



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1198*

Biased Evolution

Causes and Consequences

GERRIT BRANDIS



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6206
ISBN 978-91-554-9518-3
urn:nbn:se:uu:diva-276456

Dissertation presented at Uppsala University to be publicly examined in A1:107a, BMC, Husargatan 3, Uppsala, Monday, 9 May 2016 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Erik Böttger (Institute for Medical Microbiology, University of Zürich, Switzerland).

Abstract

Brandis, G. 2016. Biased Evolution. Causes and Consequences. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1198. 48 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9518-3.

In evolution alternative genetic trajectories can potentially lead to similar phenotypic outcomes. However, certain trajectories are preferred over others. These preferences bias the genomes of living organisms and the underlying processes can be observed in ongoing evolution.

We have studied a variety of biases that can be found in bacterial chromosomes and determined the selective causes and functional consequences for the cell. We have quantified codon usage bias in highly expressed genes and shown that it is selected to optimise translational speed. We further demonstrated that the resulting differences in decoding speed can be used to regulate gene expression, and that the use of ‘non-optimal’ codons can be detrimental to reading frame maintenance. Biased gene location on the chromosome favours recombination between genes within gene families and leads to co-evolution. We have shown that such recombinational events can protect these gene families from inactivation by mobile genetic elements, and that chromosome organization can be selectively maintained because inversions can lead to the formation of unstable hybrid operons.

We have used the development of antibiotic resistance to study how different bacterial lifestyles influence evolutionary trajectories. For this we used two distinct pairs of antibiotics and disease-causing bacteria, namely (i) *Mycobacterium tuberculosis* that is treated with rifampicin and (ii) *Escherichia coli* that is treated with ciprofloxacin. We have shown that in the slow-growing *Mycobacterium tuberculosis*, resistance mutations are selected for high-level resistance. Fitness is initially less important, and over time fitness costs can be ameliorated by compensatory mutations. The need for rapid growth causes the selection of ciprofloxacin resistance in *Escherichia coli* not only to be selected on the basis of high-level resistance but also on high fitness. Compensatory evolution is therefore not required and is not observed.

Taken together, our results show that the evolution of a phenotype is the product of multiple steps and that many factors influence which trajectory is the most likely to occur and be most beneficial. Over time, selection will favour this particular trajectory and lead to biased evolution, affecting genome sequence and organization.

Keywords: Evolution, Codon usage bias, Post-transcriptional regulation, Recombination, Inversion, EF-Tu, Frameshift suppression, Antibiotic resistance, Rifampicin, Ciprofloxacin, Compensatory evolution, Drug efflux, RNA polymerase, DNA gyrase

Gerrit Brandis, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Gerrit Brandis 2016

ISSN 1651-6206

ISBN 978-91-554-9518-3

urn:nbn:se:uu:diva-276456 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-276456>)

I was born in 1985 in Bielefeld, Germany and graduated from the University of Bielefeld with a Bachelor degree in Biochemistry in 2009. I moved to Uppsala for my Master studies and received a Master degree in Infection Biology in 2011. The same year, I joined Diarmaid Hughes's research group for my PhD studies.

e-mail: gerrit.brandis@gmail.com



To my family

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I **Brandis, G.**, Hughes, D. (2016) The Selective Advantage of Synonymous Codon Usage Bias in *Salmonella*. *PLoS Genetics*, 12(3): e1005926
- II **Brandis, G.**, Bergman, J. M., Hughes, D. (2016) Autoregulation of the *tufB* operon in *Salmonella*. *Molecular Microbiology*, doi: 10.1111/mmi.13364
- III Huseby, D. L., Praski-Alzrigat, L., **Brandis, G.**, Hughes, D. (2016) High-level intrinsic suppression of a frameshift mutation in an essential gene. *Manuscript*
- IV **Brandis, G.**, Hughes, D. (2016) Co-evolution protects the *tuf* genes from inactivation by mobile genetic elements. *Manuscript*
- V **Brandis, G.**, Hughes, D. (2016) A large chromosomal inversion causes the formation of an unstable *tufA/B* hybrid operon. *Manuscript*
- VI **Brandis, G.**, Wrände, M., Liljas, L., Hughes D. (2012) Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology*, 85(1): 142–151
- VII **Brandis, G.**, Hughes, D. (2013) Genetic characterization of compensatory evolution in strains carrying *rpoB* Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *Journal of Antimicrobial Chemotherapy*, 68: 2493–2497
- VIII Meftahi, N., Namouchi, A., Mhenni, B., **Brandis, G.**, Hughes, D., Mardassi, H. (2016) Evidence for the critical role of a secondary site *rpoB* mutation in the compensatory evolution and successful transmission of an MDR tuberculosis outbreak strain. *Journal of Antimicrobial Chemotherapy*, 71: 324–332

- IX **Brandis, G.**, Pietsch, F., Alemayehu, R., Hughes, D. (2015) Comprehensive phenotypic characterization of rifampicin resistance mutations in *Salmonella* provides insight into the evolution of resistance in *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy*, 70(3): 680–685
- X Pietsch, F.* , Garoff, L.* , Huseby, D. L.* , Lilja, T., **Brandis, G.**, Hughes, D. (2016) Experimental evolution identifies a new class of genes selected during the development of ciprofloxacin resistance in *Escherichia coli*. *Manuscript*
- XI Pietsch, F., Bergman, J. M., **Brandis, G.**, Marcusson, L. L., Zorzet, A., Huseby, D. L., Hughes, D. (2016) RNA polymerase mutations contribute to the evolution of ciprofloxacin resistance in *Escherichia coli*. *Manuscript*
- XII Pietsch, F.* , Huseby, D. L.* , **Brandis, G.*** , Tegehall, A., Garoff, L., Hughes, D. (2016) Mutation supply and relative fitness shape genotypes of ciprofloxacin-resistant *Escherichia coli*. *Manuscript*

* These authors contributed equally

Reprints were made with permission from the respective publishers.

Review article, not included in the thesis.

Hughes, D., **Brandis, G.** (2013) Rifampicin Resistance: Fitness Costs and the Significance of Compensatory Evolution. *Antibiotics*, 2(2): 206–216

Contents

Introduction	11
Evolution	11
Mutation	11
Natural selection	12
(Population) size matters	13
A biased genome	15
Biased nucleotide distribution	15
Biased gene distribution	16
Biased gene composition	19
Antibiotic resistance as a tool to study evolution	22
The emergence of antibiotic resistance	22
Fitness cost and compensatory evolution	24
Antibiotic resistance and evolution	25
Present Investigations	26
Papers I – III	26
The selective advantage of codon usage bias (Paper I)	26
Codon usage and gene regulation (Paper II)	27
Codon usage and reading frame maintenance (Paper III)	28
Papers IV – V	29
Biased gene distribution and its effects on recombination	29
Recombination between the <i>tuf</i> genes (Papers IV – V)	29
Papers VI – IX	31
Rifampicin resistance in <i>Mycobacterium tuberculosis</i>	31
Fitness-compensatory mutations in rifampicin-resistant RNA polymerase (Papers VI – VIII)	32
Selection of clinical rifampicin resistance (Paper IX)	33
Papers X – XII	33
Ciprofloxacin resistance in <i>Escherichia coli</i>	33
Identification of a novel class of ciprofloxacin resistance mutations (Paper X – XI)	34
Selection of clinical ciprofloxacin resistance (Paper XII)	35
Concluding Remarks	37
Acknowledgements	39
References	42

Abbreviations

5'-UTR	5' untranslated region
A	Adenine
bp	Base pair
C	Cytosine
CDS	Coding sequence
DNA	Deoxyribonucleic acid
EF-G	Elongation factor G
EF-Tu	Elongation factor Tu
G	Guanine
HGT	Horizontal gene transfer
IF	Initiation factor
kb	Kilo base pairs
L	Litre
nt	Nucleotide
Mb	Mega base pairs
MDR-TB	Multidrug-resistant tuberculosis
mg	Milligram
MIC	Minimal inhibitory concentration
mRNA	Messenger RNA
MTB	<i>Mycobacterium tuberculosis</i>
OriC	Origin of replication
RF	Release factor
RNA	Ribonucleic acid
RRDR	Rifampicin resistance determining region
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
s	Second
<i>S. typhimurium</i>	<i>Salmonella enterica</i> serovar Typhimurium
T	Thymine
TB	Tuberculosis
Ter	Terminus
tRNA	Transfer RNA

Introduction

“Mutation is random; natural selection is the very opposite of random.”
– Richard Dawkins

Evolution

The theory of evolution is one of the main pillars that modern biology rests upon. Evolution is the complex and never-ending interplay of mutation with genetic drift and natural selection. Mutations cause new traits to appear and natural selection is a filter that removes less fit variants, leading to the enrichment and spread of individuals with beneficial traits. These new traits become especially important under novel conditions. A population that colonizes a new niche, together with a dash of time, can be the recipe for the evolution of a new species.

Mutation

A continuous supply of mutations is the fuel that drives evolution. Without this constant variation, a population would be stuck in time in an ever-changing world, destined to go extinct. Mutations as such might appear at random, but there is a strict hierarchy in the different types of mutations. On a phenotypic level, mutations can be classified into three different categories: (i) mutations that cause no change in phenotype, (ii) gain-of-function mutations and (iii) loss-of-function mutations. Mutations that cause no change in phenotype, per definition, have no impact on evolution and will not be part of the further discussion. Gain-of-function mutations usually cause subtle changes to a protein by amino acid substitutions or small in-frame deletions or insertions (Brandis *et al.*, 2015). The number of mutations that can give rise to a specific phenotype is limited, and the frequency of the phenotype appearing is accordingly usually very small. Depending on the specific phenotype, the frequency of mutations may vary within a few orders of magnitude but is approximately 10^{-10} per generation (Drake, 1991; Abdulkarim and Hughes, 1996). Loss-of-function mutations on the other hand are found more frequently due to the larger target size. Almost any deletion or insertion throughout the entire gene that disturbs the protein reading frame will lead to the production of a non-functional protein, which sums up to an

average frequency of 10^{-7} per generation (Schaaper *et al.*, 1986; Drake, 1991). This 1,000-fold difference in mutation rate would favour the accumulation of loss-of-function mutations over gain-of-function mutations if it weren't for the functional consequences and their effects on fitness.

Natural selection

The appearance of new mutations in a genome is not necessarily a positive event. Most mutations will be deleterious for protein function, and once a protein is inactivated it is unlikely that a secondary mutation will restore its function. Over time more and more proteins would get inactivated until the population was driven to extinction from the cumulative loss of fitness. Natural selection is the force that prevents this reductive evolution from happening. Deleterious mutations might appear at all times but they will most likely reduce the fitness of their host. This reduced fitness will take care that the frequency of the mutation does not increase within the population but that the mutation will instead go extinct. On the other hand, mutations that increase fitness will have a selective advantage and increase in frequency within the population.

The strength of natural selection varies greatly depending on the selective conditions. At the upper end is a selection coefficient of one, which means that only the mutant population can survive. Examples for this condition are the selection with a lethal dose of antibiotics or an environment in which the wild-type population cannot grow due to inability to process vital nutrients. The lower end of the spectrum is much debated but estimates reach down to selection coefficients that are as low as 10^{-9} (Hartl *et al.*, 1994). For practical reasons it has not yet been possible to experimentally confirm these estimates. Independently of the precise values, the two ends of the scale represent two very different types of evolution (Figure 1).

