Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*?

Purnananda Guptasarma

Summary

Over 80% of the genes in the *E. coli* chromosome express fewer than a hundred copies each of their protein products per cell. It is argued here that transcription of these genes is neither constitutive nor regulated by protein factors, but rather, induced by the act of replication. The utility of such replication-induced (RI) transcription to the temporal regulation of synthesis of determinate quantities of low copy number (LCN) proteins is described. It is suggested that RI transcription may be necessitated, as well as facilitated, by the folding of the bacterial chromosome into a compact nucleoid. Mechanistic aspects of the induction of transcription by replication are discussed with respect to the modulation of transcriptional initiation by negative supercoiling effects, promoter methylation status and derepression. It is shown that RI transcription offers plausible explanations for the constancy of the C period of the *E. coli* cell cycle and the remarkable conservation of gene order in the chromosomes of enteric bacteria. Some experimental tests of the hypothesis are proposed.

Accepted
27 June 1995

Introduction

The *Escherichia coli* genome is thought to contain about 3,000 genes[1]. The products of over 2,100 genes can be clearly distinguished as separate spots on autoradiograms of two-dimensional polyacrylamide gels of radiolabelled cell extracts[2,3]. Densitometric analyses of the spots indicate that a mere 300 to 400 proteins, present in copy numbers ranging from a few hundred to several tens of thousands of copies/cell, account for more than 90% (w/w) of the protein content of the *E. coli* cell[2-4]. The remaining 1,700 to 1,800 proteins, which represent more than 80% of the variety of proteins synthesised by the bacterium, seem to be present in copy numbers of only a few to a hundred copies/cell. Most of these low copy number (LCN) proteins have not yet been characterised. From the little that we do know about the few (less that fifty) such proteins of known function whose intracellular abundances have been measured, however, it is clear that LCN proteins can perform functions that are crucial to the growth and survival of the cell. These include: (1) regulation of gene expression (e.g. the lac repressor[5]: 10 to 30 copies/cell); (2) control of cell morphology (e.g. the cell division aide PBP2[6]: 20 copies/cell); (3) initiation of DNA replication (e.g. the primase dnaG[7]: 50 to 100 copies/cell); (4) mediation of low frequency DNA recombination (e.g. the Tn10 transposase[8]: less than one copy/cell); and (5) repair of DNA damage (e.g. deoxyribopyrimidine photolyase[9]: 10 to 20 copies/cell). Given the nature and importance of the functions that LCN proteins can perform, surprisingly little attention has been paid to understanding how the *E. coli* cell regulates the production of these molecules in such low, but reasonably determinate, numbers per cell. It is generally assumed that the proteins must be produced in a constitutive manner[10,11], since considerations of economy argue against specific regulation of their expression by DNA-binding protein factors.

This paper begins with an evaluation of the likelihood that LCN protein synthesis is constitutive, and it is argued that it is highly unlikely that these proteins are synthesised in an unregulated fashion throughout the cell cycle. Instead, genes encoding a majority of LCN proteins probably each only support transcription during a brief period of the cell cycle, beginning with their own replication and ending soon thereafter.

The possible existence of a cause-effect linkage between

© ICSU Press 1995 pp. 987–997

replication and transcription has been discussed by a number of different authors during the last three decades, mainly from the viewpoint that replication-induced transcription could ensure the timed synthesis of proteins in the cell, when such synthesis is necessary. In this paper, I examine the possibility of such a linkage from a much broader set of perspectives, arguing that replication-induced transcription serves as a means of regulating the quantum of expression of the many hundreds of genes in the *E. coli* chromosome that code for LCN proteins.

The biosynthesis of LCN proteins: examining various possibilities

Weak transcription

The low copy number of a LCN protein can arise, in principle, from poor transcription of DNA, inefficient translation of mRNA, efficient degradation of protein products, or indeed any combination of such factors; without accurate experimental data, it is not possible to speculate which factor is most important in any particular instance of gene expression. However, consideration of the net quantum of steady state transcription occurring in *E. coli* at any given time suggests that even if translational and proteolytic controls do operate in some instances, most LCN protein-coding (LCNP) genes can each produce no more than a very small number of transcripts (<5) per cell generation.

In cultures growing with a generation time of 40 minutes, roughly 13×10^5 nucleotides of mRNA are synthesised per minute per *E. coli* cell[12]. Assuming that the length of mRNA coding for the average *E. coli* protein (300-350 amino acids long) is about 1,000 nucleotides, one reckons that the equivalent of only about 1,300 transcripts must be produced within each cell, every minute. Since there are between 300 and 400 strongly expressed genes that appear to be transcribed throughout most of the cell cycle, it seems that, at the very least, 900 to 1200 transcripts must be generated every minute from this population of continuously transcribed genes, leaving capacity for the synthesis of only 100 to 400 LCNP transcripts; given a measured, average rate of mRNA synthesis of 50 bases/seconds[12,13] (or 20 seconds per transcript) and assuming that transcriptional initiation occurs at the promoter of any strongly transcribed gene, or operon, with a time period that is at least equal to (if not always smaller than) the time taken by a single RNA polymerase elongation complex to transcribe the average gene. [In other words, it is assumed that any strongly transcribed gene is always to be found supporting transcription by at least one RNA polymerase. This is a rather conservative assumption, considering that initiation of transcription at the promoters of many known truly constitutive genes actually occurs with time periods that are as much as five or ten times shorter (2-4 seconds) than that assumed here[10,14]]. It would thus seem that most of the transcripts synthesised in the *E. coli* cell every minute originate from strongly transcribed genes and not from LCNP genes. Even if one were to assume, possibly by erring on the side of overestimation, that as many as 300 of the 1,300 transcripts made each minute code for LCN proteins, one would reckon that only 12,000 transcripts could possibly be made from the entire population of LCNP genes during the lifetime of such a cell (40 minutes), amounting to an average of only 6 to 7 transcripts per LCNP gene per cell generation!

Such poor overall levels of transcription are not incompatible with LCN protein requirements. Given the average efficiency of utilisation of mRNA of about 50 protein molecules/transcript[12] within cells growing with a generation time of 40 minutes, it would seem that up to a hundred molecules of any protein could be made from as few as two or three transcripts. Of course, if the representative calculation presented above were based on slightly different assumptions and on data drawn from cells growing with different generation times[12], the results would be slightly different. Yet, regardless of the exact numbers considered, it becomes obvious from every such calculation that: (1) Most LCNP genes are probably only able to support the synthesis of very few transcripts each, during the entire lifetime of a bacterial cell; (2) this low quantum of transcription from the average LCNP gene is enough to ensure the production of the requisite low copy numbers of LCN proteins in the cell; and (3) only a handful of the large population of LCNP genes may actually be found transcribing within the cell at any given point in time, since the transcription machinery appears to be preoccupied with the transcription of genes coding for abundant proteins.

