Phase variation of type 1 fimbriae : a single cell investigation

Adiciptaningrum, A.M.

1 Introduction

A single cell approach to study biological systems has gained popularity due to recent technological advances and a growing appreciation for non-genetic cell-to-cell diversity. One example of this heterogeneity is phase variation, a reversible and stochastic switching between two types of expression in individual cells within a clonal population. In *E.coli*, the expression of type 1 pili (fimbriae) is phase variable, using DNA inversion as the underlying molecular mechanism. In this thesis, we present the results of our investigation on the *fim* system at the single cell level together with our investigation of the cells' physiological diversity during the cell cycle. In this study the DNA replication process was monitored simultaneously with cell growth and the link between the two processes was examined.

1.1 The single cell approach

In the field of microbiology, studies of complex cell processes, interactions and behaviors have traditionally been done at the bulk or population level. Typically, the culture of interest is inoculated with a sample of an isogenic microbial population and is then further assumed to be a population of identical cells. Any individual cell response that may exist is therefore subjected to averaging which results in mean population outcomes.

Increasing interest in cell-to-cell heterogeneity has led to the recent development of single cell approaches. These approaches, such as microscopy, flow cytometry, single-cell microarray, single cell PCR, fluorescence *in situ* hybridization, etc., allow measurement of individual cell properties. Work on microbial [39,92] and mammalian [98] cells has demonstrated that isogenic populations show a wide variety in gene expression levels, which can be attributed to fluctuations in cellular components stemming from the stochastic nature of biochemical processes. Since then, other studies (for reviews: [3,15,73]) have demonstrated that cell-to-cell variation is a widespread phenomenon in biology and that investigation at the single cell level is necessary to reveal individual heterogeneity.

The recent increase in single cell research has benefited from technological developments in fluorescent dyes and proteins, automated microscopy, growing computer storage capacity and computing ability. Even so, the interest in investigating cell-to-cell diversity already arose several decades ago. In 1953, Benzer realized a problem in studying kinetics of β-galactosidase induction with batch cultures [9]. He considered whether all cells participate equally and simultaneously in the synthesis of an enzyme. Since a direct method to quantify protein level in individual cells did not exist, he used a bacteriophage to quantify the participation of each cell in a culture growing solely on lactose. The phage duplication *in vivo* depended on the host β-galactosidase concentration already present before the moment of infection. Upon infection, the bacteriophage blocked β-galactosidase synthesis, and uniquely prevented its own development. Hence, the time at which an individual host cell was lysed depended on the enzyme content before infection. Benzer quantified the amount of β-galactosidase released to the medium at various stages of lysis as a measure of culture heterogeneity, and found that the population response is essentially homogenous in respect to enzyme content. His innovative experiment illustrates one of the many advantages that single cell methods have to offer.

The general advantage of a single cell approach is the evasion of averaging effects that are characteristic of bulk-phase population-scale methods. Many cell

properties such as viability, number of proteins per cell, and the number of structures expressed on the cell surface are discrete and intrinsic states or properties of each individual cell. Single cell techniques also allow a high spatial and temporal resolution of dynamic events *in vivo*. And in several cases, they allow connections to be made between apparent macroscopic phenomena and their microscopic cellular origins. Of course, this approach also has its caveats. Several of the techniques are low throughput, so that sufficient statistics can be difficult to obtain. In addition, it is not possible to exclude the influences of the techniques on the measured cell, since an experiment and control cannot be carried out on the same cell. Therefore, while the single cell approach can be a very powerful tool of investigation, the limitations mentioned above must be kept in mind.

1.2 Phenotypic diversity and phase variation

Brehm-Stecher and Johnson categorized factors contributing to cell-to-cell diversity in four major classes: genetic diversity, biochemical or metabolic diversity, physiological diversity, and behavioral diversity [15]. Genetic diversity can arise from a number of random, semi-random, or programmed DNA modification events. Biochemical diversity is characterized by individual differences in cellular macromolecular composition or activity. Physiological diversity describes morphological differences between individual cells, including cell cycle-related characteristics such as size and shape. Behavioral diversity comes from the difference in observable cellular responses as a consequence of other cell-to-cell variation. These different classes of diversities are not strictly separated and in many cases are intertwined. For example, an apparent behavioral diversity, such as the phase varying competence state of *Bacillus subtillis*, can be traced back to the biochemical diversity underlying it, which is the noise in *comK* expression level [79]. Or, an apparent physiological diversity such as the start of DNA replication might have to do with the concentration of a certain protein required for initiating the replication. Often, the diversities also influence the best suitable methods and techniques of investigation.

Phase variation in general, refers to a reversible “all-or-none” type of expression of one or more proteins in an individual cell within a clonal population. The reversible switching between the two types of expressions is considered to be stochastic in nature, and occurs at a frequency exceeding that of a random mutation. A certain heritable genetic or epigenetic mechanism is underlying the process and in some cases, environmental factors modulate the switching rates (for excellent reviews:[48,116,117]).

