RESEARCH ARTICLE

Hypothesis: Local variations in the speed of individual DNA replication forks determine the phenotype of daughter cells

Authors

Vic Norris[1], Francois Kepes[2], Patrick Amar[3], Ina Koch[4] and Laurent Janniere[2]

Affiliations:

[1] EA 4312, Department of Science, University of Rouen, 76821 Mont Saint Aignan Cedex, France

[2]CEA, CNRS, Univ EVRY, Université Paris-Saclay, Génomique Métabolique (UMR8030), Laboratoire iSSB, Génopole Campus 1, Genavenir 6, 5 rue Henri Desbruères F-91030 Evry, France"

[3] Laboratoire de Recherches en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, 91405 Orsay, Cedex, France

[4] Molecular Bioinformatics, Goethe University, Frankfurt, Germany

Correspondence

Vic Norris

E-mail: victor.norris@univ-rouen.fr

Abstract

How a cell coordinates its thousands of different constituents to achieve coherent – but different – phenotypes is far from fully understood. It is clear though that daughter cells with different phenotypes can be generated by the cell cycle, which comprises the events of chromosome replication, chromosome segregation and cell division. In line with this, recent experiments are consistent with an intimate relationship in bacteria between the speed of chromosome replication at a fork(s) and metabolism. The process of chromosome replication progressively changes the copy number of genes and sites in a linear order. This raises the possibility that speeding up or slowing down or even pausing replication for different times at different sites in the chromosome might be combined with various mechanisms leading to local cooperation (for example, the transcription of a gene leading to more transcription of that gene) and to global competition (for example, a gene having to compete with all the other genes for the transcriptional apparatus). If so, such replication-phenotype coupling could produce different patterns of gene expression and metabolic activity. Indeed, replication-phenotype coupling may constitute a powerful and fundamental way of generating coherent phenotypes, that is, phenotypes in which the cell's constituents perform compatible functions (rather than, for example, trying to maintain growth and to shut down growth simultaneously). In this hypothesis, such coupling would involve the dynamics of the spatially extended assemblies of molecules and macromolecules termed 'hyperstructures'. As a prelude to testing this hypothesis, we discuss some of the parameters that will need to be explored by bench experimentation and computer simulation.

1. Introduction

One of the fundamental problems in biology, highlighted by Kauffman (1), is how cells integrate gene expression and environmental conditions to steer their phenotypes in a coherent, reproducible way through the vast space of possibilities apparently available to them. A possible solution is that the phenotype is decided not at the level of thousands of genes but at the level of scores of 'hyperstructures', which are large, spatially extended assemblies of ions, molecules and macromolecules, which are implicated in functions that range from DNA replication and cell division to chemotaxis and secretion, and which are sometimes 'functioning-dependent structures' (FDSs) that only form when their constituents start to function as happens, for example, when an enzyme catalysing a reaction as a result binds to another enzyme in the same metabolic pathway (2-6). This candidate solution requires updating due to the discovery that carbon metabolism in *Bacillus subtilis*, and almost certainly other bacteria, dynamically affects the initiation and elongation steps of chromosome replication, probably by changing the activity of the enzymes responsible for these steps (L. Janniere, unpublished data). In other words, metabolism appears to be exerting a direct control over the way the chromosome is replicated. This suggests to us a reciprocal relationship in which the way the chromosome is replicated also determines the phenotype. Here we explore this idea and, because chromosome replication is part of the cell cycle, also consider whether changes in the speed of replication by individual forks in particular places might result in daughter cells with different, coherent patterns of metabolic activity.

2. Hypothesis

By slowing or accelerating the elongation step of DNA replication in different regions of the genome, a bacterium generates spatio-temporally different patterns of different genetic elements and correspondingly different patterns of hyperstructures; this results in coherent, metabolically different phenotypes. This differentiation occurs via several mechanisms that are characterised by local synergistic relationships and global competitive ones.

3. Mechanisms

Sources of local positive feedback are based on *cooperation* and include:

1. the movement of genes during their transcription by RNA polymerase from the nucleoid (where they are relatively inaccessible to RNA polymerase) to a position on the periphery of the nucleoid (where they are easily accessible).
2. the bringing together of different genes or sites on nucleic acids into a hyperstructure by various factors. These factors include (i) protein binding to sites on RNA and on DNA (where the sites may be distant from one another in terms of their position on the chromosome and hence be replicated at very different times, see Corollaries) that then increases the probability of more of these proteins binding (e.g. by raising their local concentration) and (ii) chromosome folding that brings related genes and their products closer to one another.
3. the bringing together of the enzymes in a particular pathway into a 'functioning-dependent', enzymatic hyperstructure due to the affinities of the enzymes for one another that result from the catalysis of their cognate reactions.
4. nucleation phenomena such that, once a critical size has been reached, subsequent

assembly is faster (as observed in the polymerisation of eukaryotic actin *in vitro*).
5. hyperstructure-assisted hyperstructure formation whereby the formation of one hyperstructure assists the formation of other, phenotypically related, hyperstructures by the binding of constituents of the different hyperstructures to one another – as when a transcription-translation hyperstructure encoding the enzymes in a pathway assists the formation next to it of an enzymatic hyperstructure containing these enzymes – and by favouring particular patterns of folding of the chromosome.

Sources of global negative regulation are based on *competition* and include the limitations on the:

1. quantities of the transcriptional and translational machinery.
2. physical space within the cytoplasm and membrane.

Local positive feedback and global negative regulation can act *via* either activation or repression of gene expression and *via* metabolite-induced or inhibited assembly of functioning-dependent hyperstructures (as well as *via* other processes such as degradation, various modifications to macromolecules).

3.1. Activation mechanisms

Consider an activator of transcription that has two types of binding site, low affinity and high affinity (empty and filled circles, respectively) distributed as shown in Figure 1.

Figure 1

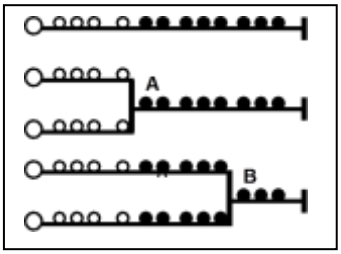

Fig. 1 Transcriptional activation via binding sites. The large circle represents the origin of replication of the chromosome and the bar represents the terminus. The small empty circles are low affinity binding sites for transcriptional activators whilst the filled circles are high affinity binding sites. **A** and **B** represent two rest-stops for replication.

