Preface

Bacteria are a large group of unicellular microorganisms. Most bacteria have a single circular chromosome that can range in size. DNA is organized into chromosomes and, in organisms other than bacteria, it is found only in the cell nucleus. Deoxyribonucleic acid(DNA)is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. Their main role is the long-term storage of information. DNA is often compared to a set of blueprints or a code, since it contains the instructions needed to construct other components of cells. This book presents current research on bacterial DNA, DNA polymerase and DNA helicases.

Chapter I - The new generation of fast and low-cost DNA sequencing methods led to a rapid accumulation of genomic data in the DNA databases after the year 2000. Hundreds of bacterial genomes have been sequenced during the last decade, however, in rare cases they were profoundly analyzed. Apparently it is much easier today to generate raw DNA sequencing data rather than to decipher their biological meaning. This study aims to shed light on the relationship between codon usage and gene expression in 158 prokaryotes belonging to the two major classes: *Archaea* and *Bacteria*. *Archaea* includes 2 families (*Crenarchaeota* and *Euryarchaeota*) and *Bacteria* 13 families (*Actinobacteria, Aquificae, Bacteroidetes, Chlamydiae, Chlorobi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Fusobacteria, Planctomycetes, Proteobacteria, Spirochaetes* and *Thermotogae*) and the total number of genera was 9 for *Archaea* and 22 for *Bacteria*. In order to study the link between codon usage and gene expression four local DNA databases were created containing respectively: a) all protein coding sequences of each organism; b) highly expressed genes; c) ribosomal proteins genes and d) low expressed genes. These databases were used to compare codon usage patterns of: a) different strains of one bacterial species; b) different subsets of genes in one species and c) different bacterial genomes. They were also employed to identify missing and atypical codons as well as to search for correlation between codon usage and the formal prokaryotic taxonomy. The obtained results indicate that the formal taxonomy of many bacterial species does not fit their codon usage pattern (as predicted by the "Genome Theory"). On the other hand there are obvious similarities in the codon usage pattern of distant (according to the formal taxonomy) organisms. The latter suggests that either the Genome Theory or the official taxonomy of microorganisms need revisions. Taking into consideration that at least two tRNAs occupy the two (A and P) functional ribosomal sites

during the translation elongation step, the authors have also studied the effect of neighbor codons (codon pairs) on gene expression. Aiming to identify preferential and rare codon pairs, the authors have developed own criteria for determining their frequency of occurrence and for evaluation of their possible effect on translation. This approach was applied to analyze the distribution of all (3904) codon pairs in the *Escherichia coli* genome and the authors found that their frequency of occurrence varied from zero to 4913 times. The predicted effect of some (particularly 3'-terminal) codon pairs on gene expression is confirmed experimentally.

Chapter II - DNA is a fragile organic molecule of finite chemical stability, capable of interacting with a variety of chemicals. In the context of the living cell this DNA reactivity is manifested as mutability. In addition to the chemical mutability, the perpetuation of DNA via imperfect replication is another source of DNA changeability. It is worth mentioning that DNA repair is error-tolerant and always leaves behind some lesions exist. Were repair mechanisms perfect, then the door of evolution would be locked. Risky rather than benign at individual level, mutations are the living things' long term strategy for survival and evolution. The creative power of mutations can explain why organisms have evolved special mechanisms to enhance mutagenesis under certain conditions. Hypermutation in bacteria under stress (SOS response) as well as hypermutation of immunoglobulin genes in response to antigen challenge in vertebrates are key examples.

The present study is focused on spontaneous mutagenesis in *Escherichia coli* K-12.

Spontaneous mutations are the net result of DNA damage and repair under normal physiological conditions. Bacterial DNA is loosely packed with proteins and occupies almost the entire cellular space. Therefore, it is extremely vulnerable to chemical attack, especially during replication and transcription, when the two DNA strands are separated. In the same arena, the cytoplasm, and at the same time, play also many small but highly reactive compounds. Although the collision of the latter with DNA seems to be unavoidable, to date the impact of ordinary cellular metabolites on spontaneous mutagenesis has not been systematically studied.