Constitutive versus *replication-induced transcription*

Considering the small numbers of transcripts produced by each LCNP gene, as well as the large number of LCNP genes and the limitation of being able to transcribe only a small fraction of such genes at any given time, it appears improbable that the *E. coli* cell manages to keep the levels of these proteins down by actively down-regulating expression from LCNP genes through use of specific protein factors.

The alternative is to assume that the genes are transcribed constitutively at frequencies of the order of one transcriptional initiation event per 10-40 minutes. Although such a possibility is not untenable, it does give rise to some unease: on the basis of what we know about the kinetics and energetics of promoter melting/recognition by RNA polymerase[14] (especially aspects relating to the special conditions that need to be brought about by supercoiling, or activator protein binding and associated looping or bending of DNA), doubts could justifiably be raised about the dependability with which initiation of transcription might occur stochastically, at such low frequencies, at promoters that seem to be unable to support recognition and/or melting by RNA polymerase for tens of minutes at a time. Since

initiation could be assumed to occur stochastically and not periodically, the time interval between any two successive initiations could equal or exceed the length of the cell cycle in many instances, with the result that certain transcripts might never be made within some cells. For proteins that are crucial to the progression of the cell cycle, as well as for proteins required to act during definite phases of the cycle, such a situation could not be desirable. Furthermore, on a time- and population-averaged basis, constitutive transcription could be expected to give rise to some variations in the copy numbers of proteins as a function of generation time (discounting adjustments in the various synthetic rates which happen to partly compensate for such changes in generation time[10]). Given that the average copy number of any gene in the cell does not change much between cells growing with generation times ranging from 1.2 hours (1.5 genome equivalents/cell) to many tens of hours (1 genome equivalent/cell)[15], it could indeed be counterproductive for cells to allow constitutive synthesis of any essential LCN protein. The scant available literature on the subject of abundances of these proteins as a function of generation time[4] indicates that, if anything, the numbers of these proteins may be reasonably constant. Altogether, it seems unlikely that constitutive transcription could be responsible for the synthesis of the LCN proteins in general, although it is possible that some LCN proteins could be expressed in this manner.

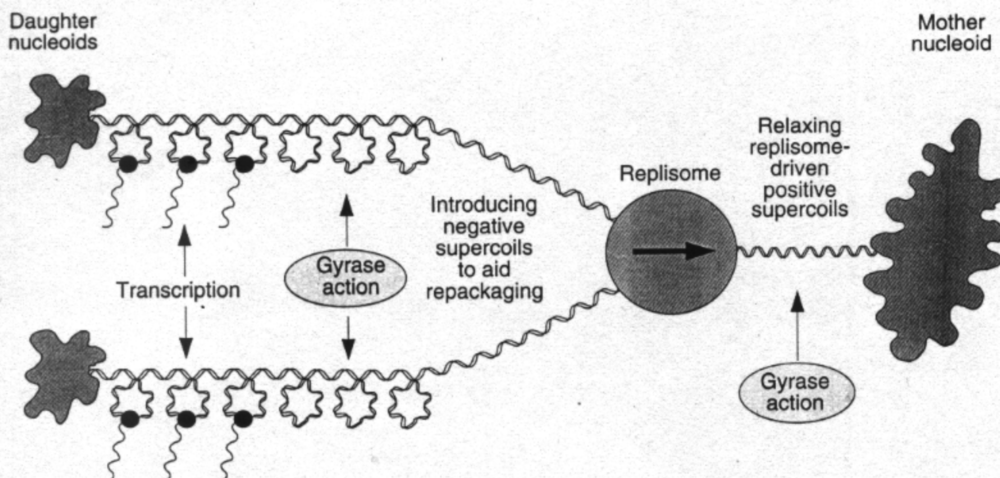
In the light of the specific issues discussed above, the advantages of replication-induced transcription can be seen to be many. (1) *Transcription would only occur under opportune conditions.* If replication were to make conditions opportune for transcriptional initiation to occur, by transiently altering the physical states of genes in some way (discussed later), all the transcripts required to be produced from an LCNP gene could be produced within the short time span during which such conditions were brought to bear on the gene, instead of stochastically (over long periods of time) at low frequencies, under unfavourable conditions. (2) *Determinate amounts of protein would be produced.* Identical conditions would apply to the passage of the replication fork, during each cell cycle. Because each gene would be replicated only once in a cell cycle, roughly the same number of transcripts would be made in each cell under most conditions of growth, leading to identical numbers of proteins being produced in all cells. (3) *The timing of synthesis would be determinate.* Since replication of any gene occurs at a more or less fixed point in time before the cell divides[16], replication-induced transcription could provide for a means of 'timed' synthesis of many LCN proteins important to the progression of the cell cycle. (4) *Determinate allocation of resources.* As argued above, only a small number of LCNP genes can support transcription at any given time. If the bulk of active RNA polymerase molecules available for the transcription of LCNP genes were to follow the replication fork on its way around the genome (effecting transcription over only a small window comprising a few tens of genes at any given time), only the genes situated immediately behind the replication fork would compete for the transcription machinery at any time, leading to more focused polymerase-promoter interactions.

Replication and the cell cycle: a refresher

The circular *E. coli* genome is replicated bidirectionally by DNA polymerase complexes known as replisomes. Two replisomes simultaneously initiate DNA replication at a site in the genome known as oriC and travel down equal lengths of the two halves of the genome in opposite directions, at identical rates, to meet at a diametrically opposed region known as *terC*, all the while synthesising both strands of the DNA duplex along each replication arm[16]. One round of replication takes between 40 and 52 minutes[12]; the precise period depends on the conditions under which growth takes place, so it varies from culture to culture. However, the time taken to replicate the genome in any two cells growing in the same culture is remarkably constant[16]. Cell division usually occurs about 20 minutes after the end of replication. Thus, the time period between cell divisions (or the length of the cell cycle) equals the time period between initiations of replication[17]. Sometimes, under very good growth conditions, the time period between initiations of replication is shorter than the time taken to replicate the genome. At such times, one, three or even five pairs of replisomes may be observed to be replicating the genome simultaneously, such that two, four or six copies of the regions neighbouring *oriC* might coexist with a single copy of *terC*, inside the same cell. Under such conditions, the generation time of the culture turns out to be shorter than the time taken to replicate the genome, and genomes of fractional ploidy (larger than unity) are distributed to daughter cells[16]. Regardless, each gene is replicated only once during each cell cycle, since the length of the cell cycle equals the time period between initiations of replication. The timing of replication of a gene also occurs at a fixed point in time before cell division because the time taken to replicate the genome as well as the time passing between termination of replication and cell division are the same in all cells[16,17]. Thus replication-induced transcription could serve as a means of producing a more or less determinate amount of any protein at a fixed point in time during the cell cycle.