This activator can form oligomers such that there is the possibility of *in cis* interactions between the activators binding to the four sites on the top daughter chromosome and, separately, between the activators binding to the four sites on the lower daughter chromosome. Replication must slow down or pause at one – and only one – of two "rest-stops". The parameters here include the number of activators, the number and proximity of sites, the association constants between activator proteins and between activator proteins and their sites (low and high affinity), and diffusion coefficients, all of which contribute to the time taken for activators to interact to form an effective structure to activate transcription. Other parameters include the length of time the replication fork remains at a rest-stop and the time between successive rest-stops (in this simple model, replication is effectively instantaneous between rest-stops and

stationary at rest-stops; a more realistic model would have regions between rest-stops where replication is relatively fast whilst the rest-stops themselves would be other regions where replication is relatively slow) as well as the position of the gene encoding the limiting factor and whether its own expression is under control of this factor. Here we assume that the activator is being produced so as to yield a constant concentration. If replication pauses at rest-stop A, the two sets of four low affinity sites are in competition with the eight high affinity sites and, depending on the choice of parameter values, this can allow activation of transcription from *both* the top and bottom chromosomes. If pausing occurs instead at step B, the two sets of low affinity sites are in competition with thirteen high affinity sites and parameter values exist that allow activation of transcription from only one set of the four low affinity sites. The important prediction here is that a broad range of parameter values exists that results in expression from both daughter chromosomes if replication pauses at rest-stop A but only from one daughter chromosome if replication pauses instead at rest-stop B.

There are numerous variants on this theme. The activator need not be a specific protein but could be a species of phospholipid in a domain or localised structures dependent on divalent ions or polyamines or polyphosphates. Indeed, a more general activation mechanism based on a similar principle is when the activator that is limiting is RNA polymerase itself.

Figure 2

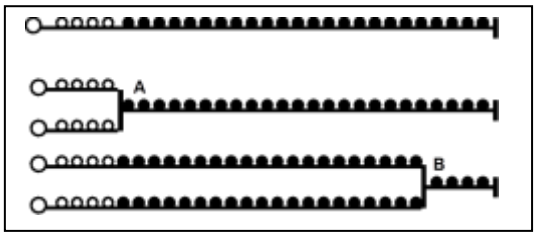

Fig. 2. Transcriptional activation via genes. The small empty circles represent genes that can form part of the same hyperstructure whilst the filled empty circles represent other genes. Other symbols as in Fig. 1.

Consider Figure 2 in which each of the small circles is a gene and in which the four empty circles on the top daughter chromosome can form part of a hyperstructure going into one half of the cell (which will become a daughter cell) whilst the other four empty circles can form part of a similar hyperstructure going into the other half of the cell. Suppose that expression of a gene within a hyperstructure favours its chance of being expressed again and suppose that RNA polymerase is limiting. The competition for RNA polymerase at rest-stop A is between the two potential hyperstructures, each containing four genes, and the rest of the genome containing 24 genes whilst the competition at rest-stop B is between the two sets of four genes in these potential hyperstructures and 44 other genes. The prediction is then as above, namely, that parameter values can be found that allow expression of both (essentially identical) hyperstructures if replication pauses at A and of only one hyperstructure if replication pauses at B.

3.2. Repression mechanisms

Now consider a repressor of transcription that has two types of binding site, high affinity and low affinity (filled and empty circles, respectively), distributed as shown in Figure

3. This repressor can form oligomers such that there is the possibility of *in cis* interactions between the repressors binding to the four high affinity sites on the top daughter chromosome and, separately (e.g., because they are in different regions of the cytoplasm due to chromosome folding), between the repressors binding to the four sites on the lower daughter chromosome.

Figure 3

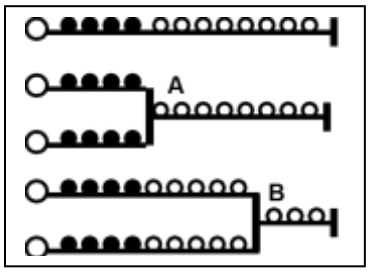

Fig. 3. Transcriptional repression via binding sites. The small empty circles are low affinity binding sites whilst the filled circles are high affinity binding sites. Other symbols as in Fig. 1.

Again, replication must pause at one – and only one – of two rest-stops. If replication pauses at rest-stop A, the two sets of four high affinity sites are in competition for repressor with the eight low affinity sites and, depending on the choice of parameter values, this could allow two discrete *repression* hyperstructures to form (in which transcription would be repressed, as when LacI binds to the operator and the auxiliary operators of one or more copies of the *lac* operon) containing the affected genes in *both* the top and bottom chromosomes. If pausing occurs instead at step B, the two sets of high affinity sites are in competition with thirteen low affinity sites and parameter values exist that allow only one repression hyperstructure to form and hence transcription to be repressed in only one of the future daughter cells. The important prediction here is that a broad range of parameter values exists that results in repression on both daughter chromosomes if replication pauses at rest-stop A but only on one daughter chromosome if replication pauses instead at rest-stop B.

As with the activator scenario, there are numerous variants on this theme. The repressor need not be a specific protein or RNA but could involve a preferential compaction or condensation of the regions containing the genes to be repressed into, for example, a cholesteric phase. To continue in this vein, perhaps the most general repression mechanism would be when the repression is via denial to the space needed for transcription and translation to occur.

Figure 4

Fig. 4. Transcriptional repression via competition for space. The small filled circles represent genes that can form part of the same hyperstructure whilst the small empty circles represent other genes. Other symbols as in Fig. 1.

Consider Figure 4 in which each of the small circles is a gene and in which the four genes (filled circles) on the top daughter chromosome can form part of a hyperstructure going into one half of the cell (which will become a daughter cell) whilst the other four genes (filled circles) can form part of a similar hyperstructure going into the other half of the cell. Suppose that expression of a gene within a hyperstructure favours its chance of being expressed again and suppose that RNA polymerase is limiting. The competition for RNA polymerase at rest-stop A is between the two potential, essentially identical, hyperstructures, each containing four genes, and the rest of the genome containing 24 genes whilst the competition at rest-stop B is between the two sets of four genes in these potential hyperstructures and 44 other genes. The prediction is then as above, namely, that parameter values can be found that allow expression of both hyperstructures if replication pauses at A and of only one hyperstructure if replication pauses at B.