The classical view on the biological role of phase variation is that it helps bacteria evade a host immune system. Phase varying mechanisms are commonly found to regulate external structures exposed to the environment. Therefore, phase variation is often associated with the regulation of virulence factors in bacterial pathogens. In support of this view, it has been shown that fimbrial phase variation in *Salmonella* is a means to evade cross-immunity between serotypes [90]. A recent theoretical study has shown that when the evasion is frequency dependent, phase variation can provide a fitness advantage by generating population heterogeneity [123].

This classical view is being challenged or complemented by a "bet-hedging" view, in which the suggested biological role of phase variation is to provide bacteria with a strategy to better survive fluctuating and unknown future environments. In a mixed population, a certain subset of the population is always prepared for any sudden experimental change, increasing the chance for overall population survival [36]. Recently, it has been shown that an irreversible decision to sporulate in a subset of a *Bacillus subtilis* population results in specialization in carbon source usage and different reproductive potential under different environments [118]. Theoretical models propose that in some fluctuating environments, a dynamically heterogeneous population could have higher net growth rate than a homogenous one [115], and that if the fluctuations are infrequent [70] or cannot be sensed precisely enough [122], stochastic switching could be preferred over a sensing mechanism.

1.3 The *fim* system of *E.coli*

As mentioned above, phase variation is a common regulatory mechanism for external structures. In *E.coli*, this includes various pili (P, S, type1, CS18, CS31A) and the Ag43 outer membrane protein. Among those pili, one of the best-studied systems is the *fim* system, involved in the expression of type 1 pili (fimbriae) widely expressed by both pathogens and commensal strains. Fimbriae are hair-like structures made of protein expressed on the cell surface (fig. 1.1.A) with lengths of about 1 μm and diameter of 6-7 nm [20,45]. A typical fimbriated *E.coli* expresses 200-500 pili on its surface. One pilus is an arrangement of more than 1000 major structural subunits (FimA) in a right-handed helical pattern (2.4 nm pitch and 3.4 subunits per turn) assembled via an usher/chaperone pathway at the outer membrane (fig. 1.1.B) [99]. At the tip of these pili, a short thin thread (tip fibrillum) made of FimF and FimG anchors an adhesin (FimH) that binds to the host mammalian receptors [54,55].

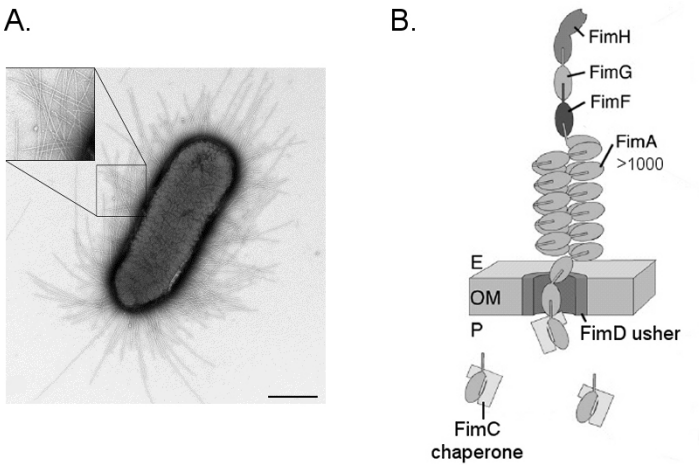

Figure 1.1. E.coli fimbriae.
(A) Electron microscopy image of fimbriated *E.coli* (adapted from [45]).
(B) Fimbrial subunits (FimA as major subunit) are expressed by the *fim* operon. The symbol E stands for extracellular, OM outer membrane and P periplasm (adapted from [99]).

In *E.coli,* the expression of fimbriae is thought to facilitate adhesion to host tissue at the early stage of infection, for example in Urinary Tract Infection (UTI) (fig. 1.2) (for a review see [106]). The expression of type 1 fimbriae, together with P fimbriae, is commonly found amongst clinical uropathogenic strains. In accordance with the classical view on phase variation mentioned earlier, the phase-varying feature in *fim* is also thought to represent the fine balance between pathogen and host interaction. While the nutritional advantage from infecting host cells is obvious, a high level of fimbriation evokes the host's inflammatory response, a threat to the bacterium's survival. Consequently, the level of fimbriation must be regulated such that those trade-offs are balanced.

Regulation for fimbrial expression lies on the invertible 314 bps DNA fragment *fimS* containing a promoter sequence. Upon inversion, the promoter sequence in *fimS* can obtain different orientations, acting as a reversible genetic switch. In the ON orientation, *fimS* drives the expression of multiple fimbrial structural genes, *fimA-fimH* (fig. 1.3). In the OFF orientation, the promoter is in an inverted configuration, and thus silent. The switching property of *fimS* is believed to be stochastic in nature, resulting in a mixed population of fimbriated and non-fimbriated cells in most environmental and physiological conditions. In contrast with other types of regulation such as repression, the stochastic inversion of *fimS*

