M.T., Robert, B., *et al.* 2003. The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nat Struct Biol*10:212–218.

Blot, N., Mavathur, R., Geertz, M., Travers, A., Mushkhelishvilli, G. 2006. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep* 7:710-805

Bongers, R.S., Hoefnagel, M.H., Starrenburg, M.J., Siemerink, M.A., Arends, J.G., Hugenholtz, J., and Kleerebezem, M. 2003. IS981-mediated adaptive evolution recovers lactate production by *ldhB* transcription activation in a lactate dehydrogenase-deficient strain of *Lactococcus lactis*. *J Bacteriol* 185:4499–4507.

Bonifield, H.R., and Hughes, K.T. 2003. Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *J Bacteriol* 185:3567–3574.

Bonnefoy, E., Rouvière-Yaniv, J. 1991. HU and IHF, two homologous histone-like proteins of Escherichia coli, form different protein-DNA complexes with short DNA fragments. *EMBO J* 10:687–696.

Boyle, E.C., and Finlay, B.B. 2003. Bacterial pathogenesis: exploiting cellular adherence. *Curr Opin Cell Biol* 15:633–639.

Breitbart, M., and Rohwer, F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* 13:278–284.

Browning, D.F., Cole, J.A., and Busby, S.J. 2000. Suppression of FNR-dependent transcription activation at the *Escherichia coli nir* promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleoprotein assembly. *Mol Microbiol* 37:1258–1269.

Bryan, A., Roesch, P., Davis, L., Moritz, R., Pellett, S., and Welch, R.A. 2006. Regulation of type 1 fimbriae by unlinked FimB- and FimE-like recombinases in uropathogenic *Escherichia coli* strain CFT073. *Infect Immun* 74:1072–1083.

Burk, R.F., and Hill, K.E. 1999. Orphan selenoproteins. *Bioessays* 21:231–237.

Buschiazzo, E., and Gemmell, N.J. 2006. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* 28:1040–1050.

Callahan, S.M., and Buikema, W.J. 2001. The role of HetN in maintenance of the heterocyst pattern in *Anabaena* sp. PCC 7120. *Mol Microbiol* 40:941–950.

Callen, B.P., Shearwin, K.E., and Egan, J.B. 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol Cell* 14:647–656.

Carbonnelle, E., Helaine, S., Nassif, X., and Pelicic, V. 2006. A systematic genetic analysis in *Neisseria meningitidis* defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. *Mol Microbiol* 61:1510–1522.

Carrasco, C.D., Holliday, S.D., Hansel, A., Lindblad, P., and Golden, J.W. 2005. Heterocyst- specific excision of the *Anabaena* sp. strain PCC 7120 *hupL* element requires *xisC*. *J Bacteriol* 187:6031–6038.

Case, R.B., Chang, Y.P., Smith, S.B., Gore, J., Cozzarelli, N.R., and Bustamante, C. 2004. The bacterial condensin MukBEF compacts DNA into a repetitive, stable structure. *Science* 305:222–227.

Casjens, S., Palmer, N., van Vugt, R., Huang, W.M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R.J., Haft, D., Hickey, E., Gwinn, M., White, O., and Fraser, C.M. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 35:490–516.

Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G.L., and Chiancone, E. 2004. DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Res* 32:5935–5944.

Ceci, P., Mangiarotti, L., Rivetti, C., and Chiancone, E. 2007. The neutrophil-activating Dps protein of *Helicobacter pylori*, HP-NAP, adopts a mechanism different from *Escherichia coli* Dps to bind and condense DNA. *Nucleic Acids Res* 35:2247–2256.

Champion, K., and Higgins, N.P. 2007. Growth rate toxicity phenotypes and homeostatic supercoil control differentiate *Escherichia coli* from *Salmonella enterica* serovar Typhimurium. *J Bacteriol* 189:5839–5849.

Champoux, J.J. 1994. Mechanism of catalysis by eukaryotic DNA topoisomerase I. *Adv Pharmacol* 29A:71–82.

Changela, A., Perry, K., Taneja, B., and Mondragon, A. 2003. DNA manipulators: caught in the act. *Curr Opin Struct Biol* 13:15–22.

Chater, K.F., and Chandra, G. 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev* 30:651–672.

Chen, C.W., Huang, C.H., Lee, H.H., Tsai, H.H., and Kirby, R. 2002. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet* 18:522–529.

Cheng, C., Kussie, P., Pavletich, N., and Shuman, S. 1998. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* 92:841–850.

Cheng, Y.S., Yang, W.Z., Johnson, R.C., and Yuan, H.S. 2000. Structural analysis of the transcriptional activation on Fis: crystal structures of six Fis mutants with different activational properties. *J Mol Biol* 302:1139–1151.

Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T., Weng, S., Cherry, J.M., and Botstein, D. 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282:2022–2028.

Choi, S.H., Baumler, D.J., and Kaspar, C.W. 2000. Contribution of *dps* to acid stress tolerance and oxidative stress tolerance in *Escherichia coli* O157:H7. *Appl Environ Microbiol* 66:3911–3916.

Choi, S.-K., and Saier, M.H., Jr. 2005. Regulation of *sigL* expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis*: potential connection between carbon and nitrogen metabolism. *J Bacteriol* 187:6856–6861.

Choi, S.-K., and Saier, M.H., Jr. 2006. Mechanism of CcpA-mediated glucose repression of the *resABCDE* operon of *Bacillus subtilis*. *J Mol Microbiol Biotechnol* 11:104–110.

Chowdhury, R.P., and Chatterji, D. 2007. Estimation of Forster's distance between two ends of Dps protein from mycobacteria: Distance heterogeneity as a function of oligomerization and DNA binding. *Biophys Chem* Mar 2 128:19–29.

Chung, Y.J., Krueger, C., Metzgar, D., and Saier, M.H., Jr. 2001. Size comparisons among integral membrane transport protein homologues in bacteria, archaea, and eucarya. *J Bacteriol* 183:1012–1021.

Cobucci-Ponzano, B., Rossi, M., and Moracci, M. 2005. Recoding in archaea. *Mol Microbiol* 55:339–348.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, J., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. 2001. Massive gene decay in the leprosy bacillus. *Nature* 409:1007–1011.

Connell, T.D., Shaffer, D., and Cannon, J.G. 1990. Characterization of the repertoire of hypervariable regions in the Protein II (*opa*) gene family of *Neisseria gonorrhoeae*. *Mol Microbiol* 4:439–449.

Cosby, W.M., Vollenbroich, D., Lee, O.H., and Zuber, P. 1998. Altered *srf* expression in *Bacillus subtilis* resulting from changes in culture pH is dependent on the Spo0K oligopeptide permease and the ComQX system of extracellular control. *J Bacteriol* 180:1438–1445.

Craig, N.L., and Kleckner, N. 1987. Transposition and site-specific recombination, pp. 1054–1070. In: Neidhardt, F.C., ed. *Escherichia coli* and *Salmonella typhimurium*. *Cellular and Molecular Biology*. Washington, DC: ASM Press.

Crampton, N., Bonass, W.A., Kirkham, J., Rivetti, C., and Thomson, N.H. 2006. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res* 34:5416–5425.

Cropp, T.A., and Schultz, P.G. 2004. An expanding genetic code. *Trends Genet* 20:625–630.

Cunha, S., Woldringh, C.L., and Odijk, T. 2001. Polymer-mediated compaction and internal dynamics of isolated *Escherichia coli* nucleoids. *J Struct Biol* 136:53–66.

Dame, R.T. 2005. The role of nucleoid associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol* 56:858–870.

Dame, R.T., Luijsterburg, M.S., Krin, E., Bertin, P.N., Wagner, R., and Wuite, G.J. 2005. DNA bridging: a property shared among H-NS like proteins. *J Bacteriol* 5:1845–1848.