3.3. Functioning-dependent assembly

Suppose replication pauses between the two sets of genes encoding the enzymes in two different enzymatic hyperstructures, ABCDE encoding enzymes A to E, and FGHIJ, encoding enzymes F to J (Figure 5A). In this case, ABCDE produces an effective repressor of FGHIJ assembly. This repressor may be the product of a gene associated with ABCDE or a metabolite produced by ABCDE or one of the ABCDE enzymes. This absence of differentiation can also be achieved if there is an absence of an activator of FGHIJ, which again could be a gene or a metabolite or an FGHIJ enzyme. Alternatively, replication pauses after replicating the genes encoding the FGHIJ enzymes (Figure 5B). In this case, ABCDE produces an ineffective repressor of FGHIJ assembly. This may be because the repressor is diluted out by the increased number of sites generated by the replication of the strands encoding the FGHIJ enzymes. Again, this repressor may be encoded by the ABCDE enzymes or be a metabolite produced by the ABCDE hyperstructure. Such differentiation may also result from the functioning-dependent start of formation of the FGHIJ hyperstructure, the probability of which is increased by the presence of two copies of the genes encoding this hyperstructure; a possible mechanism of assembly of this hyperstructure would be if the initial metabolites produced by one of the enzymes were to increase the affinities of the constituent enzymes for one another (e.g., of F for G). Important factors would include diffusion of metabolites and enzymes from both hyperstructures as well as chromosome folding and transcriptional noise.

Figure 5A

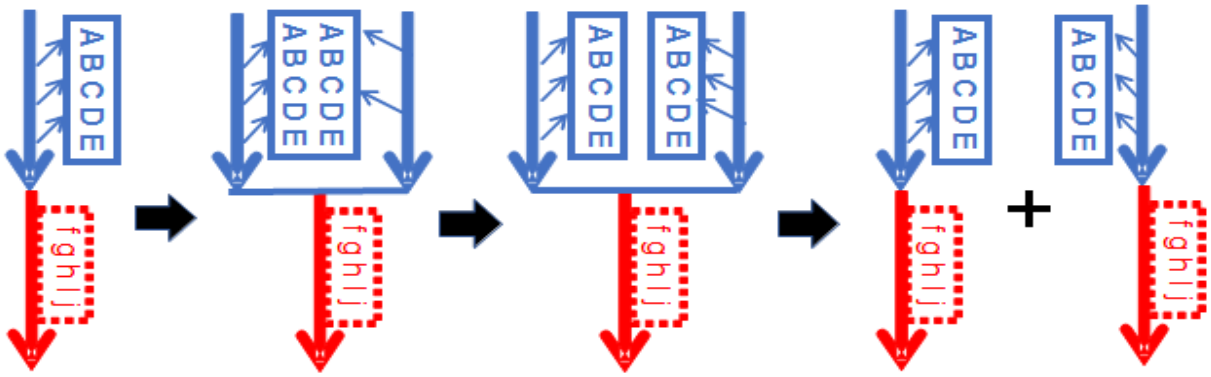

Figure 5B

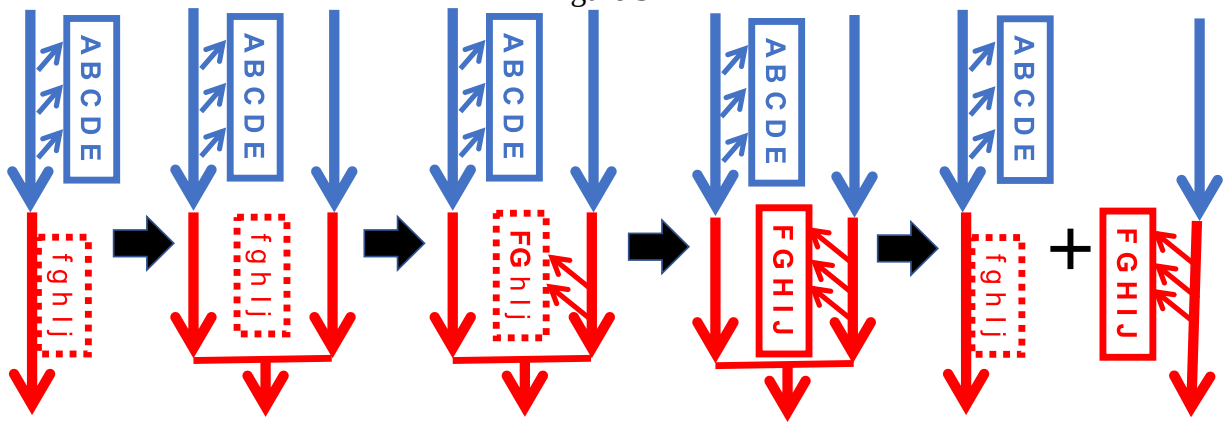

Fig. 5. Hyperstructure-based differentiation resulting from replication pausing. A/ No differentiation. Replication pauses between the two sets of genes encoding the enzymes in two different enzymatic hyperstructures. In this case, the repressor produced by the ABCDE hyperstructure is effective in inhibiting assembly of the FGHIJ hyperstructure. This repressor may be an ABCDE gene product or an ABCDE enzyme or metabolite. Alternatively, assembly of a permanent, full, FGHIJ hyperstructure can be prevented if the nascent hyperstructure fails to produce sufficient levels of an activator of FGHIJ, which again could be an FGHIJ gene or metabolite or enzyme. B/ Differentiation. Replication pauses after the genes encoding the FGHIJ hyperstructure. In this case, either ABCDE produces an ineffective repressor of FGHIJ assembly or a nascent FGHIJ hyperstructure produces an effective activator of its own assembly. Again, the repressor or activator may be a gene product, an enzyme or a metabolite. The long, thick, red and blue arrows are parts of chromosomes; the short, thin arrows represent coupled transcription-translation leading to the assembly of hyperstructures (red and blue rectangles); the letters represent enzymes; the dotted rectangles represent potential hyperstructures.

4. Evidence

4.1. Phenotypic diversity. A steadily accumulating body of evidence points to the universality of differentiation in the bacterial world (7-10). The cell cycle (of which, DNA replication is an intrinsic part) is fundamental to this differentiation. Cell division gives a stalked and a swarmer cell in *Caulobacter crescentus*, a spore and a mother cell in *B. subtilis*, and a tetrad containing chromosomes in different states in *Deinococcus radiodurans*. Even populations of *Escherichia coli* reveal a diversity – that extends to growth rates – that increases the probability that some cells will be ready to profit from new opportunities or survive new dangers (11, 12); this diversity in metabolic activity is not only intercellular but also intracellular (13)(Gangwe Nana, unpublished). The generation of such coherent diversity has been proposed as that one of the primary functions of the cell cycle which produces sister chromosomes with potentially different patterns of gene expression and hyperstructure formation (14-16).