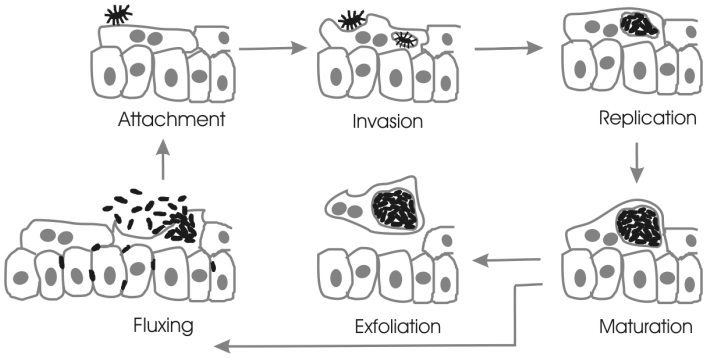

Figure 1.2. Schematic picture of Urinary Tractk Infection sequence of events.
The level of fimbriae expression has to be well regulated in order to balance the advantage of invasion (escaping starvation) and its disadvantage (exfoliation by the host immune system).

allows a partial, but immediate population response to sudden environmental changes. It also allows a unique all-or-nothing type of expression, a key feature for tuning the fine balance between a bacterial population and its host immune system.

The inversion of *fimS* is performed by two site-specific recombinases, *fimB* and *fimE*. While FimB is able to switch in both directions, FimE has an extreme bias to inverse the switch to the OFF orientation with a much higher rate (about 100-1000X more; up to 0.1 OFF switching events per cell per generation). FimB and FimE are homologous proteins (52% amino acid identity) and are members of the integrase family of tyrosine site-specific recombinases. Tyrosine recombinases break and rejoin single stranded DNA in pairs to form a Holliday junction intermediate [21,43]. The Holliday junction intermediate is formed by the interaction of multimeric protein complexes on the two Fim-recombinase binding sites flanking a 9 bps inverted-repeat on each side of *fimS* (IRR and IRL), thereby looping out the DNA between the sites (fig. 1.3). There are a total of four binding sites, called half sites, to be occupied for a successful inversion to occur. Temperature, environmental nutrients and host chemical responses all influence *fim* switching rates [41].

Besides regulating the expression of fimbrial structural genes, the orientation of *fimS* also regulates the expression of FimE. This phenomenon, called orientational control, is the expression of FimE only when *fimS* is in the ON orientation. Sequence differences mandate that stable mRNAs are only produced from the *fimS* ON orientation. In the opposite orientation, *fimE* transcription is prematurely terminated in a rho-dependent manner and more likely to be degraded [50,56,110].

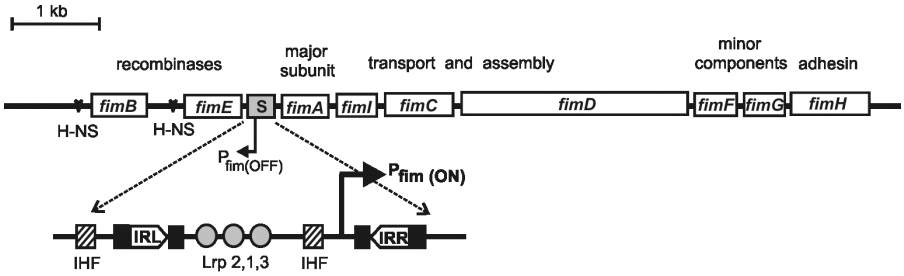

Figure 1.3 The *fim* system of *E.coli*.
Only when the switch, *fimS,* is in the ON orientation (as in the zoomed area), does the promoter sequence drive the expression of multiple fimbrial structural genes. The black boxes next to IRL and IRR are binding sites for the FimB and FimE recombinases, called half sites. There are also multiple binding sites for accessory proteins such as HNS, IHF and Lrp within and in close vicinity to the switch.

Previous models have proposed different biological reasoning for the existence of orientational control. One model proposed that the orientational control of *fim* acts as a timer to provide sufficient time for fimbrial development [121]. Other theoretical work has proposed that FimB and FimE act as antagonistic forces, each pushing towards a different steady state to enable rapid adjustment to fimbriation levels in response to environmental changes while keeping the overall fimbriation level low [25].

The regulation of the *fim* system is currently quite well understood. On the other hand, the dynamics of switching is still largely obscured. Previous investigations were done in bulk, and this approach poses certain constraints for fully understanding the system. For example, the correlation between switching events and cell processes as well as cell growth remains vague, and the effect of cell-state history on switching rates is elusive. We address these questions in the following chapters of this thesis as a report of our direct observation of individual events.

1.4 Cell cycle and DNA replication of *E.coli*

Single cell investigation also provides the possibility to correlate different factors contributing to cell-to-cell diversity. For example, in the case of the phase varying behavior of the *fim* switch of *E.coli*, it is still unknown whether a general physiological source of diversity, such as age within the cell cycle, volume and growth rate does effect the apparent behavioral diversity of individual cells. In this section, the principle behind the *E.coli* cell cycle is introduced (for a comprehensive review: [28]).