Dame, R.T., Noom, M.C., and Wuite, G.J. 2006. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* 444:387–390.

Dame, R.T., Wyman, C., and Goosen, N. 2000. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res* 28:3504–3510.

Dame, R.T., Wyman, C., Wurm, R., Wagner, R., and Goosen, N. 2002. Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. *J Biol Chem* 277:2146–2150.

Davis, B.K. 2002. Molecular evolution before the origin of species. *Prog Biophys Mol Biol* 79:77–133.

Dehio, C., Gray-Owen, S.D., and Meyer, T.F. 1998. The role of Neisserial Opa proteins in interactions with host cells. *Trends Microbiol* 6:489–495.

Demerec, M., and Hartman, P. 1959. Complex loci in microorganisms. *Annu Rev Microbiol* 13:377–406.

Denamur, E., and Matic, I. 2006. Evolution of mutation rates in bacteria. *Mol Microbiol* 60:820–827.

Deng, W., Liou, S.R., Plunkett, G., Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V., Schwartz, D.C., and Blattner, F.R. 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* 185:2330–2337.

Déziel, E., Comeau, Y., and Villemur, R. 2001. Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *J Bacteriol* 183:1195–1204.

Dole, S., Nagarajavel, V., and Schnetz, K. 2004. The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli bgl* operon downstream of the promoter. *Mol Microbiol* 52:589–600.

Doyle, M., Fookes, M., Ivens, A., Mangan, M.W., Wain, J., and Dorman, C.J. 2007. An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315:251–252.

D'Souza, C., Nakano, M.M., Frisby, D.L., and Zuber, P. 1995. Translation of the open reading frame encoded by *comS*, a gene of the *srf* operon, is necessary for the development of genetic competence, but not surfactin biosynthesis, in *Bacillus subtilis*. *J Bacteriol* 177:4144–4148.

Edwards, R.A., and Rohwer, F. 2005. Viral metagenomics. *Nat Rev Microbiol* 3:504–510.

Eggleston, A.K., and West, S.C. 1997. Recombination initiation: easy as A, B, C, D... chi? *Curr Biol* 7:R745–R749.

Elkins, C., Thomas, C.E., Seifert, H.S, and Sparling, P.F. 1991. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. *J Bacteriol* 173:3911–3913.

Espeli, O., Moulin, L., and Boccard, F. 2001. Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J Mol Biol* 314:375–386.

Esposito, D., Petrovic, A., Harris, R., Ono, S., Eccleston, J.F., Mbabaali, A., *et al.* (2002). H-NS oligomerization domain structure reveals the mechanism for high order self association of the intact protein. *J Mol Biol* 324:841–850.

Feng, L., Sheppard, K., Namgoong, S., Ambrogelly, A., Polycarpo, C., Randau, L., Tumbula-Hansen, D., and Soll, D. 2004. Aminoacyl-tRNA synthesis by pre-translational amino acid modification. *RNA Biol* 1:16–20.

Ferguson, G.P., Creighton, R.I., Nikolaev, Y., and Booth, I.R. 1998. Importance of RpoS and Dps in survival of exposure of both exponential- and stationary-phase *Escherichia coli* cells to the electrophile N-ethylmaleimide. *J Bacteriol* 180:1030–1036.

Feulner, G., Gray, J.A., Kirschman, J.A., Lehner, A.F., Sadosky, A.B., Vlazny, D.A., Zhang, J., Zhao, S., and Hill, C.W. 1990. Structure of the *rhsA* locus from *Escherichia coli* K-12 and comparison of *rhsA* with other members of the *rhs* multigene family. *J Bacteriol* 172:446–456.

Fischetti, V.A., Jones, K.F., Hollingshead, S.K., and Scott, J.R. 1988. Structure, function, and genetics of streptococcal M protein. *Rev Infect Dis* 10 (Suppl 2):S356–S359.

Fishman, S.E., and Hershberger, C.L. 1983. Amplified DNA in *Streptomyces fradiae*. *J Bacteriol* 155:459–466.

Fishman, S.E., Rosteck, P.R., Jr., and Hershberger, C.L. 1985. A 2.2-kilobase repeated DNA segment is associated with DNA amplification in *Streptomyces fradiae*. *J Bacteriol* 161:199–206.

Fitzmaurice, W.P., Benjamin, R.C., Huang, P.C., and Scocca, J.J. 1984. Characterization of recognition sites on bacteriophage HP1c1 DNA which interact with the DNA uptake system of *Haemophilus influenzae Rd. Gene* 31:187–196.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. 1995. Whole-genome random sequence of *Haemophilus influenzae* Rd. *Science* 269:496–512.

Forterre, P., Bergerat, A., and Lopez-Garcia, P. 1996. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol Rev* 18:237–248.

Forterre, P., Bouthier De La Tour, C., Philippe, H., and Duguet, M. (2000). Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* 16:152–154.

Fox, K.R., and Brown, T. 2005. An extra dimension in nucleic acid sequence recognition. *Q Rev Biophys* 38:311–320.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson, C.A., III, and Venter, J.C. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O., Venter, J.C. 1997. Genomic sequence of a Lyme disease spirochete, *Borellia burgdorferi*. *Nature* 390:580–586.

Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G., Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A., Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M., Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M.D., Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Sandusky, M., Weidman, J., Smith, H.O., and Venter, J.C. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281:375–388.

Frenkiel-Krispin, D., Levin-Zaidman, S., Shimoni, E., Wolf, S.G., Wachtel, E.J., Arad, T., *et al.* 2001. Regulated phase transitions of bacterial chromatin: a non-enzymatic pathway for generic DNA protection. *EMBO J* 20:1184–1191.

Gerstel, U., Park, C., and Romling, U. 2003. Complex regulation of *csgD* promoter activity by global regulatory proteins. *Mol Microbiol* 49:639–654.

Gilad, Y., Oshlack, A., and Rifkin, S.A. 2006. Natural selection on gene expression. *Trends Genet* 22:456–461.

Gilson, E., Clément, J.M., Perrin, D., and Hofnung, M. 1987. Palindromic units: a case of highly repetitive DNA sequences in bacteria. *Trends Genetics* 3:226–230.

Glasgow, A.C., Hughes, K.T., Simon, M.I., Berg, D.E., and Howe, M.M. 1989. Bacterial DNA inversion systems, pp. 637–660. In: Berg, D.E., and Howe, M.M., eds., *Mobile DNA*, Washington, DC: ASM Press.

Godovac-Zimmerman, J., Kleiner, O., Brown, L.R., and Drukier, A.K. 2005. Perspectives in splicing up proteomics with splicing. *Proteomics* 5:699–709.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S.G. 1996. Life with 6000 genes. *Science* 274: 546: 563–567.

Golden, J.W., Whorff, L.L., and Wiest, D.R. 1991. Independent regulation of *nifHDK* operon transcription and DNA rearrangement during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* 173:7098–7105.

Golden, J.W., and Yoon, H.S. 2003. Heterocyst development in *Anabaena Curr Opin Microbiol* 6:557–563.

Goll, M.G., and Bestor, T.H. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481–514.

Goodman, S.D., and Scocca, J.J. 1988. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* 85:6982–6986.

Grainger, D.C., Aiba, H., Hurd, D., Browning, D.F., and Busby, S.J. 2007. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucleic Acids Res* 35:269–278.

Grant, R.A., Filman, D.J., Finkel, S.E., Kolter, R., and Hogle, J.M. 1998. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 5:294–303.

Gratia, J.-P. 2005. Noncomplementing diploidy resulting from spontaneous zygogenesis in *Eschericia coli*. *Microbiology* 151:2947–2959.

Graumann, P.L. 2001. SMC proteins in bacteria: condensation motors for chromosome segregation? *Biochimie* 83:53–59.