In year 2012 we shall celebrate the 100 anniversary of the discovery of the carbonyl-amine reaction by the French chemist Louis Camille Maillard (Maillard, 1912). Here, the authors provide evidence that the Maillard reaction, referred to as non-enzymatic glycosylation or glycation in the literature, is an important endogenous source of spontaneous mutations in *E. coli*. Experimental data allowed us to estimate that glycation affects on average one per 10^5 to 10^4 nucleotides in the *E. coli* chromosome. Based on experimental evidence, the authors postulate the existence in *E. coli* of a novel DNA repair enzyme, DNA amadoriase or DNA deglycase, especially designed to combat glycation induced spontaneous mutagenesis in *E. coli*.

Chapter III - Both prokaryotes and eukaryotes store their genomic DNA in an environment, which balances the physical properties of double-stranded DNA with the mechanical effects of DNA-protein interactions. The origin of genome packing remains a mystery; it seems to go back to the very beginning of life itself. Since then, the principle of higher-order genome construction and architecture has been maintained and shared among three domains of living things. The genomes are stored as flexible higher-order structures that can shuttle between relatively active and inactive states. The major differences reside in the

structural protein components of the architecture. Bacteria and Eukaria utilize HU and histones as the most fundamental structural proteins, respectively, forming the nucleoid in bacterial cell and the nucleosomes in the nucleus. On the other hand Archaea employ either HU or histones depending on their phylogenetic origins. Nevertheless, the hierarchies of nucleoid fibers look extremely similar among different species in different domains. Focusing upon the mechanism of bacterial genome folding, recent structural investigations have revealed a step-wise folding of the genome DNA; from 10 nm to 30 nm fibers, and to 80 nm and further condensed beaded structures, depending upon the growth conditions and differential contribution of nascent single-stranded RNA. The nucleoid condensation is mainly brought about by another nucleoid protein, Dps, which occurs widely in the bacterial domain, but not in Archaea or Eukarya. The regulatory mechanisms of the *dps* gene are greatly different among bacterial species. In □-Proteobacteria such as *Escherichia coli*, the *dps* gene is governed by two regulatory systems, IHF-□s and OxyR, and is induced towards the stationary phase and under oxidative stress. The DNA topology control by *E. coli* Fis also participates in the regulation of the nucleoid condensation. In contrast, in Firmicutes such as *Staphylococcus aureus*, the gene is mainly controlled by oxidative stress-response system, PerR, and the Dps expression is directly coupled with the nucleoid condensation. Here the key features of the nucleoid dynamics will be discussed from the biochemical and structural biological points of views.

Chapter IV - *Bacillus subtilis*, which is one of the most important host microorganisms for large-scale industrial production of useful proteins, has a genome of 4.2 Mb with approximately 4106 protein-coding genes. Some of these genes are expected to be unnecessary for industrial production of proteins under controlled conditions and may be wasteful with regard to energy consumption. The authors attempted to reduce the genome size of *B. subtilis* by deleting unnecessary regions of the genome to allow the construction of simplified host cells as a platform for the further development of novel genetic systems with increased productivity.

First, the authors generated the strain MGB469 with deletion of all prophage (SPβ and PBSX) and prophage-like (pro1-7 and skin) sequences, with the exception of pro7, as well as two large operons that produce secondary metabolites (*pks* and *pps*). These cells showed normal growth, but no beneficial effects were observed with regard to recombinant protein production from plasmids carrying the corresponding genes. Second, the authors constructed several multiple-deletion mutants containing additional deletions in the MGB469 genome, resulting in total genome size reductions of 0.78 to 0.99 Mb. In most of the multiple-deletion series, extensive deletion mutants showed no beneficial improvements in traits as host strains. The strain MG1M with a total genome size reduction of 0.99 Mb showed unstable phenotypes with regard to growth rate, cell morphology, and recombinant protein productivity after successive culture, making it inappropriate for further studies. In addition, strain MGB943 derived from another lineage with genome reduction of 0.94 Mb showed reduced recombinant cellulase productivity. Finally, they generated another multiple-deletion series including the mutant MGB874 with a total genome deletion of 0.87 Mb. In comparison to wild-type cells, the metabolic network of the mutant strain was reorganized after entry into the transition state due to the synergistic effects of multiple deletions. Moreover, the levels of

production of extracellular cellulase and protease from transformed plasmids carrying the corresponding genes were markedly increased.

Our results demonstrated the effectiveness of a synthetic genomic approach with reduction of genome size to generate novel and useful bacteria for industrial uses.

Chapter V - Enterically transmitted viruses such as norovirus (NoV), rotavirus (RV), hepatitis A virus (HAV) and hepatitis E virus (HEV) are currently recognized as a major cause of food and waterborne illness in humans worldwide. Most of these viruses cannot be cultured *in vitro*, are stable in the environment for long periods of time, are generally infectious at low doses (only a few viral particles are needed to trigger illness) and would be usually present in low numbers in contaminated food and water samples. There are also increasing concerns about the possible zoonotic transmission or the emergence of new recombinant strains from animal enteric viruses closely related to human pathogenic strains. Until recently, the viral contamination of food and water remained undetected due to a lack of appropriate detection methods. Electronic microscopy (EM) and ELISA assays commonly used on stool suspensions were rather insensitive for food and water applications. Due to their increased sensitivity, conventional and now real-time quantitative reverse transcription PCR (qRT-PCR) assays were developed and widely used for the detection and identification of the nucleic acid of these challenging viruses. Along with microarrays, these new approaches have also allowed significant advances in the molecular typing of these organisms. Rapid, sensitive and reliable molecular detection and typing methods will have a significant impact in the mitigation of outbreaks and could contribute in the prevention of food/waterborne transmission. This review will provide recent trends and advances in the field of DNA polymerase-based detection and characterization methods of these 4 main food and waterborne viruses. It will focus on the use of different conventional and qRT-PCR methods and the microarray technology.

Chapter VI - Chromatin structure plays an important role in the regulation of gene transcription. Chromatin structure can be modified by various post-translational modifications, including histone acetylation, phosphorylation, methylation and ribosylation. Among those modifications, histone acetylation/deacetylation is the most important mechanism for regulating transcription and is regulated by a group of enzymes known as histone acetyltransferases/histone deacetylases (HDACs).

Recently, HDAC inhibitors have been shown to be a novel and promising new class of anti-cancer agent that can regulate the transcription of genes by disrupting the balance of acetylation/deacetylation in particular regions of chromatin. A number of HDAC inhibitors are currently in phase I and II clinical trials against a variety of cancers. Although some promising candidates have been identified (e.g., $p21^{WAF1}$ and c-Myc), the precise molecular targets remain uncertain. In this article, the authors focus on one of the DNA polymerases, telomerase, as a new candidate molecular target for HDAC inhibitors. Telomerase is composed primarily of the catalytic subunit (hTERT) and the RNA template (hTERC), and its activity correlates with levels of hTERT mRNA. hTERT expression is apparently governed by complicated regulatory pathways. Based on recent studies, the hTERT gene is likely to be targeted by histone acetylation/deacetylation.

Chapter VII - Replicative DNA polymerase (DNAP) is an enzyme that synthesizes a new DNA strand on a single-stranded template with a high processivity and a high fidelity. The

high fidelity is mainly realized via a mechanism of proofreading that is performed at the exonuclease active site spatially separate from the polymerase active site. Here the authors present a detailed account of our proposed model for the processive nucleotide incorporation and switching transition of DNA between the polymerase and exonuclease active sites by DNAP. Based on the model, the authors present detailed theoretical studies on its dynamics. The moving time of DNAP to next site after a correct incorporation and that after an incorrect incorporation are analytically studied. The experimentally measured dependence of polymerization rate on tension applied to the template is well simulated. The transfer rates of DNA from the polymerase to exonuclease active sites after a correct incorporation and after incorrect incorporations as well as the transfer rate from the exonuclease to polymerase active sites are analytically studied. Moreover, the backward motion of DNAP when large tensions are applied to the template is also analytically studied. The theoretical results are in good agreement with available experimental data. Some predicted results are presented.

Chapter VIII – Hepatitis B virus infection is a major public health problem worldwide especially in East Asia. The viral infection, if persistent, may lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma. In1998, FDA approved GlascoSmithKline's Epivir-HBV (Lamivudine), a small molecule drug that inhibits HBV DNA polymerase and interferes with viral replication. Since then, a few similar drugs have been approved and many candidates are currently at preclinical and clinical development. The use of small molecular drugs that target HBV DNA polymerase is a milestone in the treatment of chronic hepatitis B. In this review, the authors will focus on HBV DNA polymerase and discuss the preclinical and clinical development process for the HBV polymerase inhibitors. The DNA polymerase mutants associated with drug resistance will also be discussed. The mechanism of the drug resistance and further understanding of these DNA polymerase inhibitors may help the determination of better clinical regimen for HBV therapy and patient care.

Chapter IX - RecG is a DNA helicase involved in the repair of damage at a replication fork by catalyzing the reversal of the fork to create a Holliday junction. Here, based on previous structural and biochemical studies a model is presented on how RecG catalyzes the processive translocation and unwinding of DNA so as to realize the fork reversal. In the model, a power stroke induces the DNA translocation and unwinding. Moreover, in order to effectively unwind DNA of a helical structure, it is assumed that the residues (i.e., the long α helix) connecting the wedge domain and the helicase domains in which the dsDNA-binding loop is located behave elastically along the torsional direction. Thus, accompanying the power stroke, the wedge domain is forced elastically to rotate relative to the dsDNA-binding loop that binds dsDNA strongly. Using the model, the calculated DNA translocation size per ATP hydrolysis is consistent with previous experimental data. Moreover, it is showed that RecG has a strong preference for negatively supercoiled DNA, i.e., RecG shows a much higher ATPase activity and longer processivity when interacting with the negatively supercoiled DNA than with the relaxed and/or positively supercoiled DNAs. This is also in agreement with the previous experimental data. In addition, the rotational rate or the rotational angle of RecG relative to DNA per ATP hydrolysis and the torsional elastic coefficient of the residues connecting the wedge domain and the helicase domains are predicted.

Chapter X - In this chapter, the *de novo* DNA synthesis by DNA polymerase in the absence of any added template and/or primer is described. As early as the 1960s, Kornberg et al. reported that DNA polymerase I could *de novo* synthesize DNA, although it was later pointed out that the contaminated DNA or RNA may provide the seeds for DNA synthesis and the contaminated transferase that can polymerize dNDPs might play the role of template-independent DNA synthesis. The synthesized DNA polymers were characterized as homopolymer pairs such as poly(dA)/poly(dT) and poly(dI)/poly(dC), or the alternating copolymers poly(dA-dT), poly(dG-dC) and poly(dI-dC). In late 1990s, after a long silent time, Ogata et al. found that the highly purified thermophilic DNA polymerase could carry out *de novo* DNA synthesis at higher temperatures (> 65°C) and the synthesized DNA polymers were composed of repetitive palindromic sequences such as $(TACATGTA)_n$, and $(TAAT)_n$. Surprisingly, Liang et al. reported that thermophilic DNA polymerases could carry out *de novo* DNA synthesis even at room temperature. Basically, *de novo* DNA synthesis is considered to occur in two stages, although the detail mechanism is not clear. In the initial stage, dNTPs are polymerized to short DNA oligos, and then the selected repetitive sequences are elongated to long DNA in the elongation stage through self-priming and primer extension. More interestingly, the *de novo* DNA synthesis can be greatly accelerated by adding endonuclease such as restriction enzymes, which digest DNA at their recognition sites. A Cut-grow mechanism was proposed to explain both the paradox of endonuclease and the extremely high efficiency: the synthesized long DNA is cut to shorter seeds for the elongation of the next cycle so that DNA can be exponentially amplified. The possible biological and evolutionary roles of *de novo* DNA synthesis are also discussed.