1.4.1 The *E. coli* cell cycle.

The division cycle not only describes the continuous birth, elongation and division of *E.coli* cells, but further includes processes during cellular growth, such as DNA replication, nucleoid partition and septum formation. Each process has to be coordinated in space and time to ensure a successful round of cell duplication. Cooper summarized the two main principles believed to govern *E.coli* cell cycle [28]. First, to ensure survival of daughter cells, a cell will not divide unless it has at least two copies of the genome. At the moment of division, the DNA content (and other materials) of a mother cell should be twice as much as that of a daughter cell so that the cell composition is stably maintained over time. Second, DNA replication will not start unless there is enough cytoplasm. Many important species in the cytoplasm are at low molecular concentrations, so that there needs to be enough proteins, enzymes, and other species for daughter cells to continue growth in the next cell cycle. Cell growth rates vary with the growth medium: the richer the medium, the shorter the cell interdivision time.

In *E.coli*, in order to double the DNA content before division, DNA replication is initiated only once in a single cell cycle. The DNA replication starts at the origin of the replication site *oriC*, and finishes at the termination site *ter*. Replication process is bidirectional, meaning that there are two replication forks proceeding in opposite directions around each half of the chromosome simultaneously. Experimental results with fast growing cells support a model for a constant period of DNA replication (about 40 minutes), called the C period. Thus, DNA replication proceeds at a constant rate of about 500 nucleotides per second. Another constant period (about 20 minutes), called the D period was observed between the end of DNA replication and cell division. Though the physiological function of the D period is unknown, it is proposed to be the time the cell needs to prepare for cell division, but the evidence for a molecular mechanism underlying it is lacking.

According to the Cooper-Helmstetter model of DNA replication [27], in a cell with an interdivision time of 60 minutes, DNA replication begins around birth time and terminates 20 minutes before division. Due to this required D period of 20 minutes, a cell that has an interdivision time of 50 minutes will have to start replicating its DNA 10 minutes before its birth, thus in the D period of its mother cell (fig. 1.4). Since it is possible for a cell to obtain a very short interdivision time (less than 40 minutes), a new round of replication will start before the previous one has ended. So, for example, in a cell that has an interdivision time of 30 minutes, DNA replication terminates 10 minutes after cell birth, and starts around the birth time of its mother cell, while the mother cell still undergoes the last 10 minutes of its own

of DNA replication process. DNA replication patterns for fast growing cells with various interdivision times are shown in figure 1.4. Variation in growth rates was normally obtained by changing the type of medium and carbon sources in which the cells were growing.

Different experimental methods for determining the C and D period have led to a consensus on the DNA replication pattern for fast growing cells. There is less consensus, however, regarding the DNA replication pattern in slow growing cells (interdivision time more than 60 minutes). There are two main groups of observations involving slow growing cells. In the first group, several studies found that DNA replication starts around birth time and the length of the C and D periods increases (and rescales) to occupy 2/3 and 1/3 of the cell cycle [47,94] . In contrast to these observations, others found that when interdivision time increases significantly, DNA replication starts a significant time after birth. Consequently, there is a period devoid of DNA replication at the beginning of the cell cycle, called the B period [24,63,93]. An extreme case showed a constant C and D period of 40 and 20 minutes even for cells with an interdivision time of 50 hours grown in chemostat [66,68]. To summarize, different strains and different methods have lead to different conclusions. Therefore, it is useful to also review the common methods used to determine the C and D period. Short descriptions of those methods are summarized in section 1.5.

1.4.2 Models of DNA replication initiation

The pattern of DNA synthesis during division cycle at different growth rates, leads to the question of what regulates the timely firing of DNA replication initiation. Possibly, cell division triggers DNA replication initiation. But, as growth become faster and faster, DNA replication is initiated before cell division up to 2 generations earlier. So, cell division can be excluded as the sole DNA replication initiator.

The question remains then, how can a cell or a genome know when to initiate DNA replication, when a waiting time is required for one growth condition but not for others? An interesting calculation was reported by Donachie in 1968 [33]. By combining the data of the average initiation age for *E.coli* B/r [27] with measurements of the DNA/mass ratio from *Salmonella typhimurium* [104], he proposed that the average cell mass per *oriC* site at inititation (called the initiation mass) was constant and did not vary with growth rate.