4.2. Variations in the speed of replication. Studies of 'combed' chromosomes from a mutant of *E. coli* synchronised for replication reveal a heterogeneity in the pattern of replication, consistent with different replication speeds for forks in different regions (17). Different replication speeds along the arms of the *B. subtilis* chromosome were also detected in exponentially growing WT cells by a marker frequency analysis (L. Janniere, unpublished data). In vitro, individual DNA polymerases from *E. coli*

were found to have speeds of synthesis that varied ten-fold; around 70 kb was replicated by the leading strand polymerase of a single replisome whilst around 14 kb was replicated by the lagging strand polymerase, moreover these polymerases paused every 19 kb (18). Sequences that slow or halt replication have been found in both *E. coli* and *B. subtilis*. In *E. coli*, a polar DNA replication barrier is formed when the DNA-binding protein Tus forms a complex with any of the four 23-base-pair terminator (ter) sites found in the terminus region of the chromosome (in addition to Tus, other systems exist (19)). In *B. subtilis*, a replication barrier exists near the origin of replication and arrest is dependent upon the RelA protein, the action of which is correlated with high levels of the alarmone, ppGpp (20) but see (21).

Variations in the speed of DNA synthesis at the forks are implicated in many processes, including nutrient sensing via (p)ppGpp (22), replication-transcription conflict (23) (see below) and DNA repair (24). In particular, the time to replicate the chromosome of *Pseudomonas putida* is almost halved in the stress conditions of oxygen deprivation, iron limitation and solvent exposure (25). There is some evidence for the nature of the mechanisms coupling replication speed and metabolism: firstly, in *E. coli*, the velocity of the replication fork may vary from about 1000 to 200 nt/s as a function of the energy contained in the nutrients (26, 27); secondly, the *E. coli* DnaA initiator is genetically linked to two enzymes of the central carbon metabolism that convert pyruvate into acetate (Pta and AckA) (28); thirdly, in *B. subtilis*, the primase and the helicase (key enzymes in replication) appear to interact directly with metabolic enzymes such as pyruvate dehydrogenase (29) (L. Janniere, unpublished), an enzyme that might modulate the activity of the primase (30); fourthly, three *B. subtilis* enzymes that are loaded at the chromosomal replication origin early during initiation and that are known to act on the lagging-strand template in the replicating fork (the DNA polymerase DnaE, the helicase and the primase) (31, 32) are genetically connected to the five terminal reactions of glycolysis (33); fifthly, in *B. subtilis* cells grown in rich glycolytic or gluconeogenic media, at least 12 genes of the central carbon metabolism are important for replication in a medium-dependent manner (L. Jannière, unpublished data).

4.3. Location. The location of genes in *E. coli* and related bacteria is highly conserved along the chromosomal origin-to-terminus axis (34). The reasons proposed for the importance of gene location to expression (35-37) include a possible role for transcriptionally silenced regions in isolating chromosomal domains (38) and the need to restrict the diffusion of DNA-binding regulatory proteins, which depends on the location on the chromosome where the protein is synthesized (39). Evidence of just how much location can matter has been shown by several studies. Insertion into the *E. coli* chromosome of *lacZ* fusions to the *Psal* promoter, along with the cognate regulatory gene *nahR* from *Pseudomonas putida*, revealed variations that varied with location (40); using the lac promoter in a GFP-reporter construct, a 300-fold variation in expression was found to depend on the location of the construct on the *E. coli* chromosome (41); transfer of the entire origin-proximal operon encoding the global regulator FIS to the vicinity of replication terminus had major pleiotropic effects even though the lowered the *fis* gene dosage was compensated for by increased *fis* expression (37); the activity of an H-NS-regulated promoter depended on whether or not it was located in the AT-rich regions of the chromosome known to be bound by H-NS

and this activity also depended on the growth phase and growth rate (42). In *B. subtilis*, one sporulation network of genes is located close to the origin and a second sporulation network is located close to the terminus; this positioning leads to a transient gene dosage imbalance during chromosome replication that produces cell-cycle coordinated pulses of the sporulation master regulator Spo0A and allows cells to decide between sporulation and continued vegetative growth (43). Also in *B. subtilis*, variations in the replication speed recently detected by marker frequency experiments, show that different regions of the chromosome on either side of the origin of replication are present in the population in different proportions and that this gene density heterogeneity is affected in cells mutated in metabolic enzymes (Janniere, unpublished data).

4.4. Hyperstructures. The physical coupling of transcription and translation in bacteria is now widely accepted (44-49). This coupling is responsible for the colocalisation of genes and mRNA in *E. coli* and *Caulobacter crescentus*, which constitutes evidence consistent with transcription-translation hyperstructures existing at different positions along the chromosome (50). A cotranslational assembly or *transembly* hyperstructure may be consolidated by the interaction of its constituent proteins with one another as suggested by the assembly of nascent LuxB proteins during translation (51); significantly, the efficiency of this assembly is decreased if the genes (from which these luciferase subunits are synthesized via separate messenger RNAs) are located at distant chromosomal sites. Many enzymes operate in the form of *enzymatic* hyperstructures (3) in which interactions within the hyperstructure may again be important for assembly (52). An unstable, enzymatic hyperstructure is responsible for chromosome replication (17, 53, 54). This replication hyperstructure comprises a DnaE–DnaE or PolC–DnaE strand polymerisation complex along with the clamp loader, primase, helicase, and single-stranded binding proteins; the hyperstructure probably catalyses the synthesis of the deoxynucleoside triphosphates *in situ* due to the presence within it of ribonucleoside diphosphate reductase (55). The replication hyperstructure may be disrupted when the replication forks run into a barrier, as may occur when they meet RNA polymerase, which moves along the DNA twenty times slower than the forks. It is believed that replication is much more affected when transcription is in the opposite direction to replication as, for example, highly expressed, long, and essential genes are preferentially located on the leading strand (56). During the fast growth of *E. coli*, extensive ribosomal or 'nucleolar' hyperstructures form in which the constituents of ribosomes are made and probably assembled (57, 58). The collision of the replisome with a ribosomal hyperstructure might slow or halt replication and, not surprisingly, inversion of rRNA genes so as to oppose transcription and replication leads to disruption of DNA replication, loss of genome integrity, and cell death (59).