Gray-Owen, S.D. 2003. Neisserial Opa proteins: impact on colonization, dissemination and immunity. *Scand J Infect Dis* 35:614–618.

Griffith, J.D. 1976. Visualization of prokaryotic DNA in a regularly condensed chromatin-like fiber. *Proc Natl Acad Sci USA* 73:563–567.

Gromer, S., Eubel, J.K., Lee, B.L., and Jacob, J. 2005. Human selenoproteins at a glance. *Cell Mol Life Sci* 62:2414–2437.

Gualillo, O., Lago, F., Casanueva, F.F., and Dieguez, C. 2006. One ancestor, several peptides post-translational modifications of preproghrelin generate several peptides with antithetical effects. *Mol Cell Endocrinol* 256:1–8.

Gupta, S. 2005. Parasite immune escape: new views into host-parasite interactions. *Curr. Opin. Microbiol*. 8:428–433.

Haagmans, W., and van der Woude, M. 2000. Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Mol Microbiol* 35:877–887.

Haas, R., Schwarz, H., and Meyer, T.F. 1987. Release of soluble pilin antigen coupled with gene conversion in *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* 84:9079–9083.

Hall, B.G. 1998. Activation of the *bgl* operon by adaptive mutation. *Mol Biol Evol* 15:1–5.

Hastings, P.J., and Rosenberg, S.M. 2002. In pursuit of a molecular mechanism for adaptive gene amplification. *DNA Repair (Amst)* 1:111–123.

Hatfield, G.W., and Benham, C.J. 2002. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu Rev Genet* 36:175–203.

Hauck, C.R., and Meyer, T.F. 2003. 'Small' talk: Opa proteins as mediators of Neisseria-host-cell communication. *Curr OpinMicrobiol* 6:43–49.

Helene, C., Giovannangeli, C., Guieysse-Peugeot, A.L., and Praseuth, D. 1997. Sequence-specific control of gene expression by antigene and clamp oligonucleotides. *Ciba Found Symp* 209:94–102.

Henderson, I.R., Owen, P., and Nataro, J.P. 1999. Molecular switches—the ON and OFF of bacterial phase variation. *Mol Microbiol* 33:919–932.

Hengen, P.N., Bartram, S.L., Stewart, L.E., and Schneider, T.D. 1997. Information analysis of Fis binding sites. *Nucleic Acids Res* 25:4994–5002.

Hermann, A., Goyal, R., and Jeltsch, A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279:48350–48359.

Hernday, A.D., Braaten, B.A., Broitman-Maduro, G., Engelberts, P., and Low, D.A. 2004a. Regulation of the *pap* epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. *Mol Cell* 16:537–547.

Hernday, A., Braaten, B., and Low, D. 2004b. The intricate workings of a bacterial epigenetic switch. *Adv Exp Med Biol* 547:83–89.

Hersh, M.N., Ponder, R.G., Hastings, P.J., and Rosenberg, S.M. 2004. Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. *Res Microbiol* 155:352–359.

Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G., and Bremer, E. 1988a. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* 52:569–584.

Higgins, C.F., McLaren, R.S., and Newbury, S.F. 1988b. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene* 72:3–14.

Higgins, N.P. ed. 2005. *The Bacterial Chromosome*. Washington, DC: ASM Press.

Higgins, N.P., Yang, X., Fu, Q., and Roth, J.R. 1996. Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J Bacteriol* 178:2825–2835.

Hill, C.W. 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. *Res Microbiol* 150:665–674.

Hill, C.W., Sandt, C.H., and Vlazny, D.A. 1994. Rhs elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol Microbiol* 12:865–871.

Hill, S.A. 2000. Opa expression correlates with elevated transformation rates in *Neisseria gonorrhoeae*. *J Bacteriol* 182:171–178.

Hill, S.A., and Grant, C.C. 2002. Recombinational error and deletion formation in *Neisseria gonorrhoeae*: a role for RecJ in the production of *pilE* (L) deletions. *Mol Genet Genomics* 266:962–972.

Hirano, M., and Hirano, T. 2006. Opening closed arms: long distance activation of SMC ATPase by hinge-DNA interactions. *Mol Cell* 21:175–186.

Hirano, T. 2006. At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol* 7:311–322.

Holden, N.J., Totsika, M., Mahler, E., Roe, A.J., Catherwood, K., Lindner, K., Dobrindt, U., and Gally, D.L. 2006. Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic *Escherichia coli*. *Microbiology* 152:1143–1153.

Holden, N.J., Uhlin, B.E., and Gally, D.L. 2001. PapB paralogues and their effect on the phase variation of type 1 fimbriae in *Escherichia coli*. *Mol Microbiol* 42:319–330.

Hollingshead, S.K., Fischetti, V.A., and Scott, J.R. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* 207:196–203.

Hornemann, U., Otto, C.J., and Zhang, X.Y. 1989. DNA amplification in *Streptomyces achromogenes* subsp. *rubradiris* is accompanied by

a deletion, and the amplified sequences are conditionally stable and can be eliminated by two pathways. *J Bacteriol* 171:5817–5822.

Huack, C.R., and Meyer, T.F. 2003. "Small" talk: Opa proteins as mediators of *Neisseria*-host-cell communication. *Curr Opin Microbiol* 6:43–59.

Huang, CE., Milutinovich, M., and Koshland, D. 2005. Rings, bracelet or snaps: fashionable alternatives for Smc complexes. *Philo Trans R Soc Lond B Biol Sci* 360:537–542.

Huang, L., Guan, R.J., and Pardee, A.B. 1999. Evolution of transcriptional control from prokaryotic beginnings to eukaryotic complexities. *Crit Rev Eukaryot Gene Expr* 9:175–182.

Huang, S., Kang, J., and Blaser, M.J. 2006. Antimutator role of the DNA glycosylase *mutY* gene in *Helicobacter pylori*. *J Bacteriol* 188:6224–6234.

Hugouvieux-Cotte-Pattat, N., Reverchon, S., and Robert-Baudouy, J. 1989. Expanded linkage map of *Erwinia chrysanthemi* strain 3937. *Mol Microbiol* 3:573–581.

Hulton, C.S., Higgins, C.F., and Sharp, P.M. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* 5:825–834.

Humeny, A., Beck, C., Becker, C.M., and Jeltsch, A. 2003. Detection and analysis of enzymatic DNA methylation of oligonucleotide substrates by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Anal Biochem* 313:160–166.

Humphrey-Smith, I., and Blackstock, W. 1997. Proteome analysis: genomics via the output rather than the input code. *J Protein Chem* 16:537–544.

Ibba, M., and Soll, D. 2000. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* 69:617–650.

Inokuchi, Y., Hirashima, A., Sekine, Y., Janosi, L., and Kaji, A. 2000. Role of ribosome recycling factor (RRF) in translational coupling. *EMBO J* 19:3788–3798.

Iqbal, S., Parker, G., Davidson, H., Moslehi-Rahmani, E., and Robson, R.L. 2004. Reversible phase variation in the *phnE* gene, which is required for phosphonate metabolism in *Escherichia coli* K-12. *J Bacteriol* 186:6118–6123.

Ishihama, A. 1999. Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. *Genes Cells* 4:135–143.

Ives, C.L., and Bott, K.F. 1990. Characterization of chromosomal DNA amplifications with associated tetracycline resistance in *Bacillus subtilis*. *J Bacteriol* 172:4936–4944.

Jacob, F., and Monod, J. 1962. On the regulation of gene activity. *Cold Spring Harbor Symp Quant Biol* 26:193–211.

Jeong, K.C., Baumler, D.J., and Kaspar, C.W. 2006. *dps* expression in *Escherichia coli* O157:H7 requires an extended -10 region and is affected by the cAMP receptor protein. *Biochim Biophys Acta* 1759:51–59.