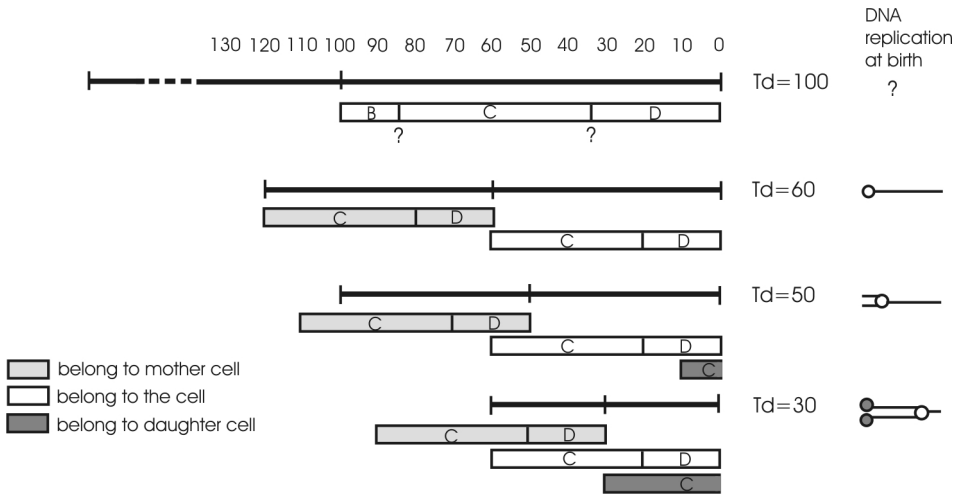

Figure 1.4. DNA replication patterns for *E.coli* with different growth rates.
As interdivision time (Td) decreases, the start of DNA replication occurs earlier. According to the Cooper-Helmstetter model of DNA replication, the time of DNA replication initiation is determined from the division time of the cell of interest (t=0). A D period of 20 minutes and a C period of 40 minutes are aligned from the end of each interdivision time. The start of the C period corresponds to the start of DNA replication. Each thin black bar represents two interdivision times, the first belongs to the mother cell and the second belongs to the cell of interest. The number of active replication rounds corresponds to the number of C periods. For simplicity, DNA replication at birth is depicted as unidirectional instead of bidirectional, and the chromosome as linear instead of circular. The replicated part of the chromosome appears as double lines, while the non-replicated part appears as a single line. The replication forks are depicted as circles with colors corresponding to the legend.

Experiments further validating the concept of initiation mass were performed following the introduction of the hypothesis. In one investigation, cells with abnormally high mass per origin started DNA replication simultaneously at all origins. First, DNA synthesis was inhibited in a thymine-requiring *E.coli* strain by thymine starvation. During this period of time, cell mass increased normally, resulting in a greater mass per origin than the normally required initiation mass for all origins in the culture. The starvation was then relieved and all origins were observed to initiate immediately. After approximately one interdivision time, the cell origin to mass ratio returned to the prestarvation level [34]. In another investigation, DNA replication rates were varied and the measured initiation mass was highly constant. When a thymine-requiring strain of *E.coli* was grown with less than the optimum thymine concentrations, the rate of DNA synthesis and thus the C period varied with thymine concentrations. It was shown that despite large

variations in the DNA replication rate, the measured initiation mass was the same [97].

The two experiments mentioned above are among the earliest investigations. They strongly suggest that there is a parallel accumulation of the cell initiation potential with cell mass, and that it is independent of the actual replication process. This leads to the long-held belief that cell mass, or something that correlates with it, is the initiator of DNA replication. A few later experiments found that initiation mass does change with growth rate. One experiment showed that initiation mass increases with increasing growth rate [26], while others found that it decreases[120]. These results indicate that initiation mass might vary, and that is important to seek the molecular entity that determines the initiation of DNA replication.

Models proposed for regulation of initiation can be classified as based on either a positive or negative regulation. It was first proposed by Cooper and Helmstetter [27] that DNA replication initiation may depend on the amount of an initiator protein which is produced continuously during the cell cycle. When a critical amount has been reached (proportional to cell mass or size), DNA replication commences (positive regulation). In order to ensure that initiation only occurs once, the hypothetical initiator is either made periodically or consumed immediately. Alternatively, a model based on negative regulation was proposed by Pritchard, Barth and Collins [96]. They proposed that there could be a burst in synthesis of an inhibitor protein, and that DNA replication occurs when the concentration of the inhibitor decreases to sub-inhibitory level as the cell volume increases during growth. To further explain the high precision and synchrony of replication initiation, both models suggested cooperation between multiple initiators/inhibitors. A more detailed description on the molecular mechanism of replication initiation is covered in section 1.6.

1.5 Methods of C and D Period Determination

To measure the C period, two major types of cell cycle analyses are commonly used: integral methods, which measure accumulation of a certain compound (in this case DNA), or differential methods, which measure the synthesis rate of a particular compound. Generally, differential methods are more successful, due to higher accuracy. Upon examination of the accumulated compound, the difference between e.g. linear and exponential accumulation could be very small. While upon comparison of the synthesis rates, the difference between a constant and exponentially increasing synthesis rate is more readily interpretable.

Generally, these methods can estimate the mean length of the C and D period quite precisely, but only a few techniques provide an estimate of the variance of these periods. The first technique is flow cytometry. By fitting theoretical curves to experimental data, both the mean and the degree of variability of B, C and D periods could be estimated. The drawback of this method is that the assumptions underlying the theoretical model could influence the data interpretation considerably [83]. Another technique used to estimate the variance of the C period is autoradiography. By measuring variability in lengths of labeled DNA strands for a given pulse duration, the maximum spread in a C period was found to be about 33% [71]. However, inherent errors exist in this method, DNA tends to shrink in quite a variable way during fixation and lengths of the tracks differ by at least a factor of two in any preparation. An alternative technique to estimate the variance of the D period combines DNA synthesis rate measurements in synchronous culture and modeling [18]. However, artifacts in the membrane elution technique could affect the data interpretation. Moreover, whether or not the assumptions incorporated in the model are also valid for slow growing cells is still unclear. So far, direct observations of these periods and their variances do not exist. Below, a summary of different methods to measure the C and D periods are described.