Activators and repressors that could mediate replication-phenotype coupling can create hyperstructures. The oligomeric DnaA protein, which plays a key role in the initiation of replication (60), also acts as a transcription factor (61). Simulations of DnaA activity in initiation have been made based on the distribution of low affinity binding sites in the origin region and higher affinity sites elsewhere (62). However, the situation with DnaA is much more complex than in our model as presented above. Not only are there several classes of binding sites but also, depending on the position of its binding sites relative to promoters, DnaA can act as an activator, a repressor or a terminator of

transcription. For example, it activates transcription from *nrd* (ribonucleoside diphosphate reductase), *glpD* (aerobic glycerol-3-phosphate dehydrogenase) and *fliC* (flagellin) whilst it represses transcription from *mioC* (biotin synthase), *rpoH* (heat shock sigma factor), *uvrB* (DNA repair), *proS* (prolyl-tRNA synthetase) and *dnaA* itself.

The heat-stable nucleoid-structuring protein (H-NS) is present in around 14000 copies in exponentially growing *E. coli*. In addition to its role in the compaction of the nucleoid (63), H-NS binds specifically to around 250 loci to cause transcriptional repression including that of its own gene; this repression involves an association with RNA polymerase (64). Another regulatory protein, the leucine-responsive regulatory protein (Lrp), interacts with H-NS to form higher order, repressive nucleoprotein structures involved in the repression of rRNA transcription (65).

Finally, the LacI repressor and its binding sites probably constitute a repression hyperstructure (3) that could in principle behave in accordance with our model. In the absence of an inducer such as lactose (or in the presence of the preferred sugar, glucose), the *lac* operon is not transcribed. This is because some of the ten copies per cell of the tetrameric LacI repressor bind with their dimers to the operator *O1* and to two auxiliary operators, *O2* and *O3*, nearby on the DNA; this on-off binding (which is an equilibrium process) increases the local concentration of LacI at these operators if they are close enough and brings them closer still to increase further the local concentration of LacI at *O1* (66). LacI binding to the operators is in competition with that of RNA polymerase to the promoter (since these sites overlap) and, importantly for our hypothesis, "there is some finite level of affinity of the protein for the "correct" site and some lower (but nonzero) and progressively decreasing affinity for other sites with decreasing degrees of homology with the correct one. To the extent that the great preponderance of wrong sites can compete with the regulatory target for protein and thus reduce the free protein concentration, the effective affinity of protein for the correct sites will also be reduced" (67).

5. Tests

The lactose operon in *E. coli* is perhaps the best understood of all operons. Its expression can be manipulated and copies of the entire operon, as well as copies of just the operators, can be inserted into different parts of the chromosome (see, for example, (41)). It might be possible to combine measurements based on the *lac* system with use of the *tus* system (as well as other systems (19)) to induce fork arrest in specific regions (68, 69) although this must be done prudently since it can prove lethal (70). Note too that the *lac* system itself might be used as a way of slowing replication (71). The combination of DNA combing and imaging by *Secondary Ion Mass Spectrometry* – the *CIS* technique – might also prove valuable. CIS offers a powerful way to determine replication speed on the scale of a few kilobases (72, 73) and it should be possible to determine replication speeds in specific regions of the chromosome by using hybridisation, or even just restriction enzyme fragments, to purify these regions for subsequent analysis by CIS.

Such experimental approaches might be combined with simulation approaches. Multi-agent programming is a promising approach to simulating the diffusion and interaction of the often large numbers of enzymes and metabolites present in biological cells (74-76). A stochastic automaton, HSIM, has been developed and used to model the dynamics of a glycolytic hyperstructure (77). To test the feasibility of the hypothesis proposed here, HSIM might be modified to represent linear

genes that can bind RNA polymerases and that can be duplicated; both genes and polymerases would be diffusible, and the effects on gene expression could be investigated by varying the parameters of pausing during duplication, varying the quantity of RNA polymerase, and varying the positive feedback weightings for cooperative RNA polymerase binding and chromosome decondensation. As HSIM takes into account geometry and spatial localisation, it could potentially take into account the effects of the greater accessibility of the region of the chromosome undergoing replication on the probability of transcription occurring. Petri nets might also be used (78). A Petri net is a directed, bipartite, labelled graph, in which events that may occur (transitions) are represented by rectangles and the conditions for these events (places) are represented by circles. Directed arcs run from a place to a transition or vice versa and state which places are pre- and/or post-conditions for the transitions. The places in a Petri net may contain a discrete number of tokens, which can be of different types in the case of a coloured Petri net. In principle, one might design a Petri net containing at least two alternative subnetworks that would be duplicated progressively. There would be a competition for tokens. One class of tokens would represent RNA polymerases/ribosomes, which make themselves, whilst another class would represent enzymes/metabolites. This would allow the investigation of the combined effects of transcriptional noise, which should lead to an increase in a gene product when the encoding gene is present in more than one copy, on catalysis, and of the subsequent reaction product on assembly of a functioning-dependent, enzymatic hyper-structure (Figure 5). Such investigation could also include the relationships between the time needed to assemble an FDS, the time for which replication pauses, and the relative positions of the genes encoding the different hyperstructures.

6. Discussion

How do cells manage to produce not only reproducible phenotypes (out of the hyper-astronomical number apparently available to them (1)) but also *coherent* phenotypes? How do cells negotiate the cell cycle? And are these questions linked? We have suggested the phenotype is decided not at the level of individual macromolecules but rather at a higher level – that of assemblies of molecules *alias* hyperstructures (79). We have also suggested that a primary function of the cell cycle is to generate coherent diversity of phenotypes within a population of cells (14, 80). The very fact of having two chemically identical chromosomes in the same cytoplasm creates a symmetry-breaking situation if the genes are in a global competition with one another for transcription by RNA polymerase and if local positive feedback circuits can operate such that a gene that is being transcribed has a greater chance of being transcribed again than one that is silent (81). This "differentiation for free" idea can be revisited in the light of: (1) a relationship between metabolism and the enzymes involved in the initiation and elongation steps in chromosome replication, both of which can change the copy numbers of genes; in an analysis of 27 mutants in 19 genes encoding enzymes of the central carbon metabolism of *B. subtilis*, it was found that at least 12 of these genes were involved in coupling growth rates with the initiation and/or elongation steps of replication (L. Janniere, unpublished data); such mutants have differences in the proportions of genes on either side of the origin of replication (L. Janniere, unpublished data); (2) the likely proximity of a gene on the chromosome to the enzymes encoded by this gene (50). If these factors affect phenotype by altering the speed of replication by forks in

different regions in the chromosome, it would help explain the importance of gene location on the chromosome (34, 43, 82-84). Reciprocally, ongoing metabolism itself might alter the speed of replication insofar as transcriptional activity reflects metabolism; for example, the replisome could be slowed down by the co-oriented transcription of the genes encoding particular subsets of metabolic enzymes.