Lifting the steric block to transcription of buried LCNP genes

The circular *E. coli* chromosome is compacted more than a thousandfold *in vivo* into a structure known as the nucleoid[18,19]. It has been calculated that the density of DNA within the nucleoid is extremely high[20]; so high, in fact, that it would be virtually impossible for an RNA polymerase molecule or a ribosome from the cytoplasm to gain access to a 'buried' gene packaged deep within its



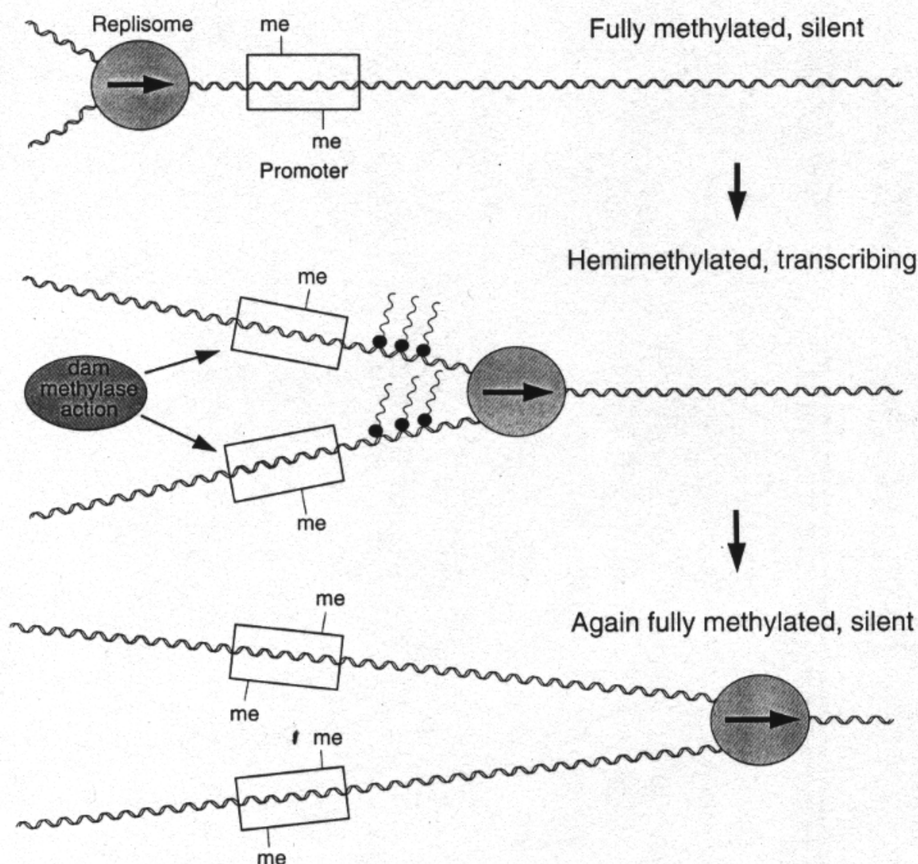
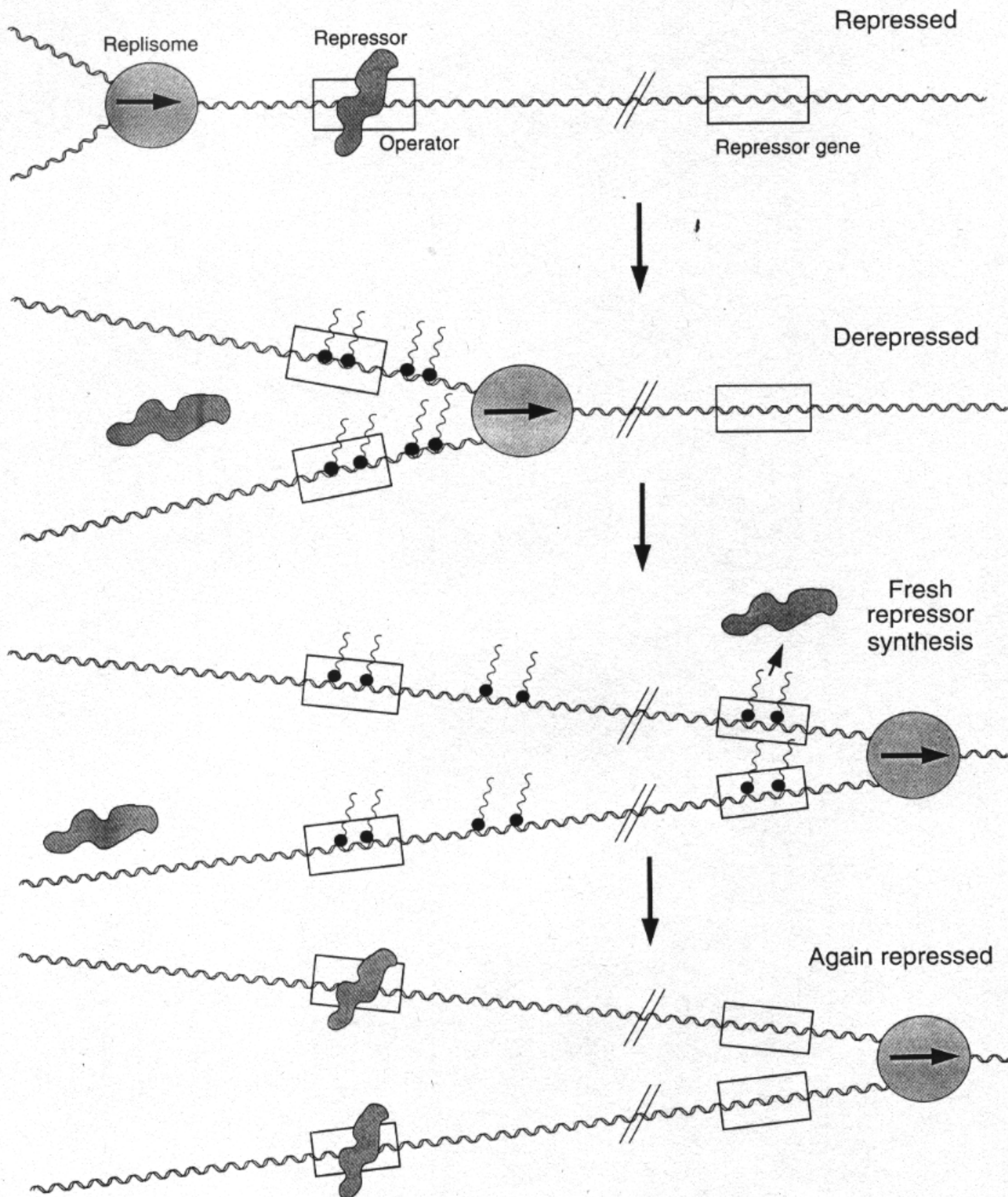