1.5.1 Methods for measuring the C period

1. Increment of DNA synthesis [16,78].

This technique is based on the assumption that upon inhibition of protein or RNA synthesis, replication initiation is prevented while elongation proceeds normally. This assumption is based on an observation that when thymine-requiring bacteria are starved for thymine, cells lose their viability in a characteristic manner. DNA still increases by about 40% or more, and during this DNA synthesis period a fraction of the population becomes immune to thymine-less death. The difference in accumulated DNA before and after protein synthesis inhibition (ΔG_t) is measured and supposed to be a function of only the C period (C) and cell interdivision time (T_d) according to the relation:

$$\Delta G_t = \frac{G_t}{G_o} = \frac{C(\ln 2)}{T_d}(1 - 2^{-C/T_d})^{-1}$$

Thus the C period can be estimated.

2. Rate stimulation [128].

This technique is based on the observation that when an exponentially growing thymine-requiring culture is starved for thymine, the rate of DNA synthesis (thymine incorporation rate) upon thymine restoration is faster than before starvation. The degree of this increase is proportional to the length of the starvation time, up to a maximum value reached after one mass-doubling time. With a model

that accounts for cell cycle parameters, DNA replication parameters, and culture age distribution, it was shown that the ratio of DNA synthesis rate after and before starvation *(R)* is a function of the mean interdivision time *(T_d)* and the C period *(C)*:

$$R = (2^{C/Td+1} - 1) / (2^{C/Td} - 1)$$

Therefore, by measuring this DNA synthesis rate (for example by a [^{14}C]thymine incorporation technique) and the culture mean interdivision time (by optical density measurements) before and after starvation, the C period can be estimated.

3. DNA-to-mass ratios [104].
Schaechter, Maaloe and Kjeldgaard first showed that mass increases exponentially as a function of growth rate, as in the case of DNA synthesis, though with lower exponent. Thus, as the growth rate increases, the amount of DNA per unit mass (DNA concentration) decreases. Cell mass is measured by the optical density at 450 μm while DNA content is determined by colorimetric technique. Interestingly, the same data showed that cell mass at initiation is constant over various growth rates. Experimental data for different cellular growth rates fit a model of DNA concentration as a function of only the C period, cell interdivision time and initiation mass. If the initiation mass is a constant, then the DNA-to-mass ratio at one growth rate can be used to predict the DNA-to-mass ratio at other growth rates. This would confirm the constancy of C. Alternatively, if the value of C and the DNA-to-mass ratio for one growth rate is known, then the C period for other growth rates can be calculated.

4. Autoradiographic analysis of DNA chain extension [22].
The basic idea of this technique is to label DNA by exposing a cell to radioactive thymidine for a short time. The cell is then gently lysed, the DNA is spread out on a slide, and subsequently processed for autoradiography. Short sequences of radioactive segments are measured, and with the knowledge of the previous labeling time, the DNA synthesis rate of a replication fork can be obtained. The size of the entire genome can also be estimated so that an approximation of the C period can be calculated.

5. Analysis of gene frequency [24].
The frequency of a certain gene depends on the number of replication points in the chromosome and its relative position from the origin of replication. In this method, two different chromosomal markers at different positions are measured for their frequency ratio (by DNA:DNA hybridization). This ratio *(a/b)* is then investigated as a function of growth rate by examining cells grown in various media. It is found that the shorter the interdivision time, the higher the frequency difference. A mathematical relationship that describes the frequency ratio as a function of the

culture's interdivision time *(T_d)*, mean replication time *(C)*, and distance between the two genes *(Δd)* are subsequently used to estimate the C period:

$$a / b = 2^{C*\Delta d/Td}$$

6. The backward membrane elution method [17].
Also called the baby-machine, exponentially grown cells are pulse-labeled (for example with [^{14}C]thymidine for DNA labelling), then filtered over a nitrocellulose membrane, and washed with fresh, prewarmed medium to remove excess label. Subsequently, the membrane is inverted, and fresh medium is pumped through the membrane so that bound cells can continue to grow and divide. During the elution, newborn cells are released into the medium (synchronized cells). These newborn cells are daughter cells from the oldest (closest to division) membrane bound cells. In a steady state exponential culture, the number of newborn cells is always twice as much as the number of dividing cells, hence the number of eluted cells over time resembles a saw tooth pattern.

The radioactivity patterns per cell during elution reflect the "backward" synthesis rate of the compound of interests (in this case DNA). For example, when slow growing cells with interdivision time longer than C+D are pulse labeled, the label will only be incorporated in the cells during DNA replication, or in other words in their C period. These cells will give rise to radioactive eluted daughter cells that would elute between D and D+C minutes, and thus the length of the C and D periods can be estimated. By investigating DNA synthesis patterns of cells with various interdivision times, a model of the DNA replication pattern can be developed.