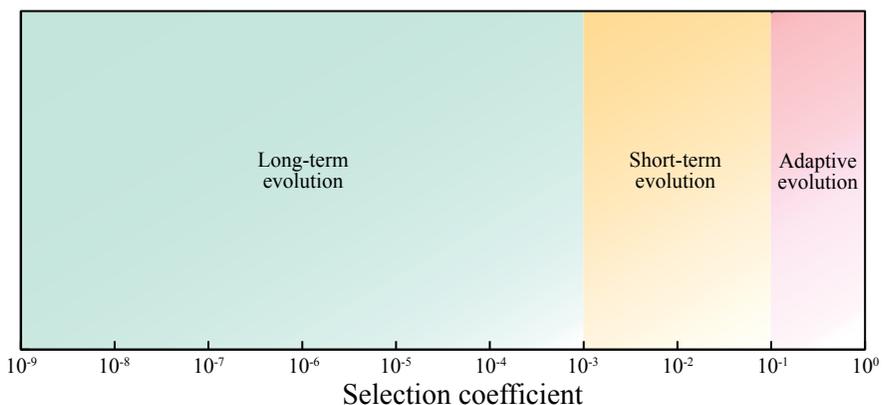


Figure 1. Overview over different types of evolution and their respective selection coefficients.

Values around a selection coefficient of one, simplified as $10^{-1} - 1$ (Figure 1), represent adaptive evolution to conditions that are novel for the population and in which only certain mutants can grow efficiently. These mutations give the sub-population new abilities, for example survival in the presence of an antibiotic or the ability to access a new nutrient source, and enable it to populate a new niche. Under certain circumstances this might be the first step in the development of a new species. Below these values, with selection coefficients in the approximate range of $10^{-3} - 10^{-1}$ (Figure 1), selection is still strong but the wild-type population is able to grow without problems. Mutations with these selection coefficients will get fixed within the population during short-term evolution and represent mutations that increase the viability under the present conditions, for example an increased metabolic efficiency. Selection at the lower end, $10^{-9} - 10^{-3}$ (Figure 1), is weak and mutations that have a selection coefficient in this range will need a long time to be fixed within the population. Long-term evolution that will favour these types of mutations requires very stable conditions that are usually not found in the environment. It is more likely that these selection coefficients represent optimization of cellular processes that are conserved and that do not change easily over time, such as the translation machinery.

The values presented above are far from precise and might vary greatly for different species and conditions, and they are intended to give an example of how the strength of selection varies under different circumstances and how even the tiniest selection can have large effects given sufficient time.

(Population) size matters

Natural selection not only helps a population to adapt to new environments but also protects it from deleterious mutations. The efficiency with which natural selection does this is highly limited by the population size (Haigh, 1978). Deleterious mutations can get fixed if the population goes through a small bottleneck (O'Brien *et al.*, 1985; Dobrynin *et al.*, 2015). The most extreme example of this is a single-individual bottleneck. Under these circumstances any mutation that appears in the individual that passes through the bottleneck will subsequently be fixed within the whole population. The population size has, also under less extreme conditions, major implications for the outcome of evolution. If a bacterial population is exposed to a new condition, such as the presence of a lethal dose of antibiotics, the population size coupled with the mutation rate will determine which types of mutations are present within the population. If the required phenotype can only be acquired by specific gain-of-function mutations, then the population size needs to be approximately 10^{10} cells so that one of the cells has the required mutation. On the other hand, if a loss-of-function mutation can give the required phenotype, a population size of 10^7 cells may be sufficient (Figure 2).

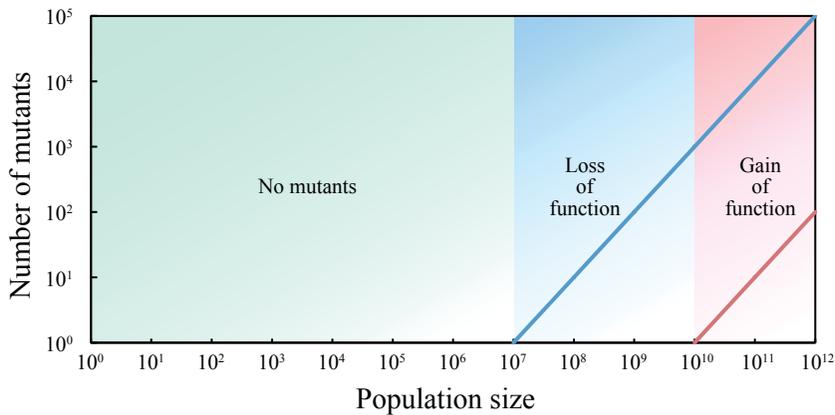


Figure 2. Effect of population size on mutant selection. The average number of loss-of-function (blue line) and gain-of-function (red line) mutations for a specific phenotype is shown as a function of population size. The critical population sizes for the appearance of loss-of-function (blue) and gain-of-function (red) mutations are indicated with background colour.

These numbers give the minimal population size required for an average of one mutant to be present in the population. Under these conditions, all cells but the resistant mutant will die, and the population goes through a single-cell bottleneck. The resistance mutation, however deleterious it might be, will be fixed within the new population. A more interesting scenario is a population size that allows multiple mutations to be present within the population. If both gain-of-function and loss-of-function mutations can give rise to the selected phenotype, and the population size is 10^{10} cells, then there should be, on average, one single gain-of-function and one thousand loss-of-function mutations present within the population. After exposure to a lethal dose of antibiotics, only these one thousand and one mutants will survive and grow into a new population. The gain-of-function mutation might be the least frequent mutation within the population at this point but may also be the fittest mutant since the loss of a protein function is often accompanied by a larger fitness cost (Marcusson *et al.*, 2009). Under these circumstances, the frequency of the gain-of-function mutant will increase as the population grows until it is the only mutant present.

The present example might be oversimplified but shows how mutation rate coupled with population size determines the number of different mutants within the population and how natural selection will subsequently favour the fittest mutation and lead to the fixation of a single mutation within the population.

A biased genome

Millions of years of evolution have left their mark on the genomes of all living organisms. Many features within the genome show strong biases that cannot have occurred by chance but rather are the result of long-term selection. These genome features can be divided into three categories: (i) the distribution of nucleotides, (ii) the distribution of genes and (iii) the composition of genes. The genome of *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) strain LT2 shows how the delicate interplay between mutation and selection has shaped a bacterial genome within each of the above-mentioned categories.

Biased nucleotide distribution

The chromosome of *S. typhimurium* LT2 is circular and approximately 4.8 Mb in length. Replication starts bi-directionally at the origin of replication (Prescott and Kuempel, 1972) and ends at the terminus (Kuempel *et al.*, 1973), which divides the chromosome into two sides (Figure 3).

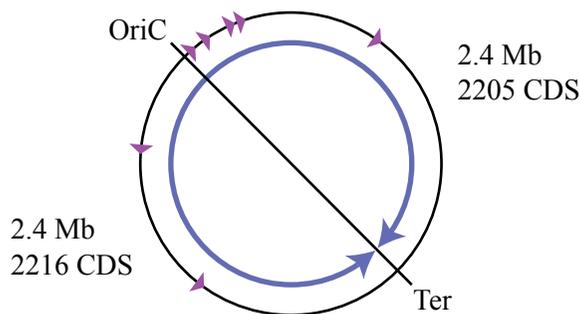


Figure 3. Structure of the *Salmonella* chromosome. The origin of replication (OriC) and the terminus (Ter) divide the circular chromosome into two equally sized halves. Blue arrows indicate the leading strand replication direction. The location and direction of the seven ribosomal RNA operons is shown with purple triangles.

Both sides are approximately 2.4 Mb in length and the size difference between them is only 1% (McClelland *et al.*, 2001). This equal distribution is most likely the results of selection to optimize replication. DNA polymerases start replication of both sides of the chromosome at the same time and will reach the terminus simultaneously only if the two sides are equally long. Large chromosomal rearrangements that result in an imbalance of these lengths cause large fitness costs (Hill and Gray, 1988; Esnault *et al.*, 2007) and show how selection is a strong force in conserving the equal length of both sides of the chromosome.

Replication is responsible for two additional genome features commonly observed in many species. In *Salmonella*, the leading strand of replication

contains 5% more guanine (G) and thymine (T) bases than adenine (A) and cytosine (C). This phenomenon is referred to as GT-bias and is thought to be the result of differences in the replication of the leading and the lagging strand (Lobry, 1996; Frank and Lobry, 1999). So far no functional consequences are known for this bias but it has been helpful to identify the replication boundaries in unannotated bacterial genomes (Grigoriev, 1998; Frank and Lobry, 2000). Additionally, many organisms show a bias in the overall GC vs. AT content. The *Salmonella* genome has only a minor bias in the GC content (52% GC) but the GC content of other organisms varies between 17% in *Carsonella ruddii* (Nakabachi *et al.*, 2006) and 75% in *Micrococcus luteus* (Muto and Osawa, 1987). GC bias is usually equally distributed throughout the genome, including non-coding sequences, which indicates that it is the result of the balance between mutation and repair (Sueoka, 1962; Muto and Osawa, 1987). It has been suggested that the GC content of a genome is, at least for some organisms, also a reflection of the temperature that the organism lives at (Zheng and Wu, 2010). GC base pairs form three hydrogen bonds, while AT base pairs form only two, and DNA sequences, as well as structured RNAs, that are rich in GC content have been shown to be more thermo-stable than others (Watson and Crick, 1953; Breslauer *et al.*, 1986). This biased nucleotide content has major implications for the evolution of a species. The GC bias within a genome reflects the bias in mutation and repair. A high GC content means that mutations from an AT bond to a GC bond are more likely than vice versa. This skew in mutational supply will favour GC-rich codons and thereby cause a bias as to which codons, and which amino acids are preferred in the population. Accordingly, species with low GC content will be biased towards amino acids with AT-rich codons. This basic feature of the genome limits the evolutionary trajectories that will be sampled by a species and can be a major contributor in determining the outcome of evolution.

Biased gene distribution

The two halves of the chromosome are not only equal in size but also contain the same amount of protein coding sequences. In *Salmonella* each half contains about 2200 coding genes and the difference between both sides is only 0.5% (Figure 3). This equal distribution is, unlike the size of the chromosome halves, most likely not the result of selection. It is more likely that it reflects the fact that the average protein is 316 amino acids long so that the gene content of the chromosome is one coding sequence (CDS) per kilo base pairs of DNA. Therefore, two equally large halves contain the same amount of protein coding sequences.

Even though the distribution of genes between the two halves is possibly not subject to selection, the location of genes is not random. The majority of genes (59%) are located on the leading strand of replication (Figure 4A).

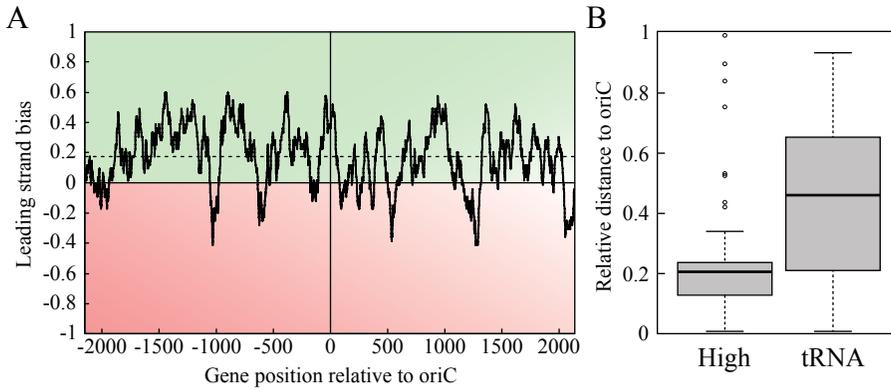


Figure 4. Gene location bias on the *Salmonella* chromosome. **(A)** Leading strand bias with a 75-gene window. Positive values (green background) represent biased location on the leading strand and negative values (red background) biased location on the lagging strand. Dashed line is the average leading strand bias throughout the chromosome **(B)** Relative distance to the origin of replication for a subset of highly expressed genes (High) compared to all tRNA genes (tRNA) that are an example of a randomly distributed gene family.

This bias is particularly strong for highly transcribed genes. Seventy-eight out of eighty-two genes (95%), that are involved in transcription and translation (elongation factors, RNA polymerase genes, as well as ribosomal RNAs and proteins) and that are among the most highly transcribed genes in *Salmonella*, are located on the leading strand (McClelland *et al.*, 2001).