Jeong, M.A., and Lim, D. 2004. A proteomic approach to study msDNA function in *Escherichia*. *J Microbiol* 42:200–204.

Jin, D.J., and Cabrera, J.E. 2006. Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli*. *J Struct Biol* 156:284–291.

Johnson, R.C., Johnson, L.M., Schmidt, J.W., and Gardner, J.F. 2005. Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome, pp. 65–132. In: Higgins, N.P., ed. *The Bacterial Chromosome*, Washington, DC: ASM Press.

Jones, K.M., Buikema, W.J., and Haselkorn, R. 2003. Heterocyst-specific expression of *patB*, a gene required for nitrogen fixation in *Anabaena* sp. strain PCC 7120. *J Bacteriol* 185:2306–2314.

Jukes, T.H., and Osawa, S. 1993. Evolutionary changes in the genetic code. *Comp Biochem Physiol B* 106:489–494.

Kaberdin, V.R., and Blasi, U. 2006. Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol Rev* 30:967–979.

Kamashev, D., Balandina, A., and Rouviere-Yaniv, J. 1999. The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J* 18:5434–5444.

Kamtekar, S., Ho, R.S., Cocco, M.J., Li, W., Wenwieser, S.V., Boocock, M.R., Grindley, N.D., and Steitz, T.A. 2006. Implications of structures of synaptic tetramers of $\gamma\delta$ resolvase for the mechanism of recombination. *Proc Natl Acad Sci USA* 103:10642–10647.

Kariyawasam, S., Johnson, T.J., and Nolan, L.K. 2006. The *pap* operon of avian pathogenic *Escherichia coli* strain O1:K1 is located on a novel pathogenicity island. *Infect Immun* 74:744–749.

Karlin, S., Mrazek, J., and Campbell, A.M. 1997. Compositional biases of bacterial genomes and evolutionary implications. *J Bacteriol* 179:3899–3913.

Kashi, Y., and King, D.G. 2006. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet* 22:253–259.

Kavenoff, R., and Bowen, B.C. 1976. Electron microscopy of membrane-free folded chromosomes from Escherichia coli. *Chromosoma* 59:89–101.

Kelly, A., Goldberg, M.D., Carroll, R.K., Danino, V., Hinton, J.C., and Dorman, C.J. 2004. A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica serovar* typhimurium. *Microbiology* 150:2037–2053.

Khemici, V., and Carpousis, A.J. 2004. The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol* 51:777–790.

Khodursky, A.B. 2007. Evolution, adaptation, and supercoiling. *J Bacteriol* 189:5789–5791.

Khopde, S., Biswas, E.E., and Biswas, S.B. 2002. Affinity and sequence specificity of DNA binding and site selection for primer synthesis by *Escherichia coli* primase. *Biochemistry* 41:14820–14830.

Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49–63.

Kim, S.H., Chao, Y., and Saier, M.H., Jr. 2006. Protein-translocating trimeric autotransporters of Gram-negative bacteria. *J Bacteriol* 188:5655–5667.

Koonin, E.V. 2007. Metagenomic sorcery and the expanding protein universe. *Nat Biotechnol* 25:540–542.

Krawiec, S., and Riley, M. 1990. Organization of the bacterial chromosome. *Microbiol Rev* 54:502–539.

Krol, A. 2002. Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis. *Biochimie* 84:765–774.

Kroll, J.S., Wilks, K.E., Farrant, J.L., and Langford, P.R. 1998. Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens. *Proc Natl Acad Sci USA* 95:12381–12385.

Kryukov, G.V., and Gladyshev, V.N. 2004. The prokaryotic selenoproteome. *EMBO Rep* 5:538–543.

Krzycki, J.A. 2004. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr Opin Chem Biol* 8:484–491.

Krzycki, J.A. 2005. The direct genetic encoding of pyrrolysine. *Curr Opin Microbiol* 8:706–712.

Kudva , I.T., Evans, P.S., Perna, N.T., Barrett, T.J., Ausubel, F.M., Blattner, F.R., and Calderwood, S.B. 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J Bacteriol* 184:1873–1879.

Kulasekara, H.D., and Blomfield, I.C. 1999. The molecular basis for the specificity of *fimE* in the phase variation of type 1 fimbriae of *Escherichia coli* K-12. *Mol Microbiol* 31:1171–1181.

Kunkel, B., Losick, R., and Stragier, P. 1990. The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes Dev* 4:525–535.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S. (and 132 more) *et al.* 1997 The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256.

Kuzminov, A. 1995. Collapse and repair of replication forks in *Escherichia coli*. *Mol Microbiol* 16:373–384.

Kuzminov, A. 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* 63:751–813.

Lahooti, M., Roesch, P.L., and Blomfield, I.C. 2005. Modulation of the sensitivity of FimB recombination to branched-chain amino acids and alanine in *Escherichia coli* K-12. *J Bacteriol* 187:6273–6280.

Lampson, B., Inouye, M., and Inouye, S. 2001. The msDNAs of bacteria. *Prog. Nucleic Acid Res Mol Biol* 67:65–91.

Lampson, B.C., Inouye, M., and Inouye, S. 2005. Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res* 110:491–499.

Lang, A.S., and Beatty, J.T. 2000. Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* 97:859–864.

Lang, A.S., and Beatty, J.T. 2001. The gene transfer agent of *Rhodobacter capsulatus* and "constitutive transduction" in prokaryotes. *Arch Microbiol* 175:241–249.

Lang, A.S., Taylor, T.A., and Beatty, J.T. 2002. Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. *J Mol Evol* 55:534–543.

Lawrence, J.G. 2002. Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* 110:407–413.

Lawrence, J.G. 2005. The dynamic bacterial genome, pp. 19–37. In: Higgins, N.P. ed. *The Bacterial Chromosome*, Washington, DC: ASM Press.

Lawrence, J.G., and Roth, J.R. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–1860.

Lawrence, J.G., and Roth, J.R. 1999. Genomic flux: genome evolution by gene loss and acquisition, p. 263–289. In Charlebois, R.L. ed. *Organization of the Prokaryotic Genome*.Washington, DC: ASM Press.

Leathart, J.B., and Gally, D.L. 1998. Regulation of type 1 fimbrial expression in uropathogenic *Escherichia coli*: heterogeneity of expression through sequence changes in the *fim* switch region. *Mol Microbiol* 28:371–381.

Leblond, P., Demuyter, P., Moutier, L., Laakel, M., Decaris, B., and Simonet, J.M. 1989. Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in *Streptomyces ambofaciens*. *J Bacteriol* 171:419–423.

Lee, S.Y., Lee, H.J., Lee, H., Kim, S., Cho, E.H., and Lim, H.M. 1998. In vivo assay of protein-protein interactions in Hin-mediated DNA inversion. *J Bacteriol* 180:5954–5960.

Leo, E., Gould, K.A., Pan, X.S., Capranico, G., Sanderson, M.R., Palumbo, M., and Fisher, L.M. 2005. Novel symmetric and asymmetric DNA scission determinants for *Streptococcus pneumoniae* topoisomerase IV and gyrase are clustered at the DNA breakage site. *J Biol Chem* 280:14252–14263.

Lescure, A., Fagegaltier, D., Carbon, P., and Krol, A. 2002. Protein factors mediating selenoprotein synthesis. *Curr Protein Pept Sci* 3:143–151.

Lewis, K. 2000. Programmed death in bacteria. *Microbiol Mol Biol Rev* 64:503–514.

Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D., and Steitz, T.A. 2005. Structure of a synaptic $\gamma\delta$ resolvase tetramer covalently linked to two cleaved DNAs. *Science* 309:1210–1215.

Liu, B., and Alberts, B.M. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 267:1131–1137.