7. Sucrose synchronized cultures [17].
In this method, cells are separated according to size by putting a concentrated cell sample on top of a linear sucrose gradient and then centrifugating them. Cells within a narrow size class, are selected to produce synchronized culture and the DNA synthesis pattern is subsequently analyzed. The smallest class of cells is assumed to be newborn cells, although the smallest cells can also originate from sick and abnormally growing cells.

8. Synchronization of DNA replication [74].
First, all ongoing replication in the cells are allowed to finish without allowing any increase in cell mass by amino acid starvation. Then the cells were starved of thymine in the presence of amino acids to induce synchronous DNA synthesis initiation. Samples were exposed to radioactive thymine at intervals and the replicated portions of the genome were determined by DNA-DNA hybridization.

The C period was found to be 40 minutes in rich medium and 52 minutes in glucose minimum medium.

9. Transductional analysis [46].
A different approach to determine the C period is used in this method. *E.coli* cells are transducted with a P1 bacteriophage and then samples are taken at discrete timesteps. Each sample is plated on different selective plates to follow the appearance of transductants with genetic markers positioned at various points on the chromosome. Apparently, the closer a certain marker is to the *ter* site, the earlier its corresponding transductants increase in numbers. The time interval between the appearances of two different transductants *(ΔT)* is a simple function of their markers chromosomal positions *(d_a* and *d_b)* and the DNA replication rate, whereby the C period *(C)* can be easily deduced:

$$\Delta T = \frac{d_a \cdot C}{50} - \frac{d_b \cdot C}{50}$$

10. DNA steptime measurement [80].
DNA steptime is the average time for adding one more nucleotide to a growing DNA chain. This technique takes advantage of DNA chain growth in the 5' to 3'direction. When DNA is hydrolyzed, nucleotides at the end of a replicating chain stay in the form of 3'-OH while interior nucleotides yield 3'-monophosphate. By incorporating a radioactive label into the cell's DNA (labeling time 6-18 seconds), the ratio of internal to end label in the DNA can be determined. By developing a mathematical model that considers the relationship between measured radioactivity values, labeling time and enzyme kinetics, the DNA steptime could be estimated. The result shows a chain growth rate of a predicted 500-700 nucleotides per second. This result combined with the length of the whole chromosome results in a very close agreement with other measured values of C.

11. Flow cytometry [108].
This method allows single-cell measurement of fluorescently labeled DNA content in individual bacteria in an exponentially growing culture by observing the cells pass over an optical system that measures cell fluorescence. Then, the experimental distribution of the cell's DNA content is fit with a theoretical distribution that considers variation in the cell population's age, interdivision time, DNA replication initiation with methodological variation. The average B, C and D periods of the population are inferred from the best theoretical fit. However, assumptions underlying the model could influence the obtained values.

1.5.2 Methods for measuring the D period

1. Residual cell division after inhibition of DNA synthesis [47].
This method assumes that except by lack of termination, inhibition of DNA replication does not disturb cell division. The D period is estimated from the residual division time. DNA replication in membrane bound cells is inhibited by either UV radiation, mitomycin addition, or thymine starvation. After 20 minutes of normal increase in the eluted cell number, there is a dramatic decrease in the eluted cell number. The data is interpreted as the amount of time needed for cells that have terminated their DNA replication to divide. This value of the D period was found to be constant and independent of interdivision time.

2. Residual cell division after inhibition of protein synthesis [67].
This particular method of measuring D period is quite simple and requires only the ability to count cells. Based on the proposal that the replication of the terminal part of chromosome requires the ability to synthesize protein [81], it is believed that cells that divided after the addition of chloramphenicol are cells that have finished DNA replication. Chloramphenicol is added to the medium to inhibit protein synthesis. Thus, by measuring the increase in cell number after chloramphenicol addition (the residual cell division) and growth rate, one can infer the length of the D period. The main critique on this method is the usage of chloramphenicol, which might have side effects on the cell division process in general (not D period related). The average D periods for exponentially growing cultures obtained with this method is constant, and does not vary with a vast range of growth rates.

1.6 DNA replication initiation

Although the model for the regulation of DNA replication initiation is still under debate, there are currently more findings supporting the positive regulation model. A protein that is specifically involved in the initiation process, DnaA, is commonly believed to be the initiator protein. Before DNA can be replicated, several discrete steps have to occur. First, the origin recognition complex (ORC) is formed at *oriC* by the intitiator DnaA protein, in either its ATP or ADP bound form. DnaA protein binds to three 9 bps binding sites within *oriC* called R1, R2 and R4. Then, the weaker recognition sites R3 and R5 are occupied in addition to the three I-boxes specific for DnaA-ATP (for reviews: [57,86,89]).

Next, accessory proteins such as IHF, HU and DiaA induce conformational changes on the DNA structure [62,101], where it becomes wrapped around the right handed DnaA-ATP filament. This complex promotes double strand separation

at adjacent AT rich regions. The separation is stabilized by extra DnaA-ATP protein binding to specific 6 bps sequences in the single stranded region. Subsequently, the 'licensing' for replication starts with the recruitment of the DnaB helicase by DnaC-ATP in the form of a B6C6 complex [6] to further open the unwound DNA. DnaC is released after hydrolysis of its ATP. The whole complex is now called the pre-replicative complex (pre-RC).