Changing the speed of replication by forks at different places on the chromosome has a considerable potential for the exploration of phenotype space if it is combined with the strand-specific segregation of hyperstructures. In the strand-specific proposal, the association of each parental strand with a particular set of hyperstructures and the continued association once replication has occurred ensures that daughter chromosomes have different phenotypes (16). The result of this combination is that whether two identical or different hyperstructures form at the same position on the new daughter chromosomes depends on the local and global speeds of replication since these speeds create the environment in which differentiation occurs. This environment is one of locally positive cooperation and globally negative competition for which a variety of mechanisms could be responsible. These mechanisms include the distribution of transcriptional activators and repressors, functioning-dependent changes in enzyme affinities, the binding of the constituents of one hyperstructure to those of another hyperstructure, and the folding of the chromosome (36, 85, 86). All these mechanisms could help produce coherent metabolic patterns in the daughter cells. A judicious marriage between simulation using programs such as HSIM and Petri nets and bench experimentation should identify the parameter values needed for such replication-based differentiation to occur.

7. Conclusion

A growing body of evidence is consistent with the hypothesis that variations in the speed of fork progression during replication play a major role in the coordination of gene expression and hyperstructure dynamics needed to create coherent phenotypes; such coordination would include metabolism feeding back into fork speed. In this hypothesis, the environment influences the assembly and activity of hyperstructures; in turn, these hyperstructures cause local changes in the speed of individual replication forks; these changes then alter the numbers of transcriptional activators and repressors and their binding sites as well as the relative numbers of particular genes; these alterations affect gene expression that, ultimately, feeds back to alter hyperstructures and, in particular, the hyperstructures that are associated with the strands so as to confer different hyperstructures and different phenotypes on the daughter cells (as proposed in the strand segregation hypothesis (16)).

Evidence for the hypothesis includes variation in the speed of the replication forks related to nutrient sensing and to replication-transcription conflict, the importance of the location of genes on their expression (and probably on hyperstructure formation), the role of metabolic enzymes in the initiation and elongation steps of replication, and the relationship between the copy number of genes and the metabolic state of the cell. Testing the replication-phenotype coupling hypothesis will require close collaborations across the disciplines and, in particular, between microbiologists studying the metabolism-replication relationship and computer scientists developing tools such as stochastic automata and Petri nets to simulate and study cellular processes.

8. Corollaries

8.1. Local positive feedback and global negative regulation can act not only via either activation or repression of gene expression but also via translation and degradation. In all cases, the end result is a hyperstructure that tends to maintain its own existence.

8.2. There is an epigenetic flavour to our hypothesis. If a bacterium replicates its chromosome, this is usually because it is growing. If it is growing, this is because it has the hyperstructures needed for growing. Hence, the mother cell already has one copy of a needed hyperstructure; if the genes in this hyperstructure are all on one strand, one of the future daughters inherits an established hyperstructure. This creates a status quo situation that is likely to affect the chance of the other daughter generating or not generating a sister hyperstructure.

8.3. Factors that affect the synthesis or degradation of the activator or repressor are clearly important in replication-phenotype coupling and include the position of the gene encoding the activator or repressor (whether it is before rest-stop A, or between rest-stops A and B, or after rest-stop B) and whether this gene is itself regulated by its own product. Given that the ratio of RNA polymerase to genes is a key parameter, these factors also include the temporal pattern of synthesis of transcriptionally active RNA polymerase and the spatial distribution of this enzyme (note that RNA polymerase itself is subject to local concentration effects).

8.4. In the light of a likely extensive inter- and intra-cellular metabolic heterogeneity, a rapid coupling between metabolism and DNA replication could affect not only the elongation step but also the initiation step, thereby generating a potentially very broad diversity of coherent phenotypes.

8.5. The effects of replication pausing on bringing genetic loci together (or on maintaining together loci that have come close to one another by diffusion etc,) will vary with the distance between these loci on the chromosome, how long pausing lasts and the concentrations of binding proteins and their affinities.

References

1. Kauffman S. The origins of order. Oxford: Oxford University Press; 1993.
2. Norris V, Gascuel P, Guespin-Michel J, Ripoll C, Saier MH, Jr. Metabolite-induced metabolons: the activation of transporter-enzyme complexes by substrate binding. Mol Microbiol. 1999;31(5):1592-5.
3. Norris V, Blaauwen TD, Doi RH, Harshey RM, Janniere L, Jimenez-Sanchez A, et al. Toward a Hyperstructure Taxonomy. Annu Rev Microbiol. 2007;61:309-29.
4. Norris V, Menu-Bouaouiche L, Becu J-M, Legendre R, Norman R, Rosenzweig JA. Hyperstructure interactions influence the virulence of the Type 3 secretion system in yersiniae and other bacteria. Appl Microbiol Biotechnol. 2012;96(1):23-36.
5. Saier MH, Jr. Microcompartments and protein machines in prokaryotes. J Mol Microbiol Biotechnol. 2013;23(4-5):243-69.
6. Norris V, Verrier C, Feuilloley M. Hybolites Revisited. Recent Pat Antiinfect Drug Discov. 2016;11(1):16-31.
7. Booth IR. Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. International Journal of Food Microbiology. 2002;78:19-30.
8. Avery SV. Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol. 2006;4(8):577-87.
9. Smits WK, Kuipers OP, Veening JW. Phenotypic variation in bacteria: the role of feedback regulation. Nat Rev Microbiol. 2006;4(4):259-71.
10. Bergmiller T, Andersson AMC, Tomasek K, Balleza E, Kiviet DJ, Hauschild R, et al. Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. Science. 2017;356(6335):311-5.
11. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science. 2004;305(5690):1622-5.
12. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, et al. Robust growth of Escherichia coli. Curr Biol. 2010;20(12):1099-103.
13. Gangwe Nana G, Gibouin D, Lefebvre F, Delaune A, Jannière L, Ripoll C, et al., editors. Intracellular and population heterogeneity in *Bacillus subtilis* revealed by Secondary Ion Mass Spectrometry. Modelling complex biological systems in the context of genomics; 2012; Evry, France: EDP Sciences.
14. Norris V. Hypothesis: chromosome separation in Escherichia coli involves autocatalytic gene expression, transertion and membrane-domain formation. Mol Microbiol. 1995;16(6):1051-7.
15. Norris V, Demarty M, Raine D, Cabin-Flaman A, Le Sceller L. Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. Biochimie. 2002;84:341-7.
16. Rocha EP, Fralick J, Vediyappan G, Danchin A, Norris V. A strand-specific model for chromosome segregation in bacteria. Mol Microbiol. 2003;49(4):895-903.
17. Breier AM, Weier HU, Cozzarelli NR. Independence of replisomes in Escherichia coli chromosomal replication. Proceedings of the National Academy of Science USA. 2005;102:3942-7.
18. Graham JE, Marians KJ, Kowalczykowski SC. Independent and Stochastic Action of DNA Polymerases in the Replisome. Cell. 2017;169(7):1201-13 e17.
19. Guijo MI, Patte J, del Mar Campos M, Louarn JM, Rebollo JE. Localized remodeling of the Escherichia coli chromosome: the patchwork of segments refractory and tolerant to inversion near the replication terminus. Genetics. 2001;157(4):1413-23.
20. Autret S, Levine A, Vannier F, Fujita Y, Seror SJ. The replication checkpoint control in Bacillus subtilis: identification of a

novel RTP-binding sequence essential for the replication fork arrest after induction of the stringent response. Mol Microbiol. 1999;31(6):1665-79.
21. Mulugu S, Potnis A, Shamsuzzaman, Taylor J, Alexander K, Bastia D. Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. Proc Natl Acad Sci U S A. 2001;98(17):9569-74.
22. Denapoli J, Tehranchi AK, Wang JD. Dose-dependent reduction of replication elongation rate by (p)ppGpp in Escherichia coli and Bacillus subtilis. Mol Microbiol. 2013;88(1):93-104.
23. Maduike NZ, Tehranchi AK, Wang JD, Kreuzer KN. Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Mol Microbiol. 2014;91(1):39-56.
24. Tan KW, Pham TM, Furukohri A, Maki H, Akiyama MT. Recombinase and translesion DNA polymerase decrease the speed of replication fork progression during the DNA damage response in Escherichia coli cells. Nucleic Acids Res. 2015;43(3):1714-25.
25. Lieder S, Jahn M, Koepff J, Muller S, Takors R. Environmental stress speeds up DNA replication in Pseudomonas putida in chemostat cultivations. Biotechnol J. 2016;11(1):155-63.
26. Helmstetter CE. Timing of synthetic activities in the cell cycle. In: Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, et al., editors. *Escherichia coli* and *Salmonella*. Washington D.C.: American Society for Microbiology; 1991. p. 1594-605.
27. Michelsen O, Teixeira de Mattos MJ, Jensen PR, Hansen FG. Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. Microbiology. 2003;149:1001-10.
28. Maciag M, Nowicki D, Janniere L, Szalewska-Palasz A, Wegrzyn G. Genetic response to metabolic fluctuations: correlation between central carbon metabolism and DNA replication in Escherichia coli. Microb Cell Fact. 2011;10:19.
29. Noirot-Gros MF, Dervyn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, et al. An expanded view of bacterial DNA replication. Proceedings of the National Academy of Science USA. 2002;99:8342-7.
30. Stein A, Firshein W. Probable identification of a membrane-associated repressor of *Bacillus subtilis* DNA replication as the E2 subunit of the pyruvate dehydrogenase complex. Journal of Bacteriology. 2000;182:2119-24.
31. Rannou O, Le Chatelier E, Larson MA, Nouri H, Dalmais B, Laughton C, et al. Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand DNA replication in Bacillus subtilis. Nucleic Acids Res. 2013;41(10):5303-20.
32. Paschalis V, Le Chatelier E, Green M, Kepes F, Soultanas P, Janniere L. Interactions of the Bacillus subtilis DnaE polymerase with replisomal proteins modulate its activity and fidelity. Open Biol. 2017;7(9).
33. Janniere L, Canceill D, Suski C, Kanga S, Dalmais B, Lestini R, et al. Genetic evidence for a link between glycolysis and DNA replication. PLoS One. 2007;2(5):e447.
34. Sobetzko P, Travers A, Muskhelishvili G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. Proc Natl Acad Sci U S A. 2012;109(2):E42-50.
35. Kepes F. Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. J Mol Biol. 2003;329(5):859-65.
36. Kepes F. Periodic transcriptional organization of the E.coli genome. J Mol Biol. 2004;340(5):957-64.
37. Gerganova V, Berger M, Zaldastanishvili E, Sobetzko P, Lafon C,

Mourez M, et al. Chromosomal position shift of a regulatory gene alters the bacterial phenotype. Nucleic Acids Res. 2015;43(17):8215-26.
38. Vora T, Hottes AK, Tavazoie S. Protein occupancy landscape of a bacterial genome. Mol Cell. 2009;35(2):247-53.
39. Kuhlman TE, Cox EC. Gene location and DNA density determine transcription factor distributions in Escherichia coli. Mol Syst Biol. 2012;8:610.
40. Sousa C, de Lorenzo V, Cebolla A. Modulation of gene expression through chromosomal positioning in Escherichia coli. Microbiology. 1997;143 (Pt 6):2071-8.
41. Bryant JA, Sellars LE, Busby SJ, Lee DJ. Chromosome position effects on gene expression in Escherichia coli K-12. Nucleic Acids Res. 2014;42(18):11383-92.
42. Brambilla E, Sclavi B. Gene regulation by H-NS as a function of growth conditions depends on chromosomal position in Escherichia coli. G3 (Bethesda). 2015;5(4):605-14.
43. Narula J, Kuchina A, Lee DY, Fujita M, Suel GM, Igoshin OA. Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. Cell. 2015;162(2):328-37.
44. Miller OL, Jr., Hamkalo BA, Thomas CA, Jr. Visualization of bacterial genes in action. Science. 1970;169(943):392-5.
45. French SL, Miller OL, Jr. Transcription mapping of the Escherichia coli chromosome by electron microscopy. J Bacteriol. 1989;171(8):4207-16.
46. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, et al. A NusE:NusG complex links transcription and translation. Science. 2010;328(5977):501-4.
47. Bakshi S, Choi H, Weisshaar JC. The spatial biology of transcription and translation in rapidly growing Escherichia coli. Front Microbiol. 2015;6:636.
48. Elgamal S, Artsimovitch I, Ibba M. Maintenance of Transcription-Translation Coupling by Elongation Factor P. MBio. 2016;7(5).
49. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P. Architecture of a transcribing-translating expressome. Science. 2017;356(6334):194-7.
50. Llopis PM, Jackson AF, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T, et al. Spatial organization of the flow of genetic information in bacteria. Nature. 2010;466(7302):77-81.
51. Shieh YW, Minguez P, Bork P, Auburger JJ, Guilbride DL, Kramer G, et al. Operon structure and cotranslational subunit association direct protein assembly in bacteria. Science. 2015;350(6261):678-80.
52. Meyer P, Cecchi G, Stolovitzky G. Spatial localization of the first and last enzymes effectively connects active metabolic pathways in bacteria. BMC Syst Biol. 2014;8(1):131.
53. Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: evidence for a factory model of replication. Science. 1998;282(5393):1516-9.
54. Migocki MD, Lewis PJ, Wake RG, Harry EJ. The midcell replication factory in *Bacillus subtilis* is highly mobile: implications for coordinating chromosome replication with other cell cycle events. Molecular Microbiology. 2004;54:452-63.
55. Sanchez-Romero MA, Molina F, Jimenez-Sanchez A. Organization of ribonucleoside diphosphate reductase during multifork chromosome replication in Escherichia coli. Microbiology. 2011;157(Pt 8):2220-5.
56. Rocha EP. The organization of the bacterial genome. Annu Rev Genet. 2008;42:211-33.
57. Cabrera JE, Jin DJ. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Molecular Microbiology. 2003;50:1493-505.
58. Jin DJ, Mata Martin C, Sun Z, Cagliero C, Zhou YN. Nucleolus-like