Chapter XI - Due to the limited processivity of replicative DNA polymerases (replicases), as well as to their incapacity to unwind parental duplex DNA to allow replication fork progression, their replication efficiency depends on the functional assistance of accessory proteins as processivity factors and helicases. In addition, the inability of DNA polymerases to start de novo DNA synthesis requires the use of a short RNA/DNA molecule to provide the 3'-OH group required to initiate DNA replication. This requisite for a primer creates a dilemma to replicate the ends of linear genomes: once the last primer for the lagging strand synthesis is removed, a portion of ssDNA at the end of the genome will remain uncopied. Bacteriophage ϕ29 has overcome these issues by means of the unique catalytic features of an outstanding enzyme, the ϕ29 DNA polymerase. This replicase belongs to the family B (eukaryotic-type) of DNA-dependent DNA polymerases and has served as model to understand the enzymology of these polymerases. As most of the family B members, ϕ29 DNA polymerase contains both 3′-5′ exonuclease and polymerization activities residing in two structurally independent domains. During two decades, site-directed mutagenesis studies of individual residues contained in regions of high amino acid similarity have provided the functional insights of this enzyme, extrapolative to other family B members. However, ϕ29 DNA polymerase is endowed with two distinctive features: high processivity and strand displacement capacity that allow it to replicate the viral genome from a single binding event, without requiring the assistance of unwinding and processivity factors. Recent crystallographic resolution of the structure of the apo and binary/ternary complexes of ϕ29 DNA polymerase, together with the biochemical studies of site-directed mutants, have given insights into the structural basis responsible for the coordination of the processive

polymerization and strand displacement. In addition, such structures have provided the mechanism of translocation of family B DNA polymerases. Another difference with respect to the rest of replicases is the ability of ϕ29 DNA polymerase to use a protein (terminal protein, TP) as primer, circumventing the end replication problem. Recent resolution of the structure of the ϕ29 DNA polymerase/TP heterodimer, together with the biochemical analysis of chimerical DNA polymerases and TPs, have given the clues of the specificity of the interaction between both proteins, suggesting a model for the transition from initiation to elongation. The authors will also discuss how the basic research on the ϕ29 DNA polymerase properties have led to the development of DNA amplification technologies based on this outstanding enzyme.

Chapter XII - Replicative DNA helicases are energy-transducing enzymes, which act as the engine of the cellular replication apparatus and unwind duplex DNA in the replication fork. These enzymes are true multifunctional enzymes that are responsible for organization and execution of DNA replication from initiation to termination. The most well studied of them all is DnaB helicase of *Escherichia coli*, which is the prototype of this class of DNA helicases. Early genetic studies of the *dnaB* gene defined important roles of DnaB protein in the elongation stage of DNA replication. DnaB protein acts as the organizer of the replisome by engaging in multiple protein-protein and protein-DNA interactions during the initiation, elongation, and termination stages of DNA replication. DnaB protein and its orthologs have homo-hexameric structure in solution and in the ssDNA-bound state. This homo-hexameric subunit structure allows for the formation of a ring-structure; ssDNA passes through the ring's central hole. Formation of this unique protein-ssDNA complex is a key element in this enzyme's ability to carry out its unique functions. Due to its stable ring structure, *E. coli* DnaB helicase requires the assistance of DnaC protein in order to form the initial DnaB•ssDNA complex. It forms a DnaB•DnaC•ATP complex that attenuates its nucleotidase activity but helps in binding ssDNA. DnaC protein is released upon slow ATP hydrolysis leading to organization of the replication proteins in the fork and replisome formation. DnaB protein engineers the assembly of the replisome through its specialized interaction with DnaG primase and DNA polymerase III holoenzyme. DnaB•ssDNA complex formation stimulates its ATPase activity of DnaB hexamer several fold. The energy of ATP hydrolysis powers translocation of DnaB protein on ssDNA in a 5′→3′ direction and helps unwind duplex DNA during translocation. The rapid and processive DNA unwinding also requires direct participation of a topoisomerase to remove the DnaB-generated super-twists without which the DNA unwinding and the replication fork would likely stall. In this review article, the authors summarize the structural and functional properties of the *E. coli* replicative DNA helicase and compare it to analogous replicative helicases from other gram-negative and gram-positive bacteria.