Fig. 1. (A) Negatively supercoiled DNA comprising the 1000-fold compacted, circular *E. coli* chromosome is shown schematically as a 'mother nucleoid' undergoing duplication by a replisome traversing one of its replication arms. Most of the positive supercoils introduced into downstream DNA by the replisome (shown to its right) are relaxed by the action of the topoisomerase, gyrase, concurrently with replication; some positive supercoils are also used for the relaxation of the negatively supercoiled DNA in the nucleoid so that it might unravel from its compacted state. Following replication (see the DNA on the left side of the replisome), negative supercoils need to be reintroduced by gyrase into the freshly generated daughter duplexes, so that these might coil up once more into



compact nucleoid structures at the same rate at which DNA replication occurs (since there is not enough space within the cell for relaxed DNA to be accommodated in, the DNA comprising daughter duplexes must tend to coil up immediately after replication). These negative supercoils, shown here in the form of both twist and writhe, could be utilised by RNA polymerase molecules to initiate and complete a few rounds of transcription from LCNP genes before such genes were to once more become buried within daughter nucleoids, to be opened up and exposed to the cytoplasm again only during the next period of replication. (B) The base A in the sequence GATC is methylated on both strands within a promoter; this keeps the gene transcriptionally silent. Replication results in the distribution of one methylated strand to each daughter duplex (the hemimethylated state), which can now support transcription for a definite period of time, i.e. until the action of dam methylase on the unmethylated strand of each hemimethylated duplex restores it to the fully methlyated, transcriptionally silent/compromised state. (C) The operator region of a gene is shown in its repressed state, with only as many repressor molecules available as could keep a single copy of the gene transcriptionally silent at all times. The repressor molecule itself is produced in very low copy numbers through RI transcription of its coding gene (located further downstream from the repressed gene). As shown, the passage of the replisome through the repressed gene results in the doubling of the copy number of operator DNA (with no accompanying increase in repressor availability), leading to titration of the available repressor amongst two operators, and hence to gene expression. The subsequent passage of the replisome through the gene encoding the repressor results in fresh synthesis of repressor and silencing of both copies.

interior[19,20], or for a DNA-bound RNA polymerase (or mRNA-bound ribosome) to effect any transcription or translation in the buried state. High-resolution autoradiographic techniques[19] show that nearly all transcription within the *E. coli* cell occurs in the cytoplasmic spaces surrounding the nucleoid and not within the nucleoid itself, suggesting that genes being expressed at any given time tend to be extruded out from the surface of the compacted chromosome (in the form of DNA loops) into less dense surroundings, where transcription and translation can occur. Furthermore, changes observed in the structure of the nucleoid upon artificially blocking transcription and/or translation[19,21] indicate that such genes probably recoil back onto the chromosomal surface during periods of inactivity, to lie dormant there until fresh reactivation of transcription causes the 'flaring out' of the relevant stretch of DNA into the cytoplasm once again[18,19]. If these observations accurately relect the transcriptional dynamics of chromosomal DNA, only those promoters/genes that are on the surface of the nucleoid at any given time would be able to support transcription.

Extending this line of thought further, all the strongly transcribed genes of the *E. coli* genome (as well as genes that are required to remain responsive to environmental stimuli) must at all times be located on the chromosomal surface. Given that there are about 3,000 genes in the *E. coli* genome[1], with the average gene being at least a kilobase in length, roughly three-quarters of the genome (length 4,400 kilobases) probably consist of coding DNA. It is unlikely that there is room enough on the chromosomal surface to accomodate the regulatory elements and/or coding regions of every one of these genes, and many weakly expressed genes must be buried in the interior of the nucleoid during most of the cell cycle. Such buried genes would, however, be briefly exposed to the cytoplasm by the act of replication, once during each cell cycle. Such freshly replicated genes would remain exposed until the moment of their burial within the nucleoids of daughter chromosomes, and could support transcription during these brief moments of their exposure to the cytoplasm.

Possible mechanisms of induction of transcription

Three possible mechanisms are summarised in Fig. 1.

Induction by negative supercoiling

The *E. coli* chromosome exists in a mildly negatively supercoiled (underwound) state[22]. It has been clear for many years that the two topoisomerases of mutually opposing action, gyrase and topoisomerase I, homeostatically regulate the state of chromosomal supercoiling. The genome can probably compact itself into a densely packed object at least in part because of its negatively supercoiled character (since bacterial DNA exists in the form of supercoiled domains that can be relaxed and uncompacted through single-strand nicks[23]). Thus, for the chromosome to coil up into the numerous supercoiled domains comprising the nucleoid, the DNA comprising each domain may need to be negatively supercoiled by the action of a topoisomerase. Since much of the genome at any given time is probably sterically inaccessible to the topoisomerases, it is likely that only those regions of each domain that are being transcribed or replicated (or are otherwise present at the surface of the chromosome), can be efficiently modified topologically.

In order for a region of the nucleoid comprising a supercoiled domain to be replicated and partitioned into daughter genomes, negative supercoils would have to be reintroduced into the freshly generated (relaxed) duplexes very soon after replication, to allow these to coil up independently once again into compact supercoiled domains in two separate chromosomes. This would require the action of gyrase, the topoisomerase thought to be capable of introducing negative supercoils into DNA (or relaxing positive supercoils, which as Liu and Wang[24] pointed out many years ago, amounts to the same thing as introducing negative supercoils). Gyrase is known to act in the vicinity of the replication fork; in fact, we know that it must act on the DNA ahead of the fork in order to resolve the positive supercoils generated by the motion of the replisome[1,24]. The argument outlined above suggests that gyrase must also act immediately behind the replication fork, in order to introduce negative supercoils into freshly replicated DNA so that it may fold (Fig. 1A).

Are such freshly introduced supercoils likely to be used only for compaction, or could they be used to facilitate other phenomena? For DNA to coil up sufficiently to become completely buried/inaccessible to transcription machinery, it would need to accumulate a certain degree of negative superhelical tension. Until such a degree of supercoiling were realised, freshly introduced negative supercoils could facilitate the occurrence of a few rounds of transcription from weak promoters exposed to the cytoplasm, much in the way that negative supercoiling facilitates transcription[25], i.e. by aiding the recognition of the promoter by RNA polymerase[26], and/or by reducing the energy requirements for promoter melting by the polymerase[25]. It is possible that nascent, incomplete transcripts (with bound RNA polymerase, ribosomes and partially synthesised polypeptide chains) would become coiled up and buried within the nucleoid during the steady introduction of negative supercoils by gyrase, only to be exposed to the cytoplasm again during the next round of replication. Recent work has shown that the replisome is able to pass stalled RNA polymerase complexes without affecting the processivity of such complexes on the daughter duplex containing the coding strand[27], so it is likely that bound transcription elongation complexes would resume transcription once more, upon being re-exposed.