DNA replication starts when two or three DNA polymerase III holoenzymes are loaded on the *oriC*, and when dNTP are available, the replication begins. Once replication begins, three processes occur to prevent extra-unwanted replication initiation. First, the conversion of DnaA–ATP into the inactive DnaA-ATP form by the complex of an Hda protein and β–clamp of the DNA polymerase III holoenzyme [60]. Second, *oriC* is sequestered by SeqA protein. Once replication has started, hemimethylated regions are produced. The hemimethylated *oriC* is a target of sequestering by the SeqA protein (for a more elaborate explanation refer to section 1.7) rendering it inaccessible to the DnaA protein. In addition, the promoter of the *dnaA* gene which is close to the *oriC* is sequestered at the same time, preventing transcription and DnaA protein synthesis [23]. Third, *in vivo* DnaA protein is titrated by the newly duplicated reservoir sites during the completion of DNA replication. There are 308 evenly distributed R-type DnaA boxes with different affinities along the *E.coli* chromosome. All three mechanisms mentioned above work together to lower the replication initiation potential of the cell, highly regulating the frequency of chromosome replication. Nonetheless, the exact timing of DNA replication initiation still varies from cell to cell. The individual DnaA protein concentration (in its ATP and ADP forms) plus other components of the complex in a cell could be responsible for this timing "noise".

1.7 DNA sequestering protein SeqA

Initiation of the DNA replication process in *E.coli* has to be well regulated to avoid multiple replication initiations in an untimely manner. One protein that is involved in negative regulation of DNA replication initiation is the DNA sequestering protein SeqA. In *E.coli*, when fully methylated parental DNA is replicated, the resulting daughter strand is hemimethylated until it is methylated by Dam methyltransferase. By binding to the hemimethylated, GATC-rich origin of replication, the DNA sequestering protein SeqA prevents re-initiation of DNA replication [13,58,76,82].

SeqA forms a homotetramer from two interacting dimers to bind stably to the hemimethylated DNA. One SeqA dimer binds to a pair of hemimethylated GATC sites separated by a maximum distance of about 31 bps. There are 1750 pairs of

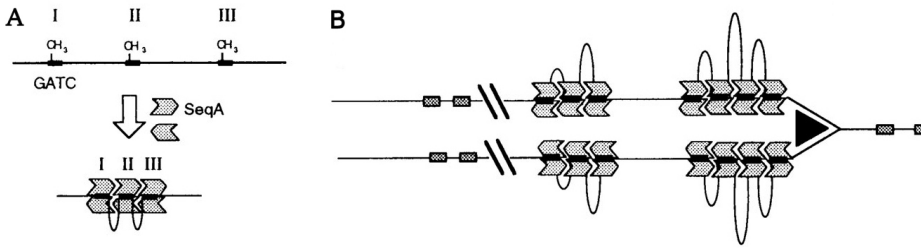

Figure 1.5 Proposed model of SeqA cluster (adapted from [19]).
(A) Hemimethylated GATC sites are targets for SeqA protein.
(B) DNA bound SeqA proteins in close proximity are able to form filaments and clusters.

GATC sites with spacings up to 31 bps rather evenly distributed along the length of the whole chromosome [19]. A recent study has shown that SeqA dimers are able to oligomerize into a filament, and mutations that disrupt filament formation lead to asynchronous DNA replication [44]. SeqA is also known to bind double-stranded methylated DNA, but with much less affinity. A model of DNA sequestering by SeqA was proposed where SeqA arranges itself into clusters along the DNA so that the DNA in between GATC sites forms loops (fig. 1.5).

In vivo, fluorescently labeled SeqA clusters are visible as discrete foci [51]. These foci are thought to be SeqA clusters on DNA plus aggregate of free SeqA proteins [59]. The number of these foci increases as the interdivision time decreases. Recently, a ChIP assay has shown that the SeqA protein binds sequentially to hemimethylated nascent DNA segments following replication fork movement in synchronized cultures of *E.coli* [125]. Therefore, the SeqA foci are assumed to represent the replication forks [85]. In this thesis the dynamics of SeqA foci *in vivo* is studied in slow growing cells.

1.8 This thesis

The outline of this thesis is presented below. The *E.coli fim* system, its cell cycle and their interconnection were studied at the single cell level using timelapse microscopy. The content of this thesis is organized as follows:

Chapter 1 introduces all subjects relevant for later chapters and some background information.
Chapter 2 covers experimental methods and materials involved in chapter 3, 4 and 5.

Chapter 3 contains the results of our single-cell level investigation into the *E.coli* cell cycle and DNA replication with a fluorescently labeled SeqA protein.
Chapter 4 contains the results of our investigation of cell-to-cell behavioral diversity within the *fim* system of *E.coli* and its relation to the cell cycle.
Chapter 5 contains the results of our investigation on the effect of switching history on *fim* OFF switching behavior.