The most likely explanation for this strand bias is the interplay of replication and transcription. Replication and transcription happen simultaneously, which inevitably leads to collisions between the DNA and RNA polymerases. The DNA polymerase moves with a speed of ~ 600 bp/s along the chromosome (Reyes-Lamothe *et al.*, 2008), which is about twenty times faster than the speed of the RNA polymerase (25 – 42 nt/s) (Proshkin *et al.*, 2010; Chen *et al.*, 2015). This speed difference leads to two types of polymerase collisions: (i) head-on collisions and (ii) collisions where the DNA polymerase catches up with the RNA polymerase (Figure 5).

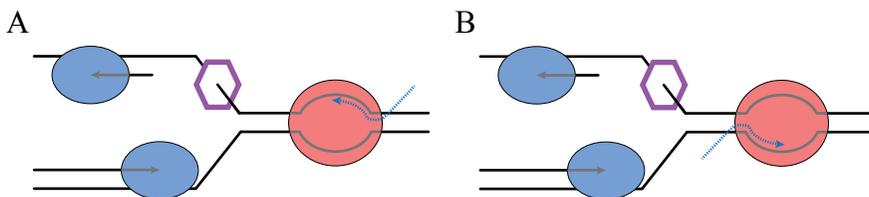


Figure 5. Schematic view of polymerase collisions during **(A)** opposite directional and **(B)** co-directional replication and transcription. DNA is shown as a black line (grey within polymerases) and RNA as a dashed blue line. Arrowheads indicate polymerase direction. DNA polymerases are shown as blue and RNA polymerase as red circles. DNA helicases are shown as a purple hexagon.

Only genes that are located on the lagging strand of replication will cause head-on collisions, which are thought to cause replication fork arrest (French, 1992; Mirkin and Mirkin, 2005) and are therefore potentially more detrimental to replication. Highly expressed genes have a higher RNA polymerase load than other genes and are therefore likely to be the main cause of polymerase collisions, which leads to a strong selection for highly expressed genes to be located on the leading strand of replication, as this would reduce the frequency of head-on collisions.

The leading strand bias is not the only bias in the location of genes. Highly expressed genes are not randomly distributed across the genome. On both sides of the chromosome, most highly expressed genes are located in a small range around one fifth of the distance between the origin and the terminus of replication (Figure 4B). The increased effective gene copy number in this region might explain this biased distribution. During rapid growth, new rounds of replication start before the previous round is finished (Cooper and Helmstetter, 1968; Helmstetter and Cooper, 1968) (Figure 6A). Genes that are located close to the origin of replication will, at least transiently, be present in multiple copies within the cell while genes close to the terminus remain single copied (Figure 6). The effective copy number of each gene is therefore a function of distance to the origin. Selection has probably caused highly expressed genes to be closer to the origin of replication in order to increase their gene dosage and by this means increase their expression levels.

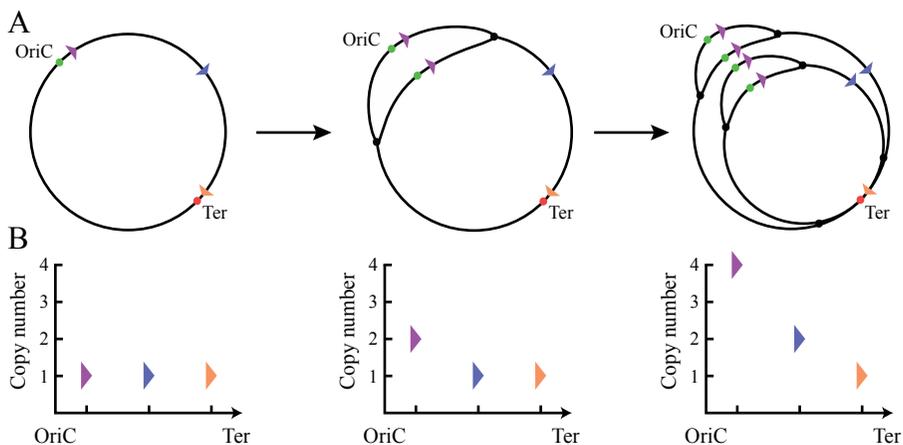


Figure 6. Effective gene copy number as function of chromosome replication and distance to the origin of replication. (A) Consecutive rounds of replication start at the origin of replication (green circle) before the previous replication forks (black circle) reach the terminus (red circle). (B) Gene copy number of three genes with different distance to the origin of replication during various states of replication shown above.

Another example of location bias is the distribution of gene families, such as the ribosomal RNA (rRNA) operons. *Salmonella* and related bacteria have seven copies of rRNA operons (McClelland *et al.*, 2001; Hayashi *et al.*, 2006). Interestingly, five of these seven operons are located on one side of the origin and only two are located on the other side. This bias might be the results of selection against detrimental recombination events. Ribosomal RNA operons contain around 5 kb of almost perfect homology, which makes them hot-spots for recombination (Anderson and Roth, 1977; Hill and Harnish, 1981). All rRNA operons are located on the leading strand of replication. Operons that are located on opposite sides of the origin are therefore inverted repeats that, if recombined, can cause an inversion of the sequence between them (Figure 3) (Hill and Harnish, 1981). This inversion will include the origin of replication and can lead to an imbalance in the size of the two chromosomal halves, which causes a major reduction in bacterial fitness (Hill and Gray, 1988; Esnault *et al.*, 2007). The bias in rRNA location may reduce the probability of these events and increase fitness of the population.

Biased gene composition

The unequal distribution on the genome is not the only bias that affects protein-coding genes. Bias in the composition of genes might be the strongest bias found in the genome and affects every step of translation.

Initiation

Translational initiation can occur at any one of the triplet nucleotide sequences AUG, GUG, UUG or CUG (Kozak, 1983; Wagner *et al.*, 1994). Nevertheless, 89% of all genes in *Salmonella* have an AUG start codon, followed by GUG (8%), UUG (2%) and CUG (<1%). This distribution might simply be a function of the affinity of initiator tRNA to the start codon but there are indications that this is a too simplified view. In *Salmonella*, translation elongation factor Tu (EF-Tu) is encoded by two near identical genes, namely *tufA* and *tufB* (Hughes, 1986). Interestingly, *tufA*, which is the most highly expressed protein-coding gene in *Salmonella*, has a GUG start codon. This would not be expected if the GUG start codon has a low affinity to the initiator tRNA. A systematic replacement of the *tuf* start codons showed another element that might contribute to the start codon selection (Brandis and Hughes, unpublished data). The CUG codon was found to be non-functional as a start codon for both *tuf* genes (Brandis and Hughes, unpublished data). This might explain why CUG is found as a start codon in only seven genes in the *Salmonella* genome of which most are phage genes. UUG was a functional start codon in both *tuf* genes but leads to a reduction in translation of both genes with an accumulative effect of a 5% fitness cost when placed in both *tuf* genes (Brandis and Hughes, unpublished data). The low frequency of UUG start codons in the *Salmonella* genome might there-

fore be due to a low affinity to the initiator tRNA. More surprising was the change of the *tufA* start codon from GUG to AUG. No change in bacterial fitness was observed, indicating that in this specific context there is no difference between the two start codons (Brandis and Hughes, unpublished data). If the two codons can be equally functional in the most highly expressed gene in *Salmonella*, it is difficult to explain the extreme bias towards the AUG start codon with initiator tRNA binding affinity. The answer might be found in the *tufB* start codon. Exchanging the AUG start codon to GUG completely abolished *tufB* expression. This effect was found to be due to the formation of a stable stem-loop structure, which excluded the ribosomal binding site (Brandis and Hughes, unpublished data). The formation of secondary structures in the 5'-UTR has been shown to be detrimental to gene expression (Kudla *et al.*, 2009) and a GC bond forms a more stable structure than an AU bond. The main selective force that drives start codon bias might therefore not be the affinity to the initiator tRNA but rather the selection to reduce secondary structures in the 5'-UTR of genes.

Elongation

The genetic code consists of 64 different codons. Excluding the three stop codons, this leaves 61 codons to encode the 20 different amino acids. Due to this excess of codons, most amino acids are encoded by more than one codon. Even though an amino acid might be encoded by up to six different codons, some codons are more frequently used than others (Table 1). This phenomenon is referred to as codon usage bias or just codon bias.

Table 1. Relative synonymous codon usage (Sharp and Li, 1987) of leucine codons within the *S. typhimurium* LT2 genome.

Leucine codon	All genes	Highly expressed genes ^a	<i>tufA</i>
UUA	0.75	0.16	0.00
UUG	0.70	0.17	0.00
CUU	0.67	0.30	0.00
CUC	0.59	0.23	0.00
CUA	0.28	0.02	0.00
CUG	3.01	5.12	6.00

^aRibosomal proteins, EF-G and EF-Tu.

Codon bias is a universal feature of genomes but the preferred codons may vary both between species and between genes within the same genome. Three main forces are thought to be responsible in shaping the codon bias within a genome, namely mutation, selection and genetic drift. Genetic drift coupled with population bottlenecks may cause a bias that has no functional basis whereas mutation bias, caused by the balance between mutation and repair and selection, acts in favour of optimal codon usage (Bulmer, 1991; Kliman and Hey, 1994; Rocha, 2004). The balance between these forces shapes the codon usage of each gene individually. The more important a

gene is for the cell, the higher is the selective pressure on codon bias in this particular gene. Highly expressed genes therefore have a higher codon bias than other genes within the same genome (Table 1) (Gouy and Gautier, 1982; Goetz and Fuglsang, 2005).

It has been shown that there is only a weak correlation between mRNA and protein levels in *Escherichia coli* (*E. coli*) and yeast (Lu *et al.*, 2006; Taniguchi *et al.*, 2010) and that gene expression of a transgene can be increased over 1,000 fold by changing the codon usage within the gene from the bias found in the original species to the bias found in the species used for expression (Gustafsson *et al.*, 2004). This shows that there is a strong positive correlation between codon bias and translational efficiency.

Translational efficiency is a combination of three factors: translational initiation, translational speed and translational accuracy. The strongest effect on codon bias is associated with translational speed. The speed with which a codon is decoded depends on the cellular concentration of the reading tRNA (Varenne *et al.*, 1984; Dong *et al.*, 1996) and codon bias can therefore be explained by the co-evolution of the codon usage frequency and the abundance of isoaccepting tRNA species. Supporting this model, it has been shown in several species that highly used codons correspond to the most abundant tRNA species (Ikemura, 1981; Ikemura, 1982; Dong *et al.*, 1996; Moriyama and Powell, 1997; Kanaya *et al.*, 1999; Duret, 2000; Rocha, 2004).

Translational accuracy determines the probability of enzymatically functional proteins to be produced. It includes both encoding of the correct amino acid sequence and the correct folding of the protein. These two factors can affect codon bias both positively and negatively. It has been observed that highly conserved amino acids show a significantly higher codon bias than non-conserved amino acids suggesting that codon bias may increase the fidelity of translation at important positions (Stoletzki and Eyre-Walker, 2007; Zhou *et al.*, 2009). Selection for the codon with the highest fidelity might increase the overall codon bias but accurate protein folding might have the opposite effect. Since co-translational protein folding is dependent on the translational speed, certain proteins might need translational pausing at particular positions to support accurate folding, which could be a strong selection for rare, or non-optimal codons at the pause sites (Thanaraj and Argos, 1996; Makhoul and Trifonov, 2002; Zhang *et al.*, 2009; Siller *et al.*, 2010; Saunders and Deane, 2010). This effect has been shown in the *cat* gene that codes for chloramphenicol acyltransferase. Replacing rare codons with 'optimal' codons (those most frequently found in highly expressed genes) within the *cat* gene resulted in a protein with significantly reduced specific activity, probably caused by the effects of altered protein folding kinetics (Komar *et al.*, 1999).