Lobanov, A.V., Kryukov, G.V., Hatfield, D.L., and Gladyshev, V.N. 2006. Is there a twenty third amino acid in the genetic code? *Trends Genet* 22:357–360.

Lobner-Olesen, A., Skovgaard, O., and Marinus, M.G. 2005. Dam methylation: coordinating cellular processes. *Curr Opin Microbiol* 8:154–160.

Loden, M., and van Steensel, B. 2005. Whole-genome views of chromatin structure. *Chromosome Res* 13:289–298.

Logsdon, J.M., and Faguy, D.M. 1999. *Thermotoga* heats up lateral gene transfer. *Curr Biol* 9:R747–R751.

Longstaff, D.G., Larue, R.C., Faust, J.E., Mahapatra, A., Zhang, L., Green-Church, K.B., and Krzycki, J.A. 2007. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci USA* 104:1021–1026.

Lu, A.L., Bai, H., Shi, G., and Chang, D.Y. 2006. MutY and MutY homologs (MYH) in genome maintenance. *Front Biosci* 11:3062–3080.

Lu, A.L., Li, X., Gu, Y., Wright, P.M., and Chang, D.Y. 2001. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys* 35:141–170.

Mahillion, J., Chandler, M. 1998. Insertion sequences. *Micro Mol Biol Rev* 62:725–774.

Mahillon, J., Leonard, C., and Chandler, M. 1999. IS elements as constituents of bacterial genomes. *Res Microbiol* 150:675–687.

Majdalani, N., Vanderpool, C.K., and Gottesman, S. 2005. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* 40:93–113.

Malik, M., Bensaid, A., Rouviere-Yaniv, J., and Drlica, K. 1996. Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations. *J Mol Biol*. 256:66–76.

Manson, J.M., and Gilmore, M.S. 2006. Pathogenicity island integrase cross-talk: a potential new tool for virulence modulation. *Mol Microbiol* 61:555–559.

Martinez, A., and Kolter, R. 1997. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol* 179:5188–5194.

Matoba, K., Yamazoe, M., Mayanagi, K., Morikawa, K., and Hiraga, S. 2005. Comparison of MukB homodimer versus MukBEF

complex molecular architectures by electron microscopy reveals a higher-order multimerization. *Biochem Biophys Res Commun* 333:694–702.

Matson, E.G., Thompson, M.G., Humphrey, S.B., Zuerner, R.L., and Stanton, T.B. 2005. Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J Bacteriol* 187:5885–5892.

Mau, B., Glasner, J.D., Darling, A.E., and Perna, N.T. 2006. Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. *Genome Biol* 7:R44.

Mazel, D. 2006. Integrons: agents of bacterial evolution. *Nat Rev Microbiol* 4:608–620.

McNulty, C., Thompson, J., Barrett, B., Lord, L., Andersen, C., and Roberts, I.S. 2006. The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. *Mol Microbiol* 59:907–922.

Mehr, I.J., and Seifert, H.S. 1998. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. *Mol Microbiol* 30:697–710.

Merickel, S.K., Haykinson, M.J., and Johnson, R.C. 1998. Communication between Hin recombinase and Fis regulatory subunits during coordinate activation of Hin-catalyzed site-specific DNA inversion. *Genes Dev* 12:2803–2816.

Merz, A.J., and So, M. 2000. Interactions of pathogenic *neisseriae* with epithelial cell membranes. *Annu Rev Cell Dev Biol* 16:423–457.

Metzgar, D., Liu, L., Hansen, C., Dybvig, K., and C. Wills, 2002. Domain-level differences in microsatellite distribution and content result from different relative rates of insertion and deletion mutations. *Genome Res* 12**:**408–413.

Meyer, T.F. 1990. Pathogenic neisseriae- a model of bacterial virulence and genetic flexibilty. *Zentralbl Bakteriol* 274:135–154.

Michaels, M.L., and Miller, J.H. 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J Bacteriol* 174:6321–6325.

Milkman, R., Jaeger, E., and McBride, R.D. 2003. Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. *Genetics* 163:475–483.

Minsky, A and Kolter, R. 2005. Stationary-phase chromosomes, pp. 155–166. In In Higgins, N.P., ed. *The Bacterial Chromosome*. Washington, DC: ASM Press.

Mizuno, M., Masuda, S., Takemaru, K., Hosono, S., Sato, T., Takeuchi, M., and Kobayashi, Y. 1996. Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the *Bacillus subtilis* genome containing the skin element and many sporulation genes. *Microbiology* 142:3103–3111.

Moore, T., Sharples, G.J., and Lloyd, R.G. 2004. DNA binding by the meningococcal RdgC protein, associated with pilin antigenic variation. *J Bacteriol* 186:870–874.

Moran, N.A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583–586.

Morgan, G.J., Hatfull, G.F., Casjens, S., and Hendrix, R.W. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Hae Neisseria and Deinococcus. *J Mol Biol* 317:337–359.

Mumm, J.P., Landy, A., and Gelles, J. 2006. Viewing single lambda site-specific recombination events from start to finish. *EMBO J* 25:4586–4595.

Murphy, L.D., and Zimmerman, S.B. 1997a. Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for in vivo nucleoid structure. *J Struct Biol* 119:336–346.

Murphy, L.D., and Zimmerman, S.B. 1997b. Isolation and characterization of spermidine nucleoids from *Escherichia coli*. *J Struct Biol* 119:321–335.

Murphy, L.D., and Zimmerman, S.B. 2001. A limited loww of DNA compaction accompanying the release of cytoplasm from cells of *Escherichia coli*. *J Struct Biol* 133:75–86.

Myers, R.S., and Stahl, F.W. 1994. Chi and the RecBC D enzyme of *Escherichia coli*. *Annu Rev Genet* 28:49–70.

Nagarajavel, V., Madhusudan, S., Dole, S., Rahmouni, A.R., Schnetz, K. 2007. Repression by binding of H-NS within the transcription unit. *J Biol Chem* 282:23622–23630.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., Ishikawa, H., Kudo, T., and Fukatsu, T. 2005. Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc Natl Acad Sci USA* 102:5477–5482.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267.

Napoli, A., Valenti, A., Salerno, V., Nadal, M., Garnier, F., Rossi, M., and Ciaramella, M. 2005. Functional interaction of reverse gyrase with single-strand binding protein of the archaeon Sulfolobus. *Nucleic Acids Res* 33:564–576.

Neidle, S. 1997. Recent developments in triple-helix regulation of gene expression. *Anticancer Drug Des* 12:433–442.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J. Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L. Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D. Pratt, M.S., Phillips, C.A., Richardson, D. Heidelberg, J., Sutton, G.G., Fleischmann, R.D. Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., and C.M. Fraser. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–329.

Nguyen, L., Paulsen, I.T., Tchieu, J., Hueck, C.J., and M.H. Saier, Jr. 2000. Phylogenetic analyses of the constituents of type III protein secretion systems. *J Mol Microbiol Biotechnol* 2:125–144.

Nicholson, B., and Low, D. 2000. DNA methylation-dependent regulation of Pef expression in *Salmonella typhimurium*. *Mol Microbiol* 35:728–742.

Nicoloff, H., Perreten, V., McMurry, L.M., and Levy, S.B. 2006. Role for tandem duplication and *lon* protease in AcrAB-TolC-dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in *marRAB* or *acrRAB*. *J Bacteriol* 188:4413–4423.

Nohmi, T., Kim, S.R., and Yamada, M. 2005. Modulation of oxidative mutagenesis and carcinogenesis by polymorphic forms of human DNA repair enzymes. *Mutat Res* 591:60–73.

Noor, M.A., and Chang, A.S. 2006. Evolutionary genetics: jumping into a new species. *Curr Biol* 16:R890–R892.

Norris, V. 1995. Hypothesis: chromosome separation in *Escherichia coli* involves autocatalytic gene expression, transertion and membrane-domain formation. *Mol Microbiol* 16:1051–1057.