Such 'resumed' transcription could also provide some of

the impetus for the action of gyrase behind the replication fork, through the operation of a feedback loop linking supercoiling and transcription, as discussed below. It is now known that the positive supercoils generated ahead of a moving RNA polymerase need to be relaxed by gyrase action (when transcription occurs on topologically constrained DNA)[28], and that negative supercoils accumulate in DNA that is preferentially acted upon by gyrase, rather than by topoisomerase I. Freshly replicated DNA is topologically constrained by folded structures formed by previously replicated DNA located further upstream; moreover, it must be acted upon preferentially by gyrase, if it accumulates negative supercoils and folds into compact structures. Thus, the positive supercoils generated in the DNA ahead of a resumed transcription event could initially attract gyrase action on freshly replicated DNA. Combining this with the idea that replication-induced transcription occurs as a result of gyrase action on newly replicated DNA, I propose that the transcription resumed in each cycle by re-exposed elongation complexes (and any rounds of transcription that follow during that cycle), are not only aided by gyrase action behind the replication fork but also partly responsible for gyrase action. It is now increasingly appreciated that supercoiling and transcription share a bidirectional cause-effect relationship that can operate in either direction. This relationship may be used to generate a feedback cause-effect loop that amplifies the accumulation of negative supercoils in freshly replicated DNA to effect DNA compaction, as well as to facilitate some rounds of transcription from LCNP genes during such compaction. (Note that such a loop could not operate in an amplifying manner on strongly expressed genes that are constantly transcribed by batteries of RNA polymerases, since the relaxation of each other's transcription-driven supercoils by successive polymerases in a battery would limit topoisomerase action on transcriptionally active DNA to a minimum).

The burial of one or more RNA polymerase elongation complexes, during RI transcription, at each of 1,700 or more sites in the genome would also explain (1) the observed association of much more mRNA with chromosomes[29] than can be explained by extant levels of transcription occurring at any given time, and (2) the fact that only 15-30% of the RNA polymerase population within cells growing under any sort of conditions, is found to be actively engaged in mRNA synthesis at any given time, though the remaining polymerases appear to be strongly associated with DNA[30].

Induction through hemimethylation

Methylation is known to regulate the initiation of transcription from promoters containing the sequence GATC[31], which functions as a substrate for *E. coli* DNA-adenine methylase (dam). For transcription to be inhibited, the adenine in GATC is usually required to be methylated on both strands. Immediately after replication, only one strand in each daughter duplex is methylated; the other strand gets methylated by the dam methylase soon after replication (Fig. 1B). Transcription can, in principle, be supported by such a 'hemimethylated' promoter, during the short time window between replication and methylation of the newly synthesised strand on both daughter duplexes, and indeed several genes – notably the ones coding for the trpR repressor and the transposase of the Tn10 transposon in *E. coli* – have been shown to transcribe immediately after replication, through such a mechanism (reported and reviewed in refs 31-35).

Induction through transient derepression

As already discussed, considerations of economy preclude large numbers of LCNP genes from being regulated by specific protein factors. However, it is not possible to rule out the 'odd' case of a strong inducible gene behaving like an LCNP gene under non-inducing (repressed) conditions. In such cases, if the repressor molecule were knocked off the template by the replication machinery, and there were not enough functional copies of it in the cell to effectively and immediately repress both copies of the replicated gene in the two daughter duplexes, it is possible that the gene would be transcribed for a brief while before it was repressed again (Fig. 1C). Such a mechanism could operate in the case of genes which, though normally maintained at the surface of the genome, happen to be only rarely required, especially inducible genes, whose products are involved in the positive self-regulation of their own synthesis. Proteins made by such genes would need to be present in a small number of copies in the cell at all times, in order to remain capable of responding to inducing signals; if small quantities of these proteins were not made afresh during most cycles of growth under repressed conditions, they would be diluted out during successive cell divisions. Under conditions of growth without induction, therefore, these small quantities could be synthesised through replication-linked synthesis – achieved through transient derepression in every cell cycle. A suggestion along slightly similar lines was made over fifteen years ago, to explain certain aspects of timed protein synthesis in eukaryotes[36].

Determinate transcriptional impedance to replisome movement

Since replication and transcription are not temporally separated in prokaryotes, a DNA polymerase complex (replisome) replicating an arm of the *E. coli* genome is likely to encounter many active RNA polymerase molecules transcribing DNA in its path. Recent studies have shown[27,37,38] that the DNA polymerase is able to pass the RNA polymerase after a slight pause without displacing it from the template strand, when transcription in the direction of motion of the replisome is encountered. When transcription is encountered occurring in the opposite direction, it seems that the DNA polymerase is slowed down considerably, and

perhaps halted for a few seconds[37,39], before the encounter is resolved through a switching of preference (by the RNA polymerase) of the DNA duplex providing the template strand used for transcription[39]. In either case, given that the replisome normally processes DNA at a rate of 750-1,000 base pairs/second[12] and takes a fixed length of time to finish replicating one arm of the *E. coli* genome, the total number of transcriptionally active genes encountered is likely to influence the C period, or the total time taken by a pair of replisomes to replicate the genome.

Available data on the distribution of replication and generation times of *E. coli* cells in culture show that although the C period can vary significantly between cultures (from 40 to 52 minutes[12], depending on the precise growth conditions used), it is in fact remarkably constant for every cell growing in the same culture[17]. (It is this constancy of the C period in all cells growing under identical conditions that allows *E. coli* to be grown in synchronously dividing cultures for several generations). The C period could only be a constant amongst cells if replisome pairs were to encounter the same quantum of transcriptional impedance in all cells in a given culture. Whereas strongly transcribed genes of both the constitutive and regulated types pose no problem in this regard, the weakly transcribed LCNP genes would pose a considerable problem if such genes were only allowed to initiate transcription stochastically at low frequencies, since there would be no control on how many such genes might transcribe in the path of the replisome within any cell. Furthermore, weakly transcribed genes appear to be oriented at random with respect to direction of passage of the replication fork in both the replication arms of the chromosome[1], suggesting that as much as half the population of weakly transcribed genes might be oriented to oppose replisome movement. Without tight control being exerted on the timing of initiation of transcription of each of these many hundred LCNP genes in every cell, it would be virtually impossible to engineer a constancy of the C period; for instance, the encountering of a mere 80-100 more oppositely oriented transcription units in one cell as compared to those encountered in a neighbouring cell could cause a difference of as much as five minutes in the C period of two cells growing in the same culture.

Since variations of this kind in the length of the C period are not observed, we are led to conclude that control is exerted on the timing of transcription of weakly transcribed genes, and the simplest way to achieve this is to link transcription of weak genes to replication. Notably, such a linkage to replication would also solve the problem of determinate impedance to replisome movement in a very simple way: if all LCNP genes were to only transcribe as an induced response to the passage of the replication fork, the replisome would never find an LCNP gene transcribing in its path, and the C period would be determined essentially by the predictable quantum of strong, constitutive transcription encountered, which would be occurring on genes coding for the abundant proteins. It would thus remain a constant.