A recent study has shown the effect of efficient translational initiation on codon bias. Translational speed was reduced within the first 30 – 50 codons of most genes. This 'translation ramp' was caused by the presence of a rela-

tively high frequency of rare codons in the beginning of most genes. It is thought that an initially slow translational speed increases the initiation efficiency and optimizes the distance between ribosomes on the mRNA in order to reduce ribosomal stalling on the message. Due to this translational ramp the codon bias observed within the first 30 – 50 codons is highly reduced, even in very highly expressed genes (Fredrick and Ibba, 2010; Tuller *et al.*, 2010).

The interplay of these three forces optimizes the translational efficiency and shapes the codon usage within each gene individually. A higher frequency of rare codons in the beginning of the gene forms a translation ramp and a high codon bias throughout the rest of the gene increases the translational speed except for those parts of the sequence where rare codons cause translational pausing needed for accurate protein folding.

Termination

Translational termination, like initiation, can occur at different codons, namely UAA, UAG and UGA. In *Salmonella* 60% of stop codons are UAA codons, followed by UGA (31%) and finally UAG (9%). These three stop codons are recognised by two distinct peptide chain release factors (RF). RF1 recognises UAA and UAG stop codons and RF2 terminates at UAA and UGA (Scolnick *et al.*, 1968). RF2 is five-fold more abundant than RF1 (Adamski *et al.*, 1994), so that the stop codon bias is a function of peptide chain release factor concentrations. This co-evolution of termination codon usage and release factor concentration is comparable to the co-evolution of codon usage and cognate tRNA abundance and is most likely selected to optimize termination rates.

Antibiotic resistance as a tool to study evolution

The emergence of antibiotic resistance

The massive release of antibiotics, both into patients and the environment, is probably the most-studied long-term evolution ‘experiment’ in modern history. Environmental levels of antibiotics are increasing and, in some areas, have reached levels that are close to clinical treatment concentrations (Bruhn, 2003; Kümmerer, 2009; He and Blaney, 2015). Alongside this trend, it has been observed that an ever increasing proportion of clinical isolates are resistant to a growing number of antibiotic classes and the number of treatment failures is rising (ECDC, 2014). The emergence of antibiotic resistance poses a threat to modern medicine. Without functional antibiotics the treatment of infectious diseases will be increasingly problematic and surgical

operations will become more dangerous due to the increased risk of life-threatening wound infections.

Resistance to antibiotics can be acquired by several different mechanisms that affect either the drug target or the drug itself (Figure 7). Three distinct mechanisms are known to alter the target in a fashion that causes antibiotic resistance. The drug target can be mutated to reduce drug affinity (Hooper, 2000), proteins can bind the drug target and thereby prevent binding of the drug (Connell *et al.*, 2003) or a different protein can be used to catalyse the same reaction as the drug target and thereby bypass the affected protein (Courvalin, 2006). Resistance mechanisms that affect the drug can either alter its effectiveness by degrading or modifying it (Bonnet, 2004; Robicsek *et al.*, 2005) or decrease the intracellular drug concentration by reducing the import or increasing the export of the drug (Levy, 1992; Sandegren and Andersson, 2009). Of these different mechanisms the ones most commonly found are target modifications, alteration of the drug and decreased intracellular concentration due to overexpression of efflux pumps.

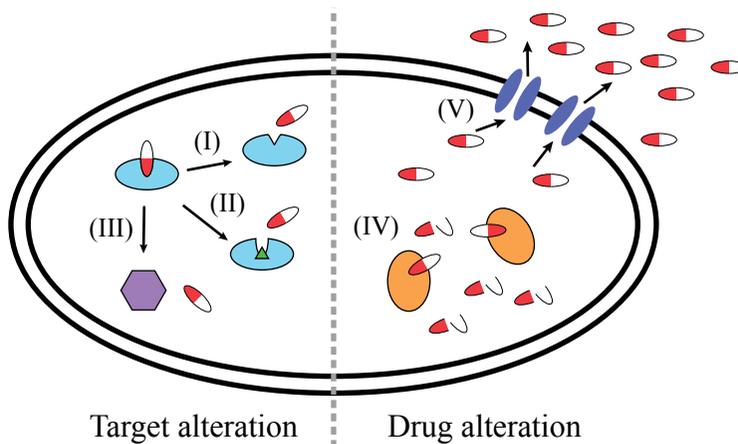


Figure 7. Resistance to antibiotics can be acquired by several different mechanisms that affect either the drug target or the drug itself. The drug target can be mutated to reduce drug affinity (I), proteins can bind the drug target and thereby prevent binding of the drug (II) and a different protein can be used to catalyse the same reaction as the drug target and thereby bypass the affected protein (III). Resistance mechanisms that affect the drug can either alter its effectiveness by degrading or modifying it (IV) or decrease the intracellular drug concentration by reducing the import or increasing the export of the drug (V).

The emergence of resistance can occur by two different mechanisms. A spontaneous mutation in the genome can give rise to resistance, or a resistance gene can be acquired by horizontal gene transfer. Spontaneous mutations can include target mutations or mutations that decrease the intracellular drug concentration by either increased drug efflux due to overexpression of non-specific efflux pumps or decreased drug uptake by deletion of pores

through which the drug enters the cell. Resistance genes that are located on mobile genetic elements such as plasmids and transposons can be acquired by horizontal gene transfer (HGT). These elements can include genes that protect the drug target, inactivate the drug or specifically export it. These two mechanisms of resistance emergence differ significantly. Resistance by mutations must arise in the strain of interest, and the balance of mutational fitness cost and drug selection pressure determines the subsequent increases in the mutant frequency in a population. Resistance by HGT often involves the acquisition of genes that have evolved in a different species, and frequently involves elements carrying several drug resistances, thus presenting a more complex resistance problem. Furthermore, resistance mechanisms on mobile genetic elements can continue to spread horizontally and can therefore easily be transferred between different pathogens.

Fitness cost and compensatory evolution

Many studies have shown that antibiotic resistance is typically associated with a fitness cost for the resistant bacteria (Andersson and Hughes, 2010). This can have different causes depending on the observed resistance mechanism. A mutation in the drug target that reduces the binding affinity for the drug might also alter the functionality of the drug target. This typically causes a fitness cost since drug targets are usually proteins or RNAs that are essential for fast growth (e. g. ribosomes or RNA polymerase) (Nagaev *et al.*, 2001; Sander *et al.*, 2002; Brandis *et al.*, 2015). Overexpressing an efflux pump to export the drug out of the cell can also cause problems for the cell. Multidrug efflux pumps are usually rather non-specific and are therefore able to export multiple drugs. On the other hand, this can lead to the export of nutrients out of the cell, which in turn might decrease the growth rate (Marcusson *et al.*, 2009, Linkevicius *et al.*, 2013). The production of enzymes that specifically cleave or modify certain drugs can involve a cost to the bacterium since these proteins need to be produced, using valuable ribosomes and nutrients that the bacterium could otherwise use for growth (Morosini *et al.*, 2000).

There are two possibilities how a bacterium can adapt to the fitness cost caused by drug resistance, namely (i) reverting to the drug-sensitive state or (ii) acquiring compensatory mutations that ameliorate the fitness cost caused by the resistance mutation. Reversion to the original, drug-sensitive state would restore the growth rate to the wild-type level but, at least for genomic mutations, is unlikely to occur since a very specific mutation is needed to achieve this goal, and in addition the bacterium loses the benefit of resistance. The more likely event, due to the larger mutational target, is the acquisition of compensatory mutations. Even high cost resistance mutations can be compensated almost back to the wild-type growth rate by single com-

pensatory mutations (Figure 8) (Brandis *et al.*, 2012; Brandis and Hughes, 2013).

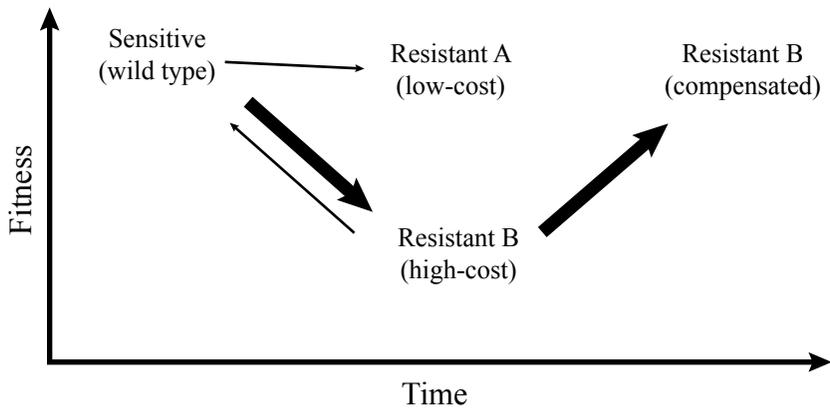


Figure 8. Evolution of drug resistance. Primary resistance mutations typically cause a fitness cost. Dependent on the mutation this cost can be very low to high. High-cost resistance mutation can then acquire secondary mutations that ameliorate the fitness cost caused by the initial resistance mutation.

Antibiotic resistance and evolution

The emergence of antibiotic resistance poses a threat to medicine but gives the possibility to study ongoing evolution on a global scale. Clinical and environmental isolates from all around the world are continuously analysed for resistance and an increasing amount are being sequenced to determine genetic changes that lead to resistance. These rapidly accumulating data are the foundation of a deeper understanding on how evolution works. Multi-resistance plasmids give insight into the evolution of plasmids and horizontal gene transfer (Sandegren *et al.*, 2011). The emergence of new proteins that can modify or export a drug, as well as proteins that can bypass the affected drug targets, give examples on how new proteins can evolve (Linkevicius *et al.*, 2016). Finally, the biased appearance of certain resistance mutations over others in clinical isolates helps to understand how a certain type of selection repeatedly leads to the same outcome and might help to predict future evolution (Komp Lindgren *et al.*, 2003; Müller *et al.*, 2013; Brandis *et al.*, 2015).

Present Investigations

Papers I – III

The selective advantage of codon usage bias (Paper I)

The preferential use of some synonymous codons over others is referred to as synonymous codon usage bias. *S. typhimurium* and other fast-growing microorganisms display an extreme codon usage bias that correlates with gene expression level (Sharp *et al.*, 1988; Sharp, 2005). It is most likely that this biased codon usage is selected for translational efficiency but whether selection is primarily for translation speed or translation accuracy is still unclear (Ikemura, 1981; Akashi, 1994; Stoletzki and Eyre-Walker, 2007). There have been two different theoretical approaches to calculate the strength of the selection for codon bias in highly expressed genes in bacteria but these approaches have resulted in two very different values. Based on selection-mutation-drift theory, a selection coefficient against non-optimal codons was estimated to be approximately 10^{-4} per codon per generation (Bulmer, 1991). In contrast, another theoretical study that was based on synonymous nucleotide polymorphisms in the *gnd* gene from natural isolates of *E. coli* arrived at a selection coefficient of around 10^{-9} per codon per generation (Hartl *et al.*, 1994). Only a limited number of experimental studies have addressed codon usage bias. These studies have found correlations between the observed selective disadvantage of synonymous codons and changes on the mRNA level, such as the formation of Shine-Dalgarno-like sequences and alterations in mRNA folding (Kudla *et al.*, 2009; Lind *et al.*, 2010; Lind and Andersson, 2013; Agashe *et al.*, 2013). Additionally, the insertion of very rare codons has been associated with increased translational error rates, with reading frame shifts, and with a reduced rate of translation (Spanjaard and van Duin, 1988; Sørensen *et al.*, 1989; Kramer and Farabaugh, 2007). It is not possible to draw general conclusions from these studies, on either the magnitude of selection or the selective basis that drives the evolution of synonymous codon usage bias and how this evolution is associated with fast growth rates and highly expressed genes.

To experimentally measure the more general selective value that is associated with synonymous codon usage bias, we decided to systematically substitute non-optimal synonymous codons for optimal codons throughout the coding sequence of two very highly expressed genes, namely *tufA* and

tufB. Using this experimental approach, we tried to address two important questions: (i) what is the magnitude of selection for optimal codons in the highly expressed *tuf* genes, and (ii) is synonymous codon usage bias selected to maximize translation speed or translation accuracy (**Paper I**).