Norris, V., den Blaauwen, T., Cabin-Flaman, A., Doi, R.H., Harshey, R., Janniere, L., Jimenez-Sanchez, A., Jin, D.J., Levin, P.A., Mileykovskaya, E., Minsky, A., Saier, M., Jr., and Skarstad, K. 2007. Functional taxonomy of bacterial hyperstructures. *Microbiol Molec Biol Rev* 71:230–253.

Norris, V., and Madsen, M.S. 1995. Autocatalytic gene expression occurs via transertion and membrane domain formation and underlies differentiation in bacteria: a model. *J Mol Biol* 253:739–748.

Ochman, H., and Davalos, L.M. 2006. The nature and dynamics of bacterial genomes. *Science* 311:1730–1733.

Odijk, T. 1998. Osmotic compaction of supercoiled DNA into a bacterial nucleoid. *Biophys Chem* 73:23–29.

O'Gara, J.P., and Dorman, C.J. 2000. Effects of local transcription and H-NS on inversion of the *fim* switch of *Escherichia coli*. *Mol Microbiol* 36:457–466.

Olavarrieta, L., Hernandez, P., Krimer, D.B., and Schvartzman, J.B. 2002. DNA knotting caused by head-on collision of transcription and replication. *J Mol Biol* 322:1–6.

Olsen, P.B., Schembri, M.A., Gally, D.L., and Klemm, P. 1998. Differential temperature modulation by H-NS of the *fimB* and *fimE* recombinase genes which control the orientation of the type 1 fimbrial phase switch. *FEMS Microbiol Lett* 162:17–23.

O'Neil, J.P., Rogan, P.K., Cariello, N., Nicklas, J.A. 1998. Mutations that alter RNA splicing of the human HPRT gene: a review of the spectrum. *Mutat Res* 411:179–214.

Oussatcheva, E.A., Pavlicek, J., Sankey, O.F., Sinden, R.R., Lyubchenko, Y.L., and Potaman, V.N. 2004. Influence of global DNA topology on cruciform formation in supercoiled DNA. *J Mol Biol* 338:735–743.

Palmer, B.R., and Marinus, M.G. 1994. The *dam* and *dcm* strains of *Escherichia coli*–a review. *Gene* 143:1–12.

Palmer, N., Fraser, C., and Casjens, S. 2000. Distribution of twelve linear extrachromosomal DNAs in natural isolates of Lyme disease spirochetes. *J Bacteriol* 182:2476–2480.

Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S., and Johnson, R.C. 1996. Variable structures of Fis–DNA complexes determined by flanking DNA—protein contacts. *J. Mol. Biol*. 264:675–695.

Paradkar, A., Trefzer, A., Chakraburtty, R., and Stassi, D. 2003. *Streptomyces* genetics: a genomic perspective. *Crit Rev Biotechnol* 23:1–27.

Park, M., Yun, S.T., Hwang, S.Y., Chun, C.I., and Ahn, T.I. 2006. The *dps* gene of symbiotic "*Candidatus* Legionella*jeonii*" in *Amoeba proteus* responds to hydrogen peroxide and phagocytosis. *J Bacteriol* 188:7572–7580.

Park, S., You, X., and Imlay, J.A. 2005. Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx-mutants of *Escherichia coli*. *Proc Natl Acad Sci USA* 102:9317–9322.

Perez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., and Latorre, A. 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* 314:312–313.

Perna, N.T., Plunkett, G. 3rd, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner F.R. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533. [Erratum in: Nature 2001 410: 240.]

Peterson, S.N., and Reich, N.O. 2006. GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence *pap* expression. *J Mol Biol* 355:459–472.

Petrushenko, Z.M., Lai, C.H., and Rybenkov, V.V. 2006a. Antagonistic interactions of kleisins and DNA with bacterial Condensin MukB. *J. Biol. Chem*. 281:34208–34217.

Petrushenko, Z.M., Lai, C.H., Rai, R., and Rybenkov, V.V. 2006b. DNA reshaping by MukB. Right-handed knotting, left-handed supercoiling. *J Biol Chem* 281:4606–4615.

Pham, P.T., Zhao, W., and Schaaper, R.M. 2006. Mutator mutants of *Escherichia coli* carrying a defect in the DNA polymerase III τ subunit. *Mol Microbiol* 59:1149–1161.

Plasterk, R.H., Simon, M.I., and Barbour, A.G. 1985. Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. *Nature* 318:257–263.

Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R. 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* 18:1766–1779.

Prosseda, G., Falconi, M., Giangrossi, M., Gualerzi, C.O., Micheli, G., and Colonna, B. 2004. The *virF* promoter in *Shigella*: more than just a curved DNA stretch. *Mol Microbiol* 51:523–537.

Qiu, D., Fujita, K., Sakuma, Y., Tanaka, T., Ohashi, Y., Ohshima, H., Tomita, M., and Itaya, M. 2004. Comparative analysis of physical maps of four *Bacillus subtilis* (*natto*) genomes. *Appl Environ Microbiol* 70:6247–6256.

Radman-Livaja, M., Biswas, T., Ellenberger, T., Landy, A., and Aihara, H. 2006. DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Curr Opin Struct Biol* 16:42–50.

Ramaswamy, K.S., Carrasco, C.D., Fatma, T., and Golden, J.W. 1997. Cell-type specificity of the *Anabaena fdxN*-element rearrangement requires *xisH* and *xisI*. *Mol Microbiol* 23:1241–1249.

Ramseier, T.M., and Saier, M.H., Jr. 1995. Cyclic AMP:cAMP receptor protein (CRP) complex: five binding sites in the control region of the *Escherichia coli* mannitol (*mtl*) operon. *Microbiology* 141:1901–1907.

Ramstein, J., Hervouet, N., Coste, F., Oberto, J., Castaing, B. 2003. Evidence of a thermal unfolding dimeric intermediate for the *Escherichia coli* histone-like HU proteins: thermodynamics and structure. *J Mol Biol* 331:102–121.

Reindel, S., Schmidt, C.L., Anemuller, S., and Matzanke, B.F. 2005. Expression and regulation pattern of ferritin-like DpsA in the archaeon *Halobacterium salinarum*. *Biometals* 18:387–397.

Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–1306.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., Mori, H., Perna, N.T., Plunkett, G., Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* 34:1–9.

Rimsky, S. 2004. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr Opin Microbiol* 7:109–114.

Rimsky, S., Zuber, F., Buckle, M., and Buc, H. 2001. A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol* 42:1311–1323.

Ritter, A., Gally, D.L., Olsen, P.B., Dobrindt, U., Friedrich, A., Klemm, P., and Hacker, J. 1997. The Pai-associated *leuX* specific tRNA5(Leu) affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. *Mol Microbiol* 25:871–882.

Rivetti, C., Walker, C., and Bustamante, C. 1998. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J Mol Biol* 280:41–59.

Roesch, P.L., and Blomfield, I.C. 1998. Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the *fim* switch to stimulate site-specific recombination in *Escherichia coli*. *Mol Microbiol* 27:751–761.

Rohwer, F. 2003. Global phage diversity. *Cell* 113:141.

Roth, J.R. 2005. Where's the beef? Looking for information in bacterial chromosomes, pp. 3–18. In: Higgins, N.P., ed. *The Bacterial Chromosome*. Washington, DC: ASP Press.

Roth, J.R., and Andersson, D.I. 2004. Amplification-mutagenesis–how growth under selection contributes to the origin of genetic diversity and explains the phenomenon of adaptive mutation. *Res Microbiol* 155:342–351.

Rowland, S.J., Boocock, M.R., and Stark, W.M. 2006. DNA bending in the Sin recombination synapse: functional replacement of HU by IHF. *Mol Microbiol* 59:1730–1743.