A functional correlate for global conservation of gene order

The absolute positions and precise order of placement of genes in the *E. coli* chromosome would appear to have some as yet unknown functional correlate, since gene order is conserved very strongly amongst the chromosomes of enteric bacteria[40], despite their capacity for extensive reshuffling by recombinational mechanisms[41]. We know that the chromosomal position of an *E. coli* gene determines the timing of its replication. In this connection, it is pertinent to note that Chandler and Pritchard[42] pointed out, almost two decades ago, that early replication timing would ensure the early availability of a larger number of templates from which transcription could occur. Thus, in the course of evolution, genes coding for proteins which are required in higher amounts (or, for a longer fraction of the cell's lifetime), could indeed have found a location closer to the chromosome's origin of replication than other genes. But while this explanation, known as the 'gene dosage hypothesis', appears to some extent to justify the positions of strongly transcribed genes, it does not provide us with a functional correlate for the conservation of position and order of occurrence of the LCNP genes. As we have already noted, replication of the *E. coli* genome takes between 40 and 52 minutes, depending on the conditions under which growth occurs. An LCNP gene which is presumed to yield, say, one transcript every 20 minutes (assuming stochastic initiation of transcription), would not be very strongly benefited or disadvantaged by obtaining (or not obtaining) two copies of itself at any one particular point of time in the cell cycle, as opposed to obtaining them at another point in the cycle, 5-10 minutes further away in time. Thus the gene dosage hypothesis cannot satisfactorily explain the positioning and ordering of the LCNP genes. Since these genes constitute over 80% of the genes in the *E. coli* chromosome, any explanation for the conservation of gene order must incorporate a plausible explanation for the conservation of LCNP gene positions on the chromosome.

So far we have considered the possibility that the expression of LCNP genes might be linked to the timing of their replication. If the cell were to require the products of LCNP genes at particular points of time during the cell cycle and/or in a particular order (or have simply grown used to their being produced in the cell, in that order!), the replication-induced transcription of the LCNP genes could indeed provide a functional correlate for the conservation of gene order. In other words, the order of placement of LCNP genes may be conserved because that is the order in which these genes express.

Temporal control of gene expression

The timed synthesis of gene repressors (and activators) in small numbers could potentially initiate chains of synthetic events involving other timed syntheses. Alternatively, timed

synthesis of regulatory factors could facilitate the passage of the replication fork through heavily transcribed, oppositely oriented operons. The *his* operon, for instance, is oriented in a direction opposite to its replication[1,43]. If the repressor of the *his* operon were synthesised somewhat in advance of the time of arrival of the replication fork at the operon, it could serve to limit transcription of the operon sufficiently to facilitate the passage of the fork through it, by appearing in the cell at twice the concentration normally needed to allow the usual attenuated transcription to occur. Gene duplication could then reduce the repressor available per copy of the gene to normal levels, to reinitiate transcription at steady-state levels. There is evidence to link the synthesis of several well-studied repressor molecules, notably the lac repressor (produced at copy numbers of 10 to 30 molecules per cell), the tryptophan operon repressor and the lambda repressor to the cell cycle[34,44,45].

For proteins performing morphogenetic tasks, or initiating chromosomal DNA replication, the need for timed synthesis of small quantities of specific molecules is too obvious to require reiteration here, and the reader is referred to the existing literature on the subject[17,46]. While it is tempting to envisage requirements of the timed synthesis of other LCN proteins during the course of the cell cycle, too little is known about other functions of the LCN proteins to warrant further speculation at this time.

Replication-induced transcription – an ambiguous history

A search of the literature shows that the idea of a linkage between replication and transcription is by no means new. In the 1960s, the synthesis of many proteins was shown to increase abruptly once in every cell cycle, in synchronously growing cultures of various bacteria and yeast[17,45-51], at a point of time corresponding approximately to the moment of replication of the coding gene. Such proteins were subclassified into 'step' and 'peak' proteins, on the basis of the shapes of the graphs plotting their synthesis *versus* time[49]. 'Step' proteins could be reasonably well explained by the doubling of gene copy number upon replication, but 'peak' proteins could not be satisfactorily explained in this manner.

It was then envisaged[51] (incorrectly, with hindsight) that RNA polymerase follows DNA polymerase with a slight lag, synthesising messenger RNA continuously in the form of one long strand and allowing it to be translated (and presumably degraded as well, at some steady rate, at its lagging end). This model suggested that all genes, without exception, are expressed in the order in which they are replicated, i.e. the order in which they occur on the chromosome. The basic tenets of this thesis were quickly brought into question, however, by a number of factors, notably: (1) the growing acceptance of the operon hypothesis as a much more plausible model of gene expression; (2) the discovery of promoter and terminator sequences flanking most coding DNA; (3) the realisation that oscillations of varying frequencies in the abundances of enzymes with time, could result from factors (e.g. feedback-linked repression) other than a simple linkage of gene expression and replication; and (4) the demonstration that synthesis of most proteins exhibiting periodic expression could also be induced at virtually any time during the cell cycle. Perhaps as a result, no further serious consideration was given to replication-linked transcription in the late 1960s, not even to the possibility that RNA might instead be discontinuously synthesised, by polymerases initiating transcription behind the replication fork, through the activation of individual promoters by some aspect of the replication process. (Such an attractive possibility appears to have been put forward – to explain how oscillations in the abundances of enzymes could be entrained by bursts of mRNA synthesis accompanying replication[52], so that the periodicities of enzyme oscillations could match the period of the cell cycle. However, despite the fact that this aspect of the periodicity of timed synthesis is still not understood[53], such a proposal did not gain wide acceptance at the time). In the context of this paper, all this is mostly of historical significance only, since every instance of such periodic transcription turned out to relate to abundant proteins that were later demonstrated to be inducible throughout the cell cycle.

The operon model initially tended to brush aside most proposals of gene regulation that did not invoke the negative modulation of transcription by repressor activity[54]. However, as positive regulation of transcription steadily gained in popularity, it became evident that there was no rigorous argument for rejecting the possible activation of transcription by some aspect of replication, even though no good guess could be hazarded at the time as to the mechanism of such activation. Following the development of two-dimensional gel techniques[2], an effort was mounted[55] to examine variations in the abundances of individual proteins with time as a function of the cell cycle, with the specific intention of evaluating the possible influence of replication on the synthesis of important (at that time, equated with abundant) proteins. Unfortunately, no definite evidence was obtained for any periodic synthesis, but the investigators emphasised that only about 700 proteins were resolved on the two-dimensional gels used by them, at a time when there seemed to be evidence for the existence of mRNA coding for as many as 2300 proteins. The lower limits of detection in the experiments performed were reckoned to lie at 1,000 molecules for a 20 kDa protein, or 80 molecules for an 80 kDa protein, indicating that a potentially large number of proteins were being missed. None of the proteins present at high copy numbers in the cell appeared to display any variations in abundance with the progression of the cell cycle, but the same could not be definitively said about the proteins present at low copy numbers, such as the lac repressor. Indeed, the investigators suggested that poorly abundant proteins might be synthesised at specific points of

time during the cell-cycle but remain undetected. It appears, therefore, that to date the only documented efforts ever made to examine variations in the general abundance of *E. coli* proteins during the course of the cell cycle, focused only on abundant proteins, because rare proteins were believed to be peripheral to the survival and growth of the *E. coli* cell at the time. Such efforts do not appear to have been repeated recently using more sophisticated methods.