We used *S. typhimurium* LT2 as a model organism for this study. *Salmonella* is a free-living, genetically tractable bacterial species with a strong selection for codon usage bias (Sharp, 2005). Previous studies have shown that synonymous codon usage is strongest in highly expressed genes (Gouy and Gautier, 1982; Goetz and Fuglsang, 2005), so we decided to use the highly expressed *tuf* genes as the targets for this study. In *Salmonella* and related species, translation elongation factor Tu is encoded by two widely separated genes, *tufA* and *tufB* (Jaskunas *et al.*, 1975; Hughes, 1986). The *tuf* genes of *Salmonella* encode identical EF-Tu proteins and are 99% identical at the nucleotide level. EF-Tu is the most highly expressed protein in *Salmonella*, accounting for approximately 9% of the protein mass of *Salmonella* growing exponentially in rich medium, and bacterial growth rate is strictly correlated to EF-Tu abundance (Tubulekas and Hughes, 1993).

We synthesized eighteen distinct synonymous *tuf* alleles. In each allele, we systematically replaced the optimal codons for one or two different amino acids, namely leucine, proline, valine, and arginine, with ten different synonymous codons. To increase the effect on bacterial fitness, we replaced multiple codons (12 to 25) for each amino acid to a less frequently used synonymous codon for that particular amino acid. We placed the eighteen synonymous *tuf* alleles in the chromosomes of *Salmonella* strains at both the native *tufA* and *tufB* positions. The resulting set of isogenic strains was used in experimental quantification of the selective advantage of codon usage bias. We measured the relative fitness cost of each of the synonymous *tuf* alleles with a high-resolution growth competition assay and determined that the average cost of a non-optimal codon is in the range of $0.2 - 4.2 \times 10^{-4}$ per codon per generation. We were able to show that reduced codon usage rather than secondary effects such as mRNA structures or Shine-Dalgarno-like sites causes the measured selective disadvantage. Finally, our data showed that the selective disadvantage per codon correlates to the increase in translational time that is expected due to the use of non-optimal codons. Taken together, our data show that the selective disadvantage of a non-optimal codon in the *tuf* genes is in the range of 10^{-4} per codon per generation, rather than 10^{-9} , and that synonymous codon usage is selected to maximize translation speed (**Paper I**).

Codon usage and gene regulation (Paper II)

In *Salmonella* and related species, elongation factor Tu is expressed from two genes, *tufA* and *tufB* (Jaskunas *et al.*, 1975; Hughes, 1986). Previous studies showed that in *Salmonella* and *E. coli* strains in which EF-TuA and

EF-TuB proteins could be distinguished electrophoretically, about two thirds of the EF-Tu were produced from *tufA* and the remaining one third from *tufB* (van der Meide *et al.*, 1982; Hughes 1990). When *tufB* was inactivated in either *Salmonella* or *E. coli* the total amount of EF-Tu dropped to 67% of the wild-type level, all of it expressed from *tufA* (van der Meide *et al.*, 1982; van der Meide *et al.*, 1983; Hughes, 1990). However, when *tufA* was inactivated in *Salmonella*, the EF-Tu level was also 67% of the wild-type level, indicating that the amount of EF-Tu from *tufB* was increased two-fold in the absence of *tufA* expression (Hughes, 1990). This shows that in *Salmonella* the expression of EF-Tu from *tufA* and *tufB* is regulated differently, and that in the absence of expression from *tufA*, expression from *tufB* can be increased significantly.

We were interested in determining how the *tufB* gene senses the deficiency of cellular EF-Tu levels and increases expression accordingly (**Paper II**). To address the mechanism of *tufB* regulation we evolved an *S. typhimurium* strain in which *tufA* was inactive (slower growing as a result of the inactivation) and selected mutants in which growth rate was improved. We observed that the selected strains had acquired point mutations upstream of the *tufB* coding sequence or synonymous mutations close to the beginning of the *tufB* gene. Based on these results we constructed a total of 121 single point mutations (67 synonymous mutations in the *tufB* coding sequence and 54 5'-UTR mutations) and measured EF-TuB expression using a *tufB-yfp* fusion protein. Next, we constructed and measured a set of 40 *tufB-yfp* fusions with different lengths of *tufB* 3'-truncations to determine the region of the *tufB* sequence that is involved in the regulation. Additionally, we showed that Rho protein function and the presence of rare codons in the early part of the *tufB* gene are involved in the regulation. Taken together, our data indicate the presence of two competing structures of stems and loops that could form in the 5'-end of the *tufB* mRNA. The 'open' structure facilitates expression of *tufB* and production of EF-TuB, while the 'closed' structure decreases *tufB* expression by a Rho-dependent transcriptional termination mechanism. Our data support a model in which translational speed in the early part of the *tufB* mRNA is used to sense the cellular EF-Tu concentration and regulate *tufB* expression by modulating the probability of Rho-dependent transcription termination (**Paper II**).

Codon usage and reading frame maintenance (Paper III)

Our previous studies (**Papers I – II**) have shown that synonymous codons are not translated with the same speed and that this difference in elongation rate can be used for different purposes. In the beginning of the coding sequence it can be used to modulate transcriptional regulation (**Paper II**) and in the rest of the gene it optimises translational speed (**Paper I**). An *E. coli* strain that we isolated during an evolution experiment shows how, under

specific circumstances, a reduced translation rate can also have drastic detrimental effects on reading frame maintenance (**Paper III**).

We isolated a frameshift mutation within the centre of the *rpoB* gene that encodes the beta subunit of RNA polymerase. This gene is essential, and yet a strain carrying the frameshift is viable. We reconstructed the frameshift mutation in a clean background and confirmed by genome sequencing that no external frameshift suppressor is required. We were able to determine the site of frameshift suppression to be a codon pair (CCC-AGG) that is moved into reading frame due to the initial frameshift mutation. The combination of the frameshift-prone CCC proline codon followed by the rare AGG arginine codon causes frameshifting at a high-level (14%). This high-level frameshift event moves translation back into the original reading frame and ensures the production of full-length RpoB protein (**Paper III**). These data show how delicate the selection of codon choice is and how the combination of the ‘wrong’ codons can drastically decrease reading frame maintenance.

Papers IV – V

Biased gene distribution and its effects on recombination

The location and orientation of highly expressed genes in the genome is greatly biased. The majority of highly expressed genes are located close to the origin of replication and are transcribed in the same direction as the leading strand of DNA replication. This bias increases gene copy number during replication (location bias) and reduces the number of clashes between DNA and RNA polymerases (orientation bias) but it also places gene families, such as the *tuf* genes and rRNA operons, in close proximity, so that they can recombine at high frequencies. If these genes are located on the same side of the origin of replication, then they are direct repeats that can form a duplication of the sequence between them (Anderson and Roth, 1977). In theory, direct repeats can also lead to a deletion of the sequence in between (Anderson and Roth, 1977) but since these regions usually contain essential genes, a deletion would be lethal. If the genes are located on opposite sides of the origin of replication, then they form inverted repeats and recombination between these inverted repeats can lead to an inversion of the sequence in between (Hill and Harnish, 1981). Recombination is frequently observed between genes in gene families and is essential for their co-evolution (Abdulkarim and Hughes, 1996).

Recombination between the *tuf* genes (Papers IV – V)

In *Salmonella* and related species, EF-Tu is encoded by two separate genes, *tufA* and *tufB* (Jaskunas *et al.*, 1975; Hughes, 1986). The two genes are lo-

cated on the opposite sides of the origin of replication, and both genes are transcribed in the same direction as the leading strand of DNA replication. *tufA* is located approximately 485 kb from the origin of replication and is the last gene in an operon encoding the ribosomal proteins S7 and S12 as well as the translation elongation factor EF-G. *tufB* is located around 275 kb from the origin and is the last gene in an operon containing 4 tRNA genes, namely *thrU*, *tyrU*, *glyT* and *thrT* (McClelland *et al.*, 2001). It has previously been shown that the *tuf* genes co-evolve and can exchange sequence by recombination, which can repair deleterious mutations within the genes (Abdulkarim and Hughes, 1996).

We were interested in testing if co-evolution can protect the *tuf* genes from inactivation by mobile genetic elements (**Paper IV**). We constructed a 3.3 kb large *cat-sacB-yfp* cassette that enabled us to detect recombination events. We inserted the cassette into seven positions within each of the two *tuf* genes to test (i) if the cassette could be removed by recombination, (ii) if recombination frequencies differed between the two *tuf* genes and (iii) if recombination frequencies differed between the various insertion sites within the genes. We were able to detect recombination events that removed the inserted cassettes in each of the tested positions in at least one of the *tuf* genes. Recombination rates were generally higher when the cassette was inserted in the *tufB* gene, compared to insertions in the *tufA* gene, but overall 10 to 100-fold lower than for the repair of single point mutations. Finally, recombination rates varied with insertion site, so that insertions towards the centre of the *tuf* genes showed higher recombination rates than insertions towards the beginning or the end. Additionally, we were able to show that recombination is RecA dependent and that at least one of the ribosomal RNA operons seems to be involved in inducing recombination. In conclusion, we were able to show that co-evolution protects the *tuf* genes not only from deleterious mutations but also from inactivation by mobile genetic elements (**Paper IV**).

Recombination between the *tuf* genes can repair deleterious mutations and insertions in the genes but can also lead to an inversion of the 760 kb between the *tuf* genes because the two genes are in inverse orientation on the chromosome, and strains carrying the inversion are viable (Hughes, 2000). Large inversions like this are infrequently observed in natural isolates so we were interested in determining the selective forces that reduce their probability of fixation within a population (**Paper V**). We constructed a *Salmonella* strain with a large chromosomal inversion between the *tufA* and *tufB* genes and found that the inversion was associated with reduced fitness (8% relative to wild type). We evolved a strain that carried the inversion, and a *recA* deletion to reduce the probability of re-inversion, with selection for improved growth. We identified putative compensatory mutations of which the majority were predicted to negatively affect the concentration of initiation factor 2 (IF2). IF2 has been shown to competitively inhibit the binding of EF-G to

the ribosome and to inhibit EF-G-dependent hydrolysis of GTP on the ribosome (Cameron *et al.*, 2002). This known competition between IF2 and EF-G suggests that the reduced fitness associated with the inversion might be caused by a reduction in the concentration of EF-G relative to IF2. Since EF-G is co-transcribed with EF-TuA, it is likely that the *tufA/B* hybrid operon created by the inversion produces less protein product than the native *tufA* operon. We concluded that the formation of unstable hybrid operons may act as a counter-selective force that reduces the probability that an inversion between the *tuf* genes gets fixed within a population (**Paper V**).

Papers VI – IX

Rifampicin resistance in *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (MTB) is a pathogenic bacterium that is the cause of the respiratory disease tuberculosis (TB). TB is the second greatest cause of death caused by a single infectious agent with 8.7 million incidents and 1.4 million deaths in the year 2011 (WHO, 2013).

Treatment of TB requires a six-month-long treatment with a combination of four different drugs: rifampicin, isoniazid, ethambutol and pyrazinamide. The increased prevalence of resistance to these antibiotics results in significant problems for effective therapy. MTB strains that are resistant to, at least, rifampicin and isoniazid are defined as multidrug-resistant tuberculosis (MDR-TB). As much as 20% of previously treated patients are now thought to have MDR-TB (WHO, 2013).

Rifampicin is a bactericidal antibiotic that binds the bacterial DNA dependent RNA polymerase at the active centre. Drug binding does not affect transcription initiation but clashes with the 2 – 3 nucleotide long RNA chain and aborts transcription (Artsimovitch and Vassylyev, 2006). Rifampicin resistance can be acquired by mutations in the gene *rpoB* causing single amino acid changes in the β -subunit of RNA polymerase. In 96% of clinical rifampicin resistant MTB isolates, point mutations, small duplications or small deletions have been reported in a stretch of 81 bp (amino acids 507 – 534) of *rpoB* called cluster I or the rifampicin resistance determining region (RRDR) (Ramaswamy and Musser, 1998). In addition, rifampicin resistance mutations can be found in cluster N (amino acids 146 and 148), cluster II (amino acids 563 – 574) and cluster III (amino acid 687) (Campbell *et al.*, 2001). The crystal structure of rifampicin bound to the RNA polymerase shows that it is mainly the amino acid residues in the RRDR that form the binding pocket for rifampicin, explaining why rifampicin resistance mutations are mainly localized in this region. Clusters N, II and III each have only a few amino acid residues that are involved in forming the rifampicin binding pocket (Ho *et al.*, 2009).