compartmentalization of the transcription machinery in fast-growing bacterial cells. Crit Rev Biochem Mol Biol. 2017;52(1):96-106.
59. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. Co-orientation of replication and transcription preserves genome integrity. PLoS Genet. 2010;6(1):e1000810.
60. Leonard AC, Grimwade JE. Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. Molecular Microbiology. 2005;55:978-85.
61. Messer W, Weigel C. DnaA initiator--also a transcription factor. Molecular Microbiology. 1997;24:1-6.
62. Hansen FG, Christensen BB, Atlung T. The initiator titration model: computer simulation of chromosome and minichromosome control. Research in Microbiology. 1991;142:161-7.
63. Dame RT, Wyman C, Goosen N. H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Research. 2000;28:3504-10.
64. Oshima T, Ishikawa S, Kurokawa K, Aiba H, Ogasawara N. Escherichia coli histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. DNA Res. 2006;13(4):141-53.
65. Pul U, Wurm R, Wagner R. The Role of LRP and H-NS in Transcription Regulation: Involvement of Synergism, Allostery and Macromolecular Crowding. J Mol Biol. 2007;366(3):900-15.
66. Müller-Hill B. The function of auxiliary operators. Molecular Microbiology. 1998;29:13-8.
67. von Hippel PH, Berg OG. On the specificity of DNA-protein interactions. Proc Natl Acad Sci U S A. 1986;83(6):1608-12.
68. Andersen PA, Griffiths AA, Duggin IG, Wake RG. Functional specificity of the replication fork-arrest complexes of Bacillus subtilis and Escherichia coli: significant specificity for Tus-Ter functioning in E. coli. Mol Microbiol. 2000;36(6):1327-35.
69. Valjavec-Gratian M, Henderson TA, Hill TM. Tus-mediated arrest of DNA replication in Escherichia coli is modulated by DNA supercoiling. Mol Microbiol. 2005;58(3):758-73.
70. Bidnenko V, Lestini R, Michel B. The Escherichia coli UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. Mol Microbiol. 2006;62(2):382-96.
71. Payne BT, van Knippenberg IC, Bell H, Filipe SR, Sherratt DJ, McGlynn P. Replication fork blockage by transcription factor-DNA complexes in Escherichia coli. Nucleic Acids Res. 2006;34(18):5194-202.
72. Cabin-Flaman A, Monnier AF, Coffinier Y, Audinot JN, Gibouin D, Wirtz T, et al. Combed Single DNA Molecules Imaged by Secondary Ion Mass Spectrometry. Anal Chem. 2011;83(18):6940-7.
73. Cabin-Flaman A, Monnier AF, Coffinier Y, Audinot JN, Gibouin D, Wirtz T, et al. Combining combing and secondary ion mass spectrometry to study DNA on chips using (13)C and (15)N labeling. F1000Res. 2016;5:1437.
74. Le Sceller L, Ripoll C, Demarty M, Cabin-Flaman A, Nyström T, Saier Jnr. M, et al. Modelling bacterial hyperstructures with cellular automata. Interjournal of Complex Systems. 2000;Paper 366:http://www.lri.fr/~pa/Hsim/InterJournal.pdf.
75. Ballet P, Zemirline A, Marcé L. The BioDyn language and simulator. Application to an immune response and *E. coli* and phage interaction. Journal of Biological Physics and Chemistry. 2004;4:93-101.
76. Soula H, Robardet C, Perrin F, Gripon S, Beslon G, Gandrillon O. Modeling the emergence of multi-protein dynamic structures by principles of self-organization through the use of 3DSpi, a Multi-Agent-

based Software. Biomed Central Bioinformatics. 2005;6:228.
77. Amar P, Legent G, Thellier M, Ripoll C, Bernot G, Nystrom T, et al. A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. BMC Syst Biol. 2008;2:27.
78. Koch I, Nothen J, Schleiff E. Modeling the Metabolism of Arabidopsis thaliana: Application of Network Decomposition and Network Reduction in the Context of Petri Nets. Front Genet. 2017;8:85.
79. Norris V, den Blaauwen T, Cabin-Flaman A, Doi RH, Harshey R, Janniere L, et al. Functional taxonomy of bacterial hyperstructures. Microbiol Mol Biol Rev. 2007;71(1):230-53.
80. Norris V. Why do bacteria divide? Front Microbiol. 2015;6:322.
81. Norris V, Madsen MS. Autocatalytic gene expression occurs via transertion and membrane domain formation and underlies differentiation in bacteria: a model. J Mol Biol. 1995;253(5):739-48.
82. Teleman AA, Graumann PL, Lin DC, Grossman AD, Losick R. Chromosome arrangement within a bacterium. Current Biology. 1998;8:1102–9.
83. Niki H, Yamaichi Y, Hiraga S. Dynamic organization of chromosomal DNA in *Escherichia coli.* Genes and Development. 2000;14:212–23.
84. Viollier PH, Thanbichler M, McGrath PT, West L, Meewan M, McAdams HH, et al. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. Proceedings of the National Academy of Science USA. 2004;101:9257-62.
85. Junier I, Martin O, Kepes F. Spatial and topological organization of DNA chains induced by gene co-localization. PLoS Comput Biol. 2010;6(2):e1000678.
86. Kepes F, Jester BC, Lepage T, Rafiei N, Rosu B, Junier I. The layout of a bacterial genome. FEBS Lett. 2012.