However, cell-cycle linked expression, mediated by replication, has recently been proposed as a mechanism of biosynthesis of several (LCN) proteins, including the lac repressor and various transposases[34,44,45], on the basis of experimental evidence for periodic transcription. A mechanism for replicational induction has also emerged in the form of alterations in the state of methylation of promoters by replication (already discussed). Thus, the possible occurrence of RI transcription is gaining acceptance, and at least one mechanism linking transcription to replication has been proposed. Two additional mechanisms, involving derepression and transient exposure of a normally buried gene to the cellular milieu, accompanied by gyrase-mediated negative supercoiling prior to repackaging and burial, are presented in this paper for the first time.

Predictions and experimental tests

A number of tests for RI transcription spring to mind from the foregoing discussions. Five relatively straightforward tests are described below.

(1) Blocking of replication should result in the halting of all LCN protein synthesis. However, synthesis of many constitutively produced proteins could be expected to continue. Cessation of synthesis of LCN proteins should be observable on two-dimensional gels in terms of a gradual reduction in the relative abundances of LCN proteins with time (measured as a fraction of total cellular protein).

(2) Changing the chromosomal location of a gene coding for an important 'timed' cellular event should result in a corresponding change in the timing of the event. I have implied that some genes responsible for cell division, such as the gene coding for the protein PBP2, may be synthesised through RI transcription (please see Introduction). The relocation of such a gene (or genes) should result in observable effects on the nature or timing of cell division in synchronous culture.

(3) Cellular abundances of LCN proteins should vary with time. As mentioned earlier, an effort was made to examine the abundances of cellular proteins as a function of time, in synchronous cultures. Since only about 700 proteins were resolved and the abundances of only 30 proteins were quantitated densitometrically (the rest were estimated by eye)[55], nothing could be said definitively about cell cycle-linked synthesis of poorly abundant proteins. Today, over 2,000 proteins can be resolved on two-dimensional gels. Repeating such an experiment with synchronous cultures might now be expected to reveal timed synthesis of LCN proteins.

(4) Blocking of transcription should result in the failure of the chromosome to form a nucleoid. It has been suggested in this paper that gyrase action accompanying transcription of LCNP genes aids in the compaction of the chromosome into the supercoiled domains forming the nucleoid. If replication is allowed to continue, but transcription is blocked, gyrase might be expected to be inhibited too. Thus electron microscopic studies should permit the observation of large amounts of relaxed DNA in cells in which transcription has not been allowed to occur for a substantial length of time.

(5) Northern hybridization studies performed as a function of time on total cellular RNA from synchronously grown bacterial cultures, using the DNA corresponding to individual LCNP genes as probes, should reveal the timed synthesis of transcripts from individual LCNP genes.

Acknowledgements

I would like to thank Drs J. Gowrishankar, D. Chatterji, D. Manna and R. Mishra and D. Balasubramanian of C.C.M.B., Hyderabad, India; H. Sharatchandra of I.I.Sc., Bangalore, India; S. Adhya of N.I.H., Bethesda, USA; A.S. Wilkins and D. Hoskin, Company of Biologists; and R.G. Solomon of the Dept of Biochemistry, University of Cambridge, for most useful discussions and comments.

Note added in proof

I thank Dr B. Liu, U.C., San Francisco, for bringing to my notice the work of E.P. Geiduschek and colleagues, U.C., San Diego, who have suggested that the replisome can act as a mobile enhancer of transcription of late T4 phage genes[56,57]. An accessory protein of the replisome during T4 replication, gp45 (the product of gene-45 of the phage genome) functions as a 'sliding clamp', attaching itself to regions of the DNA duplex generated from the discontinuously replicated (lagging) strand, in order to activate transcription of freshly replicated genes. Since the beta subunit of *E. coli* DNA pol III is known to be a structural/functional homolog of gp45[57], such 'sliding clamp' mediated transcriptional activation could turn out to constitute a fourth possible mechanism of RI transcription.

References

1 **Brewer, B.J.** (1988). When polymerases collide: Replication and the transcriptional organisation of the *E. coli* chromosome. *Cell* **53**, 679-686.
2 **O'Farrell, P.H.** (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021.
3 **Phillips, T.A., Vaughn, V., Bloch, P.L. and Neidhardt, F.C.** (1987). Gene-Protein index of *Escherichia coli* K-12, edition 2. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 919-966. Washington, DC: American Society for Microbiology.
4 **Pedersen, S., Bloch, P.L., Reeh, S. and Neidhardt, F.C.** (1978). Patterns of protein synthesis in *E.coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179-190.