Rudd, K.E. 2000. EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res* 28:60–64.

Saier, M.H., Jr. 1998. Genome sequencing and informatics: new tools for biochemical discoveries. *Plant Physiol* 117:1129–1133.

Saier, M.H., Jr. 2003. Answering fundamental questions in biology with bioinformatics. *ASM News* 69:175–181.

Saier, M.H. Jr., and Jacobson, G.R. 1984. *The Molecular Basis of Sex and Differentiation: A Comparative Study of Evolution, Mechanism and Control in Microorganisms*. New York: Springer-Verlag.

Saier, M.H. Jr., Wu, L.F., and Reizer, J. 1990. Regulation of bacterial physiological processes by three types of protein phosphorylation systems. *Trends Biochem Sci* 15:391–395.

Sandberg, R., Winberg, G., Branden, C.I., Kaske, A., Ernberg, I., and Coster, J. 2001. Capturing whole-genome characteristics in short sequences using a naïve Bayesian classifier. *Genome Res* 11:1404–1409.

Schmidt, A.L., and Anderson, L.M. 2006. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol Rev Camb Philos Soc* 81:531–543.

Schneider, D., and Lenski, R.E. 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res Microbiol* 155:319–327.

Schneider, D.A., Ross, W., and Gourse, R.L. 2003. Control of rRNA expression in *Escherichia coli*. *Curr Opin Microbiol* 6:151–156.

Schneider, R., Lurz, R., Luder, G., Tolksdorf, C., Travers, A., and Muskhelishvili, G. 2001. An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res* 29:5107–5114.

Schob, H., and Grossniklaus, U. 2006. The first high-resolution DNA "methylome." *Cell* 126:1025–1028.

Schofield, M.A., Agbunag, R., Michaels, M.L., and Miller, J.H. 1992. Cloning and sequencing of *Escherichia coli mutR* shows its identity to *topB*, encoding topoisomerase III. *J Bacteriol* 174:5168–5170.

Schumann, W. 2006. *Dynamics of the Bacterial Chromosome. Structure and Function*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Seifert, H.S. 1996. Questions about gonococcal pilus phase- and antigenic variation. *Mol Microbiol* 21:433–440.

Serres, M.H., and Riley, M. 2005. Gene fusions and gene duplications: relevance to genomic annotation and functional analysis. *BMC Genomics* 6:33.

Shakeri-Garakani, A., Brinkkotter, A., Schmid, K., Turgut, S., and Lengeler, J.W. 2004. The genes and enzymes for the catabolism of galactitol, D-tagatose, and related carbohydrates in *Klebsiella oxytoca* M5a1 and other enteric bacteria display convergent evolution. *Mol Genet Genomics* 271:717–728.

Shanado, Y., Kato, J., and Ikeda, H. 1998. *Escherichia coli* HU protein suppresses DNA-gyrase-mediated illegitimate recombination and induction. *Genes Cells* 3:511–520.

Shindo, H., Furubayashi, A., Shimizu, M., Miyake, M., and Imamoto, F. 1992. Preferential binding of *E. coli* histone-like protein HU alpha to negatively supercoiled DNA. *Nucleic Acids Res* 20:1553–1558.

Siguier, P., Filee, J., and Chandler, M. 2006. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 9:526–531.

Simon, M.I., and Silverman, M. 1983. Recombinational regulation of gene expression in bacteria, pp. 211–227. In: Davies, J., and Gallant, J.A., eds. *Gene Function in Prokaryotes*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Singh, R.K., Palm, G.J., Panjikar, S., and Hinrichs, W. 2007. Structure of the *apo* form of the catabolite control protein A (CcpA) from *Bacillus megaterium* DNA-binding domain. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63:253–257.

Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S.M., Marko, J.F., and Johnson, R.C. 2006. Mechanism of chromosome compaction and looping by the *Escherichia coli* nucleoid protein Fis. *J Mol Biol* 364:777–798.

Sonden, B., and Uhlin, B.E. 1996. Coordinated and differential expression of histone-like proteins in *Escherichia coli*: regulation and function of the H-NS analog StpA. *EMBO J* 15:4970–4980.

Soultanas, P. 2005. The bacterial helicase-primase interaction: a common structural/functional module. *Structure* 13:839–844.

Srikhanta, Y.N., Maguire, T.L., Stacey, K.J., Grimmond, S.M., and Jennings, M.P. 2005. The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes. *ProcNatl Acad Sci USA* 102:5547–5551.

Srividhya, K.V., Krishnaswamy, S. 2007. Subclassification and targeted characterization of prophage-encoded two-component cell lysis cassette. *J. Biosci.* 32:979–990.

Stentebjerg-Olesen, B., Chakraborty, T., and Klemm, P. 2000. FimE-catalyzed off-to-on inversion of the type 1 fimbrial phase switch and insertion sequence recruitment in an *Escherichia coli* K-12 *fimB* strain. *FEMS Microbio. Lett* 182:319–325.

Stephani, K., Weichart, D., and Hengge, R. 2003. Dynamic control of Dps protein levels by ClpXP and ClpAP proteases in *Escherichia coli*. *Mol Microbiol* 49:1605–1614.

Stillman, T.J., Upadhyay, M., Norte, V.A., Sedelnikova, S.E., Carradus, M., Tzokov, S., Bullough, P.A., Shearman, C.A., Gasson, M.J.,

Williams, C.H., Artymiuk, P.J., and Green, J. 2005. The crystal structures of *Lactococcus lactis* MG1363 Dps proteins reveal the presence of an N-terminal helix that is required for DNA binding. *Mol Microbiol* 57:1101–1112.

Stragier, P., Kunkel, B., Kroos, L., and Losick, R. 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science* 243:507–512.

Su, D., Li, Y., and Gladyshev, V.N. 2005. Selenocysteine insertion directed by the 3'-UTR SECIS element in *Escherichia coli*. *Nucleic Acids Res* 33:2486–2492.

Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J., and Wigley, D.B. 1997. Crystal structure of the site-specific recombinase, XerD. *EMBO J*. 16:5178–5187.

Sun, X., Mierke, D.F., Biswas, T., Lee, S.Y., Landy, A., and Radman-Livaja, M. 2006. Architecture of the 99 bp DNA-six-protein regulatory complex of the lambda att site. *Mol Cell* 24:569–580.

Suyama, M., Lathe, W.C., 3rd, and Bork, P. 2005. Palindromic repetitive DNA elements with coding potential in *Methanocaldococcus jannaschii*. *FEBS Lett* 579:5281–5286.

Swanson, J., Bergstrom, S., Robbins, K., Barrera, O., Corwin, D., and Koomey, J.M. 1986. Gene conversion involving the pilin structural gene correlates with pilus+ in equilibrium with pilus changes in *Neisseria gonorrhoeae*. *Cell* 47:267–276.

Swinger, K.K., and Rice, P.A. 2004. IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14:28–35.

Swinger, K.K., Lemberg, K.M., Zhang, Y., and Rice, P.A. 2003. Flexible DNA bending in HU–DNA cocrystal structures. *EMBO J* 22:3749–3760.

Takemaru, K., Mizuno, M., Sato, T., Takeuchi, M., and Kobayashi, Y. 1995. Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology* 141:323–327.

Talukder, A.A., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* 181:6361–6370.

Tapias, A., Lopez, G., and Ayora, S. 2000. *Bacillus subtilis* LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic Acids Res* 28:552–559.

Taylor, D.E., Rooker, M., Keelan, M., Ng, L.K., Martin, I., Perna, N.T., Burland, N.T., and Blattner, F.R. 2002. Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. *J Bacteriol* 184:4690–4698.

Tenson, T., DeBlasio, A., and Mankin, A. 1996. A functional peptide encoded in the *Escherichia coli* 23S rRNA. *Proc Natl Acad Sci USA* 93:5641–5646. [Erratum in: Proc Natl Acad Sci USA 1996 93: 8796.]

Tenson, T., Xiong, L., Kloss, P., and Mankin, A.S. 1997. Erythromycin resistance peptides selected from random peptide libraries. *J Biol Chem* 272:17425–17430.

Thanbichler, M., Viollier, P.H., and Shapiro, L. 2005. The structure and function of the bacterial chromosome. *Curr Opin Genet Dev* 15:153–162.

Tikole, S., and Sankararamakrishnan, R. 2006. A survey of mRNA sequences with a non-AUG start codon in RefSeq database. *J Biomol Struct Dyn* 24:33–42.

Tirumalai, R.S., Healey, E., and Landy, A. 1997. The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci USA* 94:6104–6109.

Tobes, R., and Pareja, E. 2006. Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics* 7:62.

Tobes, R., and Ramos, J.L. 2005. REP code: defining bacterial identity in extragenic space. *Environ Microbiol* 7:225–228.

Travers, A., and Muskhelishvili, G. 2005a. Bacterial chromatin. *Curr Opin Genet Dev* 15:507–514.

Travers, A., and Muskhelishvili, G. 2005b. DNA supercoiling—a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol* 3:157–169.

Tse-Dinh, Y.C. 1998. Bacterial and archaeal type I topoisomerases. *Biochim Biophys Acta* 1400:19–27.

Tsuru, T., Kawai, M., Mizutani-Ui, Y., Uchiyama, I., and Kobayashi, I. 2006. Evolution of paralogous genes: Reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*. *Mol Biol Evol* 23:1269–2985.

Twiss, E., Coros, A.M., Tavakoli, N.P., and Derbyshire, K.M. 2005. Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress. *Mol Microbiol* 57:1593–1607.

Uno, R., Nakayama, T., and Tomita, M. 2006. Over-representation of Chi sequences caused by di-codon increase in *Escherichia coli* K-12. *Gene* 380:30–37.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., and Jeltsch, A. 2002. The *Escherichia coli dam* DNA methyltransferase modifies DNA in a highly processive reaction. *J Mol Biol* 319:1085–1096.

van der Woude, M.W., and Baumler, A.J. 2004. Phase and antigenic variation in bacteria. *Clin Microbiol Rev*17:581–611.

van Noort, J., Verbrugge, S., Goosen, N., Dekker, C., and Dame, R.T. 2004. Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci USA* 101:6969–6974.

Varshavsky, A.J., Nedospasov, S.A., Bakayev, V.V., Bakayeva, T.G., and Georgiev, G.P. 1977. Histone-like proteins in the purified *Escherichia coli* deoxyribonucleoprotein. *Nucleic Acids Res* 4:2725–2745.

Vogel, J., and Sharma, C.M. 2005. How to find small non-coding RNAs in bacteria. *Biol. Chem*. 386:1219–1238.

von Hippel, P.H. 1998. An integrated model of the transcription complex in elongation, termination, and editing. *Science* 281:660–665.

Wang, Q., Mordukhova, E.A., Edwards, A.L., and Rybenkov, V.V. 2006. Chromosome condensation in the absence of the non-SMC subunits of MukBEF. *J Bacteriol* 188:4431–4441.

Wang, Y.D., Zhao, S., and Hill, C.W. 1998. Rhs elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*. *J Bacteriol* 180:4102–4110.

Wassarman, K.M. 2002. Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* 109:141–144.

Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G., Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C., and Blattner, F.R. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71:2775–2786. [Erratum in: Infect Immun. 2003 Jul;71(7):4223]

Welch, R.A., Burland, V., Plunkett, G. 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud,

D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., and Blattner, F.R. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 99:17020–17024.

Wells, R.D., Dere, R., Hebert, M.L., Napierala, M., and Son, L.S. 2005. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res* 33:3785–3798.

Wertz, G.W., and Moudy, R.M. 2004. Antigenic and genetic variation in human respiratory syncytial virus. *Pediatr Infect Dis J* 23 (Suppl. 1): S19–S24.

Weyand, N.J., Braaten, B.A., van der Woude, M., Tucker, J., and Low, D.A. 2001. The essential role of the promoter-proximal subunit of CAP in *pap* phase variation: Lrp- and helical phase-dependent activation of *papBA* transcription by CAP from -215. *Mol Microbiol* 39:1504–1522.

Wigley, D.B. 1995. Structure and mechanism of DNA topoisomerases. *Annu. Rev. Biophys. Biomol. Struct*. 24:185–208.

Wigley, D.B. 1998. Teaching a new dog old tricks? *Structure* 6:543–548.

Wilson, L.A., and Sharp, P.M. 2006. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: Evolution and implications for ERIC-PCR. *Mol Biol Evol* 23:1156–1168.

Woldringh, C.L. 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol Microbiol* 45:17–29.

Woldringh, C.L., Jensen, P.R., and Westerhoff, H.V. 1995. Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces? *FEMS Microbiol Lett* 131:235–242.

Wolk, C.P. 1996. Heterocyst formation. *Annu Rev Genet* 30:59–78.

Wright, B.E. 2004. Stress-directed adaptive mutations and evolution. *Mol Microbiol* 52:643–650.

Wu, L.-F., and Saier, M.H., Jr. 1991. Differences in codon usage among genes encoding proteins of different function in *Rhodobacter capsulatus*. *Res Microbiol* 142:943–949.

Xia, Y., Gally, D., Forsman-Semb, K., and Uhlin, B.E. 2000. Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. *EMBO J* 19:1450–1457.

Xie, G., Brettin, T.S., Bonner, C.A., and Jensen, R.A. 1999. Mixed-function supraoperons that exhibit overall conservation, albeit shuffled gene organization, across wide intergenomic distances within eubacteria. *Microb CompGenomics* 4:5–28.

Xie, Y., Yao, Y., Kolisnychenko, V., Teng, C.H., and Kim, K.S. 2006. HbiF regulates type 1 fimbriation independently of FimB and FimE. *Infect Immun* 74:4039–4047.

Yang, C.R., Shapiro, B.E., Hung, S.P., Mjolsness, E.D., and Hatfield, G.W. 2005. A mathematical model for the branched chain amino acid biosynthetic pathways of *Escherichia coli* K12. *J Biol Chem*280:11224–11232.

Yang, W., and Steitz, T.A. 1995. Crystal structure of the site-specific recombinase $\gamma\delta$ resolvase complexed with a 34 bp cleavage site. *Cell* 82:193–207.

Yang, Y., and Ames, G.F. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc Natl Acad Sci USA* 85:8850–8854.

Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289:751–758.

York, M.K., and Stodolsky, M. 1981. Characterization of P1*arg*F derivatives from *E. coli* K-12 transduction. I. IS1 elements flank the *argF* segment. *Mol Gen Genet* 181:230–240.

Young, B.A., Gruber, T.M., and Gross, C.A. 2002. Views of transcription initiation. *Cell* 109:417–420.

Yu, H.G., and Koshland, D. 2005. Chromosome morphogenesis: condensin-dependent cohesin removal during meiosis. *Cell* 123:397–407.

Zhang, C.C., Laurent, S., Sakr, S., Peng, L., and Bedu, S. 2006. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol Microbiol* 59:367–375.

Zimmer, C. 2007. Evolution. Jurassic genome. *Science* 315:1358–1359.

Zimmerman, S.B. 2006a. Cooperative transitions of isolated *Escherichia coli* nucleoids: implications for the nucleoid as a cellular phase. *J Struct Biol* 153:160–175.

Zimmerman, S.B. 2006b. Shape and compaction of *Escherichia coli* nucleoids. *J Struct Biol* 156:255–261.

Zimmerman S.B., and Trach, S.O. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* 222:599–620.

Editor: Michael M. Cox