Fitness-compensatory mutations in rifampicin-resistant RNA polymerase (Papers VI – VIII)

Resistance to rifampicin is acquired by altering the structure of the drug-binding pocket. Since this binding pocket is located in the bacterial RNA polymerase, an important enzyme for rapid growth, it is most likely that these mutations cause a fitness cost. It has been demonstrated in rifampicin resistant *E. coli* that, at least for some of the mutants, this initial fitness cost could be reduced by secondary mutations in the gene *rpoB* (Reynolds, 2000).

We were interested in testing if mutations in genes other than *rpoB* can compensate the fitness cost of rifampicin resistance mutations (**Paper VI**). Using *Salmonella* as a model organism we evolved a previously characterised, high-cost rifampicin resistance mutation (*rpoB* R529C) with selection for improved growth. Compensatory mutations were identified in the RNA polymerase genes *rpoA*, *rpoB*, and *rpoC*. These findings show that mutations in different components of RNA polymerase are responsible for fitness compensation of rifampicin resistance mutations (**Paper VI**).

While we were working on the compensatory evolution of the *rpoB* R529C mutation, whole genome sequences of clinical, rifampicin resistant, MTB samples, that showed secondary mutations in the genes *rpoA* and *rpoC*, were published. These mutations were proposed to be compensatory mutations, which reduce the fitness cost caused by the initial rifampicin resistance mutations, but no genetic reconstructions were performed to verify this hypothesis (Comas *et al.*, 2011; Casali *et al.*, 2012). We decided to follow up on these data (**Paper VII**). Compensatory mutations in the clinical MTB isolates were associated with the rifampicin resistance mutation *rpoB* S531L. We constructed a *Salmonella* strain with the *rpoB* S531L mutation and selected for improved growth. As before, we were able to identify secondary mutations in the genes *rpoA*, *rpoB* and *rpoC* and genetic reconstruction of these mutations showed that they compensate for the fitness cost caused by the initial rifampicin resistance mutation. Additionally, we were able to show that these mutations are found close to the putative compensatory mutations described in clinical *M. tuberculosis* samples. These data strongly support the hypothesis that the secondary *rpoA* and *rpoC* mutations, found in clinical MTB isolates, are indeed fitness compensatory mutations (**Paper VII**).

So far, all our studies were performed using *Salmonella* as a model organism and even though our results were highly compatible with genome sequencing results from clinical MTB isolates, it might be argued that *Salmonella* is too different from MTB to draw any final conclusions. We therefore decided to use the *Mycobacterium* BCG vaccine strain to analyse the role of a secondary site *rpoB* mutation in a MTB isolate associated with a major MDR-TB outbreak in Tunisia (**Paper VIII**). The MTB isolate carried the *rpoB* double mutation S531L, V615M. We engineered BCG strains that

harboured either the *rpoB* S531L mutation alone or the double mutation *rpoB* S531L and *rpoB* V615M. Compared to the *rpoB* single mutant, the double mutant displayed improved growth characteristics and fitness rates equivalent to wild-type BCG. This finding shows that secondary mutations contribute significantly to the successful transmission of the MDR-TB strain and confirmed that our previous analysis using *Salmonella* was indeed comparable to MTB (**Paper VIII**).

Selection of clinical rifampicin resistance (Paper IX)

Our previous studies confirmed that compensatory evolution is an important feature of the development of rifampicin resistance in MTB. Next, we were interested in finding out which forces drive the resistance development in the clinical setting since the presence of compensatory mutations indicate that the initial selection of resistance mutations is not mainly driven by fitness but by other factors (**Paper IX**).

As before, we used *Salmonella* as a model organism. We did an extensive literature research to identify every known *rpoB* mutation that is proposed to confer rifampicin resistance in MTB or *E. coli*. We then constructed isogenic *Salmonella* strains with each of the identified mutations. In total, 122 unique mutations (78 substitutions, 4 duplications and 40 deletions) successfully conferred rifampicin resistance in *Salmonella*. For each of these mutants, we measured the minimal inhibitory concentration (MIC) of rifampicin and the fitness cost in the presence and absence of rifampicin. We then compared the results to available sequence data on clinical MTB isolates. Our data suggest that rifampicin resistance mutations in clinical MTB isolates are primarily selected for high-level resistance (rather than low fitness cost), and that fitness-compensatory mutations that reduce the fitness cost caused by the resistance mutation are subsequently selected (**Paper IX**).

Papers X – XII

Ciprofloxacin resistance in *Escherichia coli*

Ciprofloxacin is an important broad-spectrum antibiotic that is used to treat a wide variety of infections in human medicine, including urinary tract infections caused by *E. coli* (Aidara-Kane, 2012; Castro et al., 2013). The main binding target of ciprofloxacin is the DNA gyrase and the drug-gyrase complex retains DNA cutting activity but loses the ability to re-ligate the cut DNA strand. This inhibition leads to an accumulation of double-stranded DNA breaks that cause cell death (Chen *et al.*, 1996). Due to its intensive use, resistance to ciprofloxacin has been increasing (Bow, 2011; Chen *et al.*, 2012). Unlike resistance development to rifampicin there are no single ge-

netic alterations that can raise the MIC above the clinical breakpoint of 1 mg/L (Kahlmeter *et al.*, 2003; Marcusson *et al.*, 2009). There are two known classes of mutations that increase the MIC of ciprofloxacin in *E. coli*: (i) mutations in the primary and secondary drug target genes: *gyrA* and *gyrB* (DNA gyrase) as well as *parC* and *parE* (DNA topoisomerase IV) and (ii) mutations in the genes that encode regulators of drug efflux: *marR*, *acrR* and *soxR* (Kumagai *et al.*, 1996; Hooper, 1999, Wang *et al.*, 2001; Fàbrega *et al.*, 2010; Betitra *et al.*, 2014). A combination of three to four of these mutations is required to increase the MIC to ciprofloxacin above the clinical breakpoint (Marcusson *et al.*, 2009).

Identification of a novel class of ciprofloxacin resistance mutations (Paper X – XI)

Multiple mutations are needed to increase the MIC to ciprofloxacin of *E. coli* above the clinical breakpoint but little is known about the order in which these mutations appear. We wanted to identify trajectories that can lead to clinical resistance (**Paper X**). We evolved independent lineages of *E. coli* in liquid culture and on agar in the presence of increasing concentrations of ciprofloxacin to above the clinical resistance breakpoint. Deep sequencing was used to analyse the genetic composition of populations at different stages during evolution in liquid. We found that during the evolution of resistance to ciprofloxacin, different mutational classes accumulate in successive waves, and that a novel class of mutations (mutations affecting genes involved in transcription and translation) contributed to the development of resistance. At each stage in evolution, multiple mutant alleles of genes were present but these were filtered as selection pressure increased. The initial wave of mutations affected primarily the drug target gene *gyrA* with one mutation typically going to fixation early in evolution. This was followed by a wave of mutations affecting regulators of drug efflux (*marR*, *acrR*, *soxR*) with one mutation typically going to fixation late in evolution. Third and fourth waves of mutations affected genes with roles in transcription and translation, and additional drug target genes (*gyrB*, *parC*, *parE*), respectively. Mutations affecting transcription and translation were selected both in liquid culture evolution (transfer bottleneck 3×10^8 cfu) and agar evolution (single cell bottleneck). Whole genome sequencing of individual clones from liquid and agar evolution experiments confirmed the linkage of these different mutational classes on individual genomes. The data suggest that mutations affecting transcription and/or translation functions play a role in ciprofloxacin resistance development, and that the trajectories of evolution, while variable at the allele and gene level, follow a predictable path to resistance when viewed at a higher level of function (**Paper X**).

We decided to investigate how mutations affecting transcription contribute to the development of ciprofloxacin resistance (**Paper XI**). Mutations in the *rpoB* gene were frequently observed during the evolution of ciprofloxacin resistance and we constructed isogenic *E. coli* strains that recapitulated the evolutionary trajectories of three lineages that acquired distinct *rpoB* mutations. We found that these mutations increased the MIC of ciprofloxacin and conferred a competitive growth advantage in the presence of ciprofloxacin. The *rpoB* mutations also increased the MIC for the unrelated antibiotic chloramphenicol, which indicates that increased drug efflux might be the cause for the reduced drug susceptibility. Using quantitative RT-PCR, we were able to show that the *rpoB* mutations lead to an increased transcription of the drug efflux pumps MdtK, supporting this hypothesis. These data show how mutations in genes that are responsible for transcription can contribute to the evolution of ciprofloxacin resistance (**Paper XI**).

Selection of clinical ciprofloxacin resistance (Paper XII)

We have previously shown that mutations affecting genes involved in transcription and translation can contribute to the development of ciprofloxacin resistance when selected in the lab (**Paper X – XI**) but so far we were not able to determine whether these mutations contribute to the development of clinical ciprofloxacin resistance. There are many genes that are involved in transcription and translation and many of these genes show differences on the amino acid level between phylogenetically diverse clinical isolates of *E. coli*. It is so far not possible to dissect natural variations from selected resistance mutations. We therefore decided to investigate the trajectories of ciprofloxacin-resistance evolution in clinical *E. coli* isolates using only the ‘standard’ resistance mutations (**Paper XII**).

We analysed the patterns of ciprofloxacin resistance mutations present in a large set of clinical *E. coli* isolates and found that the majority of isolates carried the mutation combination *gyrA* S83L D87N *parC* S80I. Based on the overall frequencies and combinations in which these three mutations were found in clinical isolates, we deduced that the triple mutant evolves in one particular trajectory: (i) *gyrA* S83L, (ii) *parC* S80I and (iii) *gyrA* D87N. We constructed a set of single, double, triple and quadruple mutants to experimentally test the trajectories of resistance development and found that the trajectory found in clinical isolates represents, at each step, the mutation with the highest increase in MIC to ciprofloxacin combined with the smallest effect on fitness. We then decided to model the development of ciprofloxacin resistance *in silico* to identify conditions under which the clinical trajectory is selected for. We used a small number of experimentally determined resistance-related parameters (mutation rate, MIC and relative fitness) and varied the strength of selective pressure as well as the bottleneck size and the total population size. The modelling results were highly consistent with the

experimental data such that the deduced clinical trajectory was the only trajectory that leads to the most frequent triple mutant. In addition, the modelling suggested that clinical resistance evolves preferentially under conditions where mutation supply is large. However, even small populations can evolve the clinical resistance genotype if the drug selection pressure is very low.

In conclusion, our experimental data and modelling show that high-level mutational resistance to ciprofloxacin in *E. coli* can evolve under conditions of low drug selective pressure (as found environmentally), but also under conditions of relatively high drug selective pressure as long as mutation supply is not limiting. In addition, we have shown that the interplay between susceptibility and relative fitness associated with different genotypes is the most critical factor determining which mutations are selected during the evolution of clinically successful resistant mutants (**Paper XII**).

Concluding Remarks

The aim of this thesis has been to investigate the forces that have driven evolution in the past and to demonstrate how new conditions (such as antibiotic pressure) affect evolution in the present. We were mainly interested in why evolution repeatedly leads to the same outcome and what the consequences of this biased evolution are.

The selection for optimal codon usage displays the weakest selection pressure that has been determined experimentally so far and is an example of how even the smallest advantages can have a big influence over time (**Paper I**). It also shows how biased evolution can have secondary beneficial as well as detrimental effects. Biased codon usage in combination with the associated bias in tRNA abundance leads to the presence of a variety of synonymous codons that encode identical information on the protein level but that are decoded with different speeds. These differences in decoding speed can be used by the bacteria as an additional level of regulation for optimal transcription, translation and protein folding (**Paper II**). On the other hand, when things go ‘wrong’, for example due to mutations, the different decoding velocities can have major detrimental effects on coding frame maintenance (**Paper III**). Another selective pressure that has shaped the bacterial genome over the last millions of years is the selection for gene location. As observed with the codon usage bias, biased gene location has both beneficial, as well as detrimental secondary effects. The locations of the *tuf* genes lead to recombination events at a high frequency that protect the genes from inactivation (**Paper IV**) but it can also cause the formation of a large chromosomal inversion that is detrimental for bacterial fitness (**Paper V**).

The evolution of antibiotic resistance shows how differences in the selective conditions (different species and life-styles) have major effects on the outcome of evolution. The life-styles of clinical *M. tuberculosis* and *E. coli* could not be more different. *M. tuberculosis* is very slow growing with a relatively small population size, while *E. coli* has to grow very rapidly with a large population size to be able to cause urinary tract infections. These differences lead to very different types of resistance evolution. In the slow-growing MTB, resistance mutations are selected for high-level resistance. Fitness is initially less important and over time fitness costs can be ameliorated by compensatory mutations (**Papers VI – IX**). The need for rapid growth causes ciprofloxacin resistance mutations in *E. coli* not only to be

selected on the basis of high-level resistance but also on high fitness. Compensatory evolution is therefore not required (**Papers X – XII**).

Taken together, our studies show that there are usually several trajectories that evolution can take that lead to the same result. These trajectories might represent single mutations or multi-step events. Each of these trajectories has the possibility to occur but they are far from being equal. Every step in the trajectory has a certain chance to occur and displays a variety of beneficial and detrimental effects. Many factors, such as mutation rate, the environmental conditions or the life-style of the bacterium affect the net result on how frequent and beneficial the particular step is. Each trajectory is the product of all of its single steps and has a total frequency and benefit value. It is most likely that there is one trajectory that is more likely to occur and more beneficial than the others. Over time, selection will favour this particular trajectory and lead to biased evolution.

Acknowledgements

It has been almost seven years since I first moved to Sweden and there are many people that I would like to thank for their help and support during these years.

First I would like to thank **Catharina Svensson**. When I initially applied for the Master program in Infection Biology here in Uppsala, my application was denied by studera.nu (up to this day I have no idea why) and without Catharina's help my Swedish adventure would have ended before it could begin. Unfortunately, this meant that I was admitted to the Master program only weeks before the semester started and everybody who ever tried to find housing in Uppsala might know how impossible it is to spontaneously move here. I already packed a tent to sleep in but luckily, **Åsa**, **Mattias** and **Alexander** took me into their home and eventually helped me to find an apartment in Uppsala.

My biggest thanks go to my supervisor **Diarmaid**! Thank you for taking me into your group even though I had very limited experiences with microbiology. When I first came to Sweden I could not differ between *Salmonella* and *Staphylococcus*; they both start with an S after all. I learned a lot from you during these years and I am grateful that you always let me follow my own ideas. It has been a great experience for me and I hope we will be able to stay connected.

I also want to thank my co-supervisor **Dan** for help and comments on my projects. Although you made it tough to find a committee for my defence (the list of people in the field that he hasn't published with is very short).

To past and current colleagues in Diarmaid's group: **Disa** and **Marie**, it was great to meet you even though our time together was rather short. It is nice that you both came back to BMC eventually. **Jessica**, thank you for all the help when I first started in the group. It was fun to work with you on the *tufB* regulation paper. **Doug**, thank you for all the fruitful discussions that we had, especially the ones about cheetahs. You were a lot of help and it was great to work with you. **Franzi**, I always had a lot of fun trying to talk about science to you in German. I never managed to say a single sentence without using English words. I guess I should stay away from German conferences. **Eva**, thank

you for a fun collaboration. We did not manage to finish it in time for my thesis but I am sure we will be done when it is your turn. **Linnéa**, I know that you are at the point in your PhD studies where you feel as if things are not progressing fast enough. No worries, everything will turn out fine and if something does not work just give it a jujutsu kick. **Lisa**, it is fun to have somebody else in the group that also works on gene regulation. Now you are the next in line to write your thesis. I'm sure you will do a good job. **CaoSha**, you always say that I have the most interesting projects. You are one of the few people within the group that I have not published with, yet. Maybe we should change this so that you can also be part of an interesting project. **Kavita**, it is nice that you left the "children's" office and moved over to us grown ups. Now the work can begin for real. **Tua**, the Christmas lights you hung up in the lab made the holidays cosy. It is nice that you are staying on. I also want to thank my students **Rahel**, **Andreas** and **Sara** for their contributions to my projects

A special thanks also goes to all the other members of the D7:3 corridor. **Linus**, thank you for many interesting discussions and suggestions during the seminars. **Jocke**, I am happy to see that you are starting your own group. I am sure you are going to do a great job. **Lionel**, welcome to the corridor. **Anna**, it is always fun to have you around and to discuss our projects. It was a great idea to have an aquarium in the corridor even though Dan ate almost all of the fish (I feel that I should mention here that we called one of the fish Dan). **Ulrika**, Dan's lab would collapse without you. **Karin**, it is always fun to have fika with you. **Hervé**, thank you for the many suggestions during the seminars. They were always appreciated. **Hind**, welcome to Sweden. You are Jocke's first PhD student, so no pressure. **Fredrika** and **Michael**, you make me feel like I'm back in Germany. With you there is always some German spoken in the corridor or in the radio (I am not sure if this is a good thing though). **Marius**, we have both been super-villains in the same spex movie. I guess we could use that as a plan B if the scientific carrier fails (In case somebody from VR reads this: Do the world a favour and fund us! Or else...). **Omar**, last Christmas we established that we both work too much. Your wife might have stopped complaining about it but I doubt that mine ever will. **Jessica**, too bad that you won't stay long enough for my defence. Good luck back in NM. **Hava**, I guess you are the next one in line. Good luck with your defence. **Lisa** (you guys now which one I mean), thank you for lending us all the baby equipment. I'm sure it will be well used. **Erik L**, please don't try too hard to disprove me with your work on synonymous codons. **Erik G**, thank you for all the fluorescence proteins. They made my projects a lot easier. **Jon**, thank you for your valuable comments on my projects. **Sohaib**, we started the Master course together. It is nice that you stayed for a PhD.

I would also like to thank my collaborators outside the D7:3 corridor. **Lars Liljas** thank you for your help with the RpoB structures. **Helmi Mardassi** and his group, thank you for a fruitful collaboration on the MDR-TB outbreak strain.

Ich danke meinen Eltern **Gabriele** und **Onno**, die mich stets unterstützt haben. Danke auch an meine Schwester **Lilli** für die vielen Abende, die wir - man könnte fast sagen nutzlos - mit Fernsehen und Computerspielen verbrachten. Trotzdem möchte ich diese nicht missen. Ganz besonders bedanke ich mich bei **Christina** und **Dietmar**, die immer für mich da waren.

Til min norske familie: **Magnhild**, **Peter** og **Gunnhild**, takk for at dåkke tok meg inn i familien. Takk til alle i Norge som gjer at eg kjenner meg velkommen der når me er på besøk. **Jorunn**, **Astrid** and **John**, det er alltid gøy å ha dåkke på besøk her i Uppsala.

Finally, I have to thank my wonderful wife **Lisa**. Without you none of this would have been possible. People keep on asking how we are able to live and work together without getting fed up with each other but I would not want to miss a minute of this great trip we had so far. We had to go through some tough times during the last couple of years but you have been my rock that helped me to stay positive. Soon we will be a family of three. Thank you for putting up with me during these last weeks as I am writing this thesis. I know you deserve a lot more attention during this pregnancy than I am able to give to you right now. I love you!

References

- Abdulkarim F & Hughes D (1996) Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. *J Mol Biol* **260**: 506–522.
- Adamski FM, McCaughan KK, Jørgensen F, Kurland CG & Tate WP (1994) The concentration of polypeptide chain release factors 1 and 2 at different growth rates of *Escherichia coli*. *J Mol Biol* **238**: 302–308.
- Agashe D, Martinez-Gomez NC, Drummond DA & Marx CJ (2013) Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol Biol Evol* **30**: 549–560.
- Aidara-Kane A (2012) Containment of antimicrobial resistance due to use of antimicrobial agents in animals intended for food: WHO perspective. *Rev Sci Tech* **31**: 277–287.
- Akashi H (1994) Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**: 927–935.
- Anderson RP & Roth JR (1977) Tandem genetic duplications in phage and bacteria. *Annu Rev Microbiol* **31**: 473–505.
- Andersson DI & Hughes D (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* **8**: 260–271.
- Artsimovitch I & Vassilyev DG (2006) Is it easy to stop RNA polymerase? *Cell Cycle* **5**: 399–404.
- Betitra Y, Teresa V, Miguel V & Abdelaziz T (2014) Determinants of quinolone resistance in *Escherichia coli* causing community-acquired urinary tract infection in Bejaia, Algeria. *Asian Pac J Trop Med* **7**: 462–467.
- Bonnet R (2004) Growing group of extended-spectrum β -lactamases: the CTX-M enzymes. *Antimicrob Agents Chemother* **48**: 1–14.
- Bow EJ (2011) Fluoroquinolones, antimicrobial resistance and neutropenic cancer patients. *Curr Opin Infect Dis* **24**: 545–553.
- Brandis G & Hughes D (2013) Genetic characterization of compensatory evolution in strains carrying *rpoB* Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *J Antimicrob Chemother* **68**: 2493–2497.
- Brandis G, Pietsch F, Alemayehu R & Hughes D (2015) Comprehensive phenotypic characterization of rifampicin resistance mutations in *Salmonella* provides insight into the evolution of resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* **70**: 680–685.
- Brandis G, Wrands M, Liljas L & Hughes D (2012) Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Mol Microbiol* **85**: 142–151.
- Breslauer KJ, Frank R, Blöcker H & Marky LA (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* **83**: 3746–3750.
- Bruhn ST (2003) Pharmaceutical antibiotic compounds in soils—a review. *J Plant Nutr Soil Sci* **166**: 145–167.
- Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.

- Cameron DM, Thompson J, March PE & Dahlberg AE (2002) Initiation Factor IF2, Thiostrepton and Micrococcin Prevent the Binding of Elongation Factor G to the Escherichia coli Ribosome. *J Mol Biol* **319**: 27–35.
- Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A & Darst SA (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**: 901–912.
- Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, *et al.* (2012) Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* **22**: 735–745.
- Castro W, Navarro M & Biot C (2013) Medicinal potential of ciprofloxacin and its derivatives. *Future Med Chem* **5**: 81–96.
- Chen CR, Malik M, Snyder M & Drlica K (1996) DNA gyrase and topoisomerase IV on the bacterial chromosome: quinolone-induced DNA cleavage. *J Mol Biol* **258**: 627–637.
- Chen H, Shiroguchi K, Ge H & Xie XS (2015) Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Mol Syst Biol* **11**: 781–781.
- Chen YH, Ko WC & Hsueh PR (2012) The role of fluoroquinolones in the management of urinary tract infections in areas with high rates of fluoroquinolone-resistant uropathogens. *Eur J Clin Microbiol Infect Dis* **31**: 1699–1704.
- Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, *et al.* (2011) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* **44**: 106–110.
- Connell SR, Tracz DM, Nierhaus KH & Taylor DE (2003) Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance. *Antimicrob Agents Chemother* **47**: 3675–3681.
- Cooper S & Helmstetter CE (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *Journal of Molecular Biology* **31**: 519–540.
- Courvalin P (2006) Vancomycin resistance in gram-positive cocci. *Clin Infect Dis* **42 Suppl 1**: 25–34.
- Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, *et al.* (2015) Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biol* **16**: 277.
- Dong H, Nilsson L & Kurland CG (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol* **260**: 649–663.
- Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* **88**: 7160–7164.
- Duret L (2000) tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet* **16**: 287–289.
- ECDC (2014) *Antimicrobial resistance surveillance in Europe 2013. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net)*. European Centre for Disease Prevention and Control.
- Esnault E, Valens M, Espéli O & Boccard F (2007) Chromosome structuring limits genome plasticity in Escherichia coli. *PLoS Genet* **3**: e226.
- Fàbrega A, Martin RG, Rosner JL, Tavio MM & Vila J (2010) Constitutive SoxS expression in a fluoroquinolone-resistant strain with a truncated SoxR protein and identification of a new member of the marA-soxS-rob regulon, mdtG. *Antimicrob Agents Chemother* **54**: 1218–1225.

- Frank AC & Lobry JR (1999) Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238**: 65–77.
- Frank AC & Lobry JR (2000) Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* **16**: 560–561.
- Fredrick K & Ibba M (2010) How the Sequence of a Gene Can Tune Its Translation. *Cell* **141**: 227–229.
- French S (1992) Consequences of replication fork movement through transcription units in vivo. *Science* **258**: 1362–1365.
- Goetz RM & Fuglsang A (2005) Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from Escherichia coli. *Biochem Biophys Res Commun* **327**: 4–7.
- Gouy M & Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* **10**: 7055–7074.
- Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**: 2286–2290.
- Gustafsson C, Govindarajan S & Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* **22**: 346–353.
- Haigh J (1978) The accumulation of deleterious genes in a population—Muller's Ratchet. *Theor Popul Biol* **14**: 251–267.
- Hartl DL, Moriyama EN & Sawyer SA (1994) Selection intensity for codon bias. *Genetics* **138**: 227–234.
- Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, *et al.* (2006) Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**: 1–5.
- He K & Blaney L (2015) Systematic optimization of an SPE with HPLC-FLD method for fluoroquinolone detection in wastewater. *J Hazard Mater* **282**: 96–105.
- Helmstetter CE & Cooper S (1968) DNA synthesis during the division cycle of rapidly growing Escherichia coli B/r. *J Mol Biol* **31**: 507–518.
- Hill CW & Gray JA (1988) Effects of chromosomal inversion on cell fitness in Escherichia coli K-12. *Genetics* **119**: 771–778.
- Hill CW & Harnish BW (1981) Inversions between ribosomal RNA genes of Escherichia coli. *Proc Natl Acad Sci USA* **78**: 7069–7072.
- Ho MX, Hudson BP, Das K, Arnold E & Ebright RH (2009) Structures of RNA polymerase–antibiotic complexes. *Curr Opin Struct Biol* **19**: 715–723.
- Hooper DC (1999) Mechanisms of fluoroquinolone resistance. *Drug Resist Updat* **2**: 38–55.
- Hooper DC (2000) Mechanisms of action and resistance of older and newer fluoroquinolones. *Clin Infect Dis* **31 Suppl 2**: S24–8.
- Hughes D (1986) The isolation and mapping of EF-Tu mutations in Salmonella typhimurium. *Mol Gen Genet* **202**: 108–111.
- Hughes D (1990). Both genes for EF-Tu in Salmonella typhimurium are individually dispensable for growth. *J Mol Biol* **215**(1): 41–51.
- Hughes D (2000) Co-evolution of the tuf genes links gene conversion with the generation of chromosomal inversions. *J Mol Biol* **297**: 355–364.
- Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol* **151**: 389–409.

- Ikemura T (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *J Mol Biol* **158**: 573–597.
- Jaskunas SR, Lindahl L & Nomura M (1975) Identification of two copies of the gene for the elongation factor EF-Tu in *E. coli*. *Nature* **257**: 458–462.
- Kahlmeter G, Brown DFJ, Goldstein FW, MacGowan AP, Mouton JW, Osterlund A, *et al.* (2003) European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. *J Antimicrob Chemother* **52**: 145–148.
- Kanaya S, Yamada Y, Kudo Y & Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**: 143–155.
- Kliman RM & Hey J (1994) The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. *Genetics* **137**: 1049–1056.
- Komar AA, Lesnik T & Reiss C (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett* **462**: 387–391.
- Komp Lindgren P, Karlsson A & Hughes D (2003) Mutation Rate and Evolution of Fluoroquinolone Resistance in *Escherichia coli* Isolates from Patients with Urinary Tract Infections. *Antimicrob Agents Chemother* **47**: 3222–3232.
- Kozak MM (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol Rev* **47**: 1–45.
- Kramer EB & Farabaugh PJ (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**: 87–96.
- Kudla G, Murray AW, Tollervey D & Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**: 255–258.
- Kuempel PL., Maglothin P & Prescott DM (1973) Bidirectional termination of chromosome replication in *Escherichia coli*. *Mol Gen Genet* **125**: 1–8.
- Kumagai Y, Kato JI, Hoshino K, Akasaka T, Sato K & Ikeda H (1996) Quinolone-resistant mutants of *Escherichia coli* DNA topoisomerase IV parC gene. *Antimicrob Agents Chemother* **40**: 710–714.
- Kümmerer K (2009) Antibiotics in the aquatic environment--a review--part I. *Chemosphere* **75**: 417–434.
- Levy SB (1992) Active efflux mechanisms for antimicrobial resistance. *Antimicrob Agents Chemother* **36**: 695–703.
- Lind PA & Andersson DI (2013) Fitness costs of synonymous mutations in the rpsT gene can be compensated by restoring mRNA base pairing. *PLoS ONE* **8**: e63373.
- Lind PA, Berg OG & Andersson DI (2010) Mutational robustness of ribosomal protein genes. *Science* **330**: 825–827.
- Linkevicius M, Sandegren L & Andersson DI (2013) Mechanisms and fitness costs of tigecycline resistance in *Escherichia coli*. *J Antimicrob Chemother* **68**: 2809–2819.
- Linkevicius M, Sandegren L & Andersson DI (2016) Potential of Tetracycline Resistance Proteins To Evolve Tigecycline Resistance. *Antimicrob Agents Chemother* **60**: 789–796.
- Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**: 660–665.
- Lu P, Vogel C, Wang R, Yao X & Marcotte EM (2006) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**: 117–124.

- Makhoul CH & Trifonov EN (2002) Distribution of Rare Triplets Along mRNA and Their Relation to Protein Folding. *J Biomol Struct Dyn* **20**: 413–420.
- Marcusson LL, Fridmodt-Møller N & Hughes D (2009) Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* **5**: e1000541.
- McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, *et al.* (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852–856.
- Mirkin EV & Mirkin SM (2005) Mechanisms of Transcription-Replication Collisions in Bacteria. *Mol Cell Biol* **25**: 888–895.
- Moriyama EN & Powell JR (1997) Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol* **45**: 514–523.
- Morosini MI, Ayala JA, Baquero F, Martínez JL & Blázquez J (2000) Biological cost of AmpC production for *Salmonella enterica* serotype Typhimurium. *Antimicrob Agents Chemother* **44**: 3137–3143.
- Muto A & Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci USA* **84**: 166–169.
- Müller B, Borrell S, Rose G & Gagneux S (2013) The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Trends Genet* **29**: 160–169.
- Nagaev I, Björkman J, Andersson DI & Hughes D (2001) Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. *Mol Microbiol* **40**: 433–439.
- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA & Hattori M (2006) The 160-Kilobase Genome of the Bacterial Endosymbiont *Carsonella*. *Science* **314**: 267–267.
- O'Brien SJ, Roelke ME, Marker L, Newman A, Winkler CA, Meltzer D, *et al.* (1985) Genetic basis for species vulnerability in the cheetah. *Science* **227**: 1428–1434.
- Prescott DM & Kuempel PL (1972) Bidirectional replication of the chromosome in *Escherichia coli*. *Proc Natl Acad Sci USA* **69**: 2842–2845.
- Proshkin S, Rahmouni AR, Mironov A & Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**: 504–508.
- Ramaswamy S & Musser JM (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* **79**: 3–29.
- Reyes-Lamothe R, Possoz C, Danilova O & Sherratt DJ (2008) Independent Positioning and Action of *Escherichia coli* Replisomes in Live Cells. *Cell* **133**: 90–102.
- Reynolds MG (2000) Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* **156**: 1471–1481.
- Robicsek A, Strahilevitz J, Jacoby GA, Macielag M, Abbanat D, Park CH, *et al.* (2005) Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat Med* **12**: 83–88.
- Rocha EPC (2004) Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14**: 2279–2286.
- Sandegren L & Andersson DI (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol* **7**: 578–588.
- Sandegren L, Linkevicius M, Lytsy B, Melhus A & Andersson DI (2011) Transfer of an *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-specific plasmid associated with a major nosocomial outbreak. *J Antimicrob Chemother* **67**: 74–83.

- Sander P, Springer B, Prammananan T, Sturmfels A, Kappler M, Pletschette M & Böttger EC (2002) Fitness Cost of Chromosomal Drug Resistance-Confering Mutations. *Antimicrob Agents Chemother* **46**: 1204–1211.
- Saunders R & Deane CM (2010) Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* **38**: 6719–6728.
- Schaaper RM, Danforth BN & Glickman BW (1986) Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli* lacI gene. *J Mol Biol* **189**: 273–284.
- Scolnick E, Tompkins R, Caskey T & Nirenberg M (1968) Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci USA* **61**: 768–774.
- Sharp PM (2005) Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* **33**: 1141–1153.
- Sharp PM & Li WH (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281–1295.
- Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH & Wright F (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res* **16**: 8207–8211.
- Siller E, DeZwaan DC, Anderson JF, Freeman BC & Barral JM (2010) Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *J Mol Biol* **396**: 1310–1318.
- Spanjaard RA & van Duijn J (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc Natl Acad Sci USA* **85**: 7967–7971.
- Stoletzki N & Eyre-Walker A (2007) Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol* **24**: 374–381.
- Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci USA* **48**: 582–592.
- Sørensen MA, Kurland CG & Pedersen S (1989) Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* **207**: 365–377.
- Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, *et al.* (2010) Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**: 533–538.
- Thanaraj TA & Argos P (1996) Ribosome-mediated translational pause and protein domain organization. *Protein Sci* **5**: 1594–1612.
- Tubulekas I & Hughes D (1993) Growth and translation elongation rate are sensitive to the concentration of EF-Tu. *Mol Microbiol* **8**: 761–770.
- Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaboroske J, *et al.* (2010) An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **141**: 344–354.
- van der Meide PH, Vijgenboom E, Dicke M & Bosch L (1982). Regulation of the expression of tufA and tufB, the two genes coding for the elongation factor EF-Tu in *Escherichia coli*. *FEBS Lett* **139**(2): 325–330.
- van der Meide PH, Vijgenboom E, Talens, A & Bosch L (1983). The role of EF-Tu in the expression of tufA and tufB genes. *Eur J Biochem* **130**(2): 397–407.
- Varenne S, Buc J, Lloubes R & Lazdunski C (1984) Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* **180**: 549–576.
- Wagner LA, Gesteland RF, Dayhuff TJ & Weiss RB (1994) An efficient Shine-Dalgarno sequence but not translation is necessary for lacZ mRNA stability in *Escherichia coli*. *J Bacteriol* **176**: 1683–1688.

- Wang H, Dzink-Fox JL, Chen M & Levy SB (2001) Genetic characterization of highly fluoroquinolone-resistant clinical *Escherichia coli* strains from China: role of *acrR* mutations. *Antimicrob Agents Chemother* **45**: 1515–1521.
- Watson JD & Crick FHC (1953) A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.
- WHO (2013) Global Tuberculosis Report 2012. World Health Organization.
- Zhang G, Hubalewska M & Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* **16**: 274–280.
- Zheng H & Wu H (2010) Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. *BMC Bioinformatics* **11 Suppl 11**: S7.
- Zhou T, Weems M & Wilke CO (2009) Translationally Optimal Codons Associate with Structurally Sensitive Sites in Proteins. *Mol Biol Evol* **26**: 1571–1580.

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1198*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-276456



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016