Chapter XIII - The helicase encoded by the hepatitis C virus (HCV) is shown in this chapter to catalyze homologous DNA strand exchange. Single-stranded DNA oligonucleotides complementary to either the short or long strand of a partially duplex DNA substrate affected the rate and extent of unwinding catalyzed by either the full-length HCV NS3 protein fused to a portion of HCV NS4A, or a truncated NS3 protein lacking the protease domain. The oligonucleotides did not, however, sequester HCV helicase and prevent it from separating the original duplex after a single binding cycle. Furthermore, when DNA

oligonucleotides were pre-incubated with HCV helicase, they did not prevent subsequent duplex separation, indicating that the enzyme catalyzes strand exchange. The protease portion of NS3 was not needed for this strand exchange. Fluorescent DNA substrates were further used to directly monitor both this ssDNA assisted unwinding and homologous strand exchange, and the effect of a protein trap (poly(U) RNA) on HCV helicase-catalyzed strand exchange was examined. The results demonstrate that HCV helicase can simultaneously bind at least three DNA strands and imply that HCV NS3 helicase could play an important role not only in viral RNA unwinding, but also in the folding of the HCV genome.

Short Communication - Eukaryotic DNA helicases are members of the helicase superfamily together with two other functional classes, RNA helicases and chromatin remodeling ATPases. Members of this superfamily play crucial roles in various nucleic acid- and chromatin-mediated cellular processes such as DNA replication, repair and recombination, pre-mRNA splicing, ribosome biogenesis, RNA interference, and chromatin remodeling. Since these reactions are essential for the maintenance, expression, and regulation of genetic information in the chromosome, dysfunctional helicase genes may lead to genetic diseases such as cancer. Indeed, genetic mutations of the human RecQ-like BLM helicase and WRN DNA helicase result in Bloom syndrome and Werner syndrome, which are characterized by the early development of various cancers and premature aging, respectively . Most helicases share conserved amino acid sequence motifs; their corresponding genes are classified into five superfamilies (SF1-SF5) based on the occurrence and characteristics of conserved motifs.

The genome sequences of two representative eukaryotes, the budding yeast *Saccharomyces cerevisiae* and the nematode *Caenorhabditis elegans* were determined in 1996 and 1998, respectively. Subsequently, high-throughput genome-wide analyses have been systematically performed to clarify the cellular roles of 6000 and 19000 genes discovered in each genome, including gene expression profiling , proteome analyses, comprehensive analyses of loss-of-function phenotypes, and genetic interaction analyses. A huge amount of functional data from these studies has been analyzed and deposited in public databases such as the *Saccharomyces* Genome Database (SGD) and the WormBase. However, a few decades after the initial sequencing, a large number of genes remain functionally-unknown (i.e., orphan) in both organisms; for instance, over 1000 genes are uncharacterized even in yeast. A large number of helicase-like genes have been found in the sequenced eukaryotic genomes, including many genes encoding orphan helicase-related proteins. Because of the biological importance of the helicase family, a comprehensive analysis of the functions of helicase family members in two model organisms, *S. cerevisiae* and *C. elegans*, has been performed. In this Short Communication, the authors briefly summarize the results from studies of helicase-like proteins in both organisms and describe the effectiveness and limitations of a genome-wide analysis of orphan family members.