5 Gilbert, W. and Muller-Hill, B. (1966). Isolation of the lac repressor. *Proc. Natl Acad. Sci .USA* **56**. 1891-1898.
6 Spratt, B.G. (1977). Properties of the penicillin binding proteins of *E. coli* K-12. *Eur. J. Biochem.* **72**, 341-349.
7 Rowen, L. and Kornberg, A. (1978). Primase, the DnaG protein of *E.coli*. *J. Biol. Chem.* **253**, 758-764.
8 Raleigh, E.A. and Kleckner, N. (1984) Multiple IS10 rearrangements in *E. coli*. *J. Mol. Biol.* **173**, 437-461.
9 Sancar, G.B., Smith, F.W., Lorence, M.C., Rupert, S. and Sancar, A. (1984). Sequences of the *E. coli* photolyase gene and protein. *J. Biol. Chem.* **259**, 60330-66038.
10 Maaloe, O. (1979). Regulation of the protein synthesing machinery – ribosomes, tRNA, factors and so on. In *Biological Regulation of Development*, vol. 1. (ed. R. F. Goldberger), pp. 487-541. Plenum Publishing Corporation.
11 Sompayrac, L. and Maaloe, O. (1973). Autorepressor model for the control of DNA replication. *Nature* (New Biol.) **241**. 133-135.
12 Bremer, H. and Dennis, P.P. (1987). Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 1527-1542. Washington, DC: American Society for Microbiology.
13 Bremmer, H. and Yuan, D. (1968). RNA chain growth rates in *E. coli*. *J. Mol. Biol.* **38**, 163-180.
14 Travers, A.A. (1987). Structure and function of *E. coli* promoter DNA. *CRC Crit. Rev. Biochem.* **22**, 181-219.
15 Cooper, S. and Helmstetter, C.E. (1968). Chromosome replication and the division cycle of *E. coli* B/r. *J. Mol. Biol.* **31**, 519-540.
16 Donachie, W.D. and Robinson, A.C. (1987). Cell division: Parameter values and the process. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 1578-1593. Washington, DC: American Society for Microbiology.
17 Donachie, W.D. (1979). The cell cycle of *Escherichia coli*. In *Developmental Biology of Prokaryotes* (ed. J. H. Parish), pp. 11-35. Blackwell Scientific Publications, Oxford.
18 Drlica, K. (1987). The Nucleoid. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 91-103. Washington, DC: American Society for Microbiology.
19 Ryter, A. and Chang, A. (1975). Localization of transcribing genes in the bacterial cell by means of high resolution autoradiography. *J. Mol. Biol.* **98**, 797-810.
20. Maaloe, O. and Kjeldgaard, N.O. (1966). The bacterial nucleus. In *Control of Macromolecular Synthesis* (ed. B. Davis). pp. 188-197. W.A. Benjamin Inc. NY.
21 Bjornsti, M.A., Hobot, J., Kelus, A., Villiger, W. and Kellenburger, E. (1985). New electron microscopy data on the structure of the nucleoid, and their functional consequences. In *Bacterial Chromatin* (ed. C. Gualerzi), pp. ?-?. Springer-Verlag, New York.
22 Worcel, A. and Burgi, E. (1972). On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* **71**, 127-147.
23 Sinden, R.R. and Pettijohn, D.E. (1981). Chromosomes of living *E. coli* cells are segregated into domains of supercoiling. *Proc. Natl Acad. Sci. USA* **78**, 224-288.
24 Liu, L.F. and Wang, J.C. (1987). Supercoiling of the DNA template during RNA transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024-7027.
25 Pruss, G.J. and Drlica, K. (1989). DNA supercoiling and prokaryotic transcription. *Cell* **56**, 521-523.
26 Wang, J.Y. and Syvanen, M. (1992). DNA twist as a transcriptional sensor for environmental changes. *Mol. Microbiol.* **6**, 1861-1866.
27 Liu, B., Wong, M.L., Tinker, R.L., Geiduschek, E.P. and Alberts, B.M. (1993). The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. *Nature* **366**, 33-39.
28 Tsao, Y-P., Wu, H-Y. and Liu, L.F. (1989). Transcription-driven supercoiling of DNA: Direct biochemical evidence from *in vitro* studies. *Cell* **56**, 111-118.
29 Pettijohn, D.E. and Hecht, R.M. (1973). RNA molecules bound to the folded bacterial genome stabilize DNA folds and segregate domains of supercoiling. *CSH Symp. Quant. Biol.* **38**, 31-41.
30 Shepherd, N.S., Churchward, G. and Bremmer, H. (1980). Synthesis and activity of ribonucleic acid polymerase in *Escherichia coli*. *J. Bacteriol.* **141**, 1098
31 Roberts, D., Hoopes, B.C., McClure, W. and Kleckner, N. (1985). IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**, 117-130.
32 Sternberg, N. (1985). Evidence that adenine methylation influences DNA-protein interactions in *Escherichia coli*. *J. Bacteriol.* **164**, 490-493.
33 Marinus, M.G. (1987). Methylation of DNA. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 697-702. Washington, DC: American Society for Microbiology.
34 Hoopes, B.C. and McClure, W.M. (1987). Strategies in regulation of transcription initiation. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 1231-1240. Washington, DC: American Society for Microbiology.
35 Messer, W. and Weidner, M.N. (1988). Timing and Targeting: The Biological Functions of Dam Methylation in *E. coli*. *Cell* **54**, 735 -737.
36 Tyson, J. and Sachsenmaier, W. (1979). Derepression as a model for control of DNA-division cycle in eukaryotes. *J. Theor. Biol.* **79**, 275-280.
37 French, S. (1992). Consequences of replication fork movement through transcription units *in vivo*. *Science* **258**, 1362-1365.
38 Liu, B., Wang, M.L. and Alberts, B. (1994). A transcribing RNA polymerase molecule survives DNA replication without aborting its growing RNA chain. *Proc. Natl Acad. Sci. USA* **91**, 10660-10664.
39 Liu, B. and Alberts, B. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* **267**, 1131-1137.
40 Riley, M. and Krawiec, S. (1987). Genome organisation. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 967-987. Washington, DC: American Society for Microbiology.
41 Anderson, R.P. and Roth, J.R. (1977). Tandem Genetic Duplications in Phage and Bacteria. *Ann. Rev. Microbiol.* **31**, 473-506.
42 Chandler, M.G. and Pritchard, R.H. (1975). The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol. Gen. Genet.* **138**, 127-135.
43 Bachmann, B.J. (1987). Linkage map of *Escherichia coli* K-12, edition 7. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 807-876. Washington, DC: American Society for Microbiology.
44 Halling, P.J. (1989). Do the laws of chemistry apply to living cells? *Trends Biochem. Sci.* **14**, 317-318.
45 Edelmann, P.L. and Edlin, G. (1974). Regulation of synthesis of the lactose repressor. *J. Bacteriol.* **120**, 657-665.
46 Donachie, W.D., Begg, K.G. and Sullivan, N.F. (1984). Morphogenes of *Escherichia coli*. In *Microbial Development* (ed. R. Losick and L. Shapiro), pp. 27-62. CSH Monograph Series.
47 Masters, M. and Pardee, A.B. (1965). Sequence of enzyme synthesis and gene replication during the cell cycle of *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **54**, 64- 70.
48 Kuempel, P.L., Masters, M. and Pardee, A.B. (1965) Bursts of enzyme synthesis in the bacterial duplication cycle. *Biochem. Biophys. Res. Commun.* **18**, 858-867.
49 Mitchison, J.M. (1971). Enzyme synthesis. In *The Biology of the Cell Cycle*, pp. 159-180. Cambridge University Press.
50 Donachie, W.D. and Masters, M. (1969). Temporal control of gene expression in bacteria. In *The Cell Cycle* (ed. G. M. Padilla *et al.*), pp 37-60. Academic Press, New York.
51 Tauro, P. and Halvorson, H.O. (1966). Effect of gene position on the timing of enzyme synthesis in synchronous cultures of yeast. *J. Bacteriol.* **92**, 652-661.
52 Goodwin, B.C. (1966). An entrainment model for timed enzyme synthesis in bacteria. *Nature* **209**, 479-481.
53 Tyson, J.J. (1983). Periodic enzyme synthesis and oscillatory repression: Why is the period of oscillation close to the cell cycle time? *J. Theor. Biol.* **103**, 313-328.
54 Beckwith, J. (1987). The operon: A historical account. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. F. C. Neidhardt *et al.*), pp. 1439-1443. Washington, DC: American Society for Microbiology.
55 Lutkenhaus, J.F., Moore, B.A., Masters, M. and Donachie, W.D. (1979). Individual proteins are synthesized continuously throughout the *Escherichia coli* cell cycle. *J. Bacteriol.* **138**, 352-360.
56 Herendeen, D.R., Kassavetis, G.A., Barry, J., Alberts, B.M. and Geiduschek, E.P. (1989). Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. *Science* **245**, 952-958.
57 Brody, E.N., Kassavetis, G.A., Ouhammouch, M., Sanders, G.M., Tinker, R.L. and Geiduschek, E.P. (1995). Old phage, new insights: Two recently recognized mechanisms of transcriptional regulation in bacteriophage T4 development. *FEMS Microbiol. Letters* **128**, 1-8.

Purnananda Guptasarma was at the Centre for Cellular and Molecular Biology, Hyderabad-7, AP, India. His present address is the